Estadística

Descriptiva

y

Probabilidad

(Teoría y problemas)

Estadística Descriptiva y Probabilidad

(Teoría y problemas)

3ª Edición

Autores
I. Espejo Miranda
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
A. M. Rodríguez Chía
A. Sánchez Navas
C. Valero Franco

UCA Universidad de Cádiz
Servicio de Publicaciones

Edita: Servicio de Publicaciones de la Universidad de Cádiz
C/ Dr. Marañón, 3
11002 Cádiz

http://www.uca.es/publicaciones

ISBN: 978-84-9828-058-6

Depósito legal:

Índice general

Prólogo

1. Introducción

El objetivo principal que se persigue con este libro es el de ofrecer a los alumnos de titulaciones experimentales un manual estadístico básico que, sin dejar de lado el rigor conceptual, proporcione una visión práctica e intuitiva de la estadística descriptiva y el cálculo de probabilidades, campos básicos y fundamentales de la ciencia estadística.

Los contenidos del manual Estadística Descriptiva y Probabilidad se han organizado en dos partes, en la primera de ellas se estudia la estadística descriptiva, dedicándose la segunda al cálculo de probabilidades. Los objetivos de cada una de las partes son esencialmente distintos, en efecto, mientras que la estadística descriptiva tiene interés por sí misma, poniendo de manifiesto los aspectos más relevantes de un conjunto de datos; el cálculo de probabilidades, aunque pueden tener en ocasiones una utilidad terminal, generalmente, proporciona herramientas que se usarán para resolver problemas de inferencia estadística. En cualquier caso, conviene remarcar que en un análisis inferencial también hay que hacer un estudio descriptivo de la muestra.

A lo largo del libro se introducen los conceptos desde una perspectiva unidimensional para, a continuación, generalizarlos al caso de más de una dimensión. La importancia del estudio multidimensional se debe al hecho de considerar la existencia de posibles interacciones entre las distintas variables objeto de estudio. Esto no significa que se traten técnicas multivariantes, que ciertamente quedan fuera de los objetivos que se han marcado con la realización de este manual; en cambio, sí se hace una primera, aunque tímida, incursión en el campo de la modelización, con la inclusión de un capítulo sobre ajuste y regresión.

Cada uno de los capítulos del libro comienza con una presentación del tipo de problema que se va a abordar, continúa con la exposición de los contenidos ilustrados con distintos ejemplos y, en ocasiones, acompañados por algún ejercicio que pretende profundizar en alguna cuestión de interés, para finalizar con ejercicios resueltos, que intenta globalizar los aspectos más relevantes del capítulo y una colección de ejercicios propuestos. En la parte final del libro se incluyen dos apéndices, uno sobre Combinatoria y otro conteniendo las tablas de algunas de las distribuciones.

Para aquellos estudiantes que necesiten en sus curriculas conocimientos de estadística inferencial, los autores ofrecen en el manual titulado *Inferencia Estadística. Teoría y Problemas* (Servicio de Publicaciones de la Universidad de Cádiz, 2002) los contenidos correspondientes a esta parte fundamental de la Estadística. En dicho manual, se abordan cuestiones relativas a Teoría de muestras, Estimación puntual, Estimación por intervalos, Contraste de hipótesis, incluyendo las alternativas No paramétricas y Análisis de la varianza. El manual, al igual que el que nos ocupa, está salpicado de ejemplos y ejercicios resueltos y, sin pretender ser un libro teórico, el lector que así lo desee puede encontrar respuestas fundamentadas a las cuestiones conceptuales que se van planteando.

LOS AUTORES.

2. History (Histórico)

La primera versión de este libro aparece durante el curso 90/91, en la antigua Escuela de Empresariales de Cádiz, hoy Facultad de Ciencias Económicas y Empresariales, en modo de apuntes elaborados por F. Fernández Palacín y A. Sánchez Navas. Esta primera entrega incluía contenidos genéricos de una introducción a la Estadística, así como una serie de temas relacionados con aspectos económicos, en concreto Números índices, Series temporales y Medidas de desigualdad.

En el año 1996, y en aras de ampliar su uso como herramienta de trabajo en otras titulaciones donde el departamento de Estadística e Investigación Operativa tiene docencia, se ajustan los contenidos del manual y se reparten sus contenidos en dos librillos editados por la Copistería San Rafael: “Estadística Descriptiva„ “Variable Aleatoria y Probabilidad”, en cuya elaboración además de los citados autores interviene M. A. López Sánchez.

Durante el año 2000 estos manuales se someten a una amplia revisión, se amplían los contenidos y se refunden en un solo libro, publicado por el Servicio de Publicaciones de la Universidad de Cádiz. Sus autores son F. Fernández Palacín, M.A. López Sánchez, M. Muñoz Márquez, A.M. Rodríguez Chía, A. Sánchez Navas y C. Valero Franco. Posteriormente, durante el año 2003, se actualizan los contenidos incorporándose entre los autores I. Espejo Miranda.

En el año 2005, tras una nueva revisión, los autores deciden publicar el libro bajo licencia *GNU Free Documentation License.*

Una versión electrónica de este documento está en:
`http://www.uca.es/grupos-inv/FQM270.`

3. Licencia de Documentación Libre de GNU

This is an unofficial translation of the GNU Free Documentation License (Version 1.2, Noviembre 2002) into Spanish. It was not published by the Free Software Foundation, and does not legally state the distribution terms for documentation that uses the GNU FDL – only the original English text of the GNU FDL does that. However, we hope that this translation will help Spanish speakers understand the GNU FDL better.

Ésta es una traducción no oficial de la GNU Free Document License (Versión 1.2, Noviembre 2002) a Español (Castellano). No ha sido publicada por la Free Software Foundation y no establece legalmente los términos de distribución para trabajos que usen la GFDL (sólo el texto de la versión original en Inglés de la GFDL lo hace). Sin embargo, esperamos que esta traducción ayude los hispanohablantes a entender mejor la GFDL. La versión original de la GFDL esta disponible en la Free Software Foundation. `http://www.gnu.org/copyleft/fdl.html` Esta traducción está basada en una de la versión 1.1 de Igor Támara y Pablo Reyes. Sin embargo la responsabilidad de su interpretación es de Joaquín Seoane.

Copyright (C) 2000, 2001, 2002 Free Software Foundation, Inc. 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA. Se permite la copia y distribución de copias literales de este documento de licencia, pero no se permiten cambios[1].

Preámbulo

El propósito de esta Licencia es permitir que un manual, libro de texto, u otro documento escrito sea **"libre"** en el sentido de libertad: asegurar a todo el mundo la libertad efectiva de copiarlo y redistribuirlo, con o sin modificaciones, de manera comercial o no. En segundo término, esta Licencia proporciona al autor y al editor[2] una manera de obtener reconocimiento por su trabajo, sin que se le considere responsable de las modificaciones realizadas por otros.

Esta Licencia es de tipo **"copyleft"**, lo que significa que los trabajos derivados del documento deben a su vez ser libres en el mismo sentido. Complementa la Licencia Pública General de GNU, que es una licencia tipo copyleft diseñada para el software libre.

[1]Ésta es la traducción del Copyright de la Licencia, no es el Copyright de esta traducción no autorizada.

[2]La licencia original dice **"publisher"**, que es, estrictamente, quien publica, diferente de editor, que es más bien quien prepara un texto para publicar. En castellano editor se usa para ambas cosas.

Hemos diseñado esta Licencia para usarla en manuales de software libre, ya que el software libre necesita documentación libre: un programa libre debe venir con manuales que ofrezcan la mismas libertades que el software. Pero esta licencia no se limita a manuales de software; puede usarse para cualquier texto, sin tener en cuenta su temática o si se publica como libro impreso o no.

Recomendamos esta licencia principalmente para trabajos cuyo fin sea instructivo o de referencia.

1. Aplicabilidad y definiciones

Esta Licencia se aplica a cualquier manual u otro trabajo, en cualquier soporte, que contenga una nota del propietario de los derechos de autor que indique que puede ser distribuido bajo los términos de esta Licencia. Tal nota garantiza en cualquier lugar del mundo, sin pago de derechos y sin límite de tiempo, el uso de dicho trabajo según las condiciones aquí estipuladas. En adelante la palabra **"Documento"** se referirá a cualquiera de dichos manuales o trabajos. Cualquier persona es un licenciatario y será referido como **"Usted"**. Usted acepta la licencia si copia. modifica o distribuye el trabajo de cualquier modo que requiera permiso según la ley de propiedad intelectual.

Una **"Versión Modificada"** del Documento significa cualquier trabajo que contenga el Documento o una porción del mismo, ya sea una copia literal o con modificaciones y/o traducciones a otro idioma.

Una **"Sección Secundaria"** es un apéndice con título o una sección preliminar del Documento que trata exclusivamente de la relación entre los autores o editores y el tema general del Documento (o temas relacionados) pero que no contiene nada que entre directamente en dicho tema general (por ejemplo, si el Documento es en parte un texto de matemáticas, una Sección Secundaria puede no explicar nada de matemáticas). La relación puede ser una conexión histórica con el tema o temas relacionados, o una opinión legal, comercial, filosófica, ética o política acerca de ellos.

Las **"Secciones Invariantes"** son ciertas Secciones Secundarias cuyos títulos son designados como Secciones Invariantes en la nota que indica que el documento es liberado bajo esta Licencia. Si una sección no entra en la definición de Secundaria, no puede designarse como Invariante. El documento puede no tener Secciones Invariantes. Si el Documento no identifica las Secciones Invariantes, es que no las tiene.

Los **"Textos de Cubierta"** son ciertos pasajes cortos de texto que se listan como Textos de Cubierta Delantera o Textos de Cubierta Trasera en la nota que indica que el documento es liberado bajo esta Licencia. Un Texto de

Cubierta Delantera puede tener como mucho 5 palabras, y uno de Cubierta Trasera puede tener hasta 25 palabras.

Una copia **"Transparente"** del Documento, significa una copia para lectura en máquina, representada en un formato cuya especificación está disponible al público en general, apto para que los contenidos puedan ser vistos y editados directamente con editores de texto genéricos o (para imágenes compuestas por puntos) con programas genéricos de manipulación de imágenes o (para dibujos) con algún editor de dibujos ampliamente disponible, y que sea adecuado como entrada para formateadores de texto o para su traducción automática a formatos adecuados para formateadores de texto. Una copia hecha en un formato definido como Transparente, pero cuyo marcaje o ausencia de él haya sido diseñado para impedir o dificultar modificaciones posteriores por parte de los lectores no es Transparente. Un formato de imagen no es Transparente si se usa para una cantidad de texto sustancial. Una copia que no es **"Transparente"** se denomina **"Opaca"**.

Como ejemplos de formatos adecuados para copias Transparentes están ASCII puro sin marcaje, formato de entrada de Texinfo, formato de entrada de LaTeX, SGML o XML usando una DTD disponible públicamente, y HTML, PostScript o PDF simples, que sigan los estándares y diseñados para que los modifiquen personas. Ejemplos de formatos de imagen transparentes son PNG, XCF y JPG. Los formatos Opacos incluyen formatos propietarios que pueden ser leídos y editados únicamente en procesadores de palabras propietarios, SGML o XML para los cuáles las DTD y/o herramientas de procesamiento no estén ampliamente disponibles, y HTML, PostScript o PDF generados por algunos procesadores de palabras sólo como salida.

La **"Portada"** significa, en un libro impreso, la página de título, más las páginas siguientes que sean necesarias para mantener legiblemente el material que esta Licencia requiere en la portada. Para trabajos en formatos que no tienen página de portada como tal, **"Portada"** significa el texto cercano a la aparición más prominente del título del trabajo, precediendo el comienzo del cuerpo del texto.

Una sección **"Titulada XYZ"** significa una parte del Documento cuyo título es precisamente XYZ o contiene XYZ entre paréntesis, a continuación de texto que traduce XYZ a otro idioma (aquí XYZ se refiere a nombres de sección específicos mencionados más abajo, como **"Agradecimientos"**, **"Dedicatorias"**, **"Aprobaciones"** o **"Historia"**. **"Conservar el Título"** de tal sección cuando se modifica el Documento significa que permanece una sección **"Titulada XYZ"** según esta definición[3].

[3]En sentido estricto esta licencia parece exigir que los títulos sean exactamente **"Acknowledgements"**, **"Dedications"**, **"Endorsements"** e **"History"**, en

El Documento puede incluir Limitaciones de Garantía cercanas a la nota donde se declara que al Documento se le aplica esta Licencia. Se considera que estas Limitaciones de Garantía están incluidas, por referencia, en la Licencia, pero sólo en cuanto a limitaciones de garantía: cualquier otra implicación que estas Limitaciones de Garantía puedan tener es nula y no tiene efecto en el significado de esta Licencia.

2. Copia literal

Usted puede copiar y distribuir el Documento en cualquier soporte, sea en forma comercial o no, siempre y cuando esta Licencia, las notas de copyright y la nota que indica que esta Licencia se aplica al Documento se reproduzcan en todas las copias y que usted no añada ninguna otra condición a las expuestas en esta Licencia. Usted no puede usar medidas técnicas para obstruir o controlar la lectura o copia posterior de las copias que usted haga o distribuya. Sin embargo, usted puede aceptar compensación a cambio de las copias. Si distribuye un número suficientemente grande de copias también deberá seguir las condiciones de la sección 3.

Usted también puede prestar copias, bajo las mismas condiciones establecidas anteriormente, y puede exhibir copias públicamente.

3. Copiado en cantidad

Si publica copias impresas del Documento (o copias en soportes que tengan normalmente cubiertas impresas) que sobrepasen las 100, y la nota de licencia del Documento exige Textos de Cubierta, debe incluir las copias con cubiertas que lleven en forma clara y legible todos esos Textos de Cubierta: Textos de Cubierta Delantera en la cubierta delantera y Textos de Cubierta Trasera en la cubierta trasera. Ambas cubiertas deben identificarlo a Usted clara y legiblemente como editor de tales copias. La cubierta debe mostrar el título completo con todas las palabras igualmente prominentes y visibles. Además puede añadir otro material en las cubiertas. Las copias con cambios limitados a las cubiertas, siempre que conserven el título del Documento y satisfagan estas condiciones, pueden considerarse como copias literales.

Si los textos requeridos para la cubierta son muy voluminosos para que ajusten legiblemente, debe colocar los primeros (tantos como sea razonable colocar) en la verdadera cubierta y situar el resto en páginas adyacentes.

Si Usted publica o distribuye copias Opacas del Documento cuya cantidad exceda las 100, debe incluir una copia Transparente, que pueda ser leída

---

inglés.

por una máquina, con cada copia Opaca, o bien mostrar, en cada copia Opaca, una dirección de red donde cualquier usuario de la misma tenga acceso por medio de protocolos públicos y estandarizados a una copia Transparente del Documento completa, sin material adicional. Si usted hace uso de la última opción, deberá tomar las medidas necesarias, cuando comience la distribución de las copias Opacas en cantidad, para asegurar que esta copia Transparente permanecerá accesible en el sitio establecido por lo menos un año después de la última vez que distribuya una copia Opaca de esa edición al público (directamente o a través de sus agentes o distribuidores).

Se solicita, aunque no es requisito, que se ponga en contacto con los autores del Documento antes de redistribuir gran número de copias, para darles la oportunidad de que le proporcionen una versión actualizada del Documento.

4. Modificaciones

Puede copiar y distribuir una Versión Modificada del Documento bajo las condiciones de las secciones 2 y 3 anteriores, siempre que usted libere la Versión Modificada bajo esta misma Licencia, con la Versión Modificada haciendo el rol del Documento, por lo tanto dando licencia de distribución y modificación de la Versión Modificada a quienquiera posea una copia de la misma. Además, debe hacer lo siguiente en la Versión Modificada:

A. Usar en la Portada (y en las cubiertas, si hay alguna) un título distinto al del Documento y de sus versiones anteriores (que deberán, si hay alguna, estar listadas en la sección de Historia del Documento). Puede usar el mismo título de versiones anteriores al original siempre y cuando quien las publicó originalmente otorgue permiso.

B. Listar en la Portada, como autores, una o más personas o entidades responsables de la autoría de las modificaciones de la Versión Modificada, junto con por lo menos cinco de los autores principales del Documento (todos sus autores principales, si hay menos de cinco), a menos que le eximan de tal requisito.

C. Mostrar en la Portada como editor el nombre del editor de la Versión Modificada.

D. Conservar todas las notas de copyright del Documento.

E. Añadir una nota de copyright apropiada a sus modificaciones, adyacente a las otras notas de copyright.

F. Incluir, inmediatamente después de las notas de copyright, una nota de licencia dando el permiso para usar la Versión Modificada bajo los términos de esta Licencia, como se muestra en la Adenda al final de este documento.

G. Conservar en esa nota de licencia el listado completo de las Secciones Invariantes y de los Textos de Cubierta que sean requeridos en la nota de Licencia del Documento original.

H. Incluir una copia sin modificación de esta Licencia.

I. Conservar la sección Titulada **"Historia"**, conservar su Título y añadirle un elemento que declare al menos el título, el año, los nuevos autores y el editor de la Versión Modificada, tal como figuran en la Portada. Si no hay una sección Titulada **"Historia"** en el Documento, crear una estableciendo el título, el año, los autores y el editor del Documento, tal como figuran en su Portada, añadiendo además un elemento describiendo la Versión Modificada, como se estableció en la oración anterior.

J. Conservar la dirección en red, si la hay, dada en el Documento para el acceso público a una copia Transparente del mismo, así como las otras direcciones de red dadas en el Documento para versiones anteriores en las que estuviese basado. Pueden ubicarse en la sección **"Historia"**. Se puede omitir la ubicación en red de un trabajo que haya sido publicado por lo menos cuatro años antes que el Documento mismo, o si el editor original de dicha versión da permiso.

K. En cualquier sección Titulada **"Agradecimientos"** o **"Dedicatorias"**, Conservar el Título de la sección y conservar en ella toda la sustancia y el tono de los agradecimientos y/o dedicatorias incluidas por cada contribuyente.

L. Conservar todas las Secciones Invariantes del Documento, sin alterar su texto ni sus títulos. Números de sección o el equivalente no son considerados parte de los títulos de la sección.

M. Borrar cualquier sección titulada **"Aprobaciones"**. Tales secciones no pueden estar incluidas en las Versiones Modificadas.

N. No cambiar el título de ninguna sección existente a **"Aprobaciones"** ni a uno que entre en conflicto con el de alguna Sección Invariante.

O. Conservar todas las Limitaciones de Garantía.

Si la Versión Modificada incluye secciones o apéndices nuevos que califiquen como Secciones Secundarias y contienen material no copiado del Documento, puede opcionalmente designar algunas o todas esas secciones como invariantes. Para hacerlo, añada sus títulos a la lista de Secciones Invariantes en la nota de licencia de la Versión Modificada. Tales títulos deben ser distintos de cualquier otro título de sección.

Puede añadir una sección titulada **“Aprobaciones”**, siempre que contenga únicamente aprobaciones de su Versión Modificada por otras fuentes –por ejemplo, observaciones de peritos o que el texto ha sido aprobado por una organización como la definición oficial de un estándar.

Puede añadir un pasaje de hasta cinco palabras como Texto de Cubierta Delantera y un pasaje de hasta 25 palabras como Texto de Cubierta Trasera en la Versión Modificada. Una entidad solo puede añadir (o hacer que se añada) un pasaje al Texto de Cubierta Delantera y uno al de Cubierta Trasera. Si el Documento ya incluye textos de cubiertas añadidos previamente por usted o por la misma entidad que usted representa, usted no puede añadir otro; pero puede reemplazar el anterior, con permiso explícito del editor que agregó el texto anterior.

Con esta Licencia ni los autores ni los editores del Documento dan permiso para usar sus nombres para publicidad ni para asegurar o implicar aprobación de cualquier Versión Modificada.

5. Combinación de documentos

Usted puede combinar el Documento con otros documentos liberados bajo esta Licencia, bajo los términos definidos en la sección 4 anterior para versiones modificadas, siempre que incluya en la combinación todas las Secciones Invariantes de todos los documentos originales, sin modificar, listadas todas como Secciones Invariantes del trabajo combinado en su nota de licencia. Así mismo debe incluir la Limitación de Garantía.

El trabajo combinado necesita contener solamente una copia de esta Licencia, y puede reemplazar varias Secciones Invariantes idénticas por una sola copia. Si hay varias Secciones Invariantes con el mismo nombre pero con contenidos diferentes, haga el título de cada una de estas secciones único añadiéndole al final del mismo, entre paréntesis, el nombre del autor o editor original de esa sección, si es conocido, o si no, un número único. Haga el mismo ajuste a los títulos de sección en la lista de Secciones Invariantes de la nota de licencia del trabajo combinado.

En la combinación, debe combinar cualquier sección Titulada **“Historia”** de los documentos originales, formando una sección Titulada **“Historia”**; de la misma forma combine cualquier sección Titulada **“Agradecimientos”**, y cualquier sección Titulada **“Dedicatorias”**. Debe borrar todas las secciones tituladas **“Aprobaciones”**.

6. Colecciones de documentos

Puede hacer una colección que conste del Documento y de otros documentos liberados bajo esta Licencia, y reemplazar las copias individuales de esta Licencia en todos los documentos por una sola copia que esté incluida en la colección, siempre que siga las reglas de esta Licencia para cada copia literal de cada uno de los documentos en cualquiera de los demás aspectos.

Puede extraer un solo documento de una de tales colecciones y distribuirlo individualmente bajo esta Licencia, siempre que inserte una copia de esta Licencia en el documento extraído, y siga esta Licencia en todos los demás aspectos relativos a la copia literal de dicho documento.

7. Agregación con trabajos independientes

Una recopilación que conste del Documento o sus derivados y de otros documentos o trabajos separados e independientes, en cualquier soporte de almacenamiento o distribución, se denomina un **"agregado"** si el copyright resultante de la compilación no se usa para limitar los derechos de los usuarios de la misma más allá de lo que los de los trabajos individuales permiten. Cuando el Documento se incluye en un agregado, esta Licencia no se aplica a otros trabajos del agregado que no sean en sí mismos derivados del Documento.

Si el requisito de la sección 3 sobre el Texto de Cubierta es aplicable a estas copias del Documento y el Documento es menor que la mitad del agregado entero, los Textos de Cubierta del Documento pueden colocarse en cubiertas que enmarquen solamente el Documento dentro del agregado, o el equivalente electrónico de las cubiertas si el documento está en forma electrónica. En caso contrario deben aparecer en cubiertas impresas enmarcando todo el agregado.

8. Traducción

La Traducción es considerada como un tipo de modificación, por lo que usted puede distribuir traducciones del Documento bajo los términos de la sección 4. El reemplazo de las Secciones Invariantes con traducciones requiere permiso especial de los dueños de derecho de autor, pero usted puede añadir traducciones de algunas o todas las Secciones Invariantes a las versiones originales de las mismas. Puede incluir una traducción de esta Licencia, de todas las notas de licencia del documento, así como de las Limitaciones de Garantía, siempre que incluya también la versión en Inglés de esta Licencia y las versiones originales de las notas de licencia y Limitaciones de Garantía. En caso de desacuerdo entre la traducción y la versión original en Inglés de esta Licencia, la nota de licencia o la limitación de garantía, la versión original en Inglés prevalecerá.

Si una sección del Documento está Titulada **"Agradecimientos"**, **"Dedicatorias"** o **"Historia"** el requisito (sección 4) de Conservar su Título (Sección 1) requerirá, típicamente, cambiar su título.

9. Terminación

Usted no puede copiar, modificar, sublicenciar o distribuir el Documento salvo por lo permitido expresamente por esta Licencia. Cualquier otro intento de copia, modificación, sublicenciamiento o distribución del Documento es nulo, y dará por terminados automáticamente sus derechos bajo esa Licencia. Sin embargo, los terceros que hayan recibido copias, o derechos, de usted bajo esta Licencia no verán terminadas sus licencias, siempre que permanezcan en total conformidad con ella.

10. Revisiones futuras de esta licencia

De vez en cuando la Free Software Foundation puede publicar versiones nuevas y revisadas de la Licencia de Documentación Libre GNU. Tales versiones nuevas serán similares en espíritu a la presente versión, pero pueden diferir en detalles para solucionar nuevos problemas o intereses. Vea `http://www.gnu.org/copyleft/`.

Cada versión de la Licencia tiene un número de versión que la distingue. Si el Documento especifica que se aplica una versión numerada en particular de esta licencia o **"cualquier versión posterior"**, usted tiene la opción de seguir los términos y condiciones de la versión especificada o cualquiera posterior que haya sido publicada (no como borrador) por la Free Software Foundation. Si el Documento no especifica un número de versión de esta Licencia, puede escoger cualquier versión que haya sido publicada (no como borrador) por la Free Software Foundation.

ADENDA: Cómo usar esta Licencia en sus documentos

Para usar esta licencia en un documento que usted haya escrito, incluya una copia de la Licencia en el documento y ponga el siguiente copyright y nota de licencia justo después de la página de título:

> Copyright (c) AÑO SU NOMBRE. Se concede permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre de GNU, Versión 1.2 o cualquier otra versión posterior publicada por la Free Software Foundation; sin Secciones Invariantes ni Textos de Cubierta Delantera ni Textos de Cubierta Trasera. Una copia de la licencia está incluida en la sección titulada GNU Free Documentation License.

Si tiene Secciones Invariantes, Textos de Cubierta Delantera y Textos de Cubierta Trasera, reemplace la frase **“sin ... Trasera”** por esto:

> siendo las Secciones Invariantes LISTE SUS TÍTULOS, siendo los Textos de Cubierta Delantera LISTAR, y siendo sus Textos de Cubierta Trasera LISTAR.

Si tiene Secciones Invariantes sin Textos de Cubierta o cualquier otra combinación de los tres, mezcle ambas alternativas para adaptarse a la situación.

Si su documento contiene ejemplos de código de programa no triviales, recomendamos liberar estos ejemplos en paralelo bajo la licencia de software libre que usted elija, como la Licencia Pública General de GNU (**“GNU General Public License”**), para permitir su uso en software libre.

4. GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free"in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The **"Document"**, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as **"you"**. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A **"Modified Version"** of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A **"Secondary Section"** is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The **"Invariant Sections"** are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The **"Cover Texts"** are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A **"Transparent"** copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent"is called **"Opaque"**.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The **"Title Page"** means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page"means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section **"Entitled XYZ"** means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as **"Acknowledgements"**, **"Dedications"**, **"Endorsements"**, or **"History"**.) To **"Preserve the Title"** of such a section when you modify the Document means that it remains a section "Entitled XYZ.[a]ccording to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History"in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History"section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements.ºr "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled “Endorsements.ºr to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History”in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate”if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See `http://www.gnu.org/copyleft/`.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version.[a]pplies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright ©YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Parte A

Estadística Descriptiva

Introducción a la estadística descriptiva

La primera parte de este libro está dedicada a la estadística descriptiva. Atendiendo a lo que tradicionalmente se ha entendido por descriptiva se estaría hablando de un conjunto de herramientas, formado por coeficientes y técnicas, que tratan de resumir la información contenida en un conjunto de datos. Sin embargo, la estadística descriptiva es mucho más que eso, en realidad es una parte fundamental de cualquier análisis estadístico complejo, en la que se empiezan a tomar decisiones que afectarán al conjunto de la investigación. Los coeficientes descriptivos darán información sobre la estructura de la población que se estudia, indicando, por ejemplo, si ésta es simétrica, si realmente se trata de una única población o hay una superposición de poblaciones, también pueden detectarse valores extraordinariamente raros, etc.

Desde otra óptica, la mayoría de los coeficientes descriptivos tendrán su homólogo inferencial o poblacional, que necesariamente deberán ser estudiados a la luz de aquellos. Haciendo una pequeña abstracción muchos de los coeficientes descriptivos, los más importantes, se convierten en poblacionales al sustituir frecuencias por probabilidades.

En resumen, el análisis descriptivo es una parte inseparable de cualquier análisis estadístico, que puede tener su continuidad con un análisis inferencial cuando los datos que se manejan se corresponden con una muestra probabilística extraída de una población.

Esta primera parte del libro está compuesta por tres capítulos, el primero de ellos aborda el problema unidimensional. Se trata de identificar la información que se va a analizar, bien sean variables cuantitativas o de clase, procediéndose a organizarla en distribuciones de frecuencias. Se indica que la primera toma de contacto con las peculiaridades de una distribución se obtiene a través de sus representaciones gráficas y se da al menos una representación para cada uno de los tipos de datos que se manejan. Se calculan todos los coeficientes tradicionales: medidas de centralización, de posición, de dispersión y de forma; se obtienen los momentos respecto al origen y respecto a la media, indicándose que generalizan la mayoría de las medidas anteriores. Se introduce la desigualdad de Tchebychev, poniéndose de manifiesto la relación existente entre la varianza y la media aritmética. Se estudian las transformaciones de variables, haciendo ver que el objetivo es conseguir distribuciones más regulares, que sean comparables, más simétricas; entre todas las transformaciones se dedica especial atención a la normalización o tipificación. Por último, se hace una breve incursión en el análisis exploratorio de datos, recurriendo a representaciones, como los diagramas de cajas, que resaltan las regularidades y las especificidades del conjunto de datos, entre las que cabe destacar la presencia de observaciones candidatas a ser valores extraños o anómalos.

El capítulo segundo supone una generalización al caso de que conjuntamente se tenga más de una variable, viéndose con detenimiento el caso bivariable y destacando el hecho de la posible existencia de relaciones entre dichas variables. La existencia de dependencias merece una especial atención por las consecuencias que de ella se derivan en muchas técnicas estadísticas. Se introducen coeficientes que expresarán el grado de relación entre las variables, distinguiendo los casos en que éstas sean continuas, ordenadas o de clase; lo que conduce a definir medidas de correlación, concordancia y contingencia o asociación.

En el último capítulo de esta parte se aborda el problema del ajuste y la regresión en el plano, lo que supone un primer acercamiento a la modelización estadística. El desarrollo del tema se hace planteando un modelo lineal, empleándose para la estimación de los parámetros el método de los mínimos cuadrados. El análisis de la bondad del ajuste se realiza a través del coeficiente de determinación. El método de la regresión a la media permite calibrar la calidad de los posibles ajustes a realizar. También se analizan algunas extensiones a los casos de modelos linealizables y polinomiales.

Capítulo 1

Síntesis de la información

1. Reseña histórica

1.1. Introducción

Al acercarse a una ciencia es interesante indagar en sus raíces históricas para obtener una visión de su naturaleza y de sus objetivos como disciplina científica. El estudio de dichas raíces permitirá entender el grado de desarrollo actual, la relación entre sus distintas partes, comprender su terminología -dado que el nombre de un coeficiente, de una técnica, ... suele estar asociado a su origen histórico-, e incluso prever en que dirección evolucionará. En el caso de la Estadística este estudio retrospectivo es particularmente rico en enseñanzas.

A lo largo de los tiempos han sido muchas las concepciones que se le ha dado a la ciencia Estadística, desde la que la ha entendido como un conjunto de técnicas aplicables a una serie de datos, hasta la que la ha concebido como un proceso de extrapolación de conclusiones de la muestra a la población. Actualmente, no puede entenderse la Estadística como un conjunto de conceptos y expresiones matemáticas abstractas, olvidando las motivaciones históricas sobre las que se construyó y su actual papel esencial en cualquier tipo de investigación empírica, tal y como destaca Kruskal en su Enciclopedia Internacional de Estadística.

1.2. Orígenes de la estadística descriptiva

Los orígenes históricos de la Estadística (descriptiva) hay que buscarlos en los procesos de recogida de datos, censos y registros sistemáticos, asumiendo un papel asimilable a una aritmética estatal para asistir al gobernante, que necesitaba conocer la riqueza y el número de sus súbditos con fines tributarios y políticos.

Los primeros registros de riqueza y población que se conocen se deben a los egipcios. Ramsés II en el 1400 a.C. realizó el primer censo conocido de las tierras de Egipto, no siendo éste, se supone, ni el primero ni el último que se hiciera en las tierras bañadas por el Nilo.

Posteriormente, desde el siglo III a.C., en las civilizaciones china y romana se llevan a cabo censos e inventarios de posesiones, que pueden considerarse precedentes institucionalizados de la recogida de datos demográficos y económicos de los Estados Modernos.

Hay que realizar una mención especial del período helénico, en el que las escuelas matemáticas se suceden. Centros como el de Quios, donde estudió Hipócrates (Hipócrates de Quios) el matemático, considerado como el inventor del método matemático y escuelas como las de Cirene, Megara y al final Atenas, donde se reunen los matemáticos, unos alrededor de Protágoras y otros en torno a Sócrates.

En la Edad Media se vuelve a la utilización de la Aritmética para la recogida de datos, existiendo menos interés por la elucubración matemática abstracta. Es en este período de tiempo cuando Carlomagno ordenó en su "Capitulare de villis" la creación de un registro de todos sus dominios y bienes privados.

En el siglo XVII se producen avances sustanciales, y así, en las universidades alemanas se imparten enseñanzas de "Aritmética Política", término con el que se designa la descripción numérica de hechos de interés para la Administración Pública. Destacados autores de Aritmética Política fueron los ingleses Graunt (1620-1674) y Petty (1623-1687).

Con métodos de estimación en los que cabía la conjetura, la experimentación y la deducción, Graunt llega a estimar tasas de mortalidad para la población londinense, analizando además la verosimilitud de la información de que disponía. Por su parte, Petty, cuyas aportaciones estadísticas fueron menos relevantes, tiene el mérito -en opinión de Gutiérrez Cabria- de proponer la creación de un departamento de estadística, en el que se reuniese información no sólo de carácter demográfico, sino también sobre recaudación de impuestos, educación y comercio. Surge en esta época la conciencia de la necesidad de disponer de información, conciencia que va tomando cuerpo a partir de la segunda mitad del siglo XVII en la mayor parte de las potencias europeas y americanas, considerándose como primera oficina de estadística la instituida en Suecia en 1756.

En España, el interés por las investigaciones estatales nació con la preocupación de los Reyes Católicos por mejorar el estado de las "Cosas Públicas", estableciéndose el primer censo del que se tiene referencia en 1482, elaborado por Alonso de Quintanilla. Durante el siglo XVIII se elaboraron censos como el de Ensenada en 1749 y el de Floridablanca en 1787, con una metodología con visos de modernidad. Los actuales censos de periodicidad decenal empezaron a elaborarse en 1860 a cargo de la Junta General de Estadística.

2. La organización de la información

Los datos constituyen la materia prima de la Estadística, pudiéndose establecer distintas clasificaciones en función de la forma en que éstos vengan dados. Se obtienen datos al realizar cualquier tipo de prueba, experimento, valoración, medición, observación,...

Este capítulo tiene por finalidad la descripción de un conjunto de datos, sin considerar que éstos puedan pertenecer a un colectivo más amplio y, por supuesto, sin la intención de proyectar los resultados que se obtengan al colectivo global; objeto esto último de lo que se conoce como Inferencia Estadística.

2.1. Variable y atributo

Se realiza una primera clasificación del tipo de datos en función de que las observaciones resultantes del experimento sean de tipo cualitativo o cuantitativo, en el primero de los casos se tiene un *atributo* y en el segundo una *variable*. Para hacer referencia genéricamente a una variable o a un atributo se utilizará el término *carácter*.

Ejemplo 1.1 *Como ejemplos de atributos pueden considerarse el color del pelo de un colectivo de personas, su raza o el idioma que hablan y como variables su estatura, peso o edad.*

Para poder operar con un atributo es necesario asignar a cada una de sus clases un valor numérico, con lo que se transforma en una variable, esta asignación se hará de forma que los resultados que se obtengan al final del estudio sean fácilmente interpretables.

Ejercicio 1.1 *Clasifique los siguientes datos según sean variables o atributos:*

a) *El color de ojos de un grupo de 20 personas.*

b) *La nacionalidad de un conjunto de individuos.*

c) *Las dioptrías de un grupo de personas miopes.*

d) *Los matices de color de un cuadro impresionista.*

e) *Las dianas que consigue un arquero sobre un total de 100 intentos.*

2.2. Variables discretas y continuas

Dentro del conjunto de las variables se distingue entre *discretas* y *continuas*. Se dice que una variable es discreta cuando entre dos valores consecutivos no toma valores intermedios y que es continua cuando puede tomar cualquier valor dentro de un intervalo.

Ejemplo 1.2 *La estatura de un grupo de personas sería una variable continua, mientras que el número de cabellos que tienen en la cabeza sería una variable discreta.*

En la práctica todas las variables son discretas debido a la limitación de los aparatos de medida, y así, en el ejemplo de las estaturas, quizás se podría detectar una diferencia de una cienmilésima de metro, o a lo más, de una millonésima, pero dados dos individuos que se diferencien en una millonésima no puede detectarse otro que tenga una estatura intermedia. De todas formas, en general se trata a las variables "teóricamente" continuas como tales, por razones que se pondrán de manifiesto más adelante.

Ejercicio 1.2 *Indique cuáles de las siguientes variables son continuas y cuáles discretas:*

a) *El número de moléculas de agua de un pantano.*

b) *La edad exacta de un grupo de 50 niños.*

c) *La distancia por carretera entre las capitales de provincia peninsulares españolas.*

d) *La distancia al centro de la diana de las flechas lanzadas por un arquero.*

e) *El número de docenas de huevos que se recolecta al día en una granja de gallinas.*

Si la ocasión lo requiere se tiene la posibilidad de transformar una variable discreta en continua o viceversa. Para transformar una variable discreta en continua, una vez ordenados los valores, se asigna a cada uno de ellos un intervalo que tenga por extremos el punto medio respecto al valor anterior y el punto medio respecto al valor siguiente. Esta operación tiene interés, por ejemplo, en la aproximación de distribuciones discretas a continuas, como se tendrá la oportunidad de comprobar en la segunda parte de este manual.

Para transformar una variable continua en discreta basta con hacer corresponder a cada uno de los intervalos su punto medio o *marca de clase*.

Ejercicio 1.3 *Transforme la variable continua que toma valores en los intervalos* $(0, 2]$, $(2, 3]$, $(3, 6]$, $(6, 10]$, $(10, 15]$ *en variable discreta.*

2.3. Clasificación de las series estadísticas

Además de por su naturaleza, se pueden realizar distintas clasificaciones del conjunto de los datos o serie estadística.

1. **Por su número**
   a) Finitas. Las que tienen un número finito de elementos.
   b) Infinitas. Cuando tienen infinitos elementos.
2. **Por su obtención**
   a) Objetivas. Obtenidas con métodos exactos de medición.
   b) Subjetivas. Obtenidas mediante apreciaciones personales.
3. **Por su dimensión**
   a) Unidimensionales: $x_1, x_2, x_3, \cdots, x_n$.
   b) Bidimensionales: $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$.
   c) n-dimensionales: $(x_1^1, x_2^1, \cdots, x_n^1), \cdots, (x_1^r, x_2^r, \cdots, x_n^r)$.
4. **Por su dependencia temporal**
   a) Temporales. Los valores se toman en instantes o períodos de tiempo.
   b) Atemporales. No dependen de ningún soporte temporal.

2.4. Distribución de datos

La organización de los datos constituye la primera etapa de su tratamiento, pues, facilita los cálculos posteriores y evita posibles confusiones. Realmente, la organización de la información tiene una raiz histórica y aunque actualmente con el desarrollo de los medios informáticos deja de tener importancia desde un punto de vista aplicado, desde la perspectiva de la enseñanza de la Estadística tiene un gran valor conceptual. La organización va a depender del número de observaciones distintas que se tengan y de las veces que se repitan cada una de ellas. En base a lo anterior se pueden estructurar los datos de tres maneras distintas:

1. *Tipo I:* Cuando se tiene un número pequeño de observaciones casi todas distintas, éstas se darán por extensión.

   Ejemplo 1.3 *En la serie: 2, 3, 5, 7, 7, 8, 11, 14, 16, 19, el 7 se repite dos veces y el resto de los valores está presente una vez.*

2. *Tipo II:* Cuando se tiene un gran número de observaciones pero muy pocas distintas, se organizan en una *tabla de frecuencias*, es decir, cada uno de los valores acompañado de la frecuencia con la que se presenta.

   Ejemplo 1.4 *La tabla*

Valor	*Frecuencia*
2	*4*
4	*4*
5	*3*
6	*2*
7	*3*
8	*3*
9	*1*

   indica que el valor 2 se repite 4 veces, el valor 4 se repite 4 veces, etc.. . .

3. *Tipo III:* En el caso de que haya muchas observaciones, la mayoría de ellas distintas, pueden disponerse agrupándolas en intervalos e indicando el número de observaciones que caen dentro de cada intervalo.

Ejemplo 1.5 *La tabla*

Intervalo	*Frecuencia*
(2,3]	*4*
(3,7]	*6*
(7,12]	*12*
(12,21]	*8*
(21,25]	*6*
(25,30]	*4*
(30,50]	*3*

nos dice que en el intervalo $(2,3]$ *hay 4 observaciones, que en el* $(3,7]$ *hay 6, etc...*

En cualquiera de los tres casos o tipos se tiene una *distribución de frecuencias*. A la variable que representa a la distribución se le llama genéricamente X, a cada uno de los valores que toma la variable se le denota por x_i, y a la frecuencia con que toma dicho valor por n_i. Para evitar confusiones es aconsejable ordenar los valores de la variable de menor a mayor. Los valores ordenados de una distribución se presentan con los subíndices entre paréntesis:

$$x_{(1)}, x_{(2)}, \cdots, x_{(n)}$$

de tal forma que siempre se verifica que $x_{(i)} \leq x_{(i+1)}$.

Para efectuar cálculos, sea cuál sea el tipo de distribución, se disponen los datos de la siguiente forma:

x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	n_k	$N_r = n$	f_r	$F_r = 1$

Donde:

- n representa al número total de observaciones y será igual a $\sum_{i=1}^{r} n_i$

- f_i es la frecuencia relativa, definida como $\frac{n_i}{n}$
- N_i es la frecuencia absoluta acumulada, que se obtiene como $\sum_{j=1}^{i} n_j$
- F_i es la frecuencia relativa acumulada, que viene dada por $\sum_{j=1}^{i} f_j$

Observe que si la distribución es de tipo I cada una de las frecuencias absolutas es igual a 1, y si la distribución es de tipo III los valores x_i representan a las marcas de clase o puntos medios de los intervalos[1].

3. Representaciones gráficas

En función de la naturaleza de los datos y de la forma en que éstos se presenten existen distintos tipos de representaciones. Se muestran aquí las más interesantes.

1. El *diagrama de tarta* se emplea para representar atributos.

   Ejemplo 1.6 *En una votación entre cuatro candidatos a representante de una comunidad se han obtenido los siguientes resultados:*

Candidato	*Número de votos*
A	*287*
B	*315*
C	*275*
D	*189*

   La representación gráfica mediante un diagrama de tarta sería la que se muestra en la figura 1.1.

2. Una distribución dada por extensión, se representa mediante un *diagrama de puntos.*

[1] Dado el intervalo (L_i, L_{i+1}), la marca de clase viene dada por $x_i = \frac{L_i+L_{i+1}}{2}$

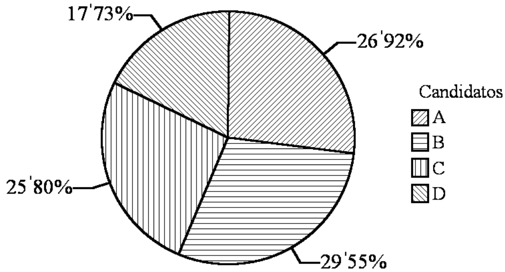


Figura 1.1: Diagrama de tarta

Ejemplo 1.7 *En un estudio sobre el peso y la estatura de un grupo de siete estudiantes se han obtenido las siguientes mediciones:*

$$(73, 1'87), (67, 1'75), (75, 1'80),$$
$$(66, 1'67), (80, 1'95), (64, 1'78), (83, 1'77).$$

La representación gráfica mediante un diagrama de puntos es la que se muestra en la figura 1.2.

A dicha representación se le suele denominar nube de puntos *o* diagrama de dispersión; *se estudiará más a fondo en el capítulo 2.*

3. Para representar una distribución del tipo II, se utiliza un *diagrama de barras*:

   Ejemplo 1.8 *La representación de la distribución del ejemplo 1.4 es la que se muestra en la figura 1.3.*

4. Por último, si se tiene una distribución del tipo III, se utiliza un *histograma*:

   Ejemplo 1.9 *El histograma correspondiente a la distribución del ejemplo 1.5 es el de la figura 1.4.*

Observe que el efecto que produce el histograma es el de relacionar el número de observaciones con el área dentro de cada rectángulo, por lo que si éstos tienen la misma base, es decir, si los intervalos son de la misma amplitud, basta con construir rectángulos con base los intervalos y altura las frecuencias asociadas a ellos. En

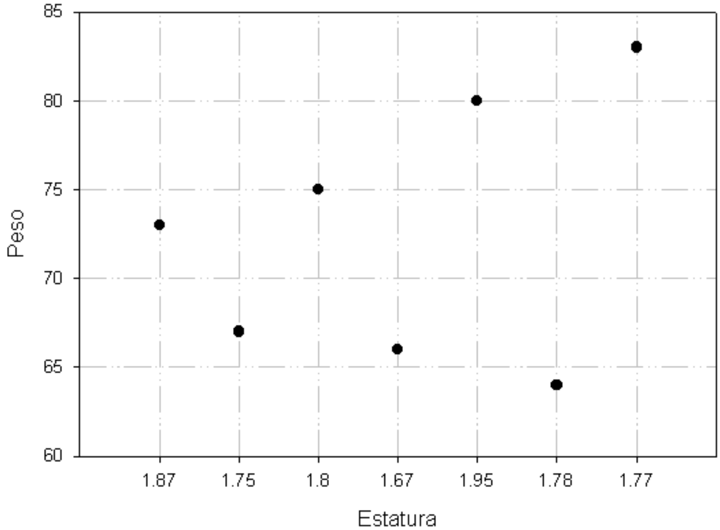


Figura 1.2: Diagrama de puntos

cambio, si las bases son distintas, o lo que es lo mismo, si los intervalos son de distinta amplitud, y se emplea el criterio anterior de asignación de alturas, se producirá una distorsión óptica. Por ello, en estos casos en vez de utilizar la frecuencia como altura de los rectángulos se utiliza la denominada *altura del histograma* o densidad de observaciones en el intervalo, definida como $h_i = \frac{n_i}{a_i}$, donde a_i es la amplitud del intervalo correspondiente.

4. Medidas centrales

Una vez organizados los datos en su correspondiente distribución de frecuencias, se procede a dar una serie de medidas que resuman toda esa información y que, "de alguna manera", representen a la distribución.

4.1. La media

La *media* es una medida de representación central que necesariamente debe cumplir tres requisitos:

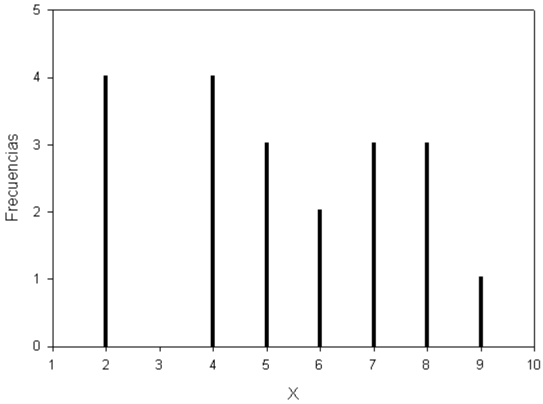


Figura 1.3: Diagrama de barras

1. Para su obtención deben utilizarse todas las observaciones.

2. Debe ser un valor comprendido entre el menor y el mayor de los valores de la distribución.

3. Debe venir expresada en la misma unidad que los datos.

Entre todas las funciones que verifican estas tres propiedades se destaca la *media aritmética*, a partir de ahora *media* simplemente, que se define de la siguiente manera:

$$\bar{x} = \frac{\displaystyle\sum_{i=1}^{r} x_i n_i}{n}.$$

Donde las x_i representan, según el caso, a los valores de la variable o a las marcas de clase de los intervalos.

Ejemplo 1.10 *La media de la distribución del ejemplo 1.4 viene dada por:*

$$\begin{aligned} \bar{x} &= \frac{2 \cdot 4 + 4 \cdot 4 + 5 \cdot 3 + \cdots + 9 \cdot 1}{4 + 4 + 3 + \cdots + 1} \\ &= \frac{105}{20} = 5'25. \end{aligned}$$

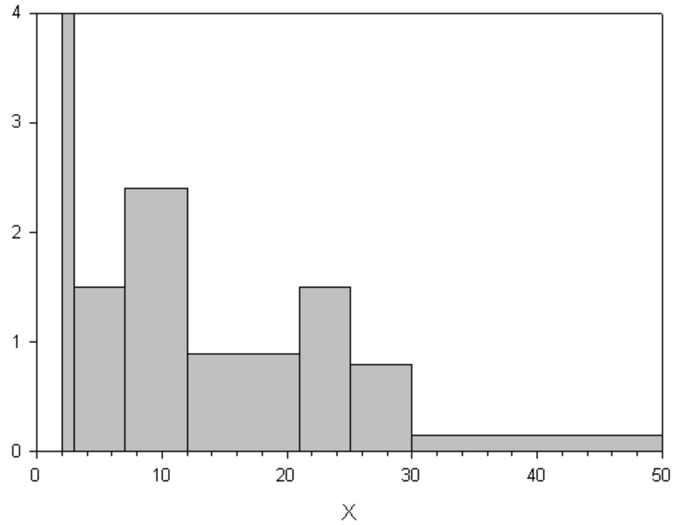


Figura 1.4: Histograma

Con el mismo esquema también se puede definir la *media geométrica* como:

$$\bar{x}_g = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_r^{n_k}}.$$

Ejemplo 1.11 *La media geométrica de la distribución del ejemplo 1.3 se obtendría como:*

$$\bar{x}_g = \sqrt[10]{2 \cdot 3 \cdot 5 \cdot 7^2 \cdot \ldots \cdot 19} = 7'483.$$

Cuando se tiene que hacer un promedio de un grupo de razones se utiliza la *media armónica*, definida como:

$$\bar{x}_a = \frac{n}{\displaystyle\sum_{i=1}^{k} \frac{n_i}{x_i}}.$$

Ejemplo 1.12 *La media armónica de la distribución del ejemplo 1.4 se obtendría como:*

$$\bar{x}_a = \frac{20}{\frac{4}{2} + \frac{4}{4} + \frac{3}{5} + \cdots + \frac{1}{9}} = 4'125.$$

Otra media que tiene interés práctico es la *media ponderada*. Esta consiste en asignar a cada valor x_i un peso w_i que depende de la importancia relativa de cada uno de estos valores bajo algún criterio. Su expresión responde a:

$$\bar{x}_p = \frac{\sum_{i=1}^{r} n_i w_i x_i}{\sum_{i=1}^{r} n_i w_i}.$$

Ejemplo 1.13 *Para superar la asignatura de estadística, un alumno debe ser evaluado en distintas pruebas referentes a la misma: test, problemas y práctica, cada una de ellas ponderada según su importancia o contribución en la nota final. Así, los pesos de cada prueba serán del* 30 %, 50 % *y* 20 % *respectivamente. Sabiendo que las notas obtenidas por el alumno en cada prueba son 7, 3 y 5 respectivamente, ¿cuál es la nota global en la asignatura?*

$$\begin{aligned} \bar{x}_p &= \frac{7 \cdot 30 + 3 \cdot 50 + 5 \cdot 20}{30 + 50 + 20} \\ &= \frac{460}{100} = 4'6. \end{aligned}$$

Propiedades de la media. Se analizan a continuación una serie de propiedades de la media que hacen de ésta una medida óptima de representación.

1. La suma de las desviaciones de los valores de la distribución respecto a la media es igual a cero, es decir:

$$\sum_{i=1}^{r} (x_i - \bar{x}) n_i = 0.$$

2. Si a cada observación de una distribución X se le suma una constante k (traslación), se tiene una nueva variable $Y = X + k$ con media igual a la de X más la constante k.

3. Si se multiplica una variable X por una constante k (homotecia), la variable resultante $Y = kX$ tendrá media igual a k por la media de X.

   Estas dos propiedades se pueden resumir en la siguiente:

$$Y = aX + b \qquad \Rightarrow \qquad \bar{y} = a\bar{x} + b.$$

4. La media es el valor ϕ que hace mínima la expresión:

$$\sum_{i=1}^{r} (x_i - \phi)^2 n_i.$$

Precisamente ese mínimo será la varianza de X, medida de dispersión que se estudia más adelante. Por otra parte, se comprobará que esta propiedad de la media garantiza su bondad como medida de representación.

Ejercicio 1.4 *Demuestre las propiedades anteriores.*

4.2. La mediana

La *mediana* es un valor que, previa ordenación, deja la mitad de las observaciones en la recta real a la izquierda y la otra mitad a la derecha. Es decir, el 50 % de los datos son menores o iguales que la mediana y el otro 50 % mayores o iguales a ésta. Para su cálculo y suponiendo que los valores están ordenados se procede de la siguiente manera:

1. Si los datos vienen dados por extensión, y hay un número impar de ellos la mediana es el elemento que se encuentra en el centro, es decir $x_{(\frac{n+1}{2})}$. Si el número de datos fuera par habría dos elementos centrales y la mediana se obtendría como la media de ambos, es decir:

$$M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}.$$

Ejemplo 1.14 *La mediana de la distribución del ejemplo 1.3 se obtendría como:*

$$M_e = \frac{x_{(5)} + x_{(6)}}{2} = \frac{7+8}{2} = 7'5.$$

2. A partir de una distribución de tipo II ordenada, se construye la columna de frecuencias absolutas acumuladas, se obtiene el valor de $\frac{n}{2}$, deslizándose por la columna de N_i hasta detectar la primera frecuencia mayor o igual que $\frac{n}{2}$; si dicha frecuencia es estrictamente mayor que $\frac{n}{2}$ la mediana toma el valor de la observación que la ostenta, si por el contrario $\frac{n}{2}$ coincide con algún N_i la mediana vale $\frac{x_i+x_{i+1}}{2}$.

Ejemplo 1.15 *Para calcular la mediana en la distribución del ejemplo 1.4 se obtiene $\frac{n}{2}$ que es igual a 10, construyendo la columna de frecuencias acumuladas:*

x_i	n_i	N_i	
2	4	4	
4	4	8	
5	3	11	$\longleftarrow$
6	2	13	
7	3	16	
8	3	19	
9	1	20	

Puesto que $N_2 < 10$ y $N_3 > 10$ entonces $M_e = 5$.

3. Por último, si la distribución viene agrupada en intervalos, se construye también la columna de N_i para fijar el intervalo donde se halla la mediana, éste queda determinado porque es el primero que verifica que la frecuencia acumulada del intervalo es mayor o igual que $\frac{n}{2}$. Una vez fijado el intervalo, la mediana adopta la siguiente expresión:

$$M_e = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i$$

donde L_{i-1} es el extremo inferior del intervalo y a_i su amplitud.

Ejemplo 1.16 *En la distribución del ejemplo 1.5,* $\frac{n}{2} = 21'5$. *La tabla de frecuencias acumuladas que se obtiene es:*

$(L_{i-1}, L_i]$	n_i	N_i	
$(2, 3]$	4	4	
$(3, 7]$	6	10	
$(7, 12]$	12	22	$\longleftarrow$
$(12, 21]$	8	30	
$(21, 25]$	6	36	
$(25, 30]$	4	40	
$(30, 50]$	3	43	

Por tanto:

$$M_e = 7 + \frac{21'5 - 10}{12} 5 = 11'79.$$

Ejercicio 1.5 *Demuestre que la mediana es el valor* ϕ *que hace mínima la expresión:*

$$\sum_{i=1}^{r} |x_i - \phi| n_i.$$

4.3. Las modas

La *moda* absoluta de una distribución es el valor que más veces se repite. Además de la moda absoluta, aquellos valores que tengan frecuencia mayor a la de los valores adyacentes serán modas relativas.

Las modas se pueden obtener fácilmente cuando los datos vienen dados en forma puntual.

Ejemplo 1.17 *En la distribución 2, 3, 3, 4, 6, 7, 7, 7, 10, la moda absoluta es 7, puesto que es el valor que se repite más veces, concretamente 3. Además, el 3 es una moda relativa, puesto que su frecuencia es 2, superior a la de los valores 2 y 4, ambas iguales a 1.*

Si las observaciones vienen agrupadas en intervalos hay que distinguir dos casos:

1. Intervalos de igual amplitud. En este caso se fija el intervalo que tenga mayor frecuencia –intervalo modal absoluto– y aquellos con frecuencia superior a la de los intervalos adyacentes –intervalos modales relativos–. Dentro de cada intervalo modal la moda corresponde al valor:

$$M_o = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} a_i.$$

Ejemplo 1.18 *En la distribución que sigue, el intervalo modal absoluto es el* $(4, 5]$*, además se tiene un intervalo modal relativo, el* $(6, 7]$*.*

$(L_{i-1}, L_i]$	n_i
$(2,3]$	2
$(3,4]$	3
$(4,5]$	7
$(5,6]$	3
$(6,7]$	6
$(7,8]$	5
$(8,9]$	3

La moda absoluta será:

$$M_o = 4 + \frac{3}{3+3} 1 = 4'5.$$

Y la moda relativa:

$$M_o = 6 + \frac{5}{5+3} 1 = 6'625.$$

2. Intervalos de distinta amplitud. En este caso el intervalo modal absoluto será aquel que tenga mayor *altura de histograma*, h_i, con idéntica discusión que antes para las modas relativas. La expresión de la moda viene dada por:

$$M_o = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} a_i.$$

Ejemplo 1.19 *Para la distribución que sigue:*

$(L_{i-1}, L_i]$	n_i	h_i
$(2,3]$	1	1
$(3,7]$	6	$1'5$
$(7,9]$	12	6
$(9,14]$	8	$1'6$
$(14,20]$	6	1
$(20,30]$	4	$0'4$

El intervalo modal, sólo existe uno, es $(7,9]$*, con lo que la moda vale:*

$$M_o = 7 + \frac{1'6}{1'6 + 1'5} 2 = 8'032.$$

Para terminar este epígrafe observe que cuando las distribuciones son de intervalos los cálculos puntuales de la mediana y la moda utilizan criterios de ponderación que suponen, como no puede ser de otra manera, la disposición uniforme de las observaciones dentro de los intervalos.

4.4. Comparación entre media, moda y mediana

Salvo en casos muy específicos, la media es la mejor de las medidas de representación, pues la moda es bastante inestable y un pequeño cambio en las observaciones puede afectarle mucho, mientras que la mediana es insensible al tamaño de los datos, permaneciendo constante si, por ejemplo, se altera arbitrariamente y en cierto sentido las observaciones extremas. Por otra parte, si se dispone de las modas y medianas de dos distribuciones hay que conocer cada uno de los datos de éstas para calcular la moda y mediana de la distribución conjunta. La media por el contrario es sensible a las alteraciones de los datos, al tamaño de éstos y si se conocen las medias de dos conjuntos de datos, basta con saber los tamaños de ambos grupos para calcular la media global.

Ejercicio 1.6 *Calcule la media, mediana y moda de la distribución:*

1, 2, 4, 7, 9, 9, 9, 11, 13, 14, 17, 21, 34

Obtenga de nuevo dichas medidas para la distribución a la que se ha añadido los valores -1 y 47. Comente los resultados en lo que se refiere a la estabilidad de las medidas obtenidas.

5. Medidas de posición

Se llaman medidas de posición o *cuantiles* de orden k a aquellas que dividen a la distribución en k partes, de tal forma que en cada una de esas partes haya el mismo número de elementos[2]. De entre todas las medidas de posición destacan los *cuartiles*, los *deciles* y los *percentiles*. Los cuartiles dividen a la distribución en cuatro partes iguales, los deciles en diez y los percentiles en cien. Habrá, por tanto, tres cuartiles (Q_1, Q_2, Q_3), nueve deciles $(D_1, D_2, \cdots, D_9)$ y, noventa y nueve percentiles $(P_1, P_2, \cdots, P_{99})$. El segundo cuartil, el quinto decil y el quincuagésimo percentil son iguales y coinciden con la mediana. En distribuciones puntuales el cálculo es idéntico al de la mediana, siendo ahora $\frac{rn}{k}$ el valor de discusión. En distribuciones por intevalos la forma general de cálculo para un cuantil, al que se denota por $C_{\frac{rn}{k}}, \quad k = 4, 10, 100, \ldots$, es la siguiente:

$$C_{\frac{rn}{k}} = L_{i-1} + \frac{\frac{rn}{k} - N_{i-1}}{n_i} a_i.$$

Siendo el intervalo i-ésimo el primero que verifica $N_i \geq \frac{rn}{k}$.

Ejemplo 1.20 *En la distribución:*

$(L_{i-1}, L_i]$	n_i	N_i	
$(2, 3]$	4	4	
$(3, 7]$	6	10	
$(7, 12]$	12	22	$\leftarrow P_{35}$
$(12, 21]$	8	30	
$(21, 25]$	6	36	$\leftarrow Q_3$
$(25, 30]$	4	40	
$(30, 50]$	3	43	

El Q_3 se obtendría calculando $\frac{3 \cdot 43}{4} = 32'5$. La primera frecuencia acumulada mayor que $32'5$ corres-

[2]La mediana es un caso particular de cuantil, que divide la distribución en dos partes iguales.

ponde al intervalo $(21, 25]$*, por lo que:*

$$Q_3 = 21 + \frac{32'5 - 30}{6} 4 = 22'66.$$

Para calcular el P_{35} *se obtiene* $\frac{35 \cdot 43}{100} = 15'05$*. El intervalo donde se encuentra el percentil buscado es el* $(7, 12]$ *y, por tanto:*

$$P_{35} = 7 + \frac{15'05 - 10}{12} 5 = 9'10.$$

6. Medidas de dispersión

A continuación se estudian una serie de medidas que por una parte indicarán el nivel de concentración de los datos que se están analizando y por otra informarán sobre la bondad de los promedios calculados como representantes del conjunto de datos.

6.1. Varianza y desviación típica

La *varianza* y su raíz cuadrada positiva, la *desviación típica*, son las más importantes medidas de dispersión, estando íntimamente ligadas a la media como medida de representación de ésta. La varianza viene dada por la expresión:

$$S^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{n}.$$

Y la desviación típica es, por tanto, $S = +\sqrt{S^2}$.

El dar dos expresiones para un mismo concepto se explica porque la varianza es un término de más fácil manejo, mientras que la desviación típica viene dada en la misma unidad que la variable. Tanto una como la otra son siempre positivas y valen cero sólo en el caso de que todos los valores coincidan con la media (representatividad absoluta de la media).

Ejemplo 1.21 *Dada la distribución:*

$(L_{i-1}, L_i]$	x_i	n_i
$(-2, 2]$	0	1
$(2, 4]$	3	3
$(4, 8]$	6	6
$(8, 12]$	10	13
$(12, 20]$	16	8
$(20, 24]$	22	6
$(24, 30]$	27	5
$(30, 40]$	35	3

Cuya media vale 15, se calcula la varianza y la desviación típica como:

$$S^2 = \frac{(0-15)^2 \cdot 1 + \cdots + (35-15)^2 \cdot 3}{45} = 82$$

$$S = \sqrt{82} = 9'055.$$

Propiedades de la varianza

1. Si se le suma una constante a una variable, la varianza de la nueva variable no cambia.

2. Si se multiplica una variable por una constante, la varianza de la nueva variable es igual a la de la antigua multiplicada por la constante al cuadrado.

Estas dos propiedades pueden resumirse en la siguiente expresión:

$$Y = aX + b \qquad \Rightarrow \qquad S_Y^2 = a^2 S_X^2.$$

Ejercicio 1.7 *Demuestre las propiedades anteriores.*

Ejemplo 1.22 *Dada la variable X con media $\bar{x} = 12$ y desviación típica $S_X = 9$, la variable $Y = 3X - 4$ tendrá de media y desviación típica:*

$$\bar{y} = 3\bar{x} - 4 = 3 \cdot 12 - 4 = 32$$

$$S_Y = \sqrt{3^2} \cdot S_X = \sqrt{9} \cdot 9 = 27.$$

6.2. Otras medidas de dispersión

6.2.1. El recorrido y el rango

Se define el primero como la diferencia entre el mayor y el menor de los valores y el segundo como el intervalo cuyos extremos son el mínimo y el máximo de la distribución. Tienen la ventaja de que son fáciles de calcular, aunque cuando hay valores aislados en las puntas de la distribución dan una visión distorsionada de la dispersión de ésta.

Ejemplo 1.23 *En la distribución del ejemplo 1.4 el recorrido vale 7, mientras que el rango es* $[2, 9]$.

6.2.2. La desviación absoluta

La desviación absoluta respecto a la media, está definida por:

$$D_m = \frac{\sum_{i=1}^{r} |x_i - \bar{x}| n_i}{n}.$$

También puede definirse respecto a la mediana, siendo ésta el valor que minimiza dicha expresión.

6.2.3. Recorrido intercuartílico

Viene dado por:

$$R_I = Q_3 - Q_1.$$

Es una medida adecuada para el caso en que se desee que determinadas observaciones extremas no intervengan, evitándose, de este modo, una visión sesgada de la variabilidad de la distribución. Como inconveniente principal tiene que en su confección sólo intervienen el 50 % de los valores centrales.

Las expresiones que se acaban de ver expresan la dispersión de la distribución en términos absolutos, se precisa definir a partir de ellas, otras que hagan posible la comparación entre varias variables y que tengan en cuenta el tamaño de las observaciones. Obsérvese que la distribución formada por los elementos {0'1, 0'2, 0'3, 0'4, 0'5} y la que constituyen {1000'1, 1000'2, 1000'3, 1000'4, 1000'5} tienen la misma varianza y, sin embargo, es evidente que en el primero de los casos los elementos están muy dispersos y en el segundo bastante concentrados, ésto es consecuencia de la primera de las propiedades de la varianza. Para evitar estas situaciones se estudia la siguiente medida.

6.3. Coeficiente de variación

Se define como el cociente entre la desviación típica y el valor absoluto de la media. Se trata de una medida adimensional, tiene en cuenta el rango de valores en el que se mueve, permite comparar la dispersión de varias distribuciones, es invariante respecto a homotecias y sensible frente a traslaciones. Además de lo anterior, el *coeficiente de variación* da información sobre la representatividad de la media; y aunque no hay valores fijos de comparación, pues depende de circunstancias tales como el número de observaciones, se puede considerar, a efectos prácticos, una cota de $0'5$ como límite para admitir que la media representa aceptablemente al conjunto de la distribución.

Ejemplo 1.24 *En el caso del ejemplo 1.21 se tiene que:*

$$C_V = \frac{S}{|\bar{x}|} = \frac{9'055}{15} = 0'60.$$

Lo que implica que la media no representa en modo alguno al conjunto de la distribución.

6.4. Recorrido semiintercuartílico respecto a la mediana

Viene dado por:

$$R_{SI} = \frac{Q_3 - Q_1}{M_e}$$

que al igual que la anterior es una medida adimensional, con las ventajas e inconvenientes mencionados para el recorrido intercuartílico.

7. Desigualdad de Tchebychev

Esta desigualdad relaciona a la media y a la varianza y tiene la expresión:

$$f(|x_i - \bar{x}| \leq aS) \geq 1 - \frac{1}{a^2} \quad , \quad a > 1.$$

Que justifica el caracter de medida de dispersión de la varianza. Así, en un intervalo de centro la media y radio 4 veces la desviación típica se encuentra, al menos, el 93'75 por ciento de la distribución.

Observación 1.1 *La desigualdad de Tchebychev proporciona una cota inferior para el porcentaje de observaciones en un determinado intervalo con centro la media de la distribución.*

Ejemplo 1.25 *Dada una distribución con media,* $\bar{x} = 25$*, y desviación típica,* $S = 4$*, el intervalo* $[\bar{x} - 3S, \bar{x} + 3S] = [13, 37]$ *garantiza la presencia en su interior de, al menos, el* $88'88\,\%$ *de la distribución.*

8. Momentos de la distribución

Las medidas que se han visto hasta el momento presentan visiones parciales de la distribución, se pretende dar ahora una herramienta eficaz que generalice esa idea, de tal forma que la mayoría de las características se puedan expresar utilizando dicha herramienta. Así, se hace referencia a los *momentos de la distribución.*

8.1. Momentos respecto al origen

Se define el *momento de orden k respecto al origen* como:

$$a_k = \frac{\sum_{i=1}^{r} x_i^k n_i}{n}.$$

Es evidente que a_0 es igual a 1 y que a_1 es igual a la media.

8.2. Momentos respecto a la media

El *momento de orden k respecto a la media* viene dado por:

$$m_k = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^k n_i}{n}.$$

Se puede comprobar que m_0 es igual a 1, que m_1 es cero y que m_2 es la varianza.

Es posible expresar los momentos respecto a la media en función de los momentos respecto al origen.

Ejercicio 1.8 *Demuestre que:*

a) $m_2 = S^2 = a_2 - a_1^2$

y que:

b) $m_3 = a_3 - 3a_2a_1 + 2a_1^3$.

Ejemplo 1.26 *En el ejemplo 1.21 el cálculo de la varianza se podría haber hecho, utilizando la fórmula anterior, de la siguiente manera:*

$$\begin{aligned} S^2 &= \sum_{i=1}^{r} \frac{x_i^2 n_i}{n} - \bar{x}^2 \\ &= \frac{0^2 \cdot 1 + 3^2 \cdot 3 + \cdots + 35^2 \cdot 3}{45} - 15^2 \\ &= 82. \end{aligned}$$

9. Medidas de forma

Este epígrafe y el siguiente se detienen a analizar la "forma" de la distribución, tratando a la variable desde un enfoque distinto al seguido hasta ahora, en primer lugar se examina la simetría y a continuación el apuntamiento.

9.1. Simetría

Los *coeficientes de simetría* indicarán si la distribución es simétrica y, caso de no serlo, el tamaño y la tendencia de su asimetría. Para ello, se distinguen dos tipos de distribuciones, las que tienen forma de campana y las que no la tienen, empleándose expresiones alternativas para su cálculo.

1. Si la distribución tiene forma de campana se utiliza la expresión:

$$A_s = \frac{\bar{x} - M_o}{S}.$$

De tal forma que cuando A_s es igual a cero la distribución es simétrica, si es menor, asimétrica negativa o tendida a la derecha, y si es mayor, asimétrica positiva o tendida a la izquierda.

Ejemplo 1.27 *Dada la distribución campaniforme:*

$(L_i, L_{i+1}]$	x_i	n_i	h_i
$(2,4]$	3	2	1
$(4,8]$	6	6	$1'5$
$(8,12]$	10	12	3
$(12,20]$	16	12	$1'5$
$(20,24]$	22	3	$0'75$

Donde $\bar{x} = 12$, $S = 5'12$ *y* $M_o = 10$, *ocurre que:*

$$A_s = \frac{12 - 10}{5'12} = 0'39.$$

La representación gráfica de la distribución viene dada en la figura 1.5.

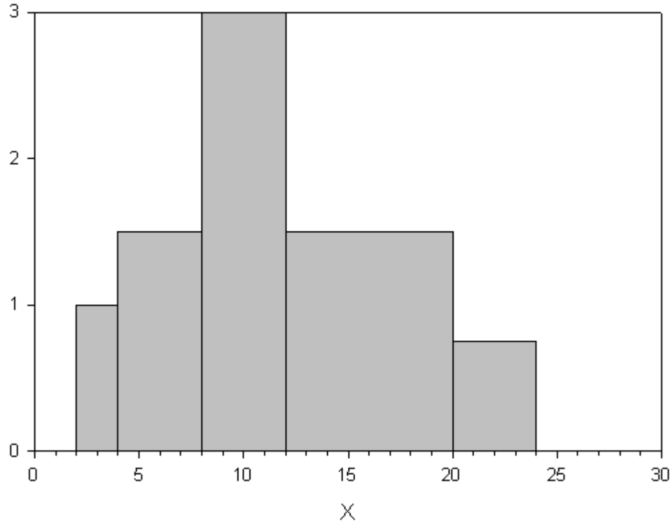


Figura 1.5: Histograma

> *Con lo que la distribución está, como puede observarse en el gráfico, inclinada, levemente, a la izquierda.*

2. Si la distribución no tiene forma de campana o se desconoce este hecho se calcula la simetría mediante el coeficiente:

$$g_1 = \frac{m_3}{S^3}.$$

Siendo la discusión igual a la del caso anterior.

Observe que cuando la distribución es simétrica coinciden la media y la mediana, y que si además tiene forma de campana ambas son iguales a la moda.

9.2. Curtosis

El grado de apuntamiento de una distribución se examina a través del *coeficiente de curtosis*, para lo cual se compara con la distribución Normal tipificada o $N(0,1)$ que se trata en el capítulo 5 (figura 1.6).

Se puede adelantar, no obstante, que tiene forma de campana y que su estructura "probabilística" viene dada por la expresión:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

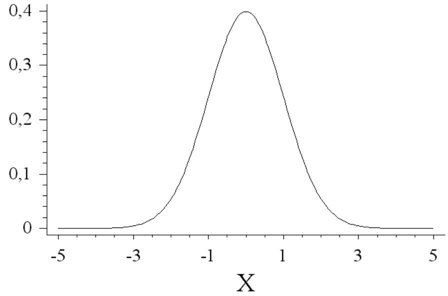


Figura 1.6: Función de densidad $N(0,1)$

El coeficiente de curtosis toma la expresión:

$$g_2 = \frac{m_4}{S^4}.$$

Cuando dicho coeficiente vale 3 coincide con el de la $N(0,1)$ y se dice que la distribución es mesocúrtica, si es menor que 3 platicúrtica y si es mayor que 3 leptocúrtica.

Ejemplo 1.28 *En la distribución de frecuencias:*

Valor	*Frecuencia*
2	*5*
4	*4*
5	*3*
6	*2*
7	*2*
8	*3*
9	*1*

claramente no campaniforme, se tiene que: $n = 20$, $\bar{x} = 5$, $S = 2'258$, $m_3 = 1'2$ *y* $m_4 = 47'1$ *por lo que* *el coeficiente de asimetría vendría dado por:*

$$g_1 = \frac{m_3}{S^3} = \frac{1'2}{11'51} = 0'104.$$

Lo que implicaría que la distribución está levísimamente inclinada hacia la izquierda.

Por lo que respecta al coeficiente de curtosis:

$$g_2 = \frac{m_4}{S^4} = \frac{47'1}{25'99} = 1'81.$$

Tratándose, por consiguiente, de una distribución claramente aplastada o platicúrtica.

10. Transformaciones

A veces se tiene el inconveniente de que la distribución que se estudia presenta muchas irregularidades, como asimetrías acentuadas, valores extremos, etc..., en otras ocasiones se debe comparar la posición de dos elementos que pertenecen a poblaciones con características distintas o del mismo elemento en situaciones distintas. En estos casos es recomendable efectuar una transformación que haga más regular la distribución y, por tanto, con mejores condiciones para su estudio. Particular importancia tiene la tipificación de una variable.

10.1. Normalización o tipificación

Dada una variable X con media $\bar{x}$ y desviación típica S, la tipificación consiste en realizar la siguiente transformación:

$$Z = \frac{X - \bar{x}}{S}.$$

A la nueva variable Z se le llama *variable normalizada* o *tipificada* y tiene media 0 y desviación típica 1. Haciendo un símil, la media y la desviación típica de una variable pueden considerarse como el centro de gravedad de la distribución y su escala, respectivamente, por lo que al tipificar distintas variables las centramos en el mismo punto y las dotamos de la misma escala; además, los valores tipificados pierden la unidad de la variable. Por lo anterior, la tipificación tiene la propiedad de hacer comparables individuos que pertenezcan a distintas distribuciones, aún en el caso de que éstas vinieran expresadas en diferentes unidades.

Ejemplo 1.29 *Dos trabajadores del mismo sector ganan 620€ y 672€, respectivamente. El primero pertenece a la empresa A, cuya retribución media y desviación típica vienen dados por:* $\bar{x}_A = 580$€ *y* $S_{x_A} = 25$€, *mientras que para la empresa del segundo trabajador se tiene:* $\bar{x}_B = 640$€ *y* $S_{x_B} = 33$€. *Tanto uno como el otro ganan salarios por encima de la media, por lo que si se quiere conocer cuál de los dos ocupa mejor posición relativa dentro de su empresa hay que tipificar sus puntuaciones, y así:*

$$z_A = \frac{620 - 580}{25} = 1'6$$

mientras que:

$$z_B = \frac{672 - 640}{33} = 0'97.$$

Por lo que, aunque en términos absolutos el trabajador de la empresa B gana más que el de A, en relación al conjunto de los empleados de cada empresa el empleado de A ocupa mejor posición.

Otras transformaciones usuales son la del logaritmo y la de la raíz cuadrada que consiguen una mayor simetría y concentración de los valores de la distribución.

11. Análisis exploratorio de datos

El análisis exploratorio de datos (AED) está formado por un conjunto de técnicas estadísticas, fundamentalmente gráficas, que pretenden dar una visión simple e intuitiva de las principales características de la distribución en estudio. El AED puede ser un fin por sí mismo o una primera etapa de un estudio más completo. Como aspectos más destacables que abarca el AED, están los que se refieren a la forma de la distribución y a la detección de valores anómalos.

11.1. Diagramas de tallo y hojas de Tukey

El *diagrama de tallo y hojas* es una representación semi-gráfica donde se muestra el rango y distribución de los datos, la simetría y si hay candidatos a valores atípicos. Para su construcción se siguen los siguientes pasos:

1. Se redondean los valores a dos o tres cifras significativas.

2. Se divide el rango de los datos en k intervalos, cada uno representado por una fila de la tabla que está dividida por una línea vertical en dos partes. En cada fila, los datos individuales son representados por uno o dos dígitos, según el rango, (llamado tallo), mientras que a la derecha de la línea vertical se coloca el último dígito del valor (llamado hoja). Si hay algún punto que se encuentra lejano de la mayoría de los valores (candidato a valor atípico), éste es colocado en hoja superior o inferior separada. La tabla de tallo y hojas se acompaña de una columna de frecuencias acumuladas creciente inferior y superiormente hasta el tallo que contiene la mediana que queda señalado entre paréntesis.

   Ejemplo 1.30 *A partir de la información recogida sobre los caballos de potencia de distintos vehículos, se representa el diagrama de tallo y hojas para dicha variable (figura 1.7).*

Su uso es recomendable siempre que el número de datos no sea muy grande (menor que 50).

11.2. Diagrama de caja ó diagrama de box-whisker

Los *diagramas de caja* son representaciones gráficas sencillas que no necesitan un número elevado de valores para su construcción. Se utilizan para estudiar tanto la dispersión como la forma de una distribución.

```
   unidad = 10,0  1|2  representa 120,0

  1  0|5
  7  0|666777
 26  0|8889999999999999999
(20) 1|00000000001111111111
 43  1|2233333
 36  1|44444444444555555
 19  1|6666666677
  9  1|889
  6  2|000
  3  2|22

  HI | 245,0
```

Figura 1.7: Diagrama de tallo y hojas

Asimismo son especialmente útiles para comparar distintas distribuciones entre sí.

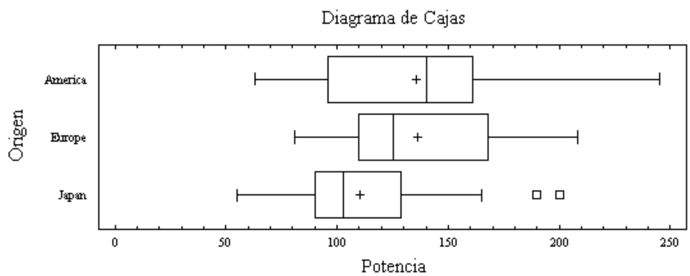


Figura 1.8: Diagrama de cajas

La caja representa el 50 % central de la distribución, la línea situada en el interior de la caja es la mediana, mientras que la cruz se corresponde con la media. Los extremos inferiores y superiores de los segmentos (también llamados bigotes) delimitan lo que se denomina como valores "normales" y coinciden, respectivamente, con el mínimo y el máximo de los valores una vez excluidos los candidatos a valores anómalos. Los candidatos a valores anómalos se etiquetan como atípicos y coinciden con aquellas observaciones que se encuentran fuera del intervalo (LI, LS), donde:

$$\begin{aligned} LI &= Q_1 - 1'5R_I \\ LS &= Q_3 + 1'5R_I, \end{aligned}$$

es decir, a una distancia de Q_1, por la izquierda, o de Q_3, por la derecha, superior a una vez y media el recorrido intercuartílico; denominándose, en este caso, atípicos de primer nivel. Cuando la distancia, por uno de los dos lados, es superior a tres recorridos intercuartílicos, el valor atípico se denomina de segundo nivel.

Los valores atípicos de primer y segundo nivel quedan normalmente identificados en el diagrama de cajas por símbolos diferenciados ($\triangle$, $\diamondsuit$, $\cdot$), debiendo considerarse la posibilidad de realizar una depuración de los mismos antes de comenzar el tratamiento de los datos.

12. Ejercicios

12.1. Ejercicio resuelto

1.1 Para realizar un determinado experimento se ha medido la anchura interorbital, en mm., de una muestra de 40 palomas, obteniéndose los siguientes datos:

12'2, 12'9, 11'8, 11'9, 11'6, 11'1, 12'3, 12'2, 11'8, 11'8
10'7, 11'5, 11'3, 11'2, 11'6, 11'9, 13'3, 11'2, 10'5, 11'1
12'1, 11'9, 10'4, 10'7, 10'8, 11'0, 11'9, 10'2, 10'9, 11'6
10'8, 11'6, 10'4, 10'7, 12'0, 12'4, 11'7, 11'8, 11'3, 11'1

Se pide:

a) Construya una distribución de frecuencias y calcule la media, desviación típica y coeficiente de variación.

b) Agrupe los datos en intervalos con la amplitud más adecuada, calculando de nuevo los parámetros anteriores y comparándolos con los resultados obtenidos a partir de los datos no agrupados. Dibuje el histograma.

En lo que sigue trabaje con la distribución por intervalos.

c) ¿En qué intervalo de centro la media se encuentra, al menos, el 75 % de la distribución?

d) Calcule la mediana y la moda.

e) Obtenga el intervalo donde se encuentra el 40 % central de la distribución.

f) Estudie la simetría y el apuntamiento de la distribución.

Solución:

a) La distribución de frecuencias sería:

x_i	n_i	x_i	n_i	x_i	n_i	x_i	n_i
$10'2$	1	$11'0$	1	$11'7$	1	$12'2$	2
$10'4$	2	$11'1$	3	$11'8$	4	$12'3$	1
$10'5$	1	$11'2$	2	$11'9$	4	$12'4$	1
$10'7$	3	$11'3$	2	$12'0$	1	$12'9$	1
$10'8$	2	$11'5$	1	$12'1$	1	$13'3$	1
$10'9$	1	$11'6$	4				

Gráficamente dicha distribución puede presentarse mediante el polígono de frecuencias de la figura 1.9.

Para calcular la media:

$$\bar{x} = \sum_{i=1}^{r} \frac{x_i n_i}{n} = \frac{459'2}{40} = 11'48 \text{ mm}$$

Es conveniente comprobar siempre que la media es un valor razonable y, en particular, dentro del rango de valores de la variable. En nuestro caso $10'2 < 11'48 < 13'3$.

La desviación típica vendría dada por:

$$\begin{aligned} S &= \sqrt{\sum_{i=1}^{r} \frac{x_i^2 n_i}{n} - \bar{x}^2} \\ &= \sqrt{\frac{5290'28}{40} - (11'48)^2} \\ &= \sqrt{0'4666} = 0'6831 \text{ mm} \end{aligned}$$

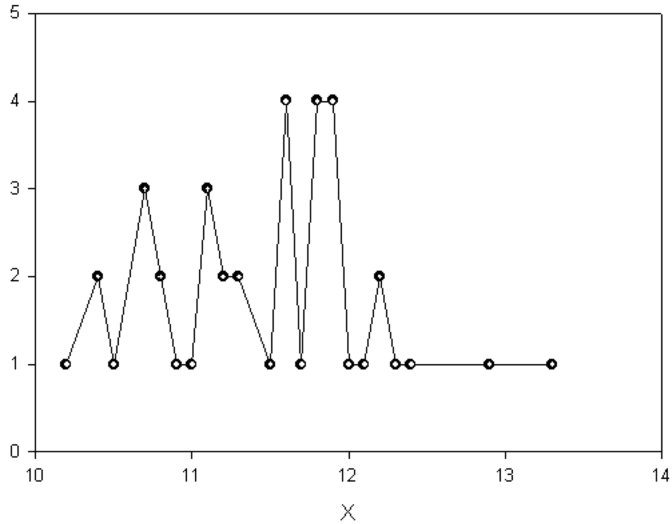


Figura 1.9: Polígono de frecuencias

Y el coeficiente de variación:

$$CV = \frac{S}{|\bar{x}|} = \frac{0'6831}{11'48} = 0'0595.$$

El bajo valor del coeficiente de variación indica que los valores están muy concentrados y que la media representa aceptablemente al conjunto de la distribución. En general, valores de CV menores a $0'1$ indican una alta concentración, entre $0'1$ y $0'5$ una concentración media y valores superiores a $0'5$ una alta dispersión y una media poco o nada representativa.

Observe que tanto la desviación típica como el coeficiente de variación son medidas positivas.

b) Para agrupar la distribución en intervalos se elige un número de éstos alrededor de $\sqrt{n}$, en nuestro caso $\sqrt{40} = 6'32 \simeq 7$. Los intervalos son de amplitud aproximada:

$$\frac{\text{Recorrido}}{\text{N}^{\text{o}} \text{ de intervalos}} = \frac{13'3 - 10'2}{7} = 0'44.$$

Buscando siempre que sea un valor fácil de manejar, en este caso se opta por una amplitud de $0'5$. La distribución en intervalos quedaría:

x_i	$[L_{i-1}, L_i)$	n_i
$10'25$	$[10, 10'5)$	3
$10'75$	$[10'5, 11)$	7
$11'25$	$[11, 11'5)$	8
$11'75$	$[11'5, 12)$	14
$12'25$	$[12, 12'5)$	6
$12'75$	$[12'5, 13)$	1
$13'25$	$[13, 13'5)$	1

donde ahora x_i representa la marca de clase.

A partir de estos datos se tiene: $\bar{x} = 11'5\text{mm}$, $S = 0'6708\text{mm}$ y $CV = 0'0583$. Con pequeñas variaciones respecto a los valores obtenidos para la distribución original, en todo caso, perfectamente asimilables y habiéndose conseguido una mayor facilidad de cálculo.

El histograma se representa en la figura 1.10. Como se puede apreciar la información visual que proporciona es mucho más clara que la que daría el polígono de frecuencias, a todas luces ininteligible. Se trata de una distribución unimodal y un poco tendida hacia la derecha, aunque esto se cuantificará más adelante.

c) Para contestar a esta cuestión se utiliza la desigualdad de Tchebychev, que dice:

$$f(|x_i - \bar{x}| \leq kS) \geq 1 - \frac{1}{k^2}.$$

Para $k = 2$, $1 - \frac{1}{k^2} = 0'75$, por lo que operando con el valor absoluto, el intervalo será:

$$[\bar{x} - 2S, \bar{x} + 2S] = [10'1138, 12'8462].$$

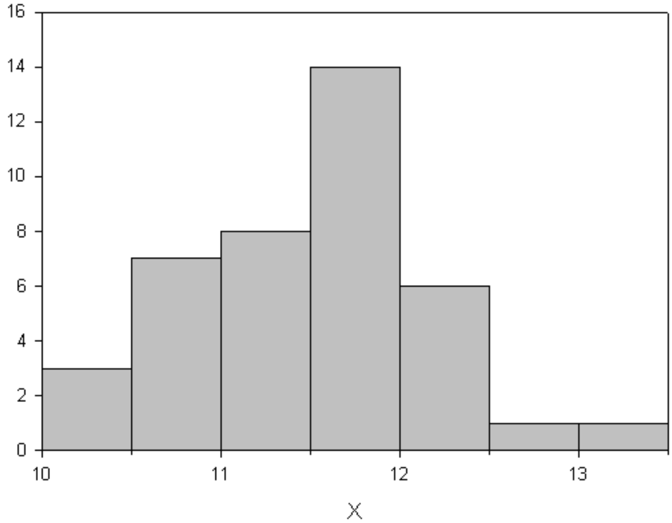


Figura 1.10: Histograma

d) Para calcular la mediana se obtiene la columna de frecuencias acumuladas:

x_i	$(L_{i-1}, L_i]$	n_i	N_i	
$10'25$	$10 - 10'5$	3	3	
$10'75$	$10'5 - 11$	7	10	
$11'25$	$11 - 11'5$	8	18	
$11'75$	$11'5 - 12$	14	32	$\leftarrow$
$12'25$	$12 - 12'5$	6	38	
$12'75$	$12'5 - 13$	1	39	
$13'25$	$13 - 13'5$	1	40	

La mediana se encuentra en aquel intervalo tal que $N_i \geq \frac{n}{2} = 20$, por tanto $M_e \in (11'5, 12]$, por lo que utilizando la fórmula apropiada, se tiene:

$$M_e = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i = 11'5 + \frac{20 - 18}{14} 0'5 = 11'5714 \text{ mm}$$

Por lo que $11'5714$ deja el $50\,\%$ de la distribución a la izquierda y el otro $50\,\%$ a la derecha.

Para calcular la moda, puesto que los intervalos son de igual amplitud, se selecciona aquel que tenga mayor frecuencia, en este caso el $(11'5, 12]$ que tiene frecuencia 14, y se aplica la fórmula correspondiente:

$$M_o = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} a_i = 11'5 + \frac{6}{6+8} 0'5 = 11'7143 \text{ mm}$$

e) El 40 % central de la distribución está contenido en el intervalo (P_{30}, P_{70}). El percentil P_{30} se encuentra en el intervalo $(L_{i-1}, L_i]$ para el que se verifica que $N_i \geq \frac{30 \cdot 40}{100} = 12$. Observando la columna de frecuencias acumuladas se ve que dicho intervalo es el $(11, 11'5]$. Por tanto:

$$P_{30} = 11 + \frac{12-10}{8} 0'5 = 11'125.$$

Operando de forma análoga:

$$P_{70} = 11'5 + \frac{28-12}{14} 0'5 = 11'8571.$$

Por lo que el intervalo pedido será el $(11'125, 11'8571)$.

f) Puesto que la distribución tiene forma de campana el coeficiente de simetría viene dado por:

$$A_s = \frac{\bar{x} - M_o}{S} = \frac{11'5 - 11'7143}{0'6708} = -0'319.$$

Por lo que la distribución está ligeramente inclinada hacia la derecha.

El coeficiente de apuntamiento:

$$g_2 = \frac{m_4}{S^4} = \frac{\frac{1}{40}\sum_{i=1}^{7}(x_i - \bar{x})^4 f_i}{(0'6708)^4} = \frac{0'400488}{0'202475} = 1'97796.$$

Al ser $g_2 < 3$ la distribución es platicúrtica, es decir, más aplastada que la distribución $N(0,1)$.

12.2. Ejercicios propuestos

1.1. Al comenzar el curso se pasó una encuesta a los alumnos del primer curso de un colegio, preguntándoles, entre otras cuestiones, por el número de hermanos que tenían, obteniéndose los siguientes resultados:

3, 3, 2, 2, 8, 5, 2, 4, 3, 1, 4, 5, 3, 3, 3, 3, 3, 2, 5
1, 3, 3, 2, 2, 4, 3, 3, 2, 2, 4, 4, 3, 6, 3, 3, 2, 2, 4
3, 4, 3, 2, 2, 4, 4, 3, 3, 4, 2, 5, 4, 1, 2, 8, 2 ,3, 3, 4

a) Represente este conjunto de datos con un diagrama de barras.

b) Calcule media, moda y mediana.

c) Estudie la dispersión de los datos.

d) Analice la simetría de la distribución.

1.2. Los pesos de un colectivo de niños son:

60, 56, 54, 48, 99, 65, 58, 55, 74, 52, 53, 58, 67, 62, 65
76, 85, 92, 66, 62, 73, 66, 59, 57, 54, 53, 58, 57, 55, 60
65, 65, 74, 55, 73, 97, 82, 80, 64, 70, 101, 72, 96, 73, 55
59, 67, 49, 90, 58, 63, 96, 100, 70, 53, 67, 60, 54

Obtenga:

a) La distribución de frecuencias agrupando por intervalos.

b) La mediana de la distribución.

c) La media de la distribución, indicando su nivel de representatividad.

d) Utilizando la agrupación en intervalos, el porcentaje de alumnos que tienen un peso menor de 65 kg y el número de alumnos con un peso mayor de 60 kg dentro del grupo de los que pesan menos de 80 kg.

1.3. En el Consejo de Apuestas del Estado se han ido anotando, durante una temporada, el número de premiados de quinielas según la cantidad de aciertos, obteniéndose la siguiente tabla:

Nº de aciertos	11	12	13	14	15
Nº de personas (miles)	52	820	572	215	41

Calcule:

a) La mediana, la moda y los cuartiles de la distribución.

b) La simetría de la distribución.

1.4. En un puerto se controla diariamente la entrada de pesqueros según su tonelaje, resultando para un cierto día los siguientes datos:

Peso(Tm.)	0-25	25-50	50-70	70-100	100-500
N° de barcos	5	17	30	25	3

Se pide:

a) El peso medio de los barcos que entran en el puerto diariamente, indicando la representatividad de dicha medida.

b) El intervalo donde se encuentra el 60 % central de la distribución.

c) El grado de apuntamiento.

d) El tonelaje más frecuente en este puerto.

1.5. El número de días de hospitalización de los enfermos que llegan en un cierto día a un servicio de urgencias, viene dado por:

N° de días	0-1	2-5	6-8	9-15
N° de enfermos	53	24	16	7

Se pide:

a) Un coeficiente que represente la distribución indicando dicho nivel de representatividad.

b) El porcentaje de enfermos que se quedan hospitalizados más de 5 días.

c) El valor que divide a la distribución en dos partes iguales.

1.6. Según un estudio se sabe que la planificación óptima de una determinada empresa exige que el 70 % sean administrativos, el 25 % jefes de departamento y el 5 % inspectores. Para realizar esta planificación se lleva a cabo un examen tipo test, obteniéndose las siguientes puntuaciones:

Puntuación	Empleados
[0,20)	70
[20,50)	115
[50,75)	95
[75,100)	5

a) ¿Cúal es la puntuación mínima para ser jefe de departamento?

b) ¿Y para ser inspector?

1.7. Para la selección de personal en dos empresas se realiza un test obteniéndose las siguientes puntuaciones porcentuales:

Factoría I		Factoría II	
Puntuación	Porcentaje	Puntuación	Porcentaje
[0,10]	0'07	[10,20]	0'08
[11,19]	0'25	[21,25]	0'16
[20,28]	0'38	[26,30]	0'20
[29,41]	0'19	[31,39]	0'28
[42,50]	0'11	[40,44]	0'23
		[45,50]	0'05

a) ¿Cuál de las dos factorías ha sido menos homogénea en los resultados?

b) ¿Qué persona ha tenido una puntuación mayor con respecto a su factoría: el que ha obtenido 35 en la factoría I o el que consiguió 38 en la II?

1.8. La vida útil de cierto tipo de bombonas de gas presenta la siguiente distribución:

Horas	Fracción de bombonas
[10,30)	0'04
[30,40)	0'27
[40,50)	0'34
[50,70)	0'26
[70,80]	0'09

Calcule:

a) La vida media de las bombonas de gas.

b) El tiempo de vida más frecuente.

c) El intervalo, con centro la media, donde se encuentre, al menos, el 85 % de la distribución.

d) El apuntamiento de la distribución.

1.9. El gasto de 100 experimentos, siendo la unidad 100€, viene dado por la siguiente tabla:

Gasto	[10,20)	[20,30)	[30,40)	[40,55]
Nº Experimentos	15	50	30	5

Calcule:

a) El gasto medio de los experimentos.

b) El porcentaje de experimentos que tienen un gasto entre 2300€ y 3500€.

c) Los precios que dividen a la distribución en cuatro partes iguales.

d) El gasto más frecuente.

1.10. En una entidad bancaria se sabe que, por término medio, el 15 % de los cheques son sin fondo. Las cantidades recogidas en dichos cheques, en euros, son las siguientes:

Importe de los cheques	Número de cheques
[0,200)	325
[200,600)	515
[600,1000)	420
[1000,3000]	270

Calcule:

a) El importe medio de los cheques sin fondo.

b) El importe más frecuente de los cheques sin fondo.

1.11. Para un determinado experimento se venía trabajando con unas temperaturas que variaban entre 100ºC y 130ºC. Estas temperaturas tenían una media de 110ºC y una desviación típica de 16ºC. Con un nuevo sistema se ha conseguido aumentar esta temperatura en 12ºC. ¿Cómo varía la dispersión relativa de dicha temperatura?

1.12. La producción de una empresa está organizada en dos factorías. La distribución de los salarios en cada una de ellas es la siguiente:

Salario en euros	Obreros Factoría A	Obreros Factoría B
[180,360)	20	20
[360,480)	23	28
[480,600)	22	14
[600,900)	15	8
[900,1200]	3	2

Se pide:

a) El salario más frecuente de la factoría A.

b) El salario que divide a la distribución de la factoría B en dos trozos iguales.

c) Los salarios medios de cada factoría.

d) El salario medio total a partir de los salarios medios de cada factoría.

1.13. El consumo de gasolina de dos coches de las marcas Citroën y Mercedes es, respectivamente, de 10 y 11 litros cada 100 km. Para el conjunto de los coches Citroën y Mercedes, se tienen, respectivamente, consumos medios de $7'4$ y $10'5$ litros y varianzas de 9 y 16 litros^2. Indique cuál de los dos coches tiene mayor consumo relativo dentro de su grupo.

1.14. En los contratos de venta de un fabricante, existe una cláusula por la que acepta la devolución de piezas defectuosas. Se consideran defectuosas aquellas cuya longitud no esté comprendida entre $(\bar{x}-l, \bar{x}+l)$. La longitud media de dichas piezas es de 22 mm. con desviación típica $0'3$. ¿Cuánto debe valer l para que el porcentaje de piezas devueltas no supere el 10 %?

1.15. Una empresa automovilística ha realizado un estudio sobre el grado de satisfacción de sus clientes (X) con la compra de vehículos pertenecientes a los segmentos medio (M) y alto (A), obteniendo los siguientes resultados:

X	n_M	n_A
0-6	4	4
7-13	6	7
14-20	9	9
21-27	12	8
28-34	9	2

a) Se considera que el grado de satisfacción es aceptable si la puntuación obtenida es superior a 19. Calcule el porcentaje de personas en cada grupo con un grado de satisfacción aceptable.

b) ¿Cuál de los dos grupos presenta mayor variabilidad?

1.16. Una población está dividida en dos subpoblaciones A y B, de las cuales se conoce lo siguiente:

$$n_A = 12,\ n_B = 9,\ \sum_A x_i = 234, \sum_B x_i = 138,\ \sum_A x_i^2 = 5036,$$

$$\sum_B x_i^2 = 2586,\ M_{O_A} = 20,\ M_{O_B} = 22,\ M_{e_A} = 19,\ M_{e_B} = 16.$$

a) ¿Se puede calcular la media global del conjunto? En caso afirmativo, calcúlela.

b) ¿Se puede calcular la moda y la mediana global del conjunto? En caso afirmativo, calcúlela.

c) ¿Cuál de las medias de las dos subpoblaciones es más representativa?

1.17. De una distribución se sabe que su media vale 5 y que el momento de orden dos con respecto al origen vale 29. Obtenga una cota inferior del porcentaje de dicha distribución que se encuentra en el intervalo $[2, 8]$.

1.18. Cuantificando a seis individuos en las variables X e Y, se dispone sólo de algunos de estos valores, se muestra a continuación la información disponible:

X	16	6	1		6	6
Y				24		

Complete la tabla sabiendo que la media de Y vale 14, que $\sum_{i=1}^{6} x_i = 48$, y que al tipificar los valores de X se obtiene el mismo resultado que al tipificar los valores de Y.

Capítulo 2

Análisis conjunto de variables

En el capítulo anterior se ha considerado un único carácter, sin embargo, es frecuente estudiar conjuntamente varios caracteres y preguntarse si existe o no algún tipo de relación entre ellos. Este capítulo se dedica al estudio de la relación entre dos caracteres, comenzando con la organización y sintetización de la información, siguiendo un esquema análogo al establecido en el capítulo anterior, para concluir con el estudio de la relación entre ambos. Cuando se analiza la relación entre dos caracteres se pueden presentar dos casos extremos: el primero de ellos será aquel en que conocido el valor de un carácter se pueda obtener el valor del otro, el segundo se presenta cuando la información sobre un carácter no arroja ninguna información sobre el otro. Entre estas situaciones extremas se dan una infinidad de casos intermedios, por ello, el objetivo del capítulo será analizar el nivel de influencia existente entre los caracteres. Hay que indicar, no obstante, que dicho análisis no establecerá cuál es la causa y cuál el efecto entre ambos, sino sólo la intensidad de la relación.

1. Distribución conjunta de dos caracteres

Cuando el investigador está interesado en el estudio de dos caracteres de una población, se obtienen dos observaciones para cada individuo, que se recogen en forma de pares de valores, que deben ser organizados

X, Y	y_1	$\cdots$	y_j	$\cdots$	y_s	
x_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1s}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{is}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_r	n_{r1}	$\cdots$	n_{rj}	$\cdots$	n_{rs}	$n_{r\cdot}$
	$n_{\cdot 1}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot s}$	n

Tabla 2.1: Distribuciones conjuntas y marginales de (X, Y)

en función de la naturaleza de dichos caracteres.

Al igual que en el caso unidimensional es interesante organizar los datos en forma de tabla de frecuencias, sin embargo, al tener que especificar los valores que toman ambos caracteres, la tabla debe ser de doble entrada o bidimensional, véase la tabla 2.1. Supongamos que X toma r valores distintos $x_1, x_2, \ldots, x_r$, e Y toma s valores distintos $y_1, y_2, \ldots, y_s$. Se define la frecuencia absoluta del par (x_i, y_j), que se denota por n_{ij}, como el número de veces que se observa dicho par de valores. Esta distribución se denomina *distribución conjunta* de (X, Y).

Se verifica que $\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n$, valor que aparece recogido en la parte inferior derecha de la tabla. Conservando la notación, f_{ij}, con $f_{ij} = \frac{n_{ij}}{n}$, es la frecuencia relativa del par (x_i, y_j) y por lo tanto $\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = 1$.

Si la distribución es de atributos, la *tabla* se llama *de contingencia* y si es de variables se denomina *de correlación*. Inicialmente, se centra el estudio en el caso en el que los caracteres sean variables, para abordar el estudio de tablas de contingencia en posteriores apartados de este capítulo.

La situación de los valores no nulos en la tabla de doble entrada da una idea intuitiva de la posible relación entre ambos caracteres, así, el que las mayores frecuencias se den alrededor de una diagonal viene a indicar la existencia de relación, mientras que el que no se dé esta circunstancia va a suponer, generalmente, la ausencia de la misma.

2. Distribuciones marginales

En la tabla 2.1, se han sumado las frecuencias que aparecen en cada una de las filas y columnas, colocándose los resultados en los márgenes, donde:

$$n_{\cdot j} = \sum_{i=1}^{r} n_{ij} \ , \quad n_{i\cdot} = \sum_{j=1}^{s} n_{ij} \quad y$$

$$n = n_{\cdot\cdot} = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij},$$

de tal forma que la primera y última columna de la tabla 2.1, constituyen la *distribución marginal de* X, y la primera y última fila la *distribución marginal de* Y. Lógicamente se verifica que:

$$\sum_{i=1}^{r} f_{i\cdot} = 1 \ , \quad \sum_{j=1}^{s} f_{\cdot j} = 1 \quad y$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij} = 1,$$

lo que garantiza la condición de ambas distribuciones.

Se interpreta $f_{i\cdot}$ como la proporción de datos que toman el valor x_i de X, independientemente del valor que tome Y. Una notación análoga se maneja para la variable Y.

Obsérvese que considerar la distribución marginal de una variable equivale a considerar la distribución de ésta independientemente de la otra.

3. Distribuciones condicionadas

Cuando se posee información previa de una de las variables en estudio, ésta puede modificar la información disponible de la otra. En particular, cuando se considera la distribución de una variable para un valor fijo de la otra se obtiene la *distribución condicionada*. Más concretamente, las frecuencias condicionadas son:

$$f_{i_{|j}} = \frac{n_{ij}}{n_{\cdot j}} \quad \forall i = 1, 2, \cdots, r$$

$$f_{_{|i}j} = \frac{n_{ij}}{n_{i\cdot}} \quad \forall j = 1, 2, \cdots, s.$$

Que son, respectivamente, la condicionada de X para el valor y_j de Y, para $j = 1, 2, \cdots, s$, y la condicionada de Y para el valor x_i de X, para $i = 1, 2, \cdots, r$.

Ejemplo 2.1 *Un alumno de Estadística está interesado en estudiar la estatura y el peso del grupo de 21 alumnos varones que pertenecen a su clase. A tal efecto y una vez provisto de los adecuados aparatos de medida, metro y balanza, se dispone a realizar las mediciones. Los resultados que obtuvo se ofrecen en la tabla 2.2. La estatura y el peso se dan con una precisión de 0'01 metros y 1 kilogramo, respectivamente.*

Estatura	Peso	Estatura	Peso	Estatura	Peso
1'78	80	1'83	78	1'73	70
1'72	68	1'80	76	1'66	66
1'88	87	1'84	94	1'80	77
1'81	85	1'77	74	1'75	70
1'73	78	1'77	77	1'69	72
1'89	84	1'82	82	1'71	77
1'80	82	1'71	67	1'68	66

Tabla 2.2: Tabla de datos

El alumno domina el análisis descriptivo univariante y no tiene dificultad en aplicarlo a cada una de las variables que ha considerado, obteniendo así información sobre las medias, dispersiones, simetrías, etc., de la estatura y el peso. Sin embargo, conside-

ra la posibilidad de que entre las dos variables exista algún tipo de relación. A la vista de los datos, y pensando que en otra situación parecida el número de individuos en estudio fuera mucho más grande, decide agrupar a éstos en clases uniformes. Después de analizar la situación establece intervalos de amplitud 0'05 metros y 5 kilogramos, para estatura y peso respectivamente, reorganizando la información obtenida en una tabla de doble entrada. El resultado que obtuvo se recoge en la tabla 2.3.

	Estatura				
Peso	1'65-1'70	1'71-1'75	1'76-1'80	1'81-1'85	1'86-1'90
65-69	2	2			
70-74	1	2	1		
75-79		2	3	1	
80-84			2	1	1
85-89				1	1
90-94				1	

Tabla 2.3: Distribución conjunta

A la vista del resultado obtenido, el investigador observa que ha perdido precisión respecto a los datos originales. Efectivamente, utilizando la tabla de doble entrada lo único que se puede decir, por ejemplo, es que hay dos individuos que midiendo entre 1'71 y 1'75 metros pesan entre 75 y 79 kilogramos, ignorándose las mediciones exactas de éstos. No obstante, entiende que aunque el volumen de datos fuera muy grande la tabla de doble entrada seguiría siendo válida con la adición, tal vez, de algunas clases extremas y, además, piensa que el error que se cometería no sería muy grande si, llegado el caso, se viera en la necesidad de asignar a cada intervalo su marca de clase. Por otra parte, y en el haber de la abstracción realizada, un me-

ro análisis visual hace entender que entre las dos variables existe cierta relación, pues los valores no nulos de la tabla se distribuyen alrededor de una diagonal, obteniéndose un resultado que como era de esperar, y a falta de algún tipo de cuantificación que se realice más adelante, hace corresponder, en general, a los individuos de estatura baja los de poco peso y a los de estatura alta los de mayor peso. No se puede decir que conocida la estatura de un individuo quede determinado su peso pero sí que se puede acotar éste e incluso hacer una previsión aproximada de su valor. Tampoco se puede decir cuál de las dos variables determina los valores de la otra.

A continuación, el alumno piensa que quizás sería interesante ofrecer los valores de la tabla como proporciones del número total de observaciones, para ello divide cada elemento de la tabla por el número de individuos en estudio, obteniendo la tabla 2.4.

	Estatura				
Peso	1'65-1'70	1'71-1'75	1'76-1'80	1'81-1'85	1'86-1'90
65-69	2/21	2/21	0	0	0
70-74	1/21	2/21	1/21	0	0
75-79	0	2/21	3/21	1/21	0
80-84	0	0	2/21	1/21	1/21
85-89	0	0	0	1/21	1/21
90-94	0	0	0	1/21	0

Tabla 2.4: Distribución relativa conjunta

Ahora su interés se centra en conocer la proporción de sus compañeros que pertenecen a una de las clases de estatura independientemente del peso que tengan. Para ello, se da cuenta de que sólo tiene que sumar cada columna, obteniendo, por ejemplo, que

	Estatura					
Peso	1’65-1’70	1’71-1’75	1’76-1’80	1’81-1’85	1’86-1’90	$f(P)$
65-69	2/21	2/21	0	0	0	4/21
70-74	1/21	2/21	1/21	0	0	4/21
75-79	0	2/21	3/21	1/21	0	6/21
80-84	0	0	2/21	1/21	1/21	4/21
85-89	0	0	0	1/21	1/21	2/21
90-94	0	0	0	1/21	0	1/21
$f(E)$	3/21	6/21	6/21	4/21	2/21	1

Tabla 2.5: Distribuciones marginales

hay tres individuos cuya estatura está comprendida entre 1’65 y 1’70 metros, seis entre 1’71 y 1’75 metros, y así sucesivamente. Realizando la misma operación con las filas obtiene los resultados para el peso. Al objeto de organizar esta información decide añadir una fila y una columna en la tabla donde almacena los resultados, véase la tabla 2.5.

Siguiendo con los datos del ejemplo, nuestro investigador se pregunta por la proporción de compañeros que poseen una cierta estatura dentro de los del grupo que pesan entre 75 y 79 Kilogramos, que sabemos constituyen $\frac{6}{21}$ *del total.*

Le resulta fácil comprobar que de entre los que tienen ese peso hay $\frac{2}{6}$ *que tienen una estatura entre 1’71 y 1’75 metros. Observe que podría haber llegado al mismo resultado de haber dividido* $\frac{2}{21}$ *entre* $\frac{6}{21}$*, es decir, la proporción de individuos con altura en la clase* $[1'71, 1'75]$ *y peso en* $[75, 79]$ *entre el correspondiente a la proporción marginal de la variable peso en la clase* $[75, 79]$.

4. Independencia

La independencia-dependencia viene a medir la información que arroja sobre una de las variables el conocimiento que se tiene de la otra variable. Así, una información total implica *dependencia funcional*, la nula información *independencia*, y una información parcial *dependencia estadística*.

Formalmente, se dice que X es independiente de Y si se verifica que:

$$f_{i_{|j}} = f_{i\cdot} \quad \forall i = 1, \cdots, r \quad j = 1, 2, \cdots, s.$$

Es decir, si la frecuencia condicionada coincide con la marginal. De la misma forma se define la independencia de Y respecto de X.

La definición de distribución condicionada da una expresión alternativa para la independencia, y así X e Y son independientes si:

$$f_{ij} = f_{i\cdot} f_{\cdot j} \quad \forall i, j,$$

que además pone de manifiesto que la independencia se establece en un doble sentido; es decir, X es independiente de Y si y sólo si Y lo es de X.

Ejemplo 2.2 *En el ejemplo que se arrastra, nuestro joven estadístico se pregunta por la posibilidad de que exista algún tipo de relación entre las variables, en el sentido de que conocido el valor de una de las variables se pueda decir algo sobre la otra. Él observa que si el peso está comprendido entre los 65 y los 70 kilos la estatura debe estar entre 1'65 y 1'75 metros, y que no hay individuos que en ese rango de pesos mida más de 1'75 metros. Es más, este ejemplo le hace ver que si uno de los cruces de las clases, por ejemplo (x_i, y_j), tiene frecuencia nula, el conocimiento de que una de las variables toma valores en la clase x_i imposibilita que la otra variable tome valores en la clase y_j, y viceversa. Pensando en su problema, llega a la conclusión de* *que existiría una dependencia total o funcional si el conocimiento del valor de una de las variables determina el valor que tomará la otra. Esto implica que si X depende funcionalmente de Y en cada fila hay una sola frecuencia distinta de cero, y si Y depende funcionalmente de X ocurre lo mismo con las columnas.*

Por otra parte, le resulta evidente que las variables son independientes si fijado cualquier valor de una de las variables la otra variable mantiene sus porcentajes iguales a los de su distribución condicionada. Entre estas dos situaciones extremas, descubre que existen muchas posibilidades intermedias.

Por otra parte, se dice que X depende funcionalmente de Y, si conocido el valor que toma Y queda determinado el valor de X.

Para acabar esta sección se comprueba con un contraejemplo que la dependencia funcional no se establece en doble sentido.

Ejemplo 2.3 *En la siguiente distribución:*

X/Y	y_1	y_2	y_3
x_1	12	0	4
x_2	0	7	0

X depende funcionalmente de Y, puesto que conocido el valor de Y queda determinado el de X, pero el recíproco no se da, puesto que si X toma el valor x_1, Y puede tomar el valor y_1 o el y_3.

5. Medidas de dependencia. Coeficientes de relación

Los términos *asociación, correlación, contingencia, concordancia* y otros similares, se suelen utilizar como equivalentes muy a menudo. No obstante, haciendo un uso más correcto de la terminología estadística, aún con significado semejante, se puede considerar:

- correlación de variables propiamente dichas, o sea, medidas en escala de intervalo.
- concordancia de ordenaciones, entendiéndose como tales las denominadas variables ordinales, y
- asociación o contingencia de variables nominales o atributos.

Así, para clasificar los coeficientes que detectan y miden el grado de relación, o dependencia estadística, se ha tenido en cuenta el tipo y la naturaleza de las variables sometidas a estudio.

5.1. Variables continuas. Correlación

5.1.1. Covarianza

Para facilitar el estudio y la notación de la covarianza, se introduce previamente el concepto de *momentos bidimensionales*.

Se define el momento de orden (h, k) respecto al origen como:

$$a_{h,k} = \sum_{i=1}^{r} \sum_{j=1}^{s} x_i^h y_j^k f_{ij}.$$

Es fácil ver que $a_{1,0}$ es la media de X y que $a_{0,1}$ es la media de Y.

Por otro lado, el momento de orden (h, k) respecto a la media viene dado por:

$$m_{h,k} = \sum_{i=1}^{r} \sum_{j=1}^{s} (x_i - \bar{x})^h (y_j - \bar{y})^k f_{ij}.$$

Constatándose que $m_{1,0}$ es cero, al igual que $m_{0,1}$, que $m_{2,0}$ y $m_{0,2}$ son las varianzas de X e Y, respectivamente, y que es posible expresar los momentos respecto a la media en función de los momentos respecto al origen. En particular se da la relación

$$m_{1,1} = a_{1,1} - a_{1,0}a_{0,1} \ .$$

A $m_{1,1}$ se le denomina *covarianza* de la distribución, denotándosele también por S_{xy}. Este coeficiente juega un importante papel en el estudio de la relación lineal entre las variables. Para analizar esta cuestión, se consideran las representaciones gráficas de la figura 2.1 que reflejan distintas situaciones, dichas representaciones reciben el nombre de *nube de puntos* o, también, *diagrama de dispersión.*

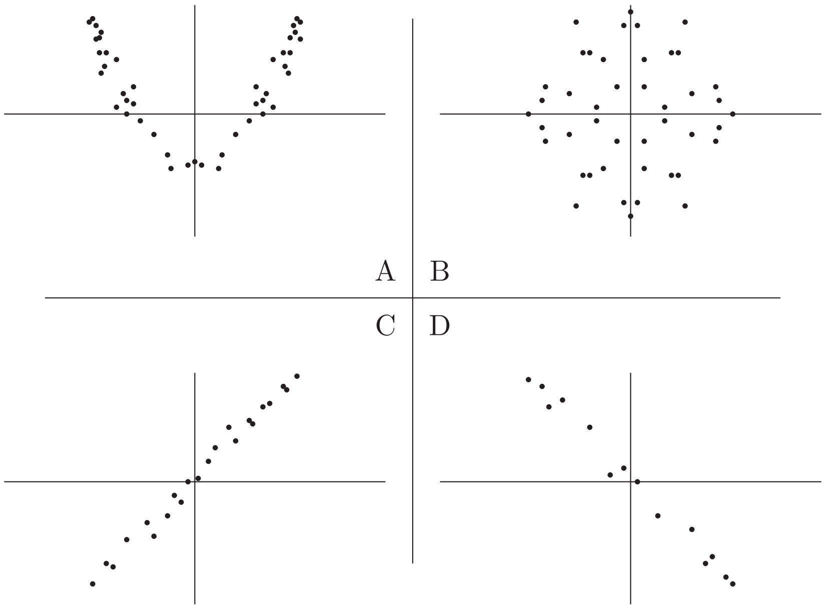


Figura 2.1: Análisis de la covarianza

El punto que viene determinado por la media de X y la media de Y constituye el *centro de gravedad* de las nubes de puntos en todos los casos.

Como se sabe, la covarianza viene dada por la expresión

$$S_{xy} = \sum_{i=1}^{r} \sum_{j=1}^{s} (x_i - \bar{x})(y_j - \bar{y}) f_{ij} \ .$$

S_{xy} es una medida simétrica y se puede leer como la suma de los productos de las desviaciones de X por las desviaciones de Y con respecto a sus medias respectivas; de tal forma, que si el signo de la desviación de X coincide con la de Y, como ocurre en el primer y tercer cuadrante, se genera un sumando positivo; y cuando el signo es distinto -segundo y cuarto cuadrante- la aportación a la covarianza es negativa. Por tanto, la concentración de valores en los distintos cuadrantes determina el signo y la cuantía de S_{xy}. Así, en los casos A y B de la figura 2.1, S_{xy} se aproxima a cero, en el caso C va a ser alta y positiva, y en el D alta y negativa. Por tanto, se está en condiciones de afirmar que la covarianza detecta la relación lineal entre las variables y el sentido de ésta, pero no distingue entre la no presencia de relación, caso B, y la existencia de alguna dependencia no lineal, caso A. De todas formas, aún para el estudio de relaciones lineales la covarianza adolece de ciertos problemas, como el de venir acompañada de las unidades de las variables y el de depender del número de observaciones.

5.1.2. Coeficiente de correlación de Pearson

Para obviar las carencias de la covarianza se introduce el *coeficiente de correlación lineal* o *coeficiente de correlación de Pearson*

$$r = \frac{S_{xy}}{S_x S_y} \ ,$$

que es una medida adimensional, ordinal, toma valores en el intervalo $[-1, 1]$ y tiene el signo de S_{xy}, por lo que cuando la relación lineal entre X e Y es exacta y directa, es decir, todos los puntos se encuentran sobre una recta con pendiente positiva, vale 1, cuando es exacta e inversa, es decir, todos los puntos se encuentran sobre una recta con pendiente negativa, vale -1 y cuando no hay relación lineal 0; con un análisis lógico para las posiciones intermedias. Cuando r vale cero, se dice que las *variables* están *incorreladas.*

En el caso lineal, al cuadrado de r se le llama *coeficiente de determinación* y se le denota por R^2, representando una medida cardinal o cuantitativa para medir la relación lineal entre las variables. Se estudia este coeficiente con más detalle en el capítulo siguiente.

Se concluye este apartado indicando que la independencia implica incorrelación, pero el recíproco no siempre es cierto. Este resultado es consecuencia de que la independencia supone la descomposición de los momentos de orden (h, k) (respecto al origen o respecto a la media) en el producto de los momentos $(h, 0)$ y $(0, k)$; así, $a_{1,1} = a_{1,0}a_{0,1}$ y por tanto $S_{xy} = m_{1,1} = a_{1,0}a_{0,1} - a_{1,0}a_{0,1} = 0$, con lo que $r = 0$ y las variables están incorreladas. En sentido contrario, la incorrelación sólo implica esa descomposición para el momento $(1, 1)$. En cierta forma, se puede decir que la incorrelación es una independencia de primer orden o lineal.

Ejercicio 2.1 *Demuestre que las variables X e Y de la siguiente distribución:*

X	Y
2	8
1	5
0	4
-1	5
-2	8

están incorreladas, pero no son independientes; es más, existe una relación funcional entre ellas. Indíquela.

Por tanto, el coeficiente de correlación de Pearson mide el grado de relación lineal entre dos variables cuantitativas indicando el sentido directo o inverso de la relación. Es el más común de todos los coeficientes porque es la base de otras muchas medidas de relación entre variables de distinta naturaleza, de hecho, a menudo se tiende a interpretar cualquier coeficiente como si del de Pearson se tratase.

5.1.3. Coeficiente de correlación biserial

Se utiliza para establecer el grado de correlación entre dos variables cuantitativas cuando una de ellas ha sido dicotomizada previamente. Se trata de una modificación del coeficiente de correlación de Pearson entre una variable continua X y otra Y que se ha dicotomizado y que en origen responde a una estructura de distribución normal[1].

[1] La distribución normal se estudiará en el capítulo 5

El coeficiente de correlación biserial se denota por r_b y se puede calcular indistintamente por cualquiera de las siguientes expresiones:

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{S_x}(\frac{pq}{y}) = \frac{\bar{X}_p - \bar{X}}{S_x}(\frac{p}{y}) ,$$

donde:

X es la variable continua

Y es la variable dicotomizada

$\bar{X}_p$ es la media de X cuando Y vale 0

$\bar{X}_q$ es la media de X cuando Y vale 1

$\bar{X}$ es la media de la distribución marginal de X

S_x es la desviación típica de la marginal de X

p es la proporción de elementos con asignación 0 en la variable Y

q es la proporción de elementos con asignación 1 en la variable Y, ($q = 1 - p$)

y es el valor de la ordenada correspondiente a un valor de x que divide el área de la distribución normal tipificada en dos partes, una igual a p y otra igual a q.

Se interpreta de forma análoga al coeficiente de correlación de Pearson en lo referente a la intensidad de la relación, no a su sentido; además, cuando la correlación es alta y el requisito de normalidad de Y no se cumple de forma estricta, el coeficiente de correlación biserial puede valer más de 1 o menos de -1.

Como variante, aunque con idéntica interpretación y similar notación y expresión, se debe tener presente el *coeficiente de correlación biserial–puntual*, que se utiliza para medir la correlación entre una variable continua y otra dicotómica por naturaleza, definido por:

$$r_{bp} = \frac{\bar{X}_p - \bar{X}_q}{S_x}\sqrt{pq} = \frac{\bar{X}_p - \bar{X}}{S_x}\sqrt{\frac{p}{q}} .$$

Observación 2.1 *Desde el punto de vista práctico, el coeficiente de correlación biserial se usa sobre todo para hacer inferencias. Su cálculo necesita conocer la distribución normal, puesto que es necesario obtener el valor y.*

Ejemplo 2.4 *Con la finalidad de buscar el mayor rendimiento de la tierra, un agricultor, preocupado por su cosecha de naranjas, está interesado en estudiar el grado de relación entre la cantidad de fruta recogida y la lluvia caída en los últimos 10 años. Para ello parte de la siguiente información, obtenida por él mismo, en la que ha clasificado los años en secos (S) o lluviosos (L):*

Naranjas (Tm)	Año	Naranjas (Tm)	Año
10'01	L	9'57	L
8'2	L	5'9	S
7'23	S	6'8	S
11'45	L	6'8	S
8'50	L	7'9	L

Para estudiar a partir de estos datos la relación entre las variables, se recurre al coeficiente de correlación biserial-puntual[2]*, realizando la división de la cosecha en dos series, la obtenida en temporada de sequía, con valor asignado 1, y la obtenida en temporada de lluvia, con asignación el valor 0. Se denota por X la cantidad de naranjas y por Y si la temporada es de lluvia o de sequía.*

$$r_{bp} = \frac{9'27167 - 6'6825}{1'70077}\sqrt{0'6 \cdot 0'4} = 0'7457.$$

Lo que indica una relación de dependencia relativamente fuerte entre las variables.

[2]Se ha utilizado el coeficiente de correlación biserial-puntual y no el coeficiente de correlación biserial, debido a que aunque la variable "lluvia caída" es en principio continua y probablemente Normal, el uso del coeficiente de correlación biserial requiere conocimientos hasta ahora no adquiridos, como se indica en la observación 2.1.

Dada la inseguridad ante las medidas de la concentración de lluvia anual por metro cuadrado que obtuvo el agricultor, éste decide prescindir por completo de sus datos y recurrir a la información que sobre el tema proporciona anualmente el instituto meteorológico, el cuál le proporciona la cantidad de lluvia caída cada año. Se denota por X la cantidad de naranjas y por Y los m^3 de lluvia.

De esta forma los datos han sido transformados en:

Nar. (Tm)	Lluvia (m^3)	Nar. (Tm)	Lluvia (m^3)
10'01	1'3	9'57	1'4
8'2	0'9	5'9	0'67
7'23	0'87	6'8	0'56
11'45	1'75	6'8	0'87
8'50	0'96	7'9	1'24

Con esta información se analiza la relación de las variables con el coeficiente de correlación de Pearson, ya que ambas son continuas.

$$\begin{aligned} r &= \frac{S_{xy}}{S_x S_y} = 0'917511 \\ R^2 &= 0'841827. \end{aligned}$$

Con esto se concluye que existe una fuerte dependencia lineal y además directa entre ambas variables, es decir, la cosecha de naranjas es mayor cuando mayor es la cantidad de lluvia caída.

5.2. Variables ordinales. Concordancia

5.2.1. Coeficiente de correlación por rangos de Spearman

Este coeficiente se utiliza para medir la relación entre dos sucesiones de valores ordinales. Es el coeficiente de correlación de Pearson para las llamadas variables cuasi–cuantitativas, discretas, o bien, para aquellas cuantitativas que han sido transformadas en ordinales (n primeros números naturales para cada variable) tiene la forma

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

donde:

r_s es el coeficiente de correlación por rangos de Spearman

d_i es la diferencia entre el valor ordinal de la variable X y el de la variable Y en el elemento i-ésimo

n es el tamaño de la muestra

Se verifica que $-1 \leq r_s \leq 1$.

Si hay un gran número de elementos con el mismo valor en alguna de las dos variables, es decir, si hay muchos empates, es conveniente recurrir a las correcciones de este coeficiente. Quedando el coeficiente como

$$r_s = \frac{x^2 + y^2 - \sum_{i=1}^{n} {d_i}^2}{2\sqrt{x^2y^2}} \ ,$$

con:

$$x^2 = \frac{n^3-3}{12} - \sum_{i=1}^{n} T_{x_i} \ , \qquad T_{x_i} = \frac{t_{x_i}^3 - t_{x_i}}{12} \ ,$$

$$y^2 = \frac{n^3-3}{12} - \sum_{i=1}^{n} T_{y_i} \ , \qquad T_{y_i} = \frac{t_{y_i}^3 - t_{y_i}}{12} \ ,$$

donde:

t_{x_i} es el número de empates en el rango i de la variable X

t_{y_i} es el número de empates en el rango i de la variable Y

Sus características e interpretación son similares a las del coeficiente de correlación de Pearson.

5.2.2. Coeficiente τ de Kendall

De forma análoga al coeficiente de Spearman, el coeficiente τ considera el orden de los n objetos o elementos tanto de una variable como de la otra e intenta medir el grado de concordancia o correspondencia entre ellos. Dicho coeficiente viene dado por

$$\tau = \frac{P-Q}{P+Q} \ ,$$

donde:

τ es el coeficiente de Kendall

P el número de coincidencias o acuerdos

Q el número de no coincidencias o desacuerdos

Nuevamente, si hay gran número de empates, conviene aplicar una corrección, quedando el coeficiente como

$$\tau = \frac{P-Q}{\sqrt{\frac{1}{2}n(n-1)-T_x}\sqrt{\frac{1}{2}n(n-1)-T_y}} \ ,$$

con:

$$\begin{aligned} T_x &= \frac{1}{2}\sum_{i=1}^{n} t_{x_i}(t_{x_i}-1) \ , \\ T_y &= \frac{1}{2}\sum_{i=1}^{n} t_{y_i}(t_{y_i}-1) \ , \end{aligned}$$

donde t_{x_i} y t_{y_i} coinciden con los definidos para el coeficiente de correlación de Spearman.

Sus características e interpretación son similares a las del coeficiente de correlación de Pearson.

X	rg(X)	Y	rg(Y)	d_i	${d_i}^2$
5	1	1	3	-2	4
6	2	3	$9'5$	$-7'5$	$56'25$
7	3	2	$6'5$	$-3'5$	$12'25$
8	4	1	3	1	1
9	5	1	3	2	4
10	6	0	1	5	25
11	7	2	$6'5$	$0'5$	$0'25$
12	8	2	$6'5$	$1'5$	$2'25$
13	9	3	$9'5$	$-0'5$	$0'25$
14	10	2	$6'5$	$3'5$	$12'25$
					$117'5$

Tabla 2.6: Cálculo del coeficiente de correlación de Spearman

5.2.3. Coeficiente γ de Goodman–Kruskal

Se utiliza para medir el grado de concordancia entre dos variables ordinales, estando especialmente indicado cuando hay muchas observaciones y pocos valores posibles, es decir, muchos empates.

Su expresión e interpretación es muy similar a la del coeficiente de Kendall, considerando la proporción de pares semejantes y la proporción de pares no semejantes entre los empatados, resultando

$$\gamma = \frac{n_s - n_d}{n_s + n_d}$$

donde:

γ es el coeficiente de Goodman-Kruskal

n_s es el números de pares semejantes o no invertidos

n_d es el número de no semejantes o invertidos

Ejemplo 2.5 *Se pretende estudiar la relación existente entre la edad (E) y el número de hermanos (H) de un grupo de 10 chicos, para ello se cuenta con los siguientes* *datos:*

E	6	12	8	11	10	7	9	14	13	5
H	3	2	1	2	0	2	1	2	3	1

Se calculará el coeficiente de correlación por rangos de Spearman, dado que se están tratando variables cuantitativas. Obtendremos primero la versión original de dicho coeficiente.

A partir de los cálculos recogidos en la tabla 2.6, se obtiene

$$\begin{aligned} r_s &= 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 117'5}{10 \cdot 99} \\ &= 1 - \frac{705}{990} = 1 - 0'7121 = 0'2879. \end{aligned}$$

No obstante, debido al elevado número de empates debería emplearse el coeficiente modificado para dicho caso, o incluso al coeficiente de Goodman-Kruskal. Se calculará el coeficiente modificado de Spearman.

$$\begin{aligned} x^2 &= \frac{10^3-3}{12} - 0 = \frac{997}{12} \\ y^2 &= \frac{10^3-3}{12} - \left(\frac{24}{12} + \frac{62}{12} + \frac{6}{12}\right) = \frac{905}{12}. \end{aligned}$$

Por tanto el coeficiente modificado queda como

$$r_s = \frac{\frac{997+905}{12} - 117'5}{2\sqrt{\frac{997 \cdot 905}{12^2}}} = 0'2589$$

que es ligeramente inferior al original. Del resultado obtenido se concluye la escasa concordancia entre la edad y el número de hermanos.

5.3. Atributos. Contingencia

5.3.1. Coeficiente χ^2

El coeficiente χ^2 se utiliza para medir el grado de asociación entre dos variables cualitativas con h y k categorías respectivamente. Este estadístico está basado en la comparación de las frecuencias observadas con las esperadas bajo una cierta hipótesis, generalmente de independencia, respondiendo a la expresión

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \ ,$$

donde:

o_{ij} son las frecuencias observadas o empíricas

e_{ij} son las frecuencias esperadas o teóricas

Cuando h y k toman el valor 2, es decir, cuando se está trabajando con una tabla de contingencia 2×2, se aplica la denominada corrección de Yates, resultando el coeficiente:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|o_{ij} - e_{ij}| - 0'5)^2}{e_{ij}}.$$

El coeficiente siempre toma valores no negativos, pero al tratarse de una medida no acotada, es de difícil interpretación por sí sola, si bien, cuanto más relacionadas estén las variables sometidas a estudio más se alejará el coeficiente del valor 0. Su valor depende del número de observaciones y de las categorías en que éstas se dividen, por tanto el coeficiente χ^2 y sus derivados no son comparables con cualquier otro coeficiente obtenido con distinto número de categorías.

Este coeficiente χ^2 es la base de otros obtenidos a partir de él y que solucionan el problema de su falta de acotación.

5.3.2. Coeficiente de contingencia

Es uno de los coeficientes derivados del χ^2, resultando útil bajo las mismas condiciones que aquel pero con mayores posibilidades de interpretación. Se denota por C y se define como

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

siendo n el tamaño muestral.

Se cumple que $0 \leq C \leq 1$ y mide la intensidad de la relación sin indicar su sentido.

5.3.3. Coeficiente de Cramer

Es otro de los coeficientes derivados del χ^2. Se caracteriza por V y su expresión es

$$V = \sqrt{\frac{\chi^2}{n(m-1)}}$$

siendo:

n el tamaño muestral

m el mínimo entre h y k

h el número de categorías de la variable X

k el número de categorías de la variable Y

Se verifica que $0 \leq V \leq 1$ y se interpreta igual que el coeficiente de contingencia, teniendo en cuenta que sólo proporciona información sobre la relación entre las variables y no sobre el sentido de la misma.

5.3.4. Coeficiente φ

Se trata de un coeficiente especialmente indicado para medir la asociación entre dos variables dicotómicas. Su expresión es

$$\varphi = \frac{n_{11}n_{22} - n_{21}n_{12}}{\sqrt{n_{1}.n_{2}.n_{.1}n_{.2}}}$$

donde:

n_{11} es el número de veces que se da el par $(X = 0, Y = 0)$

n_{12} es el número de veces que se da el par $(X = 0, Y = 1)$

n_{21} es el número de veces que se da el par $(X = 1, Y = 0)$

n_{22} es el número de veces que se da el par $(X = 1, Y = 1)$

En cuanto a su interpretación, el coeficiente toma valores en el intervalo $[-1, 1]$, midiendo de forma similar al coeficiente de Pearson la intensidad de la asociación entre las dos variables; salvo que alguna de las frecuencias n_{ij} sea nula, en cuyo caso el coeficiente vale 1 ó -1.

En el caso en que se estudie el grado de correlación entre dos variables cuantitativas dicotomizadas, X e Y, siempre y cuando éstas respondan a variables continuas bajo una ley normal (que se estudiará más adelante), el coeficiente φ suele denominarse *coeficiente de correlación tetracórica.*

Ejemplo 2.6 *De cara a la planificación del próximo curso sería conveniente analizar la relación entre el nivel de estudios del padre y la orientación del alumno hacia las ciencias. Se cuenta para ello con la información obtenida en el centro*

	Estudios padre			
Orientación	*Nulo*	*Básico*	*Medio*	*Superior*
Orientado	*23*	*12*	*34*	*32*
No orientado	*18*	*42*	*16*	*27*

Como se trata de una tabla de contingencia, se calcula el coeficiente χ^2 y sus derivados para hacer posible la interpretación.

	Nulo	*Básico*	*Medio*	*Superior*	
Orientado	*23*	*12*	*34*	*32*	*101*
No orientado	*18*	*42*	*16*	*27*	*103*
	41	*54*	*50*	*59*	*204*

e_{ij}	1	2	3	4
1	$\frac{101 \cdot 41}{204}$	$\frac{101 \cdot 54}{204}$	$\frac{101 \cdot 50}{204}$	$\frac{101 \cdot 59}{204}$
2	$\frac{103 \cdot 41}{204}$	$\frac{103 \cdot 54}{204}$	$\frac{103 \cdot 50}{204}$	$\frac{103 \cdot 59}{204}$

e_{ij}	1	2	3	4
1	$20'30$	$26'73$	$24'75$	$29'21$
2	$20'70$	$27'26$	$25'24$	$29'79$

$$\begin{aligned}
\chi^2 &= \frac{(23-20'30)^2}{20'30} + \frac{(12-26'73)^2}{26'73} \\
&\quad + \frac{(34-24'75)^2}{24'75} + \frac{(32-29'21)^2}{29'21} \\
&\quad + \frac{(18-20'70)^2}{20'70} + \frac{(42-27'26)^2}{27'26} \\
&\quad + \frac{(16-25'24)^2}{25'24} + \frac{(27-29'79)^2}{29'79} \\
&= 0'36 + 8'12 + 3'46 + 0'26 \\
&\quad + 0'35 + 7'97 + 3'38 + 0'26 \\
&= 24'16
\end{aligned}$$

$$C = \sqrt{\frac{24'16}{24'16+204}} = 0,3254$$

$$V = \sqrt{\frac{24'16}{204 \cdot 1}} = 0,3441.$$

Luego podemos concluir que el grado de asociación entre las variables es pequeña.

Ejemplo 2.7 *En el conservatorio de música de una ciudad se pretende estudiar la relación existente entre el sexo del alumnado y su afición por los instrumentos de viento. Para ello, controlados los 482 estudiantes se tiene:*

	Hombre	*Mujer*
Aficionado	*150*	*97*
No aficionado	*123*	*112*

Dada la naturaleza dicotómica de las variables, se recurre al coeficiente φ

$$\varphi = \frac{150 \cdot 112 - 123 \cdot 97}{\sqrt{247 \cdot 235 \cdot 273 \cdot 209}} = \frac{4869}{57548'8} = 0'08.$$

Con esto se pone de manifiesto la inexistencia de relación entre el sexo y la preferencia por los instrumentos de viento.

Ejemplo 2.8 *Volviendo al ejemplo planteado en el estudio de variables continuas, véase ejemplo 2.4, y considerando un caso aún más general, se supone que la información que conservó el agricultor después de la cosecha de cada año es tan sólo el recuerdo de si fue buena o mala. Así los datos con los que se cuenta para el estudio de las variables son:*

	Mala	*Buena*
Seco	*0*	*4*
Lluvioso	*5*	*1*

Haciendo uso ahora, dado que las variables aparecen dicotomizadas, del coeficiente de correlación tetracórica

$$r_t = \frac{5 \cdot 4 - 0 \cdot 1}{\sqrt{6 \cdot 4 \cdot 5 \cdot 5}} = \frac{20}{24'4948} = 0'8165.$$

> *Poniendo nuevamente de manifiesto la relación entre la cantidad de naranjas y la lluvia. Hay que tener en cuenta que el signo que acompaña al coeficiente depende de la asignación de valores a la hora de dicotomizar las variables, por consiguiente, es interpretable la intensidad de la relación, no el sentido de la misma.*

Son varios los coeficientes de relación que a lo largo de esta sección se han ido enumerando, coincidiendo con los que por sus características, naturaleza y facilidad de cálculo son más utilizados y, por consiguiente, conocidos en los distintos campos donde su aplicación tiene cabida.

6. Ejercicios

6.1. Ejercicio resuelto

2.1 Se ha clasificado el peso de los huevos, Y, de un cierto tipo de pez en función del peso de la madre, X, obteniéndose los resultados de la tabla adjunta.

X \ Y	[25,27)	[27,29)	[29,31)	[31,33)
[500,550)	15	11	18	0
[550,600)	12	14	0	12
[600, 650)	0	3	7	18

Calcule:

a) La distribución del peso del huevo.

b) La distribución del peso de la madre cuando el huevo tiene su peso comprendido entre $[25, 27)$.

c) La media, la mediana y la moda del peso de los huevos.

d) El nivel de representatividad de la media del peso de la madre cuando el huevo está comprendido entre $[25, 27)$.

e) Estudiar si las variables son independientes.

f) El grado de dependencia lineal entre estas variables.

Solución:

a) En realidad el primer apartado lo que está pidiendo es la distribución marginal de la variable Y. Por tanto,

$$\begin{aligned} n_{[25,27)} &= n(y \in [25,27)) = n(x \in [500,550), y \in [25,27)) + \\ &\quad n(x \in [550,600), y \in [25,27)) + n(x \in [600,650), y \in [25,27)) \\ &= 15 + 12 + 0 = 27 \end{aligned}$$

procediendo de igual forma con el resto de intervalos donde Y toma valores, se obtiene que:

Y	n_i
[25,27)	27
[27,29)	28
[29,31)	25
[31,33)	30

b) Se pide la distribución de la variable X condicionada a que la variable Y tome valores en el intervalo $[25,27)$, es decir,

$$\begin{aligned} f_{|[25,27)}[500,550) &= f(x \in [500,550)/y \in [25,27)) \\ &= \frac{f(x \in [500,550), y \in [25,27))}{f(y \in [25,27))} \\ &= \frac{\frac{15}{110}}{\frac{27}{110}} = \frac{15}{27} = \frac{5}{9} \end{aligned}$$

procediendo de igual forma, se tiene:

$X/Y \in [25,27)$	f_i
$[500,550)$	$\frac{5}{9}$
$[550,600)$	$\frac{4}{9}$
$[600,650)$	0

c) Se calcula la media de variable Y,

$$\bar{y} = \frac{26 \cdot 27 + 28 \cdot 28 + 30 \cdot 25 + 32 \cdot 30}{110} = 29'05.$$

Para calcular la mediana se tiene en cuenta el apartado *a*), donde se ve que el primer intervalo cuya frecuencia absoluta acumulada supera el 50 % de los datos, es decir, 55, es el intervalo [29,31). Por tanto la mediana viene dada por

$$Me = 29 + \frac{55 - 55}{25} 2 = 29.$$

Para calcular la moda, se observa que todos los intervalos tienen igual amplitud y que el intervalo con mayor frecuencia es el [31,33), por tanto la moda es

$$Mo = 31 + \frac{0}{25 + 0} 2 = 31.$$

d) Para calcular el nivel de representatividad de la media se utiliza el coeficiente de variación, para ello, se calcula previamente la media y la desviación típica de la variable requerida. La distribución de esta variable se ha calculado en el apartado *b*), por tanto

$$\bar{x}/y \in [25, 27) = 525\frac{5}{9} + 575\frac{4}{9} = \frac{4925}{9} \qquad y$$

$$S^2_{x/y\in[25,27)} = 525^2\frac{5}{9} + 575^2\frac{4}{9} - \left(\frac{4925}{9}\right)^2 = \frac{50000}{81}$$

Con lo que el coeficiente de variación es

$$CV = \frac{\sqrt{\frac{50000}{81}}}{\frac{4925}{9}} = 0'045.$$

lo que supone que la media es muy representativa debido a que es muy pequeño el coeficiente de variación.

e) Para tratar la independencia se considera un par $(x, y) \in [500, 550) \times [25, 27)$. Se sabe que

$$f(x \in [500, 550), y \in [25, 27)) = \frac{15}{110},$$

además, se tiene que

$$f(x \in [500, 550)) = \frac{44}{110}$$

y que

$$f(y \in [25, 27)) = \frac{27}{110}$$

con lo cual,

$$f(x \in [500, 550), y \in [25, 27)) = \frac{15}{110} \neq \frac{1188}{12100} =$$

$$= f(x \in [500, 550))f(y \in [25, 27)).$$

Por tanto, se tiene que las variables no son independientes.

f) Para cuantificar el grado de dependencia lineal entre dos variables se calcula el coeficiente de determinación

$$R^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2}.$$

Se necesita calcular S_X^2, S_Y^2 y S_{XY}:

$$S_X^2 = \frac{525^2 \cdot 44 + 575^2 \cdot 38 + 625^2 \cdot 28}{110} - \left(\frac{62450}{110}\right)^2 = 1583'47$$

$$S_Y^2 = \frac{26^2 \cdot 27 + 28^2 \cdot 28 + 30^2 \cdot 25 + 32^2 \cdot 30}{110} - \left(\frac{3196}{110}\right)^2 = 5'14$$

$$\begin{aligned} S_{XY} &= \frac{1}{110}(26 \cdot 525 \cdot 15 + 26 \cdot 575 \cdot 12 + 26 \cdot 625 \cdot 0 + 28 \cdot 525 \cdot 11 \\ &+28 \cdot 575 \cdot 14 + 28 \cdot 625 \cdot 3 + 30 \cdot 525 \cdot 18 + 30 \cdot 575 \cdot 0 \\ &+30 \cdot 625 \cdot 7 + 32 \cdot 525 \cdot 0 + 32 \cdot 575 \cdot 12 + 32 \cdot 625 \cdot 18) \\ &- \frac{199590200}{12100} = 44'03. \end{aligned}$$

Con lo cual

$$R^2 = \frac{44'03^2}{5'14 \cdot 1583'47} = 0'24,$$

de donde se deduce que el grado de dependencia lineal es bastante bajo.

6.2. Ejercicios propuestos

2.1. Durante el año 1993 se han observado la población y el número de viviendas de renta libre unifamiliares en 32 municipios de la provincia de Cádiz. Los datos obtenidos se han tabulado, obteniéndose:

Y \ X	0-10	10-30	30-70	70-150	150-250
[0-2)	3				
[2-5)	3				
[5-10)	1	3	1	1	
[10-30)	2	2	6	1	
[30-80)			2	1	2
[80-180)		1	1	1	1

donde:
X = Número de viviendas
Y = Población en miles de personas

a) Obtenga las distribuciones marginales de X e Y.

b) Indique qué distribución es más homogénea.

c) Obtenga la distribución de las viviendas unifamiliares para los municipios entre dos mil y treinta mil habitantes.

d) Calcule los momentos: a_{01}, a_{02}, a_{10}, a_{11}, m_{02}, m_{20}, m_{21}.

e) Entre las poblaciones de más de 10.000 habitantes, indique cuál es el número de viviendas libres construidas más frecuente.

f) Obtenga la covarianza y el coeficiente de correlación de las variables X e Y e interprételo.

2.2. De la variable bidimensional (X, Y) se conoce su coeficiente de correlación, $r = 0'83$, y sus varianzas, $S_x^2 = 5'32$ y $S_y^2 = 8'41$. Si se multiplican por 3 los valores de X y por 2 los valores de Y, ¿que repercusión tienen estas transformaciones en la covarianza y en el coeficiente de correlación?

2.3. La tabla 2.7 muestra una serie histórica sobre el Olivar Español que recoge la superficie, rendimiento y producción, durante el periodo 1965-1979. donde:

Año	X	Y	Z
1965	73'6	69'8	8'5
1966	98'1	62'5	6
1967	99'8	98'5	8'7
1968	107'7	102'5	6
1969	107'7	97'4	3'7
1970	122	113'8	8'9
1971	127	118	7'9
1972	138'1	128'1	10'1
1973	152'1	145'8	6'8
1974	144'8	139'8	5
1975	160'7	152'9	11'1
1976	150'2	143'4	9'8
1977	152'1	146	9'5
1978	167'3	162'1	10'8
1979	165	160'2	10

Tabla 2.7: Datos ejercicio 2.3

X = Superficie en miles de Ha.
Y = Rendimiento en Qm/Ha..
Z = Producción en miles de Tm..

Se pide:

a) El diagrama de dispersión de las variables X e Y.

b) Las medidas más representativas para cada una de las variables, indicando su representatividad

c) El estudio de la relación entre las variables XY, XZ e YZ.

2.4. La siguiente tabla muestra la relación existente entre la lluvia caída, en l/m^2, en el periodo octubre–mayo y la producción obtenida en kilogramos por olivo.

X	300	400	500	600	700
Y	13	26	40	57	64
Y	24	21	31	45	69
Y	17	17	38	51	57
Y	11	26	34	58	76
Y	20	30	27	44	74

donde X representa la lluvia caída e Y la producción obtenida en kilogramos por olivo.

a) Represente el diagrama de dispersión.

b) Indique si existe alguna tendencia.

c) Cuantifique y comente la relación existente entre las dos variables.

2.5. Dada la siguiente tabla de doble entrada con valores porcentuales:

Y X	2	3	4
0	$0'22$	$0'13$	$0'04$
1	$0'16$	$0'11$	$0'05$
2	$0'08$	$0'16$	$0'05$

a) Obtenga la distribución marginal de X. Calcule su media, moda y mediana.

b) Calcule la media de Y cuando X toma el valor 3.

c) Estudie la dependencia de las variables X e Y.

2.6. Estudiar la coherencia de los siguientes resultados correspondientes a una variable bidimensional:

$$S_{xy} = -179'5,\ S_x^2 = 36'8,\ S_y^2 = 525,\ M_e(X) = -12'3,\ \bar{Y} = 0$$

2.7. De los modelos de una determinada marca de automóviles se considera el consumo medio y el tiempo de aceleración de 0 a 100

Km./h., obteniéndose los siguientes resultados:

	Cons. (lit.)				
Acel. (seg.)	[5, 6)	[6, 7)	[7, 8)	[8, 9)	[9,12)
[7, 9)			1	3	2
[9,11)		1	2	4	4
[11,14)	1	5	1	3	3
[14,18)	3	2	1		

a) Dibuje y comente el diagrama de dispersión.
b) Obtenga el consumo medio de carburante.
c) Obtenga el tiempo de aceleración medio.
d) Indique cuál de las dos medias es más representativa.
e) Estudie la relación existente entre las dos características.

2.8. A un grupo de estudiantes se les preguntó por el tiempo que tardan en llegar desde su hogar hasta la Facultad, X (minutos), el tiempo que le dedican diariamente al estudio, Y (horas), y las calificaciones obtenidas en la asignatura de Estadística, Z, obteniéndose las siguientes respuestas:

(40, 4, 4), (45, 3, 3), (30, 4, 5), (40, 4, 5), (80, 2, 5), (20, 3, 5)
(10, 1'5, 6), (10, 4, 6), (20, 4, 6), (45, 3, 3), (20, 4, 4), (30, 4, 7)
(30, 3, 7), (20, 4, 6), (30, 1, 6), (10, 5, 5), (15, 5, 5), (20, 6, 5)
(20, 3, 7), (20, 4, 5), (20, 5, 6), (60, 2, 3), (60, 5, 5)

a) Obtenga el diagrama de dispersión correspondiente al tiempo dedicado al estudio y las calificaciones obtenidas en Estadística.
b) ¿Se aprecia alguna tendencia?
c) Estudie las relaciones exitentes entre XY, XZ e YZ.

2.9. Al mismo grupo del ejercicio anterior se le ha pedido que escriba un dígito al azar entre 0 y 9 así como el número de hermanos que tiene, obteniéndose los siguientes pares de valores:

(7, 4), (0, 1), (2, 1), (2, 0), (9, 4), (7, 4), (6, 3), (8, 5)
(7, 3), (3, 2), (7, 3), (2, 1), (7, 4),(7, 3), (8, 4), (8, 5)
(5, 3), (3, 1), (4, 2), (4, 2), (5, 3), (2, 0), (4, 2)

¿Existe alguna relación entre las variables?, ¿de qué tipo?

2.10. Sea la variable bidimensional (X, Y) de la que se han obtenido 25 pares de valores, con los siguientes resultados:

$$r = 0'65, \qquad \sum_{i=1}^{25} x_i = 238, \qquad \sum_{i=1}^{25} y_i = 138$$

$$\sum_{i=1}^{25} x_i^2 = 12678, \qquad \sum_{i=1}^{25} y_i^2 = 2732$$

a) Calcule medias, varianzas y covarianza de X e Y.
b) Indique qué variable es más homogénea.

2.11. En cada uno de los estanques, A y B, se tienen 100 ejemplares de una variedad de dorada todas ellas afectadas por un parásito. La alimentación es idéntica en ambos estanques salvo en un producto encaminado a eliminar dichos parásitos, suministrado únicamente a los del estanque A. Posteriormente, se encuentra que en 71 ejemplares del A y en 58 del B han desaparecido los parásitos. Halle el coeficiente de contingencia y el coeficiente de Cramer e interprete los resultados.

2.12. Demuestre que el coeficiente de Cramer está comprendido entre 0 y 1.

2.13. Demuestre que el valor máximo del coeficiente de contingencia de una tabla $k \times k$ es $\sqrt{\frac{(k-1)}{k}}$.

2.14. Antes de un campeonato de fútbol las apuestas indican que las posiciones que ocuparán al finalizar éste cinco de los equipos participantes es $A > B > C > D > E$. Un jugador apuesta que el orden final será $A > D > B > E > C$. Mida el grado de similitud entre ambas ordenaciones.

2.15. Se mide el tiempo que 10 estudiantes tardan en realizar dos experimentos en los que predominan el cálculo mental y la capacidad espacial, respectivamente. Si los valores obtenidos son:

Estudiante	1	2	3	4	5	6	7	8	9	10
Tarea 1	32	37	45	50	48	56	78	69	77	79
Tarea 2	41	33	46	47	40	71	70	65	75	83

Estudie la relación entre los resultados obtenidos en ambas tareas.

2.16. Dos grupos de estudiantes deciden clasificar a 11 profesores. Los resultados se muestran a continuación:

Prf.	Es.	Og.	Mt.	Ig.	Fs.	Ge.	Cá.	FQ.	Oc.	Bi.	In.
Gr.I	7	4	2	8	9	10	11	6	1	3	5
Gr.II	8	2	1	5	3	11	9	10	7	4	6

Compare ambas clasificaciones.

2.17. En un grupo de 100 personas se estudian los atributos Color del Cabello (Moreno, Rubio, Castaño) y Color de los Ojos (Negro, Marrón, Azul y Verde), obteniéndose la siguiente tabla de contigencia:

Ojos \ Cabello	Moreno	Rubio	Castaño
Negro	20	8	4
Marrón	16	2	11
Azul	5	8	8
Verde	10	5	3

¿Están relacionados dichos atributos?

Capítulo 3

Ajuste y regresión bidimensional

1. Introducción

Considerada una serie estadística $(x_1, y_1), \ldots, (x_n, y_n)$, procedente de una distribución (X, Y), el problema que se plantea en este capítulo consiste en encontrar alguna relación que exprese los valores de una variable en función de los de la otra. Para hacer esto, y una vez establecida cual será la variable dependiente, se tienen dos opciones: prefijar una clase funcional[1], o estimar un valor de la variable dependiente para cada valor de la variable independiente. En el primer caso, se está realizando un ajuste, y en el segundo, se tiene un problema de regresión. La regresión, viene determinada, por tanto, por un conjunto de puntos, de tal forma que uniendo los puntos contiguos por segmentos rectilíneos se obtiene la poligonal de regresión. La regresión sólo tendrá sentido cuando en la serie bidimensional haya muchos valores de la variable dependiente para cada uno de la independiente, pues en caso contrario, es casi idéntica al diagrama de dispersión y no aporta nada nuevo. Ambas técnicas, ajuste y regresión, son complementarias, siendo la poligonal de regresión una buena referencia de la clase funcional a elegir para el ajuste, a la vez que, como se verá a lo largo del capítulo, fija el techo para la bondad de éste.

[1] Por ejemplo una recta, una parábola, una función exponencial, etc.. . .

En el caso del ajuste, la cuestión será elegir la mejor clase funcional y determinar los parámetros que identifiquen la función dentro de la clase. Esto se consigue imponiendo a dicha función que verifique condiciones de adaptación a la nube de puntos, para ello, se empleará el criterio de los mínimos cuadrados, que se desarrollará en el punto siguiente.

Los puntos de la regresión se obtienen utilizando distribuciones condicionadas. Se empleará el método de la regresión a la media, que consiste en asociar a cada valor de la variable independiente, el valor medio de la distribución de la variable dependiente condicionada a cada uno de dichos valores.

El objeto del ajuste es doble. Por una parte interpolador: puesto que se está trabajando con un cierto número de observaciones, es de esperar que si éstas son representativas del fenómeno en estudio, los elementos del colectivo se comporten de forma parecida y la función de ajuste sea también válida para ellos. Y por otra parte extrapolador: bajo la suposición de que la relación funcional entre variable dependiente e independiente permanece constante, al menos en un entorno de las observaciones, es posible hacer una previsión del valor que tomará la variable dependiente para un valor determinado de la independiente en dicho entorno. Esta característica puede ser utilizada, por ejemplo, para hacer previsiones de ventas a corto o medio plazo, estimar el volumen de cosecha en función de la lluvia caída, etc...

La elección de la familia de funciones sobre la que se hará el ajuste es uno de los problemas principales a los que se deberá hacer frente. En un principio, la observación de la nube de puntos puede dar una idea de la evolución de los valores de la variable dependiente (a partir de ahora Y) en función de los de la independiente (X). Como ya se comentó arriba, en algunos casos, puede ser de mucha utilidad el construir la poligonal de regresión.

El tema se ha dividido en dos partes, la primera dedicada al ajuste y la segunda a la regresión.

2. Ajuste. Criterio de los mínimos cuadrados

Fijada la familia de funciones que se utilizará para ajustar los valores de una serie estadística bidimensional, ésta dependerá de unos parámetros. El método que se usará para la estimación de dichos parámetros es el de los *mínimos cuadrados*, que consiste en hacer mínima la suma de las diferencias al cuadrado entre los valores observados y los correspondientes valores ajustados.

Formalmente, sean $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, los valores observados y $g(x, \alpha, \beta, \cdots, \theta)$ la función de ajuste. Los valores de los parámetros se obtienen imponiendo la condición de hacer mínima la función H, donde

$$H(\alpha, \beta, \cdots, \theta) = \sum_{i=1}^{n}[y_i - f(x_i, \alpha, \beta, \cdots, \theta)]^2.$$

Para ello, se calculan las derivadas parciales de H respecto de cada uno de los parámetros y se igualan a cero:

$$\frac{\partial H}{\partial \alpha} = 0, \frac{\partial H}{\partial \beta} = 0, \cdots, \frac{\partial H}{\partial \theta} = 0.$$

Con esto se genera un sistema de tantas ecuaciones como parámetros, llamado sistema de ecuaciones normales, que resuelto da los valores de los parámetros. A continuación, se hace un estudio más detallado para algunas funciones de uso habitual.

2.1. Caso lineal

Sean $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ los valores observados[2] y sea $f(x, a, b) = a + bx$ la recta de ajuste de los valores de Y en función de los de X. Se obtienen los valores a y b minimizando la función error cuadrático, H, dada por

$$H(a, b) = \sum_{i=1}^{n}[y_i - (a + bx_i)]^2 \ .$$

[2]Obsérvese que no se dan las observaciones con sus respectivas frecuencias, sino que se hace por extensión para facilitar la nomenclatura.

Derivando respecto a los dos parámetros

$$
\begin{aligned}
\frac{\partial H(a,b)}{\partial a} &= -2\sum_{i=1}^{n}[y_i - (a + bx_i)] \\
\frac{\partial H(a,b)}{\partial b} &= -2\sum_{i=1}^{n}[y_i - (a + bx_i)]x_i
\end{aligned}
$$

e igualando a cero, queda el siguiente sistema de ecuaciones, que se conoce como sistema de *ecuaciones normales del modelo*:

$$
\begin{aligned}
\sum_{i=1}^{n} y_i &= na + b\sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} y_i x_i &= a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2.
\end{aligned}
$$

Utilizando la notación

$$
\begin{aligned}
y_i^* &= f(x_i) = a + bx_i \\
e_i &= y_i - y_i^* \qquad (3.1)
\end{aligned}
$$

el sistema de ecuaciones normales puede expresarse de la forma

$$
\begin{aligned}
\sum_{i=1}^{n} e_i &= 0 \\
\sum_{i=1}^{n} e_i x_i &= 0 \qquad (3.2)
\end{aligned}
$$

donde e_i representa el residuo o error de la observación i-ésima.

Admitiendo que se verifican las condiciones suficientes de mínimo se pueden obtener fácilmente los valores de a y b:

$$
\begin{aligned}
a &= \bar{y} - b\bar{x} \\
b &= \frac{S_{xy}}{S_x^2}.
\end{aligned}
$$

Si se quiere obtener la línea de ajuste de X respecto a Y, llamando ahora a' y b' a los coeficientes de la recta, se obtiene

$$\begin{aligned} a' &= \bar{x} - b'\bar{y} \\ b' &= \frac{S_{xy}}{S_y^2}. \end{aligned}$$

Los valores b y b' que son las pendientes de las rectas de ajuste, reciben el nombre de *coeficientes de regresión* y representan los incrementos de las variables dependientes para aumentos unitarios de las independientes.

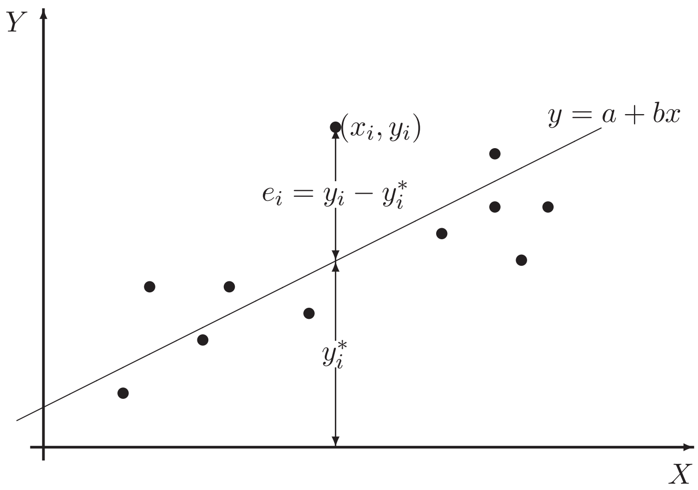


Figura 3.1: Criterio de los mínimos cuadrados

Ejemplo 3.1 *Dada la distribución bidimensional*

X	1	2	3	4	5	6
Y	2	5	9	13	17	21

Los coeficientes de la recta de ajuste de Y en función de X son

$$\begin{aligned} b &= \frac{S_{xy}}{S_x^2} = \frac{11'25}{2'91} = 3'87 \\ a &= \bar{y} - b\bar{x} = 11'166 - 3'87 \cdot 3'5 = -2'36 \end{aligned}$$

Y los coeficientes de la recta de X en función de Y son

$$\begin{aligned} b' &= \frac{S_{xy}}{S_y^2} = \frac{11'25}{43'46} = 0'26 \\ a' &= \bar{x} - b'\bar{y} = 3'5 - 0'26 \cdot 11'166 = 0'61. \end{aligned}$$

Es decir, cuando X aumenta en una unidad Y lo hace en 3'87 unidades, mientras que cuando se incrementa Y en una unidad X crece 0'26 unidades.

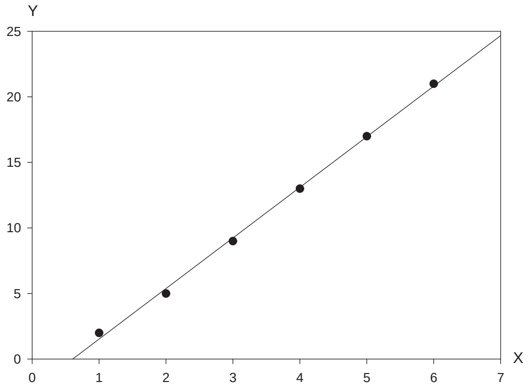


Figura 3.2: Recta de ajuste del ejemplo 3.1

Se han definido dos nuevas variables, por una parte Y^*, que representa los valores ajustados de la variable Y y que tiene por media y varianza:

$$\begin{aligned} \bar{y}^* &= \bar{y} \\ S_{y^*}^2 &= bS_{xy} \end{aligned}$$

y por otra parte, la variable residuo e, con media y varianza:

$$\begin{aligned} \bar{e} &= 0 \\ S_e^2 &= \frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{\sum_{i=1}^{n} (y_i - y_i^*)^2}{n}. \end{aligned} \tag{3.3}$$

S_e^2 recibe el nombre de varianza residual, y puede expresarse también de la forma

$$S_e^2 = \frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i}{n} \; .$$

Estas dos variables están incorreladas. En efecto, multiplicando la primera expresión de (3.2) por a y la segunda por b, se tiene

$$0 = a\sum_{i=1}^{n} e_i + b\sum_{i=1}^{n} e_i x_i = \sum_{i=1}^{n} e_i(a + bx_i) = \sum_{i=1}^{n} e_i y_i^*$$

y puesto que $\bar{e} = 0$, resulta

$$S_{ey^*} = \frac{\sum_{i=1}^{n} e_i y_i^*}{n} - \bar{e}\bar{y}^* = 0 \tag{3.4}$$

como se quería demostrar.

Ejercicio 3.1 *Demuestre las siguientes propiedades:*

a) *Las rectas de regresión de X sobre Y y de Y sobre X se cortan en el centro de gravedad de la distribución.*

b) *b, b', r y S_{xy} tienen siempre el mismo signo.*

c) *Las dos rectas de regresión coinciden sólo cuando $r = 1$ ó $r = -1$.*

d) *Se pueden expresar las rectas de regresión de Y sobre X y de X sobre Y, respectivamente, como:*

$$\begin{aligned} y - \bar{y} &= \frac{S_{xy}}{S_x^2}(x - \bar{x}) \\ x - \bar{x} &= \frac{S_{xy}}{S_y^2}(y - \bar{y}). \end{aligned}$$

e) *El coeficiente de correlación lineal puede obtenerse como:*

$$r = \pm\sqrt{bb'} \; .$$

Todo lo que se ha dicho hasta ahora es generalizable a funciones linealizables, sin más que hacer los cambios y transformaciones pertinentes, como los que se muestran en la tabla 3.1.

Función	Transformación	Cambio
$y = ax^b$	$\ln y = \ln a + b \ln x$	$\begin{cases} y' = \ln y \\ x' = \ln x \end{cases}$
$y = ab^x$	$\ln y = \ln a + x \ln b$	$\begin{cases} y' = \ln y \\ x' = x \end{cases}$
$y = a + \dfrac{b}{x}$	$y = a + \dfrac{b}{x}$	$\begin{cases} y' = y \\ x' = \dfrac{1}{x} \end{cases}$

Tabla 3.1: Linealización de funciones

2.2. Caso parabólico

Se considera la función de ajuste $f(x, a, b, c) = a+bx+cx^2$, parábola de segundo grado. Para obtener los valores de a, b y c se utiliza el método de los mínimos cuadrados y se minimiza la función H definida por

$$H(a, b, c) = \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)]^2 \ .$$

Para ello se calculan las derivadas parciales de H respecto de a, b y c y se igualan a cero:

$$\begin{aligned}
\frac{\partial H}{\partial a} &= -2 \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)] = 0 \\
\frac{\partial H}{\partial b} &= -2 \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)] x_i = 0 \\
\frac{\partial H}{\partial c} &= -2 \sum_{i=1}^{n} [y_i - (a + bx_i + cx_i^2)] x_i^2 = 0.
\end{aligned}$$

Despejando los términos independientes, se obtiene el sistema de ecuaciones normales:

$$\sum_{i=1}^{n} y_i = an + b \sum_{i=1}^{n} x_i + c \sum_{i=1}^{n} x_i^2$$

$$\begin{aligned}\sum_{i=1}^{n} y_i x_i &= a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3\\ \sum_{i=1}^{n} y_i x_i^2 &= a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4.\end{aligned}$$

De igual forma si se definen $y^* = a + bx_i + cx_i^2$ y $e_i = y_i - y_i^*$ el sistema de ecuaciones normales se puede expresar como

$$\begin{aligned}\sum_{i=1}^{n} e_i &= 0\\ \sum_{i=1}^{n} e_i x_i &= 0\\ \sum_{i=1}^{n} e_i x_i^2 &= 0.\end{aligned}$$

La varianza residual vale ahora

$$S_e^2 = \frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i - c\sum_{i=1}^{n} x_i^2 y_i}{n}.$$

3. Análisis de la bondad del ajuste

Una vez obtenidos los valores de los parámetros, y por tanto la función de ajuste, se va a dar la medida del grado de aproximación entre los valores observados y los ajustados. Según (3.3), la varianza del error o varianza residual vale

$$S_e^2 = \frac{\sum_{i=1}^{n}(y_i - y_i^*)^2}{n}$$

que coincide con el error cuadrático medio (E.C.M.). Al venir dada dicha varianza por una suma de términos al cuadrado será siempre positiva, salvo que todos los términos sean nulos, en cuyo caso el E.C.M. valdrá cero. Los términos serán nulos sólo si coinciden los valores observados con los ajustados, con lo cual, el ajuste será perfecto sólo si la varianza residual vale cero. Si se ajustan los mismos datos a dos modelos diferentes, el más afortunado será aquel que tenga un E.C.M. menor.

A continuación se relaciona la varianza residual, la varianza de Y y la varianza de Y^*. De la relación (3.1) se deduce que

$$y_i = y_i^* + e_i \ ,$$

además por (3.4) se tiene que y^* y e son incorreladas, es decir, la varianza de la suma es la suma de las varianzas, con lo cual se tiene que

$$S_y^2 = S_{y^*}^2 + S_e^2 \ .$$

Por tanto, la varianza de Y es igual a la varianza de los valores ajustados más la varianza residual. Este importante resultado se puede expresar también de la forma

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i^* - \bar{y})^2 + \sum_{i=1}^{n}(y_i - y_i^*)^2 \ ,$$

sin más que multiplicar los dos miembros de la igualdad por n.

Al ser las varianzas números positivos las relaciones siguientes son evidentes:

$$\begin{array}{ccc} 0 \leq & S_e^2 & \leq S_y^2 \\ 0 \leq & S_{y^*}^2 & \leq S_y^2 \end{array}$$

Si $S_e^2 = 0$, entonces $y_i = y_i^*$ para todo i, con lo que el modelo propuesto explica perfectamente las variaciones de la variable Y, siendo inmejorable. Si, por el contrario, $S_{y^*}^2 = 0$, $y_i^* = \bar{y}$ para todo i, con lo que el modelo de ajuste es constante y no explica, en forma alguna, las variaciones de Y. Estas son las situaciones extremas, pero el caso general se caracteriza por ser $S_e^2 > 0$ y $S_{y^*}^2 > 0$.

Evidentemente, cuanto más próximo S_e^2 esté de cero mejor será el ajuste, aunque, al ser un valor que viene acompañado de la unidad de la variable, no se puede dar una medida exacta de la bondad de éste ni hacer comparaciones entre dos distribuciones que vengan dadas en unidades distintas. Se hace necesario, pues, construir un coeficiente abstracto, de modo que se pueda expresar la bondad del ajuste en forma de porcentaje. Con esa intención se redefine[3] el *coeficiente de determinación* de la siguiente forma:

$$R^2 = \frac{S_{y^*}^2}{S_y^2} = \frac{S_y^2 - S_e^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2} \ .$$

Al ser $S_{y^*}^2 < S_y^2$, R^2 varía entre 0 y 1, multiplicado por 100 expresa el porcentaje de la variabilidad de Y que queda explicado por el ajuste. Cuando $S_{y^*}^2$ vale 0, R^2 vale 0, y el ajuste explica el 0 % de la variabilidad de Y. Cuando S_e^2 vale 0, R^2 vale 1 y el ajuste explica el 100 % de la variabilidad de Y. En general, y teniendo en cuenta que

$$1 = \frac{S_{y^*}^2 + S_e^2}{S_y^2} = \frac{S_{y^*}^2}{S_y^2} + \frac{S_e^2}{S_y^2},$$

se puede decir que el $100\frac{S_{y^*}^2}{S_y^2}$ de la variabilidad de Y queda explicada por el ajuste, y el resto, es decir el $100\frac{S_e^2}{S_y^2}$, no se explica por éste.

Ejemplo 3.2 *El coeficiente de correlación lineal al cuadrado para la distribución del ejemplo 3.1, vienen dados por:*

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = \frac{126'56}{2'914 \cdot 43'468} = 0'9991 \ .$$

Por lo que el $99'91$ % de la variabilidad de Y queda explicada por X a través de la recta de ajuste y el resto, es decir el $0'09$ % restante se explicaría bien por otra función de ajuste mejor o bien por otras variables distintas[4] *a X. Por último, puesto*

[3] Para el caso lineal ya se había introducido en el capítulo anterior.

[4] Observe que si para un mismo valor de X se tuvieran dos valores distintos de Y, X no podría explicar esa variabilidad, y de hecho ninguna función de X podría pasar por los dos puntos.

que los coeficientes de regresión son positivos, se tendrá que

$$r = +\sqrt{R^2} = 0'9995 .$$

Resultado. Cuando el ajuste realizado es lineal, R^2 coincide con el cuadrado del coeficiente de correlación lineal, es decir:

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} = r^2 .$$

4. Regresión. Método de regresión a la media

Este método consiste en definir la línea de regresión como la poligonal que pasa por la media aritmética de los puntos de igual abscisa, (curva de regresión de Y respecto a X), ó de igual ordenada, (X respecto a Y). Si sólo hubiera un valor de Y para cada valor de X, ó un valor de X para cada uno de Y, la correspondiente curva de regresión sería la poligonal que uniera todos los puntos. La regresión tiene interés, por tanto, sólo cuando hay un gran número de observaciones y a cada valor de una de las variables le corresponden muchos valores de la otra. El nombre de regresión tiene su origen en los primeros estudios que relacionaban las estaturas de grupos de padres e hijos; se observó entonces que, en general, padres de pequeña estatura tenían hijos bajos pero no tanto como ellos, y padres de talla elevada tenían hijos altos pero no tanto como ellos, produciéndose una tendencia o "regresión" hacia los valores intermedios.

Se llama *curva de regresión* de Y respecto de X, a la función que asocia a cada x_i de X, la media condicionada de Y respecto de x_i. Es decir:

$$\phi(x_i) = \bar{y}|_{X=x_i} = \bar{y}_i$$

Ejemplo 3.3 *La regresión a la media de la siguiente distribución:*

X/Y	1	2	3	4	5	6	7	8	9
1	2	4	3	7	1	8	9	3	2
2	3	6	1	8	5	9	6	4	3
3	0	0	3	5	7	8	5	4	2
4	0	0	0	2	4	6	7	4	3
5	0	0	0	0	5	8	7	4	4
6	0	0	0	2	8	9	9	6	6
7	0	0	0	0	0	0	2	3	3

viene dada por

X	1	2	3	4	5	6	7
$\phi(x)$	$5'25$	$5'11$	$5'79$	$6'61$	$6'85$	$6'67$	$8'12$

Gráficamente la curva de regresión sería la que se muestra en la figura 3.3.

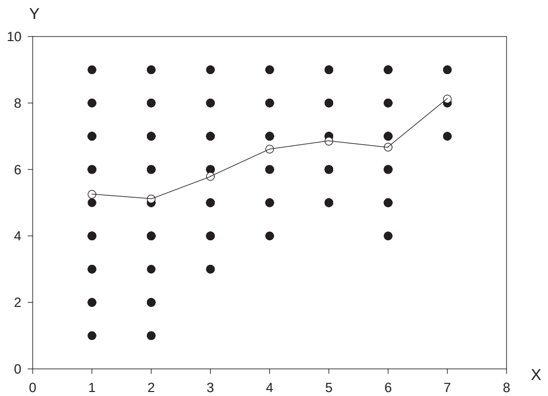


Figura 3.3: Poligonal de regresión

5. Análisis de la bondad de la regresión

A continuación se dará una medida de la proximidad de la curva de regresión a la distribución. Se considera el error cuadrático medio, que en este caso viene dado por

$$E.C.M. = \frac{\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{s}[y_{ij} - \bar{y}_i]^2 n_{ij}}{n} = \sum_{i=1}^{r} V_i[Y]\, f_{i.}\ ,$$

donde

$$V_i[Y] = \frac{\displaystyle\sum_{j=1}^{s}[y_{ij} - \bar{y}_i]^2 n_{ij}}{n_{i.}}$$

es la varianza de Y condicionada a $X = x_i$.

El E.C.M. está medido en la misma unidad que la variable con los consiguientes problemas. Se hace necesario, pues, dar una medida adimensional de la bondad de la regresión. La varianza de Y se puede expresar como

$$V[Y] = \overbrace{\sum_{i=1}^{n} V_i[Y]\, f_{i.}}^{I} + \overbrace{\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2 f_{i.}}^{II}\ ,$$

quedando descompuesta en la suma de la media ponderada de las varianzas condicionadas (I), más la varianza ponderada de las medias condicionadas (II). Es decir, la heterogeneidad de Y, resulta de la heterogeneidad debida a las distribuciones condicionadas por cada modalidad x_i, más la heterogeneidad existente entre las distintas modalidades.

Se define la *razón de correlación* de Y con respecto a X como:

$$\eta^2_{y,x} \quad = \quad \frac{\displaystyle\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2 f_{i.}}{V[Y]}$$

$$= \frac{V[Y] - \sum_{i=1}^{n} V_i(Y)\, f_{i.}}{V[Y]}$$

$$= 1 - \frac{\sum_{i=1}^{n} V_i[Y]\, f_{i.}}{V[Y]}.$$

De forma análoga se definiría la razón de correlación de X respecto a Y.

En general las dos razones de correlación son diferentes y verifican:

$$0 \leq \eta^2_{y,x}, \eta^2_{x,y} \leq 1.$$

La razón de correlación $\eta^2_{y,x}$ vale 0 si la varianza de las medias condicionadas es nula. Es decir, si todas las medias condicionadas son idénticas, en cuyo caso la curva de regresión es paralela al eje X, y se dice que Y está incorrelada con X. El recíproco también es cierto, la ausencia de correlación de Y con X, implica que la razón de correlación vale cero.

La razón de correlación $\eta^2_{y,x}$ vale 1 cuando la media de las varianzas condicionadas $V_i(Y)$ es cero; ahora bien, una suma de términos positivos es nula sólo si todos son nulos, lo cual implica que todas las $V_i(Y)$ son cero, es decir, al valor x_i de X le corresponde un único valor de Y, y, por consiguiente, Y depende funcionalmente de X. Recíprocamente, la dependencia funcional de Y respecto a X, implica que la razón de correlación vale 1.

Los distintos casos que se pueden presentar se recogen en la tabla 3.2. De igual forma que se hacía con el coeficiente de determinación, se puede multiplicar la razón de correlación por 100 y se obtendrá, en forma de porcentaje, la parte de la variación de Y que queda explicada por la curva de regresión, que será

$$\frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2 f_{i.}}{V[Y]} 100\,\% \,,$$

el resto hasta 100, es decir:

$$\frac{S_e^2}{V[Y]} 100\,\%$$

Razones de correlación	$\eta^2_{x,y} = 0$	$0 < \eta^2_{x,y} < 1$	$\eta^2_{x,y} = 1$
$\eta^2_{y,x} = 0$	Ausencia recíproca de correlación	Ausencia de correlación de Y respecto de X	Dependencia funcional de X respecto de Y Ausencia de correlación de Y respecto de X
$0 < \eta^2_{y,x} < 1$	Ausencia de correlación de X respecto de Y	Caso general	Dependencia funcional no recíproca de X respecto de Y
$\eta^2_{y,x} = 1$	Dependencia funcional de Y respecto de X Ausencia de correlación de X respecto de Y	Dependencia funcional no recíproca de Y respecto de X	Dependencia funcional recíproca

Tabla 3.2: Estudio de las razones de correlación

quedará sin explicar por ella.

Propiedad 3.1 *La curva de regresión es la "función" óptima para el criterio de los mínimos cuadrados.*

De este resultado se puede deducir que el E.C.M. de la regresión de la media es menor o igual que el E.C.M. del ajuste mínimo cuadrático, y como consecuencia, que el coeficiente de determinación, es siempre menor o igual que la menor de las razones de correlación, sea cual sea la función elegida. La comparación entre R^2 y $\text{mín}(\eta^2_{y,x}, \eta^2_{x,y})$ puede indicar si existe una función mejor para ajustar los datos.

6. Notas y conclusiones

1. A lo largo del capítulo se ha comentado que la regresión a la media tiene sentido cuando se cuenta con un amplio número de observaciones de la variable dependiente para un valor de la independiente. Sin embargo, en ciertos casos se puede suavizar esta condición. Así, si la variable X es continua, aún con un alto número de observaciones, es de esperar que haya pocos valores que se repitan, pero siempre se puede agrupar los datos en intervalos; éstos deben ser

lo suficientemente pequeños como para que los elementos pertenecientes a un mismo intervalo puedan considerarse, a los efectos pertinentes, iguales, y lo bastante grandes como para que caiga un número mínimo de observaciones en la mayoría de las clases.

2. Cuando la nube de puntos no proporcione una idea de la clase funcional que se debe elegir para realizar el ajuste, una posible solución puede ser el calcular la línea de regresión y representarla gráficamente.

3. Se ha utilizado una función para efectuar el ajuste, el calcular la previsión para un valor de la variable independiente se reduce a sustituir dicho valor en la función. Si debe emplearse la poligonal de regresión, la previsión para un valor x de X será el valor correspondiente a la clase a la que pertenece x. Esto hace que la previsión tenga dos matices diferentes: mientras que si se utiliza la poligonal de regresión, sólo se pueden hacer previsiones para valores encuadrados en alguna de las clases predefinidas —en realidad sería una interpolación—; cuando las previsiones se basan en una función de ajuste, no sólo pueden hacerse estimaciones para valores intermedios, sino que se pueden extrapolar y sacar conclusiones para valores exteriores; aunque, en este caso, la fiabilidad de la previsión depende de que las condiciones en que se realizó el ajuste, permanezcan constantes, lo que ocurrirá, normalmente, en un entorno de los puntos utilizados para realizarlo.

7. Ejercicios

7.1. Ejercicio resuelto

3.1 Dada la siguiente distribución bidimensional:

X	$0'7$	1	2	3	3	4	5	6	7	8
Y	$2'2$	$2'2$	$2'5$	$2'7$	$2'8$	3	$3'3$	$3'4$	4	4

a) Calcule las dos rectas de ajuste.

b) Compruebe que dichas rectas se cortan en el centro de gravedad.

c) Estime el valor de Y para X=10.
d) Interprete el significado de b y de b'.
e) Interprete el significado de $1 - R^2$.

Solución:

a) Para calcular las dos rectas se procede a obtener las ecuaciones normales, y de aquí, o bien, se utilizan las expresiones obtenidas en el capítulo para a y b, es decir, $a = \bar{y} - b\bar{x}$ y $b = \frac{S_{XY}}{S_X^2}$, o bien, se resuelve directamente el sistema. En este caso el sistema de ecuaciones normales viene dado por

$$\begin{aligned} 30'1 &= 10a + 39'7b \\ 134'14 &= 39'7a + 213'49b \ . \end{aligned}$$

Resolviendo por Cramer se tiene que

$$b = \frac{\begin{vmatrix} 10 & 30'1 \\ 39'7 & 134'14 \end{vmatrix}}{\begin{vmatrix} 10 & 39'7 \\ 39'7 & 213'49 \end{vmatrix}} = 0'262 \ ,$$

y despejando en la primera ecuación normal se obtiene $a = 1'97$. Con lo que la ecuación queda:

$$y = 1'97 + 0'262x.$$

Para obtener la otra recta se puede hacer lo mismo, no obstante, para ofrecer otra posibilidad de resolución se utiliza $b' = \frac{S_{xy}}{S_y^2}$ y $a' = \bar{x} - b'\bar{y}$. Así, puesto que $S_{xy} = 1'464$, $S_y^2 = 0'391$, $\bar{x} = 3'97$ y $\bar{y} = 3'01$, se tiene que

$$b' = \frac{1'464}{0'391} = 3'74 \quad y \quad a' = 3'97 - 3'74 \times 3'01 = -7'28$$

Quedando esta otra recta como:

$$x = -7'28 + 3'74y.$$

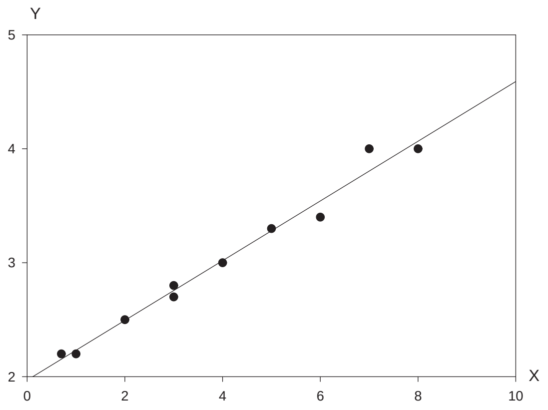


Figura 3.4: Recta de ajuste de Y en función de X

b) Para comprobar que ambas rectas se cortan en el centro de gravedad, basta con sustituir $\bar{x}$ y $\bar{y}$ en las rectas respectivas, y así:

$$\begin{aligned} y &= 1'97 + 0'262 \cdot 3'97 = 3'01 \\ x &= -7'28 + 3'74 \cdot 3'01 = 3'98. \end{aligned}$$

Como puede verse, se cometen pequeños errores de redondeo, subsanables operando con más cifras decimales.

c) Para realizar la estimación, se sustituye el valor $x = 10$ en la recta de ajuste

$$Y = 1'97 + 0'262 \times 10 = 4'59\ .$$

La calidad de la previsión depende de que ésta se haga en un entorno próximo a los valores de la variable, en nuestro caso se puede pensar que el valor 10 está en dicho entorno, y, por otra parte, de que la recta se ajuste suficientemente bien a los puntos, para lo que se debe calcular el coeficiente de determinación R^2. Para ello, se utiliza que $R^2 = bb' = 0'262 \times 3'74 = 0'98$, es decir, la recta explica a través de X el 98 %

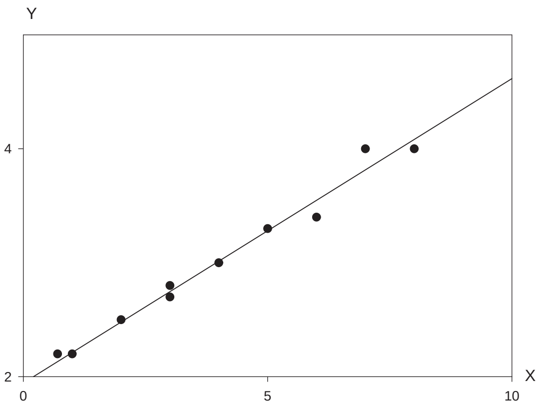


Figura 3.5: Recta de ajuste de X en función de Y

de la variabilidad de Y. A la vista de lo anterior, se puede concluir que $4'59$ es una buena previsión para Y dado que X vale 10.

d) b y b' son las pendientes de la rectas y representan los incrementos-decrementos, según que sean positivas o negativas, de la variable dependiente para incrementos unitarios de la variable independiente. En nuestro caso cuando X aumenta una unidad Y aumenta 0'262 unidades, mientras que cuando lo hace Y en una unidad X crece 3'74 unidades.

e) $1 - R^2 = 0'02$. Lo que indica que el 2 % de la variabilidad de Y, no se explica por la recta en función de X, y que puede haber una mejor función de ajuste u otras variables distintas a la X no contempladas en el modelo. En general, ocurren ambas cosas a la vez, aunque en todo caso esto habrá que discutirlo a partir de los valores de las razones de correlación.

7.2. Ejercicios propuestos

3.1. Dada la distribución:

X	1	$1'5$	2	$2'5$	3	$3'75$	$4'5$	5
Y	1	$1'5$	$2'95$	$5'65$	$8'8$	15	25	32

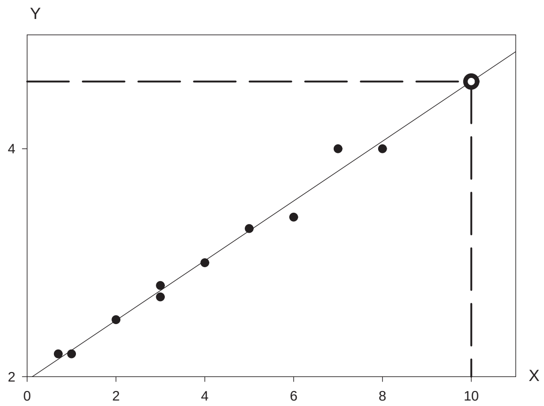


Figura 3.6: Previsión

a) Elija la mejor clase funcional para ajustar la distribución y estime sus parámetros.

b) Establezca la bondad del ajuste.

c) Calcule la previsión para Y cuando $X = 7$. Analice dicha previsión.

3.2. Dada la distribución:

$Y \quad X$	1	2	3	4	5	6	7	8	9
1	2	3	3	1	1	0	1	0	0
2	1	5	6	3	1	0	0	0	0
3	1	2	7	5	2	2	0	0	0
4	0	1	7	5	3	3	0	0	0
5	0	1	5	8	5	4	3	0	0
6	0	0	4	6	7	5	4	0	0
7	0	0	2	3	8	8	7	4	0
8	0	0	1	2	5	6	7	6	1
9	0	0	1	2	3	4	6	4	5

a) Obtenga la poligonal de regresión de Y respecto a X.

b) Calcule sus razones de correlación.

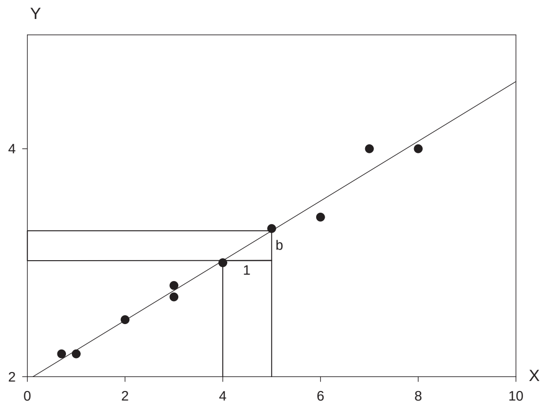


Figura 3.7: Interpretación gráfica de la pendiente de la recta

c) A la vista de la poligonal ajuste la distribución a una función, justificando dicha elección.

d) Calcule R^2, analícelo y compárelo con las razones de correlación.

3.3. Dada la distribución:

X	$2'5$	$3'75$	5	$7'5$	10	$12'5$	20
Y	8	14	$23'75$	40	62	90	165

a) Utilice una ecuación del tipo aX^b para ajustar la distribución.

b) Dé una medida de la bondad del ajuste.

3.4. Dada la distribución:

X	1	$1'5$	2	3	4	5	6	7
Y	1	$1'75$	$2'65$	$4'7$	7	$9'5$	12	15

a) Ajuste la distribución utilizando una función del tipo aX^b.

b) Analice la bondad del ajuste.

3.5. Dada la distribución:

X	5	6	8	10	13	18	20
Y	$1'5$	$1'25$	$0'93$	$0'7$	$0'46$	$0'23$	$0'15$

a) Estime los parámetros de la clase funcional $ab^{-0'2X}$ para ajustar la distribución.

b) Estudie la bondad del ajuste.

c) ¿Sería posible plantear un ajuste del tipo ab^{cX}? Justifíquelo.

3.6. Dada la distribución:

Y X	2	3	4	5	6	7
2	6	2	1	0	0	0
5	4	7	5	0	0	0
10	0	1	7	3	0	0
15	0	0	6	3	1	0
20	0	0	4	8	2	0
25	0	0	3	8	5	0
30	0	0	0	5	9	0
35	0	0	0	1	7	4
40	0	0	0	0	3	4

a) Obtenga la curva de regresión y calcule la razón de correlación de Y respecto de X.

b) Ajuste los datos a una recta y a una parábola y discuta los resultados.

3.7. Se ha obtenido utilizando el criterio de mínimos cuadrados, que la recta de ajuste de Y sobre X es $Y = 2X + 9$. Sabiendo que $\bar{x} = 5$, $\bar{y} = 10$, $S_X = 2$, $S_Y = 3$ y $S_{XY} = 0'8$. Calcule:

a) La varianza de los valores estimados en Y.

b) La recta de ajuste de Y sobre X si se le suma 5 a todos los valores de X.

c) La recta de ajuste de Y sobre X si se le suma 3 a todos los valores de Y y se multiplica por 2 todos los valores de X.

3.8. Sabiendo que las ecuaciones de las rectas de ajuste entre la temperatura (Y) y la profundidad (X) vienen dadas por

$$y^* = \frac{-x}{5} + 2 \qquad x^* = -4y + 11 \ ,$$

y además $|S_{XY}| = 8$. ¿Qué variable es más homogénea?

Parte B

Probabilidad

Introducción a teoría de la probabilidad

La segunda parte de este manual está dedicada a las herramientas por excelencia de la Estadística: la función de probabilidad y la variable aleatoria. Ello no quiere decir que en ocasiones no se planteen problemas específicos de probabilidad, pero lo cierto es que el enfoque con más proyección de la Estadística, el inferencial, no existiría sin dichas herramientas.

La existencia de fenómenos o experimentos no determinísticos, donde el conocimiento de las condiciones en las que éstos se desarrollan no garantizan los resultados, hace imprescindible el uso de una función que asigne niveles de certidumbre a cada uno de los desenlaces del fenómeno, y ahí es donde aparece la probabilidad. Los experimentos o fenómenos que poseen la característica anterior se denominan aleatorios. Intuitivamente, la concreción numérica del fenómeno mediante la asignación de valores con un cierto criterio, da origen a la variable aleatoria. Una correcta proyección de estos conceptos es lo que va a permitir estudiar grandes colectivos a partir de pequeñas partes de ellos, llamadas muestras, dando lugar a lo que se conoce como inferencia estadística.

Esta segunda parte está formada por otros tres capítulos, en el primero de ellos se introduce el concepto de probabilidad, haciéndose una breve incursión por la teoría de conjuntos, dada la existencia de una correspondencia total entre los elementos de esta teoría y los resultados de un experimento o fenómeno aleatorio. Se hace un recorrido por la evolución que a lo largo del tiempo ha tenido la probabilidad comentando sus distintas definiciones; se estudian sus propiedades y se introducen los importantes conceptos de probabilidad condicionada y de independencia. Termina el capítulo con los teoremas de la probabilidad total y de Bayes.

El segundo capítulo está dedicado a la variable aleatoria, que se presenta desde un punto de vista intuitivo y conceptual. El paso desde la función de cuantía en el caso discreto a la función de densidad en el continuo se hace de una forma gráfica y natural. La función de distribución se plantea como una alternativa a la función de densidad, al igual que las funciones generatriz de momentos y la característica, aunque indicando sus utilidades específicas. Desde una óptica más local, la esperanza matemática permitirá definir coeficientes que expresen las singularidades de la distribución, entre los que destacan la media y la varianza. Desde una perspectiva univariable se termina el tema con una alusión al cambio de variables y se da la expresión de la desigualdad de Tchebychev para el caso probabilístico. Para aquellos estudiantes que quieran ver como se generalizan algunos de estos conceptos, se estudia el caso multidimensional, con desarrollo expreso del bidimensional. Tanto para variables discretas como continuas, se dan las expresiones de las funciones de densidad y de distribución. Se analiza de nuevo la dependencia e independencia entre variables. La función esperanza toma aquí su versión n-dimensional y a partir de ella se pueden calcular una infinidad de coeficientes que aportan una visión puntual de la distribución, entre los que cabe destacar los de covarianza y correlación, proyecciones de los que se ven en descriptiva. Por último, se desarrolla el cambio de variables para el caso de dos dimensiones.

En el tercer capítulo se estudian distintas estructuras probabilísticas que modelizan una gran cantidad de situaciones reales, dedicándose especial atención a las distribuciones binomial y Poisson, en el caso discreto, y normal, en el continuo; de forma más somera también se analizan otras distribuciones que se derivan de aquellas. El teorema central del límite servirá para comprobar el papel que juega la distribución normal dentro de la Estadística. Al final del capítulo se estudian algunas distribuciones multivariables.

Capítulo 4

Teoría de la probabilidad

1. Evolución histórica

Como en la mayoría de los descubrimientos, la noción de probabilidad se ha ido desarrollando a lo largo del tiempo en función de la necesidad, de los recursos y de la aportación de los grandes genios que son capaces, en un momento de inspiración, de dar un paso de un siglo.

Es difícil establecer históricamente el nacimiento de la probabilidad, aunque su conceptualización como disciplina matemática es reciente; parece, no obstante, que su origen tiene relación con los juegos de azar. Se consideran como remotos precursores de la teoría de la probabilidad la abundante presencia del hueso astrágalo de oveja o ciervo (antecedente inmediato del dado), en excavaciones arqueológicas con una antiguedad de más de 40.000 años. En épocas más recientes, en las culturas griega, egipcia y romana la afición a los juegos de azar, especialmente mediante la tirada de dados y tablas, estaba ampliamente extendida. En estas civilizaciones el azar se explicaba mediante la voluntad divina.

En el siglo XV, Dante obtiene algunas probabilidades en juegos sencillos de lanzamientos de dados. En el siglo XVI, Cardano con su tratado *Liber de ludo aleae* (Libro de los juegos de azar) y Galileo Galilei en su *Considerazione sopra el giuoco dei dadi* (Consideraciones sobre el juego de dados), describen ciertos juegos de dados y diversos problemas combinatorios, incluso este último, publica un tratado sobre la probabilidad de error.

La mayoría de los autores consideran que la génesis del cálculo de probabilidades, como disciplina matemática, se encuentra en la resolución del problema planteado por el caballero de Mére, un jugador empedernido de la Francia del siglo XVII, a su amigo y matemático B. Pascal (1623-1662), quien mantuvo una abundante correspondencia sobre dicho problema con su colega P. Fermat (1601-1665). El problema consistía en cómo deberían repartirse el dinero de las apuestas depositado en la mesa si los jugadores se vieran obligados a finalizar la partida sin que existiera un ganador. Dicho problema se detalla en el ejercicio 4.1

Ejercicio 4.1 *Dos jugadores de cartas, A y B, apuestan 250€ cada uno, en un juego que vencerá aquel que llegue primero a tres partidas ganadas. El juego se interrumpe cuando A lleva ganadas dos partidas y B una, ¿cómo deberían repartirse el dinero?*

Además de la correspondencia a la que se hacía referencia en el párrafo anterior, se considera fundamental en el nacimiento del cálculo de probabilidades la obra del matemático holandés C. Huygens (1629-1695), quien introduce el concepto de esperanza matemática, como generalización de la media aritmética, en su obra *De ratiocinüs in ludo aleae*, (Del raciocinio en los juegos de azar) en la que además resuelve varios problemas planteados por Pascal y Fermat.

En 1713 Jacques Bernouilli (1654-1705) lega uno de los tratados clave en la construcción de la teoría de la probabilidad, publicado tras su muerte lleva por título *Ars conjectandi* (El arte de conjurar), en él se introduce el término estocástico y se detalla la ley conocida como de ensayos de Bernouilli (primer teorema límite de la teoría demostrado con todo rigor), que enunciado de una forma sencilla dice así: "la frecuencia relativa de un suceso tiende a estabilizarse en torno a un número, a medida que el número de pruebas del experimento crece indefinidamente". En esta época hay que destacar las importantes contribuciones de autores como Abraham de Moivre (1667-1754), Daniel Bernouilli (1700-1782) o Thomas Bayes (1702-1761), entre otros.

A partir del citado periodo, el cálculo de probabilidades adquiere una mayor interrelación con otras ciencias y no se circunscribe exclusivamente a los juegos de azar.

Uno de los matemáticos artífices del asentamiento de las bases de lo que hoy se conoce como probabilidad clásica es Pierre Simon, marqués de Laplace (1749-1827). El momento cumbre fue la publicación en 1812, del importante y extenso tratado *Theorie analitique des probabilités*; en él, aparece la primera definición del concepto de probabilidad, conocida hoy como definición clásica.

Contemporáneo de Laplace, merece ser destacado C. F. Gauss, (1777-1855), quien dedicó parte de su actividad a estudiar la teoría de errores, dando lugar a la ley normal, de la que estimó sus parámetros.

Después de este periodo, el interés por el cálculo de probabilidades fue disminuyendo, llegando prácticamente a desaparecer como disciplina matemática durante el siglo XIX. Esto se debió, por una parte, a la aparición de algunas contradicciones como la que puso de manifiesto el matemático francés Bertrand, y por otra, a una relativa apatía por la probabilidad debido al carácter esencialmente "deterministíco" del siglo XIX. La evolución de la teoría se desplaza hacia el Este, en particular hacia la escuela de San Petesburgo. En ella resaltan las contribuciones de P.L. Tchebychev (1821-1894) y de su discípulo A. Markov (1856-1922).

Sin embargo, no es hasta principios del siglo XX, más concretamente 1933, fecha de la publicación de la obra *Fundamentos* por el matemático ruso A. N. Kolmogorov, cuando la probabilidad pasa a convertirse en una rama más de las matemáticas. En esta obra, Kolmogorov apoyándose en la teoría de conjuntos y en la teoría de la medida, da una definición axiomática del cálculo de probabilidades, como generalización y síntesis de los conocimientos que de la probabilidad se tenían hasta entonces.

2. Conjuntos. Operaciones

Un *conjunto* es una colección en un todo de objetos bien definidos que poseen una o varias propiedades. Cada uno de los objetos del conjunto se llama *elemento*.

La idea de conjunto sólo hace referencia a la presencia de sus elementos y no a ninguna ordenación o repetición de éstos. Los conjuntos pueden venir dados por:

Extensión: Se especifica cada uno de los elementos que pertenece al conjunto.

Descripción: Se da una o varias propiedades que deben cumplir todos los elementos del conjunto. Cualquier elemento que verifique esas propiedades pertenece al conjunto.

Ejemplo 4.1

$$A = \underbrace{\{x \;/\quad x \geq 0, \quad x^2 - x - 2 = 0\}}_{\text{Descripción}} \equiv \overbrace{\{2\}}^{\text{Extensión}}$$

Se dice que dos *conjuntos* son *iguales* cuando están formados por los mismos elementos.

Ejemplo 4.2 *Los conjuntos* $A = \{x \mid x^2 - x - 2 = 0\}$ *y* $B = \{-1, 2\}$ *son iguales.*

2.1. Subconjunto

B es un *subconjunto* de A, ó bien A contiene a B, si todo elemento de B es elemento de A.

Ejemplo 4.3 *Si* $A = \{x \mid x^2 - x - 2 = 0\}$ *y* $B = \{-1\}$ *entonces* $B \subset A$.

2.2. Operaciones entre conjuntos

Unión. La unión de conjuntos es otro conjunto que está formado por todos los elementos de dichos conjuntos.

Ejemplo 4.4 *Sea* $A = \{1, 2, 4\}$, $B = \{2, 3\}$ *entonces se tiene que* $A \cup B = \{1, 2, 3, 4\}$.

Intersección. La intersección de conjuntos es otro conjunto que está formado por los elementos comunes a todos los conjuntos.

Ejemplo 4.5 *Dados* $A = \{1, 2, 4\}$ *y* $B = \{2, 3\}$ *se tiene* $A \cap B = \{2\}$.

2.3. Conjuntos característicos

Conjunto vacío. Es aquel conjunto que no tiene ningún elemento. Se denota por $\emptyset$. Este conjunto se considera como subconjunto de cualquier otro conjunto.

Ejemplo 4.6 *Tomando* $A = \{1, 2, 4\}$ *y* $B = \{3, 5\}$ *queda* $A \cap B = \emptyset$.

Conjunto universal. Es aquel formado por la totalidad de los elementos del mismo tipo. Todo conjunto se puede considerar como subconjunto de él. Se denota por Ω.

Conjuntos disjuntos. Si dos conjuntos A y B no poseen elementos comunes se dicen que son disjuntos, es decir, $A \cap B = \emptyset$.

Partición. Se dice que los conjuntos $A_1, A_2, A_3, \ldots, A_n$ forman una partición o un sistema completo de sucesos, si son disjuntos dos a dos y la unión de todos ellos es el conjunto universal, es decir:

1. $A_i \cap A_j = \emptyset \quad \forall i \neq j$
2. $A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n = \Omega$

Conjunto complementario. El conjunto complementario de un conjunto A, es el conjunto formado por todos los elementos del conjunto universal que no están en A. Se denota por $\bar{A}$.

2.4. Propiedades de las operaciones entre conjuntos

1. La unión e intersección de conjuntos son operaciones conmutativas y asociativas.

   a) $A \cup B = B \cup A, \quad A \cap B = B \cap A$

   b) $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$

2. Se verifica la distributiva de cada operación respecto a la otra.

   a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

   b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

3. El operador complementario verifica:

   a) $A \cap \bar{A} = \emptyset$ y $A \cup \bar{A} = \Omega$

   b) $\bar{\Omega} = \emptyset$ y $\bar{\emptyset} = \Omega$

   c) $\overline{\overline{A}} = A$

4. Se cumplen las llamadas Leyes de Morgan:

   a) $\overline{A \cup B} = \overline{A} \cap \overline{B}$

   b) $\overline{A \cap B} = \overline{A} \cup \overline{B}$

3. Álgebra de sucesos

Hay determinados experimentos en los que las situaciones o causas determinan perfectamente los resultados o efectos, como por ejemplo ciertas situaciones físicas. Sin embargo, existen otros experimentos en los que en las mismas condiciones se obtienen resultados diferentes, como ocurre en los juegos de azar. Los primeros experimentos reciben el nombre de *determinísticos*, mientras que los segundos se conocen como *aleatorios*.

Se parte del experimento de lanzar un dado al aire y observar el número de puntos que figura en la cara superior. Los números 1, 2, 3, 4, 5, 6 son los sucesos elementales posibles. También se puede considerar sucesos compuestos como conseguir par, trío,..., formados por la unión de sucesos elementales. Si se prolonga indefinidamente esta sucesión de pruebas con sus resultados, se llega al conjunto potencialmente infinito de todas las pruebas asociadas a un experimento aleatorio que se llama universo, población o colectivo asociado al mismo.

Para el estudio de un fenómeno determinista se hace preciso la constatación de ciertas regularidades. En el caso de fenómenos aleatorios estas regularidades aparecen al considerar un gran número de pruebas.

Ejemplo 4.7 *Si obtengo m veces el valor 3 en n tiradas de un dado la frecuencia del suceso 3 será $\frac{m}{n}$.*

El hecho de que la frecuencia de un suceso tienda a aproximarse a un número fijo al aumentar el número de pruebas se ha denominado “ley de azar” o “ley de estabilidad” de las series estadísticas.

La noción de probabilidad como valor límite ideal de estas frecuencias es la base del modelo matemático apropiado para el estudio de estos fenómenos. Su teoría constituye el cálculo de probabilidades.

3.1. Espacio muestral. Sucesos

Se llama *espacio muestral* al conjunto de todos los resultados posibles de un experimento aleatorio. Se denota por Ω.

Los espacios muestrales pueden ser:

Finitos. Como, por ejemplo, el espacio muestral correspondiente al experimento aleatorio del lanzamiento de un dado una única vez, que consta de seis elementos que corresponden a los seis resultados posibles.

Infinitos numerables. Como, por ejemplo, el resultante de la contabilización del número de veces que hay que tirar una moneda hasta que aparezca por primera vez cara, donde el espacio muestral está compuesto por los números naturales, $\{1, 2, 3, 4, \dots, n, \dots\}$

Continuos. Como, por ejemplo, el que se obtiene al medir, en radianes, el ángulo que forma una aguja, que se lanza sobre una superficie plana, con una determinada dirección prefijada, donde el espacio muestral es $[0, 2\pi]$.

Se denomina *suceso* a todo subconjunto del espacio muestral, es decir A es un suceso si $A \subseteq \Omega$. Los *sucesos elementales* son aquellos que constan de un único elemento.

Al realizar un experimento aleatorio se dice que se ha verificado el suceso A, si el resultado obtenido pertenece a A.

3.2. Relaciones entre sucesos

Implicación. Un suceso A implica otro suceso B cuando siempre que se verifique A se verifica B.

Ejemplo 4.8 *En el lanzamiento de un dado se consideran los sucesos: A =\{Obtener un 4\} y B =\{Obtener un múltiplo de 2 \}, entonces A implica B.*

Suceso contrario. Dado un suceso A, se define el *suceso contrario* de A y se denota por $\bar{A}$, como aquel que se verifica si y sólo si no se verifica A.

Ejemplo 4.9 *Si en el lanzamiento de un dado $A = \{1, 2\}$, entonces el suceso contrario viene dado por $\bar{A} = \{3, 4, 5, 6\}$.*

Unión de sucesos. Se dice que el suceso $C = A \cup B$ se verifica, si y sólo si se verifica A , B o ambos.

Ejemplo 4.10 *En el lanzamiento de un dado se describen los siguientes sucesos:*

A = \{Obtener una puntuación menor o igual que tres \}

B = \{Obtener una puntuación par\}

Entonces $A \cup B$ = \{Obtener 1, 2, 3, 4, 6\}

Suceso seguro. El *suceso seguro*, que se denota por Ω, es aquel que siempre se verifica. Para cualquier suceso A siempre se cumple que $\Omega = A \cup \bar{A}$.

Intersección de sucesos. Se dice que el suceso $C = A \cap B$ se verifica, si y sólo si se verifican A y B.

Ejemplo 4.11 *En el lanzamiento de un dado, si se tiene:*
$A = \{$Obtener una puntuación menor o igual que tres$\}$
$B = \{$Obtener una puntuación par$\}$
Entonces $A \cap B = \{$Obtener un 2$\}$.

Suceso imposible. *Suceso imposible* es aquel que no se puede verificar nunca, se denota $\emptyset$.

Sucesos incompatibles. Dos *sucesos* son *incompatibles* cuando al verificarse uno de ellos no se verifica el otro, o equivalentemente, cuando su intersección es el suceso imposible.

3.3. Álgebra de Boole

Una familia $\mathcal{A}$ de subconjuntos de Ω tiene estructura de *Álgebra* o *Álgebra de Boole* sobre Ω, si verifica:

1. $\Omega \in \mathcal{A}$
2. $\forall A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
3. $\forall A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$

El par $(\Omega, \mathcal{A})$ se dirá *espacio medible finito*.

Una familia $\mathcal{A}$ de subconjuntos de Ω se dice que tiene estructura de *σ-álgebra* sobre Ω, si

1. $\Omega \in \mathcal{A}$

2. $\forall \{A_i\}_{i \in I\!N} \Rightarrow \bigcup_{i \in I\!N} A_i \in \mathcal{A}$

3. $\forall A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$

El par $(\Omega, \mathcal{A})$ se dirá *espacio medible.*

Obsérvese que con esta definición la intersección de sucesos del álgebra y el conjunto $\emptyset$ pertenecen al álgebra.

Se establece una correspondencia biunívoca entre conjuntos y sucesos que viene dada por la tabla 4.1.

Cálculo de Probabilidades	Teoría de Conjuntos
Suceso Seguro (Espacio muestral)	Conjunto Universal
Suceso Elemental	Punto del Conjunto Universal
Suceso	Subconjunto
Sucesos Incompatibles	Conjuntos Disjuntos
Unión de Sucesos	Unión de Conjuntos
Suceso Imposible	Conjunto Vacío
Suceso Contrario	Conjunto Complementario
Intersección de Sucesos	Intersección de Conjuntos
Sistema Completo	Partición

Tabla 4.1: Cálculo de probabilidades y teoría de conjuntos

4. Distintas definiciones del concepto de probabilidad

Continuando con el estudio de un experimento aleatorio y una vez que se han definido los sucesos se aprecia la necesidad de definir alguna medida que cuantifique la incertidumbre o la asiduidad de que un determinado suceso se obtenga al realizar un experimento aleatorio, a tal medida se le denomina *probabilidad.*

La dificultad de dar una definición del concepto de probabilidad sin objeciones o limitaciones, queda reflejada por los diferentes intentos realizados a lo largo de la historia para encontrar una definición de dicho concepto.

De la introducción histórica que se ha elaborado se desprende la existencia de tres definiciones del concepto de probabilidad que a continuación se discuten.

Definición clásica, debida a Laplace: La probabilidad de un suceso A es el cociente entre el número de casos favorables al suceso y el número de casos posibles.

Los inconvenientes de definir la probabilidad de esta forma son:

- No es válida cuando los sucesos elementales no son equiprobables.
- A veces no es posible contar.

Definición frecuentista, debida a Bernouilli: La probabilidad de un suceso es el valor límite de su frecuencia relativa al repetir indefinidamente la experimentación.

Los inconvenientes de definir así la probabilidad son los siguientes:

- Desde el punto de vista del análisis no puede interpretarse el límite anterior por la imposibilidad de fijar el número de repeticiones.
- En algunas ocasiones no es posible realizar una experimentación indefinida.
- Las condiciones bajo las cuales se realiza la experimentación pueden variar a lo largo del tiempo y, con ellas, las frecuencias relativas.

Para evitar los inconvenientes de ambas definiciones, además de un gran número de paradojas y dificultades surgidas a comienzos del presente siglo, se hizo necesaria una profunda revisión del concepto de probabilidad utilizando las herramientas más precisas del momento: *La teoría de conjuntos*, desarrollada principalmente por Borel, y la potente *Teoría de la medida*, debida a Lebesgue. A la luz de estas teorías, la probabilidad empieza a entenderse como una medida de la incertidumbre, con propiedades similares a las medidas de longitud, tiempo, etc.

Una concepción más operativa es definir la probabilidad como una medida personal de la incertidumbre de un suceso, basada en aquellos experimentos previos, que con la información disponible, se consideren indistinguibles o intercambiables. En situaciones repetitivas, cuando exista una amplia experiencia, la probabilidad viene determinada por la frecuencia relativa, mientras que, en otros casos, depende de distintos tipos de información.

Todo lo anterior llevó a Kolmogorov a introducir axiomáticamente el concepto de Probabilidad.

Definición axiomática de probabilidad, debida a Kolmogorov: Dado $(\Omega, \mathcal{A})$ un espacio medible finito, una función sobre $\mathcal{A}$, $P : \mathcal{A} \longrightarrow \mathbb{R}$ se dice *medida de probabilidad finita* o simplemente *probabilidad*, si cumple los siguientes axiomas:

A1. La probabilidad de un suceso es siempre mayor o igual a cero,

$$P(A) \geq 0$$

A2. La probabilidad del suceso seguro es igual a la unidad,

$$P(\Omega) = 1$$

A3. La probabilidad de la unión de dos sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos,

$$\forall A, B \in \mathcal{A} \mid A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

Definición axiomática de probabilidad: Dado $(\Omega, \mathcal{A})$ un espacio medible, una función sobre $\mathcal{A}$, $P : \mathcal{A} \longrightarrow \mathbb{R}$ se dice *medida de probabilidad infinita, de Kolmogorov* o simplemente *probabilidad*, si cumple los siguientes axiomas:

A1. La probabilidad de un suceso es siempre mayor o igual a cero,

$$P(A) \geq 0$$

A2. La probabilidad del suceso seguro es igual a la unidad,

$$P(\Omega) = 1$$

A3. La probabilidad de la unión de sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos,

$$\forall \{A_i\}_{i\in \mathbb{N}} \mid A_i \cap A_j = \emptyset, \forall i \neq j \qquad P(\bigcup_{i\in \mathbb{N}} A_i) = \sum_{i\in \mathbb{N}} P(A_i)$$

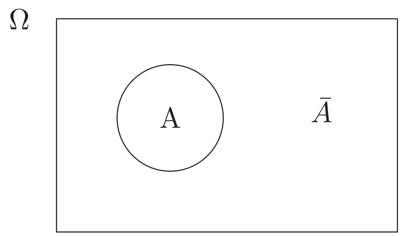
El gran inconveniente de estas definiciones es que no dan un método para el cálculo de probabilidades, por lo que en la práctica hay que basarse en las definiciones clásica y frecuentista.

5. Propiedades de la función de probabilidad

Como consecuencia de los axiomas se pueden deducir una serie de propiedades de la función de probabilidad, destacando las que se describen a continuación.

1. La probabilidad del suceso $\bar{A}$ es igual a uno menos la probabilidad del A,

$$P(\bar{A}) = 1 - P(A)$$

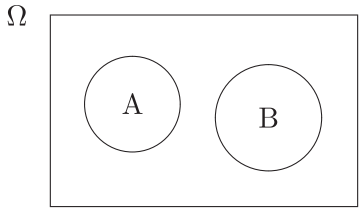


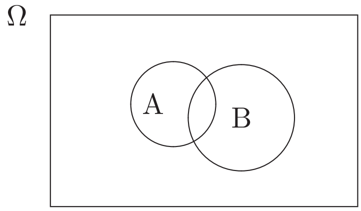
2. La probabilidad del suceso imposible es cero,

$$P(\emptyset) = 0$$

3. Si A y B son dos sucesos cualesquiera se verifica que:

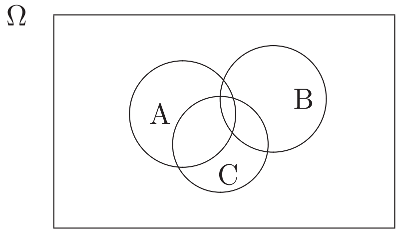
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



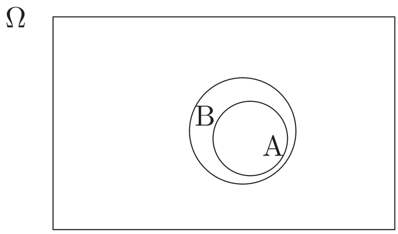


Si se tienen tres sucesos, A, B y C, la propiedad anterior tiene la forma siguiente:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$



4. Si $A \subseteq B$ entonces $P(A) \leq P(B)$



Ejercicio 4.2 *Demuestre las propiedades anteriores.*

6. Probabilidad condicionada. Independencia

La probabilidad de un determinado suceso en un experimento aleatorio puede verse modificada si se posee alguna información antes de la realización del experimento. Por ejemplo, si se consideran los alumnos de una clase de Estadística, la probabilidad de sacar aleatoriamente una alumna rubia será diferente de la de sacar una alumna rubia del grupo de las alumnas, es decir, si se parte del conocimiento de que el alumno escogido sea del sexo femenino. Para modelar este tipo de situaciones en las que se parte de una información a priori, se define el concepto de probabilidad condicionada.

Si $P(B) > 0$, la *probabilidad condicionada* de que se realice A si B se realiza, $P(A/B)$, viene definida por el cociente:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Análogamente, se define la probabilidad condicionada de B respecto a A. Utilizando conjuntamente ambos resultados se obtiene que:

$$P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$$

Ejemplo 4.12 *El 60 % de los alumnos de una clase de Estadística son chicas y se sabe que el 30 % de las chicas son rubias. ¿Cuál es la probabilidad de escoger un alumno de la clase que sea chica y rubia?*

Para resolver este ejemplo se consideran los siguientes sucesos:

$$\begin{aligned} F &= \{\textit{ser de sexo femenino}\} \\ M &= \{\textit{ser de sexo masculino}\} \\ R &= \{\textit{tener pelo rubio}\}. \end{aligned}$$

La probabilidad pedida es:

$$P(R \cap F) = P(R/F)P(F) = 0'3 \cdot 0'6 = 0'18$$

Teorema 4.1 *Si $P(B) > 0$, entonces $P(\cdot/B)$ es una probabilidad:*

A1. $P(A/B) \geq 0 \quad \forall A$

A2. $P(\Omega/B) = 1$

A3. *Para cualquier sucesión de sucesos disjuntos* $\{A_i\}_{i=1}^{n/\infty}$, *se verifica:*

$$P(\bigcup_{i=1}^{n/\infty} A_i/B) = \sum_{i=1}^{n/\infty} P(A_i/B)$$

Teorema 4.2 *Dados n sucesos, $A_1, A_2, \ldots, A_n$, se tiene:*

$$\begin{aligned} P(A_1 \cap A_2 \cap ... \cap A_n) = & \\ = \; & P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot \ldots \\ & \ldots \cdot P(A_n/A_1 \cap A_2 \cap ... \cap A_{n-1}) \end{aligned}$$

Ejemplo 4.13 *Siguiendo con el ejemplo 4.12 se sabe que el 40 % de las chicas rubias usan gafas. ¿Cuál es la probabilidad de que al escoger aleatoriamente un alumno de la clase resulte ser una chica rubia con gafas?*

Se define el suceso $G = \{\text{usar gafas}\}$ y la probabilidad pedida es:

$$\begin{aligned} P(F \cap R \cap G) &= P(G/(R \cap F))P(R/F)P(F) \\ &= 0'4 \cdot 0'3 \cdot 0'6 = 0'072. \end{aligned}$$

Ejercicio 4.3 *Demuestre los teoremas anteriores.*

7. Dependencia e independencia

En el epígrafe anterior se introdujo el concepto de probabilidad condicionada, debido a que la probabilidad de un determinado suceso se ve alterada por la información de que se dispone a priori. Sin embargo, puede suceder que dicha información no altere la probabilidad de ocurrencia de ese suceso, es decir, el que ocurra el suceso es independiente de la información.

Se dice que dos sucesos A y B son *independientes* si el conocimiento de la ocurrencia de uno no modifica la probabilidad de aparición del otro. O sea, $P(B/A) = P(B)$.

Ejercicio 4.4 *Compruebe que dos sucesos A y B son independientes si y sólo si $P(A \cap B) = P(A)P(B)$.*

La definición de independencia se puede generalizar a un número finito de sucesos. $A_1, A_2, \ldots, A_n$, se dicen mutuamente independientes si:

$$\begin{aligned}
&P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i, \forall j \neq i \\
&P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \quad \forall i, \forall j \neq i, \forall k \neq i, j \\
&\qquad\vdots \\
&P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)
\end{aligned}$$

Ejemplo 4.14 *Continuando con el ejemplo 4.12, si se sabe que el 10 % de los alumnos de la clase escriben con la mano izquierda. ¿Cuál es la probabilidad de escoger aleatoriamente una chica que escriba con la mano izquierda?*

Sea el suceso D definido por D={escribir con la mano izquierda}. Admitiendo que los sucesos D y F son independientes, la probabilidad solicitada es:

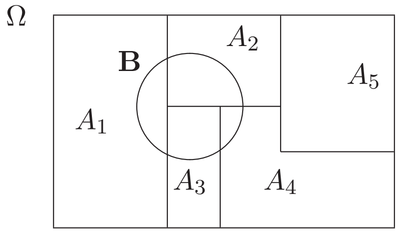
$$P(F \cap D) = P(F)P(D) = 0'6 \cdot 0'1 = 0'06.$$

8. Teorema de la probabilidad total. Teorema de Bayes

8.1. Teorema de la probabilidad total

Se considera un experimento que se realiza en dos etapas, en la primera se supone que los posibles sucesos, $A_1, A_2, \ldots, A_n$, constituyen un sistema completo, de tal forma que son conocidas las *probabilidades a priori*, $P(A_1), P(A_2), \ldots, P(A_n)$. Mientras que en la segunda etapa los resultados posibles, $\mathrm{B_j}$, tienen probabilidades desconocidas que dependen de lo que ocurre en la primera etapa. Si se conocen las probabilidades condicionadas $P(B/A_i)$ para un cierto suceso B y cada A_i se verifica que:

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$



La demostración de la igualdad anterior se basa en que al ser $A_1, A_2, \ldots, A_n$, una partición de Ω y B un elemento cualquiera del álgebra de sucesos $\mathcal{A}$, se tienen las siguientes igualdades:

$$\begin{aligned} B &= B \cap \Omega \\ &= B \cap (A_1 \cup A_2 \cup .... \cup A_n) \end{aligned}$$

por las propiedades distributiva y conmutativa:

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup .... \cup (A_n \cap B)$$

Como los sucesos $\{A_i\}_{i=1}^{n}$ son incompatibles, también lo son los sucesos $\{A_i \cap B\}_{i=1}^{n}$, por lo tanto se puede aplicar el axioma A3:

$$\begin{aligned} P(B) &= P((A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)) \\ &= P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_n \cap B) \\ &= \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(B/A_i)P(A_i) \end{aligned}$$

Se obtiene de esta forma la igualdad buscada.

Ejemplo 4.15 *Continuando con el ejemplo 4.12; si se sabe que el 20 % de los chicos son rubios. ¿Cuál es la probabilidad de escoger aleatoriamente una persona rubia?*

La probabilidad solicitada es:

$$\begin{aligned} P(R) &= P(R/M)P(M) + P(R/F)P(F) \\ &= 0'2 \cdot 0'4 + 0'3 \cdot 0'6 = 0'26. \end{aligned}$$

8.2. Teorema de Bayes

En los epígrafes anteriores hemos dado la idea intuitiva de probabilidad condicionada como la probabilidad de que ocurra un suceso sabiendo que ha ocurrido con anterioridad otro determinado suceso. Sin embargo, también se puede plantear la probabilidad de que se haya dado un determinado suceso sabiendo que como resultado final del experimento se ha obtenido otro determinado suceso.

En las mismas hipótesis del teorema anterior se tiene que:

$$P(A_k/B) = \frac{P(B/A_k)P(A_k)}{\displaystyle\sum_{i=1}^{n} P(B/A_i)P(A_i)}, \qquad k = 1, \ldots, n$$

Observe que se intenta calcular una probabilidad "antinatura", pues se pretende expresar lo que ocurre antes, A_k, en función de lo que ocurre después, B. De todas formas, lo anterior tiene sentido porque en algunas ocasiones se conoce el resultado final de un experimento, pero se desconocen algunos de los pasos intermedios, en los que se está interesado. El teorema de Bayes resuelve esta cuestión, llevando el cálculo de las probabilidades a un terreno más natural, expresando las probabilidades a posteriori, $P(A_i/B)$, en función de las *verosimilitudes*, $P(B/A_i)$.

Aplicando la definición de probabilidad condicionada:

$$P(A_k/B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B/A_k)}{P(B)}$$

Y por el teorema de la probabilidad total:

$$= \frac{P(A_k)P(B/A_k)}{\sum_{i=1}^{n} P(A_i)P(B/A_i)}$$

con lo que queda demostrado el teorema.

Ejemplo 4.16 *En el ejemplo 4.12, si se sabe que se ha elegido a una persona rubia, ¿cuál es la probabilidad de que sea chica?*

La probabilidad solicitada, a la vista del resultado del ejemplo 4.15 es

$$P(F/R) = \frac{P(R/F)P(F)}{P(R)} = \frac{0'3 \cdot 0'6}{0'26} = 0'69.$$

9. Ejercicios

9.1. Ejercicio resuelto

4.1 En una determinada ciudad se ha cometido un asesinato. De la investigación se encarga un detective, que tiene 5 sospechosos entre los que se encuentra el asesino. Se sabe que el detective trabaja con un pequeño margen de error, de forma que la probabilidad de creer inocente al verdadero asesino es de 0’05 y la probabilidad de creer culpable a una persona inocente es de 0’08. Si el detective cree que una persona es culpable, ¿cuál es la probabilidad de que esa persona sea el asesino?

Solución: Para la resolución del problema se definen los siguientes sucesos:

$$\begin{aligned} A &= \{\text{ser asesino}\} \\ I &= \{\text{ser enjuiciado inocente}\} \\ C &= \{\text{ser enjuiciado culpable}\} \end{aligned}$$

De esta forma se tiene que:

a) Hay un asesino de 5 sospechosos, por tanto, la probabilidad de que una persona elegida al azar sea el asesino es:

$$P(A) = \frac{1}{5} \qquad P(\bar{A}) = \frac{4}{5}$$

b) La probabilidad de creer inocente al verdadero asesino, es la probabilidad de ser enjuiciado inocente condicionada a que es el asesino, es decir, $P(I/A) = 0'05$. Además, se sabe que $P(C/A) = 1 - P(I/A) = 0'95$.

c) La probabilidad de creer culpable a una persona inocente, es la probabilidad de ser enjuiciado culpable condicionada a que no es el asesino, es decir, $P(C/\bar{A}) = 0'08$.

En el problema se pide la probabilidad de que una persona asesina haya sido enjuiciada culpable, es decir, la probabilidad de que una persona sea asesina condicionada a que ha sido enjuicida culpable. Por tanto, la probabilidad requerida es $P(A/C)$, para calcular dicha probabilidad se recurre al teorema de Bayes,

$$\begin{aligned} P(A/C) &= \frac{P(C/A)P(A)}{P(C/A)P(A) + P(C/\bar{A})P(\bar{A})} \\ &= \frac{0'95 \cdot 0'2}{0'95 \cdot 0'2 + 0'08 \cdot 0'8} \\ &= \frac{0'19}{0'254} = 0'748 \end{aligned}$$

9.2. Ejercicios propuestos

4.1. En una encuesta sobre las preferencias entre dos productos, realizada sobre un conjunto de 300 mujeres y 400 hombres, se han obtenido los siguientes resultados:

Producto	Hombres	Mujeres	Total
A	225	180	405
B	175	120	295

a) Represente la situación utilizando un diagrama de Venn.

b) Imagine que la encuesta ofrece información referida a dos conjuntos de edad, los menores y los mayores de 50 años. ¿Sería posible la representación incluyendo esta nueva información? De ser afirmativa la respuesta, represéntela.

4.2. Un estudiante de Estadística se dispone a realizar un estudio sobre el tipo y las condiciones de la comida que su madre le sirve a diario. Para ello establece las siguientes clasificaciones:

Estado de sal	Salada, normal, sosa
Temperatura	Caliente, fría
Tipo de alimento	Carne, pescado, verduras, pastas

Obtenga, utilizando un diagrama de árbol, el espacio muestral del tipo y las condiciones de las comidas.

4.3. Imagine que tenemos los sucesos A, B y C, exprese en lenguaje de la teoría de conjuntos las siguientes operaciones entre ellos:

a) Ocurren A y al menos uno de los otros dos.
b) Ocurren A y uno sólo de los otros dos.
c) Ocurre uno de los tres, pero no dos a la vez.
d) Ocurre, al menos, uno de los tres.
e) Ocurre C, pero no lo hacen ni A ni B.
f) Ocurren al menos dos de los tres.
g) Ocurren exactamente dos de los tres.
h) No ocurre ninguno de los tres.

4.4. Los alumnos de una determinada carrera se encuentran distribuidos en 5 cursos, de forma que en cada uno de los dos últimos cursos hay la mitad de alumnos que en cada uno de los tres primeros. Se pide que se calcule la probabilidad de que al escoger al azar a un alumno:

a) éste sea de cuarto.
b) le queden menos de tres cursos para acabar.

4.5. Represente el espacio muestral resultante al lanzar dos dados de distinto color y calcule la probabilidad de obtener una suma de siete puntos.

4.6. Represente el espacio muestral resultante al lanzar dos dados del mismo color y calcule la probabilidad de obtener una suma de siete puntos. Compare el resultado con el obtenido en el ejercicio anterior.

4.7. Un jugador lanza tres veces una moneda, si obtiene tres caras gana 100€ si obtiene una o dos caras gana 10€ y si no obtiene ninguna cara pierde 160€. ¿Es justo el juego?

4.8. Juan y Pedro juegan a una variante del juego de los chinos. Cada uno de ellos tiene tres chinos pudiendo seleccionar en una mano ninguno, uno, dos o los tres. A una señal los dos muestran los chinos seleccionados. Juan gana 10€ si sus chinos coinciden con los de Pedro o hay una diferencia de un único chino, mientras que Pedro gana 15€ en el resto de casos.

a) Calcule la probabilidad de que gane Juan.

b) ¿Qué cantidades deben ganar cada uno para que el juego sea justo?

4.9. Calcule la probabilidad de que tres alumnos seleccionados aleatoriamente en una clase cumplan años en meses consecutivos.

4.10. Calcule la probabilidad que tiene un ladrón que ha robado una tarjeta de un cajero automático de acertar con la clave, sabiendo que ésta tiene cuatro dígitos y que si no acierta en tres intentos el cajero se tragará la tarjeta.

4.11. Imagine que se encuentra un procedimiento que genera aleatoria e indefinidamente letras y signos de puntuación. ¿Cuál es la probabilidad de que un cierto momento escriba la novela *"El Quijote"*?

4.12. Se parte de que $P(A) = 0'3$, $P(B) = 0'4$ y $P(A \cap B) = 0'1$, obtenga:

a) $P(\bar{A} \cap \bar{B})$
b) $P(\bar{A} \cap B)$
c) $P(A - B)$
d) $P(A/B)$

4.13. Se considera el conjunto universal $\Omega = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$ y los sucesos $A_1 = \{(x, y) \in \Omega : 0 \leq x \leq \frac{3}{4}\}$, $A_2 = \{(x, y) \in \Omega : \frac{1}{2} \leq x \leq 1\,; \frac{1}{4} \leq y \leq \frac{3}{4}\}$ y $A_3 = \{(x, y) \in \Omega : 0 \leq x \leq \frac{1}{2}\}$.

a) Pruebe que la función $P(A) = \text{Área}(A)$, $\forall A \subseteq \Omega$ es una función de probabilidad.

b) Calcule $P(A_1 \cup A_2 \cup A_3)$, $P(\bar{A}_1 \cap \bar{A}_2)$ y $P(A_1 \cap (A_2 \cup A_3))$.

4.14. Sea una clase de estadística en la que un 20 % de los varones son rubios y un 50 % de las mujeres rubias. Si se sabe que el 30 % de la clase son varones, se pide:

a) La probabilidad de escoger aleatoriamente de la clase un varón rubio.

b) La probabilidad de escoger aleatoriamente una persona rubia de entre todos los alumnos.

c) La probabilidad de que una persona que se ha elegido aleatoriamente sea varón sabiendo que su pelo es rubio.

4.15. ¿Cuál es la probabilidad de que al tirar tres dados honrados salgan números diferentes?

4.16. Se tienen dos barajas de cartas de forma que la primera tiene 30 cartas rojas, 10 blancas y 2 negras y la segunda tiene 20 cartas rojas, 10 blancas y 12 negras. Se lanza una moneda, si sale cara se escogen tres cartas de la primera y una de la segunda; si sale cruz se escoge una carta de la primera y tres de la segunda. Calcule la probabilidad de que de las cuatro cartas extraídas dos sean blancas y las otras dos rojas.

4.17. Un estudio sobre los niveles de audiencia de diferentes cadenas de radio arrojó que el 50 % de la población escuchaba Radio A, el 40 % Radio B y el 30 % Radio C. Además, se obtuvo que el 20 % escuchaba Radio A y Radio B, el 10 % Radio A y Radio C y el 5 % Radio B y Radio C, finalmente sólo el 2 % escuchaba las tres cadenas.

a) ¿Qué porcentaje de la población escuchaba alguna cadena?

b) ¿Qué porcentaje de la población escuchaba una sola cadena?

4.18. En un programa de televisión existe una prueba que consiste en ordenar cronológicamente cinco inventos. El número de aciertos es el número de coincidencias entre las posiciones correctas y las ordenadas por el concursante. ¿Cuál es la probabilidad de que un concursante tenga al menos un acierto, sabiendo que realiza la ordenación de los inventos al azar?

4.19. Se consideran dos sucesos cualesquiera A y B, se pide:

a) Pruebe que si A y B son independientes entonces $\bar{A}$ y B, A y $\bar{B}$, y, $\bar{A}$ y $\bar{B}$ también lo son.

b) Demuestre que si $P(A) = 0$ entonces A y B son independientes.

c) Si A y B son dos sucesos disjuntos, ¿lo son también $\bar{A}$ y $\bar{B}$?

4.20. Se considera un equipo deportivo en octavos de final de una competición, que tiene una probabilidad de pasar a las siguientes fases de $\frac{4}{5}$, $\frac{3}{4}$ y $\frac{2}{3}$, respectivamente, y de $\frac{1}{2}$ de ganar la final si accede a ella, ¿cuál es la probabilidad de que gane la competición?

4.21. Un jugador de baloncesto tiene una probabilidad de encestar un lanzamiento desde una cierta posición de $\frac{1}{4}$.

a) ¿Cuál es la probabilidad de encestar tres lanzamientos consecutivos?

b) ¿Cuál es la probabilidad de que en cinco lanzamientos enceste al menos tres?

4.22. De una urna con tres bolas blancas y dos negras se extrae una bola y a continuación se lanza un dado, de forma que se introducen en la urna tantas bolas del mismo color que la extraída como el resultado obtenido al lanzar el dado. ¿Cuál es la probabilidad de que, una vez realizada esta operación, al extraer dos nuevas bolas, éstas tengan el mismo color?

4.23. Dos amigos son alumnos de la asignatura de Estadística de forma que cuando uno falta le pasa los apuntes al otro. Se sabe que el primero va a asistir a un 80 % de las clases y el segundo a un 40 %, de forma independiente. ¿Cuál es la probabilidad de que los amigos tengan todos los apuntes de clase?

4.24. Se considera una urna en la que hay 4 dados, de forma que en el primero 3 caras son unos y las restantes son doses, en el segundo 4 caras son unos y el resto doses, en el tercero 5 caras son unos y la otra un dos y en el cuarto 2 caras son unos y el resto doses.

a) ¿Cuál es la probabilidad de que al elegir un dado al azar y lanzarlo se obtenga un uno?

b) Se coge al azar un dado de la urna y al lanzarlo se obtiene un uno, ¿cuál es la probabilidad de que sea el cuarto dado?

4.25. De una urna que contiene cinco bolas blancas y tres negras, se extraen al azar cuatro bolas que se introducen en otra urna vacía, de esta urna se sacan aleatoriamente dos bolas que resultan ser una blanca y una negra. ¿Cuál es la probabilidad de que de las cuatro bolas pasadas, dos fueran blancas y las otras dos negras?

4.26. En una piscina de una piscifactoría se han introducido alevines de dos variedades de una especie en las siguientes cantidades y proporciones de machos y hembras:

Variedad	Cantidad	% machos
A	1000	7
B	1500	6

A continuación se escoge un alevín, ¿cuál es la probabilidad de que pertenezca a la variedad A, sabiendo que es hembra?

4.27. Una factoría produce un cierto artículo en tres cadenas de montaje. La cadena A fabrica el 50 % del total, la cadena B el 30 % y la C el 20 %, con porcentajes de defectuosos 0'03, 0'04 y 0'05 respectivamente. Un cliente decide analizar la calidad del producto para lo que selecciona una unidad al azar, ¿qué probabilidad hay de que dicha unidad resulte ser defectuosa?

4.28. Un niño guarda tres cajas con chocolatinas, en la primera tiene dos chocolatinas negras y una blanca, en la segunda dos negras y dos blancas y en la tercera dos blancas y una negra. En un despiste suyo, su hermana pequeña le ha cogido una chocolatina blanca, ¿cuál es la probabilidad de que la haya cogido de la primera caja?

Capítulo 5

Variable aleatoria

1. Concepto

En el capítulo anterior se introducía el concepto de probabilidad definido sobre el álgebra de Boole de los sucesos. Sin embargo, este concepto presenta el inconveniente de que no es susceptible de un buen manejo matemático, debido fundamentalmente a la diversidad de las categorías de los resultados de un experimento, de ahí que sea necesario realizar una abstracción cuantificada de dicho experimento que permita agrupar los sucesos según determinadas características comunes y consecuentemente se facilite su manejo y la aplicación del análisis matemático. Dicha abstracción se realiza asignando un número real a cada suceso del espacio muestral. La función mediante la cual se realiza esta correspondencia recibe el nombre de *variable aleatoria*. Más formalmente, se define la variable aleatoria como cualquier función medible que asocia a cada suceso un número real.

Profundizando en la idea de variable aleatoria como abstracción de los resultados de un experimento aleatorio, y puesto que cada suceso tiene una determinada probabilidad de ocurrencia, se puede trasladar dicha probabilidad al valor correspondiente de la variable aleatoria, por lo que se puede hablar de la probabilidad de que una variable aleatoria tome un determinado valor. Así,

1. Al lanzar una moneda, se le puede asociar el valor 1 al suceso elemental "cara" y el valor 0 al suceso "cruz".

2. Al lanzar dos dados al aire se puede asociar a cada resultado la suma de los puntos obtenidos.

Es importante observar que la asignación de valores a los resultados del experimento no es única, de hecho basta con fijar valores distintos a resultados distintos, para obtener una infinidad de funciones. No obstante, la idea es que dicha asignación sea lo más natural posible para que una vez manipulada la variable aleatoria los resultados sean fácilmente interpretables en términos del experimento de partida.

2. Variables discretas y continuas

Una variable aleatoria, en lo sucesivo v.a., se denomina *discreta* si toma valores aislados o puntuales.

Ejemplo 5.1 *La variable X="Número de hijos varones de una familia de 2 hijos", puede tomar los valores 0, 1 y 2. Si se desea calcular la probabilidad de que X tome cada uno de sus valores posibles, suponiendo que la probabilidad de que un hijo sea varón es 0'49, y en el supuesto de que los sucesos sean independientes, se tiene que:*

$$\begin{aligned} P(X=2) &= 0'49 \cdot 0'49 = 0'2401 \\ P(X=1) &= 0'49 \cdot 0'51 + 0'51 \cdot 0'49 \\ &= 0'4998 \\ P(X=0) &= 0'51 \cdot 0'51 = 0'2601\ . \end{aligned}$$

Siendo la suma de estas probabilidades, como era de esperar, la unidad.

Una v.a. es *continua* si puede tomar cualquier valor dentro de uno o varios intervalos determinados. Así, la variable X que asocia a cada individuo de un colectivo su estatura es continua; en este caso, la probabilidad de que la variable tome exactamente un valor determinado, 160 cms. por ejemplo, es cero. Esto tiene su justificación intuitiva, desde el punto de vista de la regla de Laplace, en que el número de casos favorables es uno sólo, mientras que el número de casos posibles es infinito. De hecho, esa es la principal característica de la variable continua y obligará a darle un tratamiento especial.

Para terminar, se puede considerar una v.a. *mixta*, la cual toma valores dentro de uno o varios intervalos y algunos valores puntuales más fuera de él o de ellos.

Lo dicho hasta ahora vale tanto para experimentos simples –lanzar un dado y anotar el número de la cara superior–, como para experimentos complejos –medir el peso, la estatura y la edad en años de un grupo de personas–. En el primer caso la variable asociada será unidimensional –discreta en el ejemplo–, y en el segundo multidimensional, tridimensional, continua para las dos primeras dimensiones y discreta para la edad.

3. Variables unidimensionales

3.1. Caracterización de variables aleatorias

Como se ha visto, la v.a. es una abstracción numérica que se hace de los resultados de un experimento aleatorio, y puesto que cada suceso tiene una determinada probabilidad de ocurrencia se traslada dicha probabilidad al valor correspondiente de la v.a. Si la variable es discreta y toma pocos valores distintos, como en el ejemplo de los hijos varones, es factible, e incluso conveniente, dar todos esos valores con sus probabilidades de una forma explícita. Pero si la variable es discreta y toma muchos valores diferentes (tal vez infinitos) o si es continua, lo anterior es poco recomendable o incluso imposible. Por ello es necesario apoyarse en una serie de funciones, relacionadas íntimamente con dichas probabilidades, que nos permitan resolver el problema. Estas funciones son la *función de cuantía* en el caso discreto, la de *densidad* en el continuo; así como, la de *distribución*, la *característica* y la *generatriz de momentos*, entre otras, en ambos casos. Este apartado se centra en las tres primeras.

3.1.1. Caso discreto

Si una v.a. toma valores $x_1, x_2, \ldots, x_n, \ldots$, (finitos o infinitos) la regla que asocia a cada uno de ellos las probabilidades $p_1, p_2, \ldots, p_n, \ldots$, respectivamente, donde $p_i = P(X = x_i)$ se denomina *función de cuantía.* Como la suma de todas las probabilidades de los sucesos elementales es uno, se tiene que:

$$\sum_{i=1}^{(n/\infty)} p_i = 1 \ .$$

Una v.a. discreta queda perfectamente determinada cuando se conoce su función de cuantía, pudiéndose expresar ésta de dos formas, por extensión o a través de una función.

Ejemplo 5.2 *Utilizando la distribución del ejemplo 5.1, la función de cuantía puede darse como:*

x	$P(X=x)$
0	$0'2601$
1	$0'4998$
2	$0'2401$

O bien:

$$P(X = x) = \binom{2}{x} 0'49^x \cdot 0'51^{2-x}$$

para $x = 0, 1, 2.$

Otra forma de caracterizar una v.a. es a través de la llamada *Función de Distribución*, definida por:

$$F(x) = P(X \leq x) \ .$$

La función de distribución en un valor x, es la probabilidad de que X tome valores menores o iguales a x. Es decir, es una función que acumula toda la probabilidad entre menos infinito y el punto donde está definida.

Propiedades de la función de distribución. La función de distribución de una variable aleatoria discreta cumple las siguientes propiedades (figura 5.1):

1. Está definida en toda la recta real.
2. Es no decreciente y no negativa.
3. Toma el valor cero en menos infinito y el valor uno en más infinito.
4. Sólo tiene discontinuidades de salto -precisamente en los puntos donde la función de cuantía es distinta de cero-.

Ejemplo 5.3 *Algunos valores de la función de distribución para el ejemplo que se arrastra son:*

$$
\begin{aligned}
F(-0'5) &= P(X \leq -0'5) = 0 \\
F(0) &= P(X = 0) = 0'2601 \\
F(1) &= P(X = 0) + P(X = 1) = 0'7599 \\
F(2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= 1 \\
F(2'5) &= 1 \; .
\end{aligned}
$$

Y de aquí es fácil ver que:

$$
F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 0'2601 & \text{si} \quad 0 \leq x < 1 \\ 0'7599 & \text{si} \quad 1 \leq x < 2 \\ 1 & \text{si} \quad x \geq 2 \; . \end{cases}
$$

Esta función se representa gráficamente en la figura 5.1.

3.1.2. Caso continuo

Como se dijo en su presentación, la característica principal de la variable continua es que la probabilidad de que tome un determinado valor es cero. Sin embargo, aún siendo ello cierto, no todos los valores tienen "la misma ocurrencia". Se puede comprobar que hay bastantes

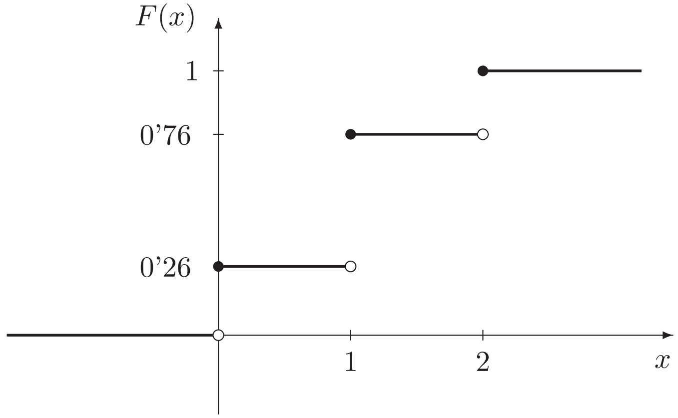


Figura 5.1: Función de distribución discreta.

individuos que tienen estatura alrededor de 160 cms, muchos menos alrededor de 210 cms y ninguno en el entorno de los 350 cms. Lo anterior hace que interese, no ya la probabilidad de tomar exactamente un valor, sino la probabilidad de que la variable se encuentre en el entorno de un punto, es decir, que tome valores dentro de un intervalo. Así, se considera un intervalo pequeño centrado en 160 cms. y de amplitud Δ, sea éste:

$$[160 - \Delta/2, 160 + \Delta/2]$$

con lo que la probabilidad de que X tome algún valor del intervalo será pequeña, pero no nula. Si a continuación se levanta sobre el intervalo considerado un rectángulo de altura $h(160)$, tal que su área sea igual al valor de la probabilidad anterior, se tiene:

$$\text{Área} = h(160) \cdot \Delta = P(160 - \Delta/2 \leq X \leq 160 + \Delta/2) \; .$$

De ahí que la altura del rectángulo pueda asociarse a una densidad de probabilidad, por ser igual al cociente entre la probabilidad del intervalo y su tamaño.

Si se seleccionan ahora valores x_i de la v.a. X, distantes entre sí una distancia Δ y se determinan las alturas $h(x_i)$ de los infinitos intervalos así obtenidos se obtiene un histograma y uniendo los centros de las caras superiores de los rectángulos, el correspondiente polígono de frecuencias (figura 5.2).

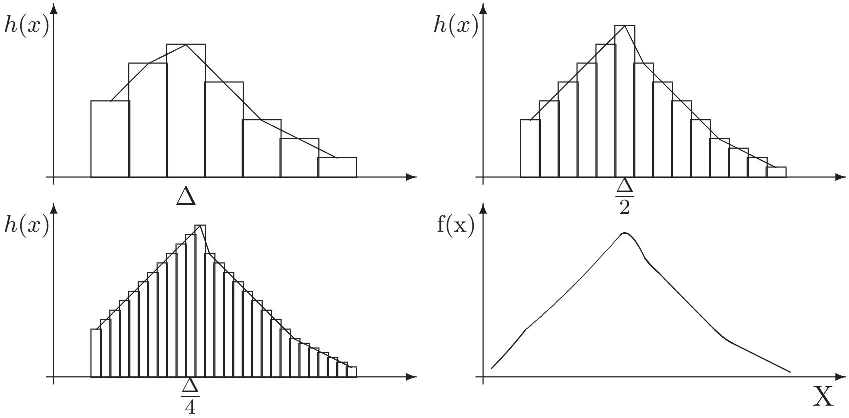


Figura 5.2: Obtención esquemática de la función de densidad.

La suma de las áreas de los rectángulos debe ser uno, por ser igual a la probabilidad de que X tome cualquier valor. Si se hace que la anchura de los rectángulos tienda a cero, el polígono de frecuencias se transforma en una curva (figura 5.2). Dicha curva es la representación gráfica de la función f, denominada *función de densidad* de la v.a. X, obtenida como límite de los valores h:

$$\lim_{\Delta \to 0} h(x) = \lim_{\Delta \to 0} \frac{P(x - \Delta/2 \leq X \leq x + \Delta/2)}{\Delta} = f(x) \; .$$

Si se desea calcular la probabilidad de que la variable se encuentre entre dos valores dados, ésta es igual al área bajo la curva f entre a y b (figura 5.3). Es decir:

$$P(a \leq X \leq b) = \int_a^b f(x) \, \mathrm{d}\, x \; .$$

Lógicamente se verifica que:

$$P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(x) \, \mathrm{d}\, x = 1 \; . \tag{5.1}$$

Obviamente f es no negativa. Cualquier función no negativa que verifique la condición dada en (5.1) será una función de densidad.

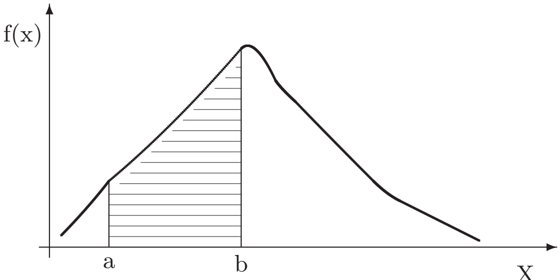


Figura 5.3: Probabilidad de que X tome valores entre a y b.

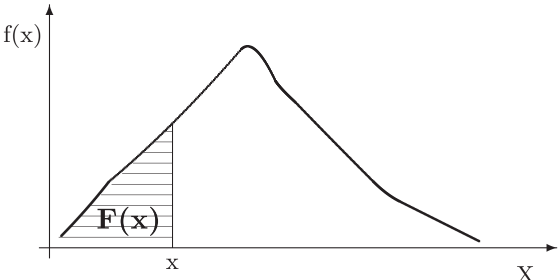


Figura 5.4: Función de distribución de una variable continua.

Al igual que en el caso discreto, la *Función de Distribución*, F, de una variable aleatoria X, se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) \, \mathrm{d}\, u \ .$$

La función de distribución en x da la probabilidad de que X tome valores menores o iguales a x, dicho valor es el área de la superficie rayada en la figura 5.4.

Las propiedades de la función de distribución en el caso continuo son idénticas a las del caso discreto con la única excepción de que ahora ésta no presenta discontinuidades de salto. Para ciertas variables continuas de uso frecuente, los valores de F se encuentran tabulados, lo cual facilita considerablemente el cálculo de probabilidades. Para terminar este epígrafe, se muestra la relación existente entre las funciones definidas:

$$P(a < X < b) = P(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}\,x = F(b) - F(a)\ .$$

Ejemplo 5.4 *Si*

$$f(x) = \begin{cases} 1 - x & \text{si} \quad 0 \le x \le 1 \\ k & \text{si} \quad 2 \le x \le 3 \\ 0 & \text{en caso contrario.} \end{cases}$$

1. *Para que f sea función de densidad debe verificar:*

   a) $f(x) \ge 0, \quad \forall x$

   b) $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\,x = 1\ .$

   La primera propiedad se cumple siempre que $k \ge 0$. Se cuestiona para qué valor de k se verifica la segunda propiedad.

   $$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\,x &= \int_0^1 (1-x)\,\mathrm{d}\,x + \int_2^3 k\,\mathrm{d}\,x \\ &= \left(x - \frac{x^2}{2}\right)\Bigg|_0^1 + kx|_2^3 \\ &= 1 - \frac{1}{2} + 3k - 2k\ . \end{aligned}$$

   Por tanto, para que f sea función de densidad, debe verificarse que $1 - \frac{1}{2} + 3k - 2k = 1$ o equivalentemente $k = \frac{1}{2}$.

2. *La función de distribución de la variable anterior es:*

   - *Si $x < 0$ entonces $F(x) = P(X \le x) = \int_{-\infty}^{x} 0\,\mathrm{d}\,x = 0$.*
   - *Si $0 \le x < 1$ entonces $F(x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}\,x = \int_{-\infty}^{0} 0\,\mathrm{d}\,x + \int_0^x (1-x)\,\mathrm{d}\,x = x - \frac{x^2}{2}$.*

- *Si* $1 \leq x < 2$ *entonces* $F(x) = \int_{-\infty}^{0} 0 \,\mathrm{d}\, x + \int_0^1 (1-x) \,\mathrm{d}\, x + \int_1^x 0 \,\mathrm{d}\, x = \frac{1}{2}$.
- *Si* $2 \leq x < 3$ *entonces* $F(x) = \int_{-\infty}^{0} 0 \,\mathrm{d}\, x + \int_0^1 (1-x) \,\mathrm{d}\, x + \int_1^2 0 \,\mathrm{d}\, x + \int_2^x \frac{1}{2} \,\mathrm{d}\, x = \frac{x-1}{2}$.
- *Si* $x \geq 3$ *entonces* $F(x) = 1$.

Resumiendo:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ x - \frac{x^2}{2} & \text{si} \quad 0 \leq x < 1 \\ \frac{1}{2} & \text{si} \quad 1 \leq x < 2 \\ \frac{x-1}{2} & \text{si} \quad 2 \leq x < 3 \\ 1 & \text{si} \quad x \geq 3 \; . \end{cases}$$

3.2. La función esperanza matemática

Si X es una v.a. discreta que toma los valores $x_1, x_2, \ldots$, con probabilidades $p_1, p_2, \ldots$, se define la *esperanza* de X como:

$$E[X] = \sum_{i \geq 1} x_i p_i \; .$$

Si g es una función de la v.a. X, se define la esperanza de $g(X)$ como:

$$E[g(X)] = \sum_{i \geq 1} g(x_i) p_i \; .$$

Ejemplo 5.5 *Para calcular el valor medio de la variable del ejemplo 5.1.*

$$E[X] = 0 \cdot 0'2601 + 1 \cdot 0'4998 + 2 \cdot 0'2401 = 0'98 \; .$$

La esperanza de la variable $Y = X^3$ *vale:*

$$E[Y] = 0^3 \cdot 0'2601 + 1^3 \cdot 0'4998 + 2^3 \cdot 0'2401 = 2'42 \; .$$

Si X es una v.a. continua con función de densidad f, se define la esperanza de X como:

$$E[X] = \int_{-\infty}^{+\infty} x f(x) \, \mathrm{d}\, x \ ,$$

Si $g(X)$ es una función de la v.a. X, se define la esperanza de $g(X)$ como:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) \, \mathrm{d}\, x \ .$$

En los dos últimos casos la definición es válida sólo si existe la integral. A partir de ahora, y salvo que la situación así lo requiera, tan sólo se expresan los resultados para variables continuas.

Ejemplo 5.6 *Para la variable del ejemplo 5.4 la esperanza vale:*

$$\begin{aligned} E[X] &= \int_0^1 (x - x^2) \, \mathrm{d}\, x + \int_2^3 \frac{x}{2} \, \mathrm{d}\, x \\ &= \left(\frac{x^2}{2} - \frac{x^3}{3} \right) \Bigg|_0^1 + \frac{x^2}{4} \Bigg|_2^3 = \frac{17}{12} \ . \end{aligned}$$

Propiedad 5.1 *Si a es una constante cualquiera, se verifica:*

1. *E[aX] = aE[X].*
2. *E[a+X] = a + E[X].*

Ambas propiedades son fácilmente demostrables sin más que aplicar la definición de esperanza.

Ejemplo 5.7 *Continuando con el ejemplo 5.4:*

1. *Si se considera $Y = 12X$, entonces $E[Y] = 17$.*
2. *Si ahora $Y = \frac{7}{12} + X$, entonces $E[Y] = 2$.*

3.3. Caracterización parcial de variables aleatorias

A partir de la función esperanza se introducen una serie de funciones que ofrecen visiones parciales de la distribución en estudio. Si X es una v.a. se definen sus momentos de orden k respecto al origen y respecto a la media, respectivamente, como:

$$\alpha_k = E[X^k] \qquad\qquad \mu_k = E[(X - E[X])^k] \ .$$

- *La media y la varianza.* Especial importancia tienen el momento de orden 1 respecto al origen y el momento de orden 2 respecto a la media:

  $$\alpha_1 = E[X] = \mu \quad V[X] = \mu_2 = E[(X-\mu)^2] = \sigma^2 \ .$$

  El primero de ellos es la *media*, esperanza o valor esperado de X y el segundo es la *varianza* de X, coincidiendo su raíz cuadrada positiva con la desviación típica.

  Igual que en estadística descriptiva, se pueden expresar los momentos respecto a la media en función de los momentos respecto al origen. En particular la varianza puede expresarse como:

  $$V[X] = \mu_2 = \sigma^2 = E[X^2] - E[X]^2 = \alpha_2 - \alpha_1^2.$$

Ejemplo 5.8 *Dada la variable:*

X	0	1	2	3
$p(x)$	$0'2$	$0'2$	$0'3$	$0'3$

Se sabe que $E[X] = 0'2 + 0'6 + 0'9 = 1'7$, *por tanto la varianza vale:*

$$\begin{aligned} V[X] &= 0 \cdot 0'2 + 1 \cdot 0'2 + 4 \cdot 0'3 + 9 \cdot 0'3 \\ &\quad -1'7^2 = 1'21 \ . \end{aligned}$$

Ejemplo 5.9 *Para una v.a. cuya función de distribución es:*

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 1 \\ x-1 & \text{si} \quad 1 \leq x < 2 \\ 1 & \text{si} \quad x \geq 2 \ . \end{cases}$$

Su varianza viene dada por:

$$V[X] = E[X^2] - (E[X])^2$$

Y puesto que:

$$\begin{aligned} E[X^2] &= \int_1^2 x^2 \, \mathrm{d}\, x = \left. \frac{x^3}{3} \right|_1^2 \\ &= \frac{8}{3} - \frac{1}{3} = \frac{7}{3} \,. \\ E[X] &= \int_1^2 x \, \mathrm{d}\, x = \left. \frac{x^2}{2} \right|_1^2 \\ &= 2 - \frac{1}{2} = \frac{3}{2} \,. \end{aligned}$$

se tiene que:

$$\begin{aligned} V[X] &= \frac{7}{3} - \left(\frac{3}{2}\right)^2 \\ &= \frac{7}{3} - \frac{9}{4} = \frac{28-27}{12} = \frac{1}{12} \,. \end{aligned}$$

Ejercicio 5.1 *Demuestre las siguientes propiedades de la varianza:*
a) $V[aX] = a^2 V[X]$
b) $V[a + X] = V[X]$.

En este punto es interesante analizar como quedan definidos algunos de los conceptos que se vieron en estadística descriptiva:

- *La moda.* Es el valor que más se repite, es decir, el de mayor probabilidad si la variable es discreta, o el de mayor densidad si es continua. En el primer caso la moda es el valor x_i, tal que p_i es mayor o igual que p_j para todo j distinto de i. Si la variable es continua, la moda coincide con el valor de X que maximiza la función de densidad, debiéndose utilizar los procedimientos analíticos de obtención de puntos óptimos. Como en el caso descriptivo, también aquí son válidas las nociones de modas múltiples y, por tanto, de modas relativas.

Ejemplo 5.10 *Para las distribuciones de los ejemplos 5.8 y 5.9 las modas obtenidas son:*

a) *Para la variable discreta existen dos modas, 2 y 3, pues la probabilidad, en ambos casos, es 0'3, mayor a la de cualquier otro valor.*

b) *En el caso continuo, $f(x) = 1$ en el intervalo $[1, 2]$, por tanto, la moda es todo el intervalo $[1, 2]$.*

- *La mediana.* Es el punto central de la distribución y coincide con el valor de x tal que $F(x) = 0'5$. En el caso de variable discreta la similitud es total con la definición dada en descriptiva.

  Ejemplo 5.11 *Siguiendo con los mismos ejemplos, para el caso discreto, trivialmente, $Me = 2$. Y para el caso continuo:*
  $$F(x) = \tfrac{1}{2} \Leftrightarrow x - 1 = \tfrac{1}{2} \Leftrightarrow Me = \tfrac{3}{2} \ .$$

- *Coeficientes de simetría y de curtosis.* Se definen respectivamente como:
  $$\gamma_1 = \frac{\mu_3}{\sigma^3}; \qquad \gamma_2 = \frac{\mu_4}{\sigma^4} \ .$$
  Siendo la discusión sobre ambos coeficientes la misma que en el caso descriptivo.

- *Normalización o tipificación.* Se dice que una v.a. está tipificada o normalizada cuando su media vale cero y su desviación típica uno. Para tipificar una variable X se le resta su media y se divide por su desviación típica:
  $$Z = \frac{X - \mu}{\sigma} \ .$$

3.4. Función característica y generatriz de momentos

A partir de la esperanza matemática, se pueden definir dos funciones que caracterizan totalmente a la distribución de probabilidad. Estas funciones son la función característica y la función generatriz de momentos.

1. *Función característica.*

$$\varphi_X(t) = E[e^{iXt}] = \int_{-\infty}^{+\infty} e^{ixt} f(x)\,\mathrm{d}\,x \ .$$

2. *Función generatriz de momentos.*

$$m_X(t) = E[e^{Xt}] = \int_{-\infty}^{+\infty} e^{xt} f(x)\,\mathrm{d}\,x \ .$$

Ejemplo 5.12 *Para la v.a. continua del ejemplo 5.9:*

- *La función característica es:*

$$\begin{aligned} \varphi_X(t) &= E[e^{itx}] = \int_1^2 e^{itx}\,\mathrm{d}\,x \\ &= \left.\frac{e^{itx}}{it}\right|_1^2 = \frac{e^{2it}}{it} - \frac{e^{it}}{it} = \frac{e^{it}}{it}(e^{it}-1) \ . \end{aligned}$$

- *La función generatriz de momentos es:*

$$\begin{aligned} m_X(t) &= E[e^{tx}] = \int_1^2 e^{tx}\,\mathrm{d}\,x \\ &= \left.\frac{e^{tx}}{t}\right|_1^2 = \frac{e^{t}}{t}(e^{t}-1) \ . \end{aligned}$$

Esta última función además de caracterizar a la distribución, permite, como su propio nombre indica, obtener fácilmente cualquier momento. Así:

$$\alpha_k = \frac{d^k m_X(t)}{dt^k}\ |_{t=0} \ .$$

En sentido inverso, el conocimiento de los momentos permite obtener la función generatriz de momentos.

3.5. Cambio de variable

Sea X una v.a. con función de densidad f e $Y = h(X)$ una función de X estrictamente monótona. Entonces, la función de densidad de la nueva variable Y viene dada por:

$$g(y) = f(x) \left| \frac{\mathrm{d}\, x}{\mathrm{d}\, y} \right| .$$

Ejemplo 5.13 *Sea X una v.a. con función de densidad:*

$$f(x) = \begin{cases} 2(x-1) & \text{si} \quad 1 < x < 2 \\ 0 & \text{en el resto} \ . \end{cases}$$

Dada $Y = \frac{X+3}{5}$, puesto que:

$$X = 5Y - 3 \Rightarrow \frac{\mathrm{d}\, x}{\mathrm{d}\, y} = 5 \ ,$$

la función de densidad de Y se obtiene como:

$$g(y) = \begin{cases} 50y - 40 & \text{si} \quad \frac{4}{5} < y < 1 \\ 0 & \text{en el resto} \ . \end{cases}$$

3.6. Desigualdad de Tchebychev

Sea X es una v.a. y k una constante positiva, se verifica que:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2} \ .$$

Esta expresión permite conocer la proporción mínima de valores de X que distan de la media, un máximo k veces la desviación típica. Para k igual a tres, por ejemplo, la desigualdad garantiza que al menos el 89 % de la distribución está en el intervalo $\mu \pm 3\sigma$.

Ejemplo 5.14 *Un fabricante de frigoríficos utiliza una máquina para introducir gas en dichos aparatos. La máquina no es perfecta y por tanto no introduce cantidades iguales en cada frigorífico, aunque se conoce que la* *cantidad media de gas que se introduce por aparato es igual a 15 litros y que la varianza es de 4 litros*^{2}.

Si se define por X *la variable aleatoria que mide la cantidad de gas introducido en un frigorífico, se tiene por la desigualdad de Tchebychev que*

$$P(X \in [15 - 2k, 15 + 2k]) \geq 1 - \frac{1}{k^2} \; .$$

Por tanto, si se desea hallar un intervalo, centrado en la media, en el que se encuentre el contenido de gas de, al menos, un 90 % *de los frigoríficos fabricados por él, sólo hay que resolver la siguiente ecuación;*

$$0'9 = 1 - \tfrac{1}{k^2}.$$

Obteniéndose que $k = \sqrt{10}$ *o* $k = -\sqrt{10}$*, como* k *tiene que ser positivo se toma* $k = \sqrt{10}$*. Por tanto el intervalo requerido es* $[8'67, 21'32]$.

4. Variables multidimensionales

4.1. Distribuciones conjunta y marginales

En este epígrafe se extienden las nociones vistas para una variable unidimensional al caso de variables n-dimensionales. En particular se trata el estudio de variables bidimensionales, siendo generalizables los resultados obtenidos aquí a cualquier dimensión finita.

En un principio se distingue entre variables discretas y continuas, aunque más adelante sólo van a considerarse las continuas.

4.1.1. Variable discreta

Sea (X, Y) v.a. discreta con su correspondiente probabilidad para cada par de valores. Se define:

$$P(X = x, Y = y) = f(x, y).$$

La función f, denominada función de cuantía, verifica:

1. $f(x,y) \geq 0 \qquad \forall (x,y)$

2. $\sum_x \sum_y f(x,y) = 1$.

Si X toma los valores $x_1, \ldots, x_m$, e Y los valores $y_1, \ldots, y_n$, entonces se puede expresar la función de cuantía conjunta a través de la tabla:

X/Y	y_1	$\ldots$	y_j	$\ldots$	y_n	Marginal X
x_1	$f(x_1,y_1)$	$\ldots$	$f(x_1,y_j)$	$\ldots$	$f(x_1,y_n)$	$f_1(x_1)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	$f(x_i,y_1)$	$\ldots$	$f(x_i,y_j)$	$\ldots$	$f(x_i,y_n)$	$f_1(x_i)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	$f(x_m,y_1)$	$\ldots$	$f(x_m,y_j)$	$\ldots$	$f(x_m,y_n)$	$f_1(x_m)$
Marginal Y	$f_2(y_1)$	$\ldots$	$f_2(y_j)$	$\ldots$	$f_2(y_n)$	1

Donde en los márgenes derecho e inferior se han obtenido las distribuciones marginales de X e Y respectivamente. Así, la probabilidad de que X tome un valor genérico x_i es:

$$P(X = x_i) = \sum_{j=1}^{n} P(X = x_i, Y = y_j) = \sum_{j=1}^{n} f(x_i, y_j) = f_1(x_i) \ .$$

De igual forma se hace para Y. Siendo f_1 y f_2 las funciones de cuantía marginales de X e Y respectivamente. La función de distribución viene dada por:

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u,v) \ .$$

4.1.2. Variable continua

Sea (X, Y) una v.a. continua, se dice que f es su función de densidad conjunta si

$$P(X \leq x, Y \leq y) \quad = \quad \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \, \mathrm{d}\, v \, \mathrm{d}\, u \ .$$

La función f debe verificar:

1. $f(x, y) \geq 0 \qquad \forall (x, y)$
2. $\displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(u, v) \, \mathrm{d}\, v \, \mathrm{d}\, u = 1 \ .$

$z = f(x, y)$ representa una superficie de densidad, de tal forma que el área encerrada entre la superficie z y el plano XY vale la unidad.

La probabilidad de que la v.a. tome valores dentro del rectángulo $(a, b) \times (c, d)$ viene dada por:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_{a}^{b} \int_{c}^{d} f(x, y) \, \mathrm{d}\, y \, \mathrm{d}\, x.$$

Si A representa cualquier suceso y R_A la región del plano XY que se corresponde con A, se define su probabilidad como:

$$P(A) = \int_{R_A} f(x, y) \, \mathrm{d}\, x \, \mathrm{d}\, y \ .$$

La función de distribución conjunta viene dada por:

$$\begin{aligned} F(x, y) &= P(-\infty < X \leq x, -\infty < Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \, \mathrm{d}\, v \, \mathrm{d}\, u \ . \end{aligned}$$

La relación entre F y f es:

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y) \ .$$

Las funciones de distribución marginales son:

$$\begin{aligned} F_1(x) &= P(X \le x) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}\,v\,\mathrm{d}\,u \\ F_2(y) &= P(Y \le y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{y} f(u,v)\,\mathrm{d}\,v\,\mathrm{d}\,u\ . \end{aligned}$$

Derivando se obtienen las correspondientes funciones de densidad marginales:

$$\begin{aligned} f_1(x) &= \frac{\partial F_1(x)}{\partial x} = \int_{-\infty}^{+\infty} f(x,v)\,\mathrm{d}\,v \\ f_2(y) &= \frac{\partial F_2(y)}{\partial y} = \int_{-\infty}^{+\infty} f(u,y)\,\mathrm{d}\,u\ . \end{aligned}$$

Ejemplo 5.15 *Sea (X,Y) una v.a. bidimensional cuya función de densidad conjunta es:*

$$f(x,y) = \begin{cases} k(\frac{xy}{2}+1) & \text{si} \quad 0<x<1,\ -1<y<1 \\ 0 & \text{en caso contrario.} \end{cases}$$

Para calcular k, de tal forma que $f(x,y)$ sea función de densidad, se procede de la siguiente forma:

1. *$f(x,y) \ge 0$ si y sólo si $k \ge 0$.*
2. *Como $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}\,x\,\mathrm{d}\,y = 1$*
   Entonces:

$$\begin{aligned} 1 &= k\int_0^1\int_{-1}^{1}(\frac{xy}{2}+1)\,\mathrm{d}\,y\,\mathrm{d}\,x \\ &= k\int_0^1 \left(\frac{xy^2}{4}+y\right)\Big|_{-1}^{1}\mathrm{d}\,x = k2x|_0^1 = 2k\ . \end{aligned}$$

Por lo tanto $k=\frac{1}{2}$ y la función de densidad conjunta viene dada por:

$$f(x,y) = \begin{cases} \frac{xy+2}{4} & \text{si} \quad 0<x<1,\ -1<y<1 \\ 0 & \text{en caso contrario.} \end{cases}$$

Se calcula ahora la función de distribución:

$$\begin{aligned}
F(x,y) &= \\
&= P(X \leq x, Y \leq y) = \int_0^x \int_{-1}^y \frac{uv+2}{4} \,\mathrm{d}\,v \,\mathrm{d}\,u \\
&= \int_0^x \left(\frac{uv^2+4v}{8}\right)\Big|_{v=-1}^{y} \mathrm{d}\,u \\
&= \frac{u^2(y^2-1)}{16} + \frac{u(y+1)}{2}\Big|_{u=0}^{x} \\
&= \frac{x^2(y^2-1)}{16} + \frac{x(y+1)}{2} \ .
\end{aligned}$$

Para valores de (x, y) dentro del campo de definición.

Las funciones de distribución marginales quedan:

$$\begin{aligned}
F_1(x) &= P(X \leq x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(u,v) \,\mathrm{d}\,v \,\mathrm{d}\,u \\
&= \int_0^x \int_{-1}^1 \frac{uv+2}{4} \,\mathrm{d}\,v \,\mathrm{d}\,u \\
&= \int_0^x \frac{uv^2+4v}{8}\Big|_{v=-1}^{1} \mathrm{d}\,u = u\,|_{u=0}^{x} = x \ .
\end{aligned}$$

Para valores de X dentro de su campo de definición.

$$\begin{aligned}
F_2(y) &= P(Y \leq y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f(u,v) \,\mathrm{d}\,v \,\mathrm{d}\,u \\
&= \int_0^1 \int_{-1}^y \frac{uv+2}{4} \,\mathrm{d}\,v \,\mathrm{d}\,u \\
&= \int_0^1 \frac{uv^2+4v}{8}\Big|_{v=-1}^{y} \mathrm{d}\,u \\
&= \frac{u^2(y^2-1)}{16} + \frac{u(y+1)}{2}\Big|_{u=0}^{1} \\
&= \frac{y^2-1}{16} + \frac{y+1}{2} = \frac{y^2+8y+7}{16} \ .
\end{aligned}$$

Para valores de Y dentro de su campo de definición.

Resumiendo se tiene:

$$F(x,y) = $$
$$= \begin{cases} 0 & \text{si} \quad x < 0 \text{ ó } y < -1 \\ \frac{x^2(y^2-1)+8x(y+1)}{16} & \text{si} \quad 0 \leq x < 1,\ -1 \leq y < 1 \\ F_1(x) & \text{si} \quad 0 \leq x < 1,\ y \geq 1 \\ F_2(y) & \text{si} \quad x \geq 1, -1 \leq y < 1 \\ 1 & \text{si} \quad x \geq 1,\ y \geq 1\ . \end{cases}$$

Por último, se pueden obtener las funciones de densidad marginales:

1. *A partir de la función de densidad conjunta:*

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}\,y \\ &= \int_{-1}^{1} (\frac{xy}{4} + \frac{1}{2})\,\mathrm{d}\,y = \left.\frac{xy^2+4y}{8}\right|_{-1}^{1} = 1\ . \end{aligned}$$

Para valores de X dentro de su campo de definición.

$$\begin{aligned} f_2(y) &= \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}\,x \\ &= \int_{0}^{1} (\frac{xy}{4} + \frac{1}{2})\,\mathrm{d}\,x = \left.\frac{x^2y+4x}{8}\right|_{0}^{1} = \frac{y+4}{8}\ . \end{aligned}$$

Para valores de Y dentro de su campo de definición.

2. *O bien, derivando parcialmente las funciones de distribución marginales:*

$$\frac{\partial F_1(x)}{\partial x} = f_1(x), \qquad \frac{\partial F_2(y)}{\partial y} = f_2(y)\ .$$

4.2. Distribuciones condicionadas. Independencia

Este epígrafe se limita a analizar los resultados para una v.a. continua. Los correspondientes a v.a. discreta se obtienen de forma inmediata intercambiando los conceptos con sus homólogos. Se define la función de densidad condicionada de X para un valor dado de Y como:

$$f(x/y) = \frac{f(x,y)}{f_2(y)} \ ,$$

siempre que $f_2(y)$ sea distinto de cero. De la forma definida, el lector puede comprobar que efectivamente se trata de una función de densidad. Su correspondiente función de distribución tiene la forma

$$F(x/y) = \frac{\int_{-\infty}^{x} f(x,y)\,\mathrm{d}\,x}{f_2(y)} \ .$$

De igual manera se obtienen las condicionadas de Y con respecto a X.

De las relaciones siguientes (todas ellas inmediatas):

$$\begin{aligned} f(x,y) &= f(y/x)f_1(x) \\ f(x/y) &= \frac{f(y/x)}{f_2(y)}f_1(x) \\ f_2(y) &= \int_{-\infty}^{+\infty} f(y/x)f_1(x)\,\mathrm{d}\,x \ , \end{aligned}$$

se puede deducir que:

$$f(x/y) = \frac{f(y/x)f_1(x)}{\int_{-\infty}^{+\infty} f(y/x)f_1(x)\,\mathrm{d}\,x} \ ,$$

que es la expresión del *Teorema de Bayes* para funciones de densidad.

Ejemplo 5.16 *Continuando con el ejemplo anterior se calcula $f(x/y)$.*

$$f(x/y) = \frac{f(x,y)}{f_2(y)} = \frac{\frac{xy}{4}+\frac{1}{2}}{\frac{y+4}{8}} = \frac{2xy+4}{y+4} \ .$$

Independencia. Se dice que las variables X e Y son independientes, si las distribuciones condicionadas coinciden con las marginales, es decir:

$$f(x/y) = f_1(x) \Longleftrightarrow f(y/x) = f_2(y) \ .$$

Igual que siempre la condición de independencia se establece en un doble sentido. El que las variables sean independientes implica que el conocimiento de que una variable toma un determinado valor o esté contenida en un rango de valores no da ninguna información sobre los valores de la otra variable.

De la definición de condicionada puede deducirse que dos variables X e Y son independientes si y sólo si:

$$f(x,y) = f_1(x)f_2(y) \ .$$

Esta definición de independencia se puede extender a cualquier conjunto finito de variables, y así, las variables $X_1, X_2, \ldots, X_n$, son independientes si se verifica:

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n) \ .$$

La independencia conjunta implica la de cualquier subconjunto de variables; en contra de lo que pasaba en la relación entre sucesos, donde para que se diera la independencia conjunta era necesario exigir la independencia para cualquier combinación de sucesos.

Ejemplo 5.17 *Para comprobar si las variables X, Y del ejemplo 5.15 son independientes se puede utilizar*

- *la definición:*

$$\begin{aligned} f(x/y) &= \frac{f(x,y)}{f_2(y)} = \frac{\dfrac{xy}{4} + \dfrac{1}{2}}{\dfrac{y+4}{8}} \\ &= \frac{2xy+4}{y+4} \neq f_1(x) = 1 \ . \end{aligned}$$

- *la caracterización:*

$$\begin{aligned} f(x,y) &= \frac{xy+2}{4} \neq f_1(x)f_2(y) \\ &= 1 \cdot \frac{y+4}{8} = \frac{y+4}{8} \ . \end{aligned}$$

Por lo tanto X e Y no son independientes.

4.3. Esperanza. Covarianza y correlación

Sea $g(X,Y)$ una función de la v.a. (X,Y) de función de densidad f y función de distribución F. Se define la esperanza matemática de $g(X,Y)$ como:

$$\begin{aligned} E[g(X,Y)] &= \int_{-\infty}^{+\infty} g(x,y)\,\mathrm{d}\,F(x,y) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,\mathrm{d}\,x\,\mathrm{d}\,y \ , \end{aligned}$$

siempre que dicha integral exista. Especial importancia tienen los casos en que g toma las formas:

- $g(X,Y) = X^hY^k$ donde la esperanza recibe el nombre de *momento de orden (h,k) respecto al origen*. Se denota α_{hk}.
- $g(X,Y) = (X-\alpha_{10})^h(Y-\alpha_{01})^k$ en cuyo caso se tiene el *momento de orden (h,k) respecto a la media* y que se identifica por μ_{hk}.

Ejercicio 5.2 *Compruebe que:*

a) $\alpha_{10} = E[X]; \quad \alpha_{01} = E[Y]$
b) $\mu_{10} = \mu_{01} = 0$
c) $\mu_{20} = \sigma_X^2; \quad \mu_{02} = \sigma_Y^2$.

Además $\mu_{11} = \sigma_{XY}$ es la covarianza de (X,Y), definiéndose el coeficiente de correlación lineal como:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \ .$$

La definición de esperanza se puede generalizar al caso de un número finito de variables, siendo su expresión:

$$E[g(X_1, X_2, \dots, X_n)] =$$
$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) \, \mathrm{d}\, x_1 \, \mathrm{d}\, x_2 \dots \mathrm{d}\, x_n.$$

Como siempre, si existe dicha integral.

Propiedad 5.2

1. $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$.
2. *Si las variables* $X_1 \dots X_n$ *son independientes, entonces:*
$$E[X_1 X_2 \dots X_n] = E[X_1]E[X_2] \dots E[X_n] \ .$$
3. *Si dos variables son independientes, su covarianza es nula. El recíproco no es cierto en general.*
4. *Si* $Z = aX + b$ *y* $T = cY + d$*, se tiene que* $\sigma_{Z,T} = ac\sigma_{XY}$.
5. $|\rho_{XY}| \leq 1$.
6. *Si* $Y = aX + b$ *entonces* $|\rho_{XY}| = 1$*, siendo su signo igual al de* a.
7. *Si* $Z = X \pm Y$*, entonces* $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}$.

Se llama *vector de medias* de la v.a. $X = (X_1, X_2, \dots X_n)$ al vector cuyas componentes son las esperanzas de cada una de las X_i, que se representan por $\bar{\mu} = E[X]$.

Se llama *matriz de varianzas y covarianzas* de la v.a. X a la matriz cuadrada de orden n dada por:

$$\Sigma_X = E[(X - \bar{\mu})^t (X - \bar{\mu})].$$

Dicha matriz contiene en la diagonal a las varianzas de las variables y en el resto a las covarianzas entre ellas. La matriz de varianzas y covarianzas es, por tanto, simétrica y además semidefinida positiva; y es definida positiva, si ninguna de las variables es combinación lineal del resto.

Ejemplo 5.18 *Las medias α_{10} y α_{01} (momentos de orden 1 respecto del origen) de la variable aleatoria bidimensional del ejemplo 5.15 se obtienen como:*

$$\begin{aligned}
\alpha_{10} &= E[X] = \tfrac{1}{2}\int_0^1 \int_{-1}^1 \left(\tfrac{x^2y}{2} + x\right) \mathrm{d}\, y \, \mathrm{d}\, x \\
&= \tfrac{1}{2}\int_0^1 \left(\tfrac{x^2y^2}{4} + xy\right)\Big|_{-1}^1 \mathrm{d}\, x \\
&= \tfrac{1}{2}x\Big|_0^1 = \tfrac{1}{2} \\
\alpha_{01} &= E[Y] = \tfrac{1}{2}\int_0^1 \int_{-1}^1 \left(\tfrac{xy^2}{2} + y\right) \mathrm{d}\, y \, \mathrm{d}\, x \\
&= \tfrac{1}{2}\int_0^1 \left(\tfrac{xy^3}{6} + \tfrac{y^2}{2}\right)\Big|_{-1}^1 \mathrm{d}\, x \\
&= \tfrac{1}{2}\tfrac{x^2}{6}\Big|_0^1 = \tfrac{1}{12}\ .
\end{aligned}$$

De la misma forma los momentos de orden 2 vienen dados por:

$$\alpha_{20} = E[X^2] = \tfrac{1}{3} \qquad \alpha_{02} = E[Y^2] = \tfrac{1}{3}.$$

De donde:

$$\begin{aligned}
\mu_{20} &= E[(X-\alpha_{10})^2] = \alpha_{20} - \alpha_{10}^2 \\
&= \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12} \\
\mu_{02} &= E[(Y-\alpha_{01})^2] = \alpha_{02} - \alpha_{01}^2 \\
&= \tfrac{1}{3} - \tfrac{1}{144} = \tfrac{47}{144}\ .
\end{aligned}$$

Para calcular la covarianza se necesita previamente:

$$\begin{aligned}
&\alpha_{11} = \\
&\quad = E[XY] = \tfrac{1}{2}\int_0^1 \int_{-1}^1 \left(\frac{x^2y^2}{2} + xy\right) \mathrm{d}\, y \, \mathrm{d}\, x \\
&\quad = \tfrac{1}{2}\int_0^1 \left[\frac{x^2y^3}{6} + \frac{xy^2}{2}\right]_{-1}^1 \mathrm{d}\, x = \frac{1}{6}\,\frac{x^3}{3}\Big]_0^1 = \frac{1}{18}\ .
\end{aligned}$$

Y por tanto:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \sigma_{x,y} = \mu_{11} = E[XY] - E[X]E[Y] \\
&= \tfrac{1}{18} - \tfrac{1}{2}\tfrac{1}{12} = \tfrac{1}{72}\ .
\end{aligned}$$

De donde el coeficiente de correlación es:

$$\rho_{XY} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} = \frac{\frac{1}{72}}{\sqrt{\frac{1}{12}}\sqrt{\frac{47}{144}}} = 0'0842 \ .$$

4.4. Cambio de variables

Sea (X, Y) una v.a. con función de densidad f. Sean $Z = g_1(X, Y)$ y $T = g_2(X, Y)$ transformaciones continuas y biunívocas definidas por dos funciones g_1 y g_2 que admiten derivadas parciales continuas. Entonces la función de densidad de la nueva variable (Z, T) existe en todos aquellos puntos donde el determinante

$$J = \frac{\partial(z,t)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial t}{\partial x} & \frac{\partial t}{\partial y} \end{vmatrix}$$

sea distinto de cero. Siendo dicha función igual a:

$$g(z,t) = f(h_1(z,t), h_2(z,t))|J_1| \ .$$

Donde h_1 y h_2 son las funciones inversas de g_1 y g_2, y $J_1 = \frac{\partial(x,y)}{\partial(z,t)}$.

Es inmediata la generalización al caso de n variables.

Ejemplo 5.19 *En el ejemplo 5.15 se considera la transformación:*

$$\left.\begin{array}{r} Z = X - Y \\ T = X + 2Y \end{array}\right\}$$

Para obtener la función de densidad de la nueva variable (Z, T) se calcula en primer lugar:

$$J = \begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} = 2 + 1 = 3 \neq 0 \ .$$

Por lo tanto la función de densidad $g(z,t)$ existe para cualquier punto. A continuación se despeja (X, Y) en función de (Z, T) y se calcula J_1.

$$\left.\begin{array}{r} X = \frac{2Z+T}{3} \\ Y = \frac{T-Z}{3} \end{array}\right\} \longrightarrow J_1 = \begin{vmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{vmatrix} = \frac{1}{3} \ .$$

Por tanto:

$$g(z,t) \;=\; \frac{\frac{2z+t}{3}\,\frac{t-z}{3}}{4}\cdot\frac{1}{3}=\frac{(2z+t)(t-z)}{108}\,,$$

resultando:

$$g(z,t) \;=\; \begin{cases} \dfrac{(2z+t)(t-z)}{108} & \begin{array}{c} 0<2z+t<3,\ y \\ -3<t-z<3 \end{array} \\ 0 & \text{en caso contrario.} \end{cases}$$

5. Ejercicios

5.1. Ejercicio resuelto

5.1 La función de distribución asociada a la producción de una máquina, en miles de unidades, es del tipo:

$$F(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ x(2-x) & \text{si} \quad 0\le x\le k \\ 1 & \text{si} \quad x>k. \end{cases}$$

a) Determine k para que F sea función de distribución.

b) Calcule la función de densidad de la variable producción.

c) Obtenga media, moda, mediana y varianza de la producción.

d) ¿Cuál es la probabilidad de que la producción sea inferior a 500 unidades? ¿Y la de que sea superior a 250 unidades?

e) Si el beneficio de la máquina viene dado, en función de la producción, por $B = 6X - 3$, calcule la función de densidad y de distribución del beneficio.

Solución:

a) Para calcular k hay que utilizar las propiedades que se conocen de la función de distribución. Se sabe que la función de distribución cuando se trata de una v. a. continua es continua en todo $\mathbb{R}$. Por tanto:

$$\begin{aligned} 1 &= \lim_{x\to k^+} F(x) = \lim_{x\to k^-} F(x) \\ &= \lim_{x\to k^-} x(2-x) = k(2-k) \end{aligned}$$

$$\Rightarrow \quad k^2 - 2k + 1 = 0$$
$$\Rightarrow \quad k = \frac{2 \pm \sqrt{4-4}}{2} = 1.$$

En consecuencia $k = 1$.

b) Para calcular la función de densidad a través de la función de distribución basta con derivar F:

$$f(x) = \begin{cases} 2 - 2x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en el resto.} \end{cases}$$

c) Por definición se tiene que:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}\, x \\ &= \int_0^1 x(2-2x) \, \mathrm{d}\, x \\ &= \left(x^2 - \frac{2x^3}{3} \right)\Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3} \,. \end{aligned}$$

Para calcular la moda hay que ver el valor que hace máxima la función de densidad. Derivando se obtiene, $f'(x) = -2$. La derivada de la función de densidad es negativa, por lo que se trata de una función decreciente y toma el valor máximo en el extremo inferior del intervalo $[0, 1]$. La moda es por tanto $x = 0$.

La mediana de una distribución es aquel valor que deja el 50 % de la distribución a la derecha y el otro 50 % a la izquierda. Expresado en términos de la función de distribución el valor mediana, Me, verifica, $F(Me) = \frac{1}{2}$.

$$2Me - Me^2 = \frac{1}{2} \implies 4Me - 2Me^2 = 1$$
$$2Me^2 - 4Me + 1 = 0 \implies Me = \frac{4 \pm \sqrt{16-8}}{4} = \frac{2 \pm \sqrt{2}}{2} \,.$$

De las dos soluciones se rechaza aquella que es mayor que uno, pues para ese valor la función de distribución vale 1. Por lo tanto:

$$Me = 1 - \frac{\sqrt{2}}{2} \,.$$

La varianza se puede calcular como:

$$V[X] = E[X^2] - E[X]^2 \ .$$

Se sabe que $E[X] = \frac{1}{3}$, entonces lo único que hay que calcular es $E[X^2]$:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} x^2 f(x) \, \mathrm{d}\, x \\ &= \int_0^1 2x^2(1-x) \, \mathrm{d}\, x = \frac{1}{6} \ . \end{aligned}$$

Y por tanto:

$$V[X] = \frac{1}{6} - \frac{1}{9} = \frac{1}{18} \ .$$

d)

$$\begin{aligned} P(X < 0'5) &= \int_{-\infty}^{0'5} f(x) \, \mathrm{d}\, x \\ &= \int_0^{0'5} (2-2x) \, \mathrm{d}\, x \\ &= \left(2x - x^2\right)\Big|_0^{0'5} = 0'75 \ . \end{aligned}$$

De la misma forma se obtiene $P(X > 0'25)$, sabiendo que:

$$\begin{aligned} P(X > a) &= 1 - P(X \le a) \\ P(X > 0'25) &= 1 - P(X \le 0'25) = 1 - \int_0^{0'25} (2-2x) \, \mathrm{d}\, x \\ &= 1 - \left(2x - x^2\right)\Big|_0^{0'25} = 0'5625 \ . \end{aligned}$$

e) Para calcular la función de densidad de $B = 6X - 3$, se aplica la fórmula del cambio de variable:

$$\begin{aligned} g(b) &= f(x) \left| \frac{\mathrm{d}\, x}{\mathrm{d}\, b} \right| \\ b &= 6x - 3 \Longrightarrow x = \frac{b+3}{6} \\ g(b) &= \left(2 - 2\left(\frac{b+3}{6}\right)\right) \frac{1}{6} = \frac{3-b}{18} \ . \end{aligned}$$

Se calcula el dominio de definición de esta nueva función de densidad.

Si $b = 6x - 3$ y $0 \leq x \leq 1 \Longrightarrow -3 \leq b \leq 3$, entonces

$$g(b) = \begin{cases} \frac{3-b}{18} & \text{si } -3 \leq b \leq 3 \\ 0 & \text{en el resto.} \end{cases}$$

La función de distribución viene dada por:

$$\begin{aligned} F(b) &= \int_{-\infty}^{b} g(x)\,\mathrm{d}\,x \\ &= \int_{-3}^{b} \frac{3-x}{18}\,\mathrm{d}\,x \\ &= \left(\frac{x}{6} - \frac{x^2}{36})\right)\Big|_{-3}^{b} = \frac{6b - b^2 + 27}{36} \end{aligned}$$

cuando $b \in (-3, 3)$, por lo que:

$$F(b) = \begin{cases} 0 & \text{si} \quad b < -3 \\ \frac{6b-b^2+27}{36} & \text{si} \quad -3 \leq b \leq 3 \\ 1 & \text{si} \quad b > 3\ . \end{cases}$$

5.2. Ejercicios propuestos

5.1. Una v.a. tiene como función de densidad:

$$f(x) = \begin{cases} \frac{1}{3} & \text{si} \quad 0 < x < 1 \\ kx & \text{si} \quad 2 < x < 4 \\ 0 & \text{en el resto.} \end{cases}$$

a) Encuentre k para que f sea función de densidad.
b) Calcule la función de distribución.
c) Obtenga la media, moda y mediana de la variable X.

5.2. Una v.a. toma los valores $-2, -1, 0, 1, 2$. Se sabe que cada valor tiene la misma probabilidad que su opuesto, que $P(X = 0) = 0'2$ y que la probabilidad del valor 1 es el doble que la del 2.

Calcule:

a) La función de cuantía y la función de distribución.
b) $P(-1 \leq X \leq 1)$, $P(-1 \leq X < 1)$, $P(X \geq 2)$.
c) La función de cuantía de la variable $Y = 2X^2$.
d) El valor esperado de la variable Y^2.

5.3. Debido al aniversario de un centro comercial se regala por cada 30€ de compra una llave para abrir un cofre de los 50 que hay. Entre ellos hay 2 que contienen premio, uno de 600€. y otro de 400€. ¿Cuál es la ganancia esperada para una persona que tiene dos llaves?

5.4. Observando una determinada v.a. de un experimento se obtiene que su función de densidad viene dada por:

$$f(x) = \begin{cases} \frac{1}{2} + x & \text{si} \quad 0 < x \leq a \\ 0 & \text{en el resto.} \end{cases}$$

a) Encuentre a para que f sea función de densidad.
b) Obtenga su función de distribución.
c) Calcule su media, moda y mediana.
d) Calcule $P(X \leq 3)$, $P(0 \leq X \leq 0'5)$, $P(X \leq -0'5)$
e) En un experimento posterior se comprobó que una nueva variable venía dada en función de ésta por: $Y = \operatorname{Ln} X$. Calcule su función de densidad.

5.5. ¿Existe algún valor de k para el cual la función:

$$f(x) = \begin{cases} k(1-x) & \text{si} \quad -1 \leq x \leq 2 \\ 0 & \text{en el resto,} \end{cases}$$

sea función de densidad? En caso afirmativo calcule su función de distribución.

5.6. El tiempo de vida, en meses, de determinados insectos viene dado por una v.a. cuya función de distribución es:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ K(1 - e^{-2x}) & \text{si} \quad x \geq 0 \ . \end{cases}$$

Calcule:

a) El valor de K para que F sea función de distribución.
b) Su función de densidad.
c) La probabilidad de que un insecto viva más de dos meses.
d) La probabilidad de que un insecto que ha vivido 4 meses viva un mes más.

5.7. Sea X una v.a. continua cuya función de densidad es:

$$f(x) = \begin{cases} (x+1)^2 & \text{si} \quad -2 \leq x \leq 0 \\ kx & \text{si} \quad 1 \leq x \leq 2 \\ 0 & \text{en el resto.} \end{cases}$$

a) Encuentre k para que f sea función de densidad.
b) Calcule la función de distribución.
c) Calcule $P(0 \leq X \leq 1'25)$ y $P(X \geq -1)$.
d) Calcule el valor medio de la variable y su varianza.
e) Si $Y = 2X^2 - 2$, calcule el valor medio de Y.

5.8. Se posee una moneda trucada de tal forma que la probabilidad de sacar cara es el triple que la de sacar cruz. Se lanza esta moneda tres veces y se cuenta el número de veces que ha salido cara.

Calcule:

a) La función de cuantía y la función de distribución de la variable definida.
b) El valor esperado y la varianza.
c) La probabilidad de que el número de caras sea mayor o igual a dos.

5.9. La longitud de ciertos caracoles marinos, en centímetros, se distribuye según la función de densidad:

$$f(x) = \begin{cases} k(x-3)(7-x) & \text{si} \quad x \in (3,7) \\ 0 & \text{en el resto.} \end{cases}$$

a) Determine k para que f sea función de densidad.

b) Calcule su función de distribución.

c) Para un estudio interesan aquellos caracoles marinos cuya longitud esté comprendida entre 3'5 y 6 centímetros ¿Qué porcentaje de caracoles hay para ese estudio?

5.10. De una v.a. discreta se sabe que los valores que toma son: -3, -2, 0, 2, 3, que $P(X = 0) = 0'5$, $P(X = 3) = 0'125$, que el valor esperado de la variable es 0 y la varianza 3'25.

Calcule la función de cuantía de $Y = X^2$.

5.11. Se sabe que la función de distribución de una variable aleatoria es

$$F(x) = \begin{cases} 0 & \text{si} & x < 0 \\ 0'15 & \text{si} & 0 \leq x < 1 \\ 0'6 & \text{si} & 1 \leq x < a \\ 0'75 & \text{si} & 2'5 \leq x < 3 \\ 1 & \text{si} & 3 \leq x\ . \end{cases}$$

Calcule:

a) La función de cuantía.

b) La esperanza, la mediana y la moda de la variable aleatoria.

c) $P(-1'5 \leq X < 3'5)$.

5.12. Se considera la siguiente función:

$$f(x) = \begin{cases} kx^2 & \text{si } x = 1,\ 2,\ 3. \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule k para que f sea función de cuantía de una variable aleatoria X.

b) Calcule la función de distribución de X.
c) Calcule media, moda, mediana y varianza.
d) $P(0'5 \leq X < 2'5)$.
e) Sea $Y = \text{Ln}(X+1)$, calcule la función de cuantía de Y.

5.13. Se considera un experimento aleatorio con un dado y con una moneda, de forma que se lanza el dado y a continuación la moneda. Si la moneda sale cara, el resultado del experimento es la puntuación obtenida al lanzar el dado y si sale cruz el resultado es la puntuación del dado más uno. Calcule la función de cuantía de este experimento.

5.14. Se lanza cuatro veces una moneda y se cuenta el número de caras que se han obtenido. Calcule la función de cuantía de este experimento.

5.15. Se lanza una moneda hasta la primera vez que sale cara. Calcule la función de cuantía de este experimento.

5.16. Se considera la función:

$$f(x) = \begin{cases} \frac{1}{N} & \text{si } x = 1, 2, \ldots, N \\ 0 & \text{en el resto.} \end{cases}$$

a) Pruebe que la función f es una función de cuantía de una variable aleatoria X.
b) Calcule su función de distribución.
c) Obtenga la media y la varianza.

5.17. Se considera la función:

$$f(x) = \begin{cases} ke^{2x-1} & \text{si } x = 0, 1, 2 \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule k para que sea función de cuantía.
b) Calcule la función de distribución correspondiente.

5.18. Sea f definida por

$$f(x) = \begin{cases} (x+1)^2 & \text{si} \quad 0 \leq x < 0'25 \\ ke^{2x} & \text{si} \quad 0'75 \leq x < 1 \\ \frac{\mathrm{Ln}(x)}{2} & \text{si} \quad 1 \leq x \leq 2 \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule k para que sea función de densidad.
b) Calcule la función de distribución correspondiente.
c) Calcule la mediana de la distribución.
d) Sea $Y = 2X + 1$, calcule $E[Y]$.

5.19. Sea f definida por:

$$f(x) = \begin{cases} \frac{1}{3} & \text{si} \quad 0 \leq x < 1 \\ k(3x^2 + 4x) & \text{si} \quad 1 \leq x < 2 \\ \frac{2}{21}x & \text{si} \quad 2 \leq x \leq 3 \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule k para que f sea función de densidad.
b) Calcule su función de distribución.
c) Sea $Y = \mathrm{Ln}(2X)$, calcule la función de densidad de Y.

5.20. Sea f definida por:

$$f(x) = \begin{cases} 1 - x & \text{si} \quad 0 \leq x < a \\ x - 1 & \text{si} \quad a \leq x \leq 2 \ . \end{cases}$$

Calcule a para que f sea función de densidad.

5.21. Sea F la función definida por

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ x - \frac{x^2}{2} & \text{si} \quad 0 \leq x < 1 \\ k(\frac{x^2}{2} - x + 1) & \text{si} \quad 1 \leq x \leq a \\ 1 & \text{si} \quad x > a \ . \end{cases}$$

a) Calcule a y k para que F sea la función de distribución de X.

b) Calcule la función de densidad de X.
c) Calcule la mediana de la distribución.
d) Calcule la distribución de $Y = e^X$.

5.22. Sea F una función de distribución dada por

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ kx^2 & \text{si} \quad 0 \leq x < 1 \\ k(4x - x^2 - 2) & \text{si} \quad 1 \leq x < 2 \\ b & \text{si} \quad x \geq 2\ . \end{cases}$$

a) Calcule k y b para que F sea función de distribución de una variable aleatoria continua.
b) Calcule $E[X]$ y $V[X]$.
c) Obtenga $P(\frac{3}{4} < X < \frac{3}{2})$.

5.23. Se considera (X, Y) la v.a. bidimensional con

$$f(x, y) = \begin{cases} k(x - y) & \text{si} \quad 2 \leq x \leq 4,\ 0 \leq y \leq 1 \\ 0 & \text{en el resto.} \end{cases}$$

a) Determine k para que f sea función de densidad.
b) Calcule su función de distribución.
c) Obtenga las distribuciones marginales y condicionadas.
d) Calcule $P(X \leq 3, Y \geq 0'5)$, $F(2, 0)$ y $F(1, 0'25)$.

5.24. Sea f la función definida por:

$$f(x, y) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{en el resto.} \end{cases}$$

Demuestre que es una función de densidad.

5.25. De las siguientes funciones, calcule el valor de k para que sean funciones de densidad de una variable aleatoria bidimensional.
a)

$$f(x, y) = \begin{cases} kx^2 & \text{si } 0 \leq x \leq 1,\ y \leq x \\ 0 & \text{en el resto.} \end{cases}$$

b)

$$f(x,y) = \begin{cases} k(2x+y) & \text{si } 0 \leq x \leq 1,\ y \leq x^2 \\ 0 & \text{en el resto.} \end{cases}$$

c)

$$f(x,y) = \begin{cases} k & \text{si } 0 \leq x \leq 1,\ |y| \leq x^2 \\ 0 & \text{en el resto.} \end{cases}$$

d)

$$f(x,y) = \begin{cases} k(x^2+y^2) & \text{si } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{en el resto.} \end{cases}$$

5.26. Sea

$$f(x,y) = \begin{cases} kxy & \text{si } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{en el resto.} \end{cases}$$

Calcule

a) El valor de k para que la función f sea función de densidad.

b) La función de distribución correspondiente.

c) Las distribuciones marginales.

d) $P(0'5 \leq X \leq 0'75, 0 \leq Y \leq 0'25)$.

e) La distribución de $X+Y$.

5.27. Sea

$$f(x,y) = \begin{cases} k & \text{si } \ 0 \leq x \leq 1,\ 0 \leq y \leq 2 \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule k para que f sea función de densidad.

b) Obtenga la función de densidad de la variable bidimensional (U,V), donde $U = X+Y$ y $V = X-Y$.

c) Calcule las marginales de (X,Y).

d) Determine las condicionadas de (U,V).

Capítulo 6

Algunos modelos probabilísticos

La identificación del patrón probabilístico por el cual se rige un determinado experimento aleatorio facilita enormemente su estudio. En efecto, cuando se desea estudiar una determinada situación real, el reconocimiento de un modelo probabilístico en dicha situación supone una abstracción matemática que permite el uso tanto de herramientas matemáticas como estadísticas para obtener su completa caracterización.

De esta forma, a lo largo de este capítulo se realizará un estudio pormenorizado de los modelos probabilísticos más importantes, de forma que permitan identificarse cuando se presenten en un determinado experimento aleatorio. Se comienza estudiando modelos discretos y continuos en el caso unidimensional para posteriormente pasar a estudiar algunos modelos en el caso bidimensional.

1. Distribución uniforme discreta

El primer modelo probabilístico que se estudia es el uniforme discreto, que consiste en distribuir a partes iguales la masa de probabilidad entre un número finito de valores.

Sea X una variable aleatoria uniforme discreta, que toma los valores $x_1, \ldots, x_n$. La función de cuantía viene dada por la siguiente expresión

$$P(X = x_i) = \frac{1}{n} \quad i = 1, \ldots, n.$$

Suponiendo que $x_1 \leq x_2 \leq \ldots \leq x_n$, su función de distribución viene dada por

$$F(x) = \begin{cases} 0 & \text{si } x < x_1 \\ \frac{1}{n} & \text{si } x_1 \leq x < x_2 \\ \frac{2}{n} & \text{si } x_2 \leq x < x_3 \\ \vdots & \vdots \\ 1 & \text{si } x_n \leq x. \end{cases}$$

Por otro lado, se tiene que $E[X] = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ y que, $E[X^2] = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^2$.

El ejemplo clásico de una uniforme discreta es el experimento de observar los resultados posibles al lanzar un dado "honesto".

2. Experimento de Bernouilli

La realización de pruebas repetidas e independientes constituyen un experimento de Bernouilli cuando cada una de ellas tiene dos posibles resultados E y F, con probabilidades de ocurrencia, invariantes a lo largo de todo el proceso, p y q, respectivamente. Obviamente se verifica que tanto p como q son estrictamente mayores que cero y además $p + q = 1$. Los dos posibles resultados suelen identificarse a menudo como *éxito* y *fracaso*, y la v.a. que se asocia al experimento toma los valores 1 para el éxito y 0 para el fracaso.

2.1. Distribución binomial

La distribución de probabilidad por la cual se rige el número de éxitos al realizar n pruebas Bernouilli independientes y con probabilidades de éxito iguales, se le denomina distribución binomial.

Ejemplo 6.1 *El número de caras en 15 lanzamientos de una moneda sigue una distribución binomial.*

Se consideran n variables aleatorias Bernouilli independientes, X_i con $i = 1, \ldots, n$, tal que, $X_i = 1$ si en el experimento Bernouilli se ha obtenido éxito y $X_i = 0$ en caso contrario. Obsérvese que si se define $X = \sum_{i=1}^{n} X_i$, se está contando el número de éxitos en n experimentos Bernouilli independientes. Por tanto, una v.a. con distribución binomial puede expresarse como suma de variables aleatorias independientes Bernouilli.

Para calcular la distribución de probabilidad de la variable binomial, se considera la ejecución n veces del experimento Bernouilli en el que han aparecido k éxitos y $n-k$ fracasos, –por ejemplo en la secuencia $EE \ldots EFF \ldots F$–, la probabilidad de este suceso viene dada por $p^k q^{n-k}$. Sin embargo, no es esa la única forma de obtener k éxitos, pues hay que considerar todas las ordenaciones de los elementos E y F, lo que da un total de:

$$P_n^{k,n-k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

La probabilidad de obtener k éxitos en n pruebas es, por tanto:

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, 2, \ldots, n.$$

Observe que se trata de una verdadera distribución de probabilidad, pues:

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1.$$

De hecho, esta expresión justifica el nombre de la distribución, pues lo anterior no es sino el desarrollo del binomio $(p+q)^n$. A la distribución se le suele notar por $B(n, p)$, es decir por su letra o letras iniciales y los parámetros que la caracterizan. De igual forma se procede en el resto de los casos. Es fácil comprobar que:

$$E[X] = np, \quad V[X] = npq, \quad M_X(t) = (pe^t + q)^n.$$

Para el caso particular de que n sea igual a 1 se obtiene la distribución de Bernouilli, $B(p)$.

Propiedad 6.1 *Si $X_1, \ldots, X_k$ son variables aleatorias independientes tal que $X_i \sim B(n_i, p)$ con $i = 1, \ldots, k$, se tiene que* $\sum_{i=1}^{k} X_i \sim B(\sum_{i=1}^{k} n_i, p)$.

Ejemplo 6.2 *De una población de animales, se sabe que el 60 % son machos. Si se extrae un conjunto de 10 animales, ¿cuál es la probabilidad de que en ese conjunto haya 7 hembras?*

Sea X = "Número de hembras en un conjunto de 10 animales", se ve claramente que $X \sim B(10, 0'4)$ (figura 6.1).

La probabilidad de que haya 7 hembras es:

$$P(X = 7) = \binom{10}{7} 0'4^7 0'6^3 = 0'04.$$

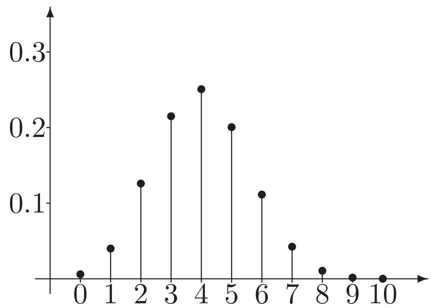


Figura 6.1: Binomial $B(10, 0'4)$

2.2. Distribución geométrica

Una variable aleatoria X definida como el número de fracasos antes de obtener el primer éxito en sucesivas realizaciones de experimentos Bernouilli independientes y con idénticas probabilidades de éxito, se dirá que sigue una distribución geométrica. Dicha variable toma los valores $k = 0, 1, 2, \ldots$, con probabilidad:

$$P(X = k) = pq^k.$$

Sus características son:

$$E[X] = \frac{q}{p} \quad V[X] = \frac{q}{p^2} \quad M_X(t) = \frac{p}{(1-qe^t)}.$$

Propiedad 6.2 *Sea $X \sim Ge(p)$, para cualesquiera n y m se verifica que*

$$P(X > m+n/X > m) = P(X \geq n).$$

Recíprocamente, si X es una variable aleatoria que toma valores no negativos y se cumple que

$$P(X > m+n/X > m) = P(X \geq n)$$

para todo m y n, entonces se tiene que X se distribuye según una distribución geométrica.

Esta propiedad se conoce como de pérdida de memoria.

Ejemplo 6.3 *Para la misma población del ejemplo 6.2, se extraen animales hasta obtener la primera hembra, ¿cuál es la probabilidad de extraer 5 machos antes que la primera hembra?*

Si se considera como éxito extraer una hembra y se define la variable X como: "Número de machos antes de obtener la primera hembra", se tiene una distribución geométrica, es decir, $X \sim Ge(0'4)$.

La probabilidad que se pide es:

$$P(X = 5) = 0'6^5 \cdot 0'4 = 0'36.$$

2.3. Distribución binomial negativa

Una v.a. X definida como el número de fracasos antes del r-ésimo éxito en sucesivas realizaciones de un experimento Bernouilli independientes con igual probabilidad, se dirá que sigue una distribución binomial negativa. La distribución geométrica es un caso particular de binomial negativa cuando $r = 1$. La v.a. X toma los valores $k = 0, 1, 2, \ldots$, con probabilidad:

$$P(X = k) = \binom{k + r - 1}{k} p^r q^k.$$

Sus principales rasgos son:

$$E[X] = \frac{rq}{p}, \quad V[X] = \frac{rq}{p^2}.$$

Propiedad 6.3 *Sean $X_1, \ldots, X_r$ variables aleatorias independientes e idénticamente distribuidas según una distribución geométrica de parámetro p. Se tiene que $\sum_{i=1}^{r} X_i$ sigue una distribución binomial negativa de parámetros r y p.*

Propiedad 6.4 *Sean $X_1, \ldots, X_k$ variables aleatorias independientes e idénticamente distribuidas según una binomial negativa de parámetros r_i y p con $i = 1, \ldots, k$. Se tiene que $\sum_{i=1}^{k} X_i$ sigue una distribución binomial negativa de parámetros r y p, siendo $r = \sum_{i=1}^{k} r_i$.*

Ejemplo 6.4 *Siguiendo con la población del ejemplo 6.2, ¿cuál es la probabilidad de extraer 5 hembras antes del tercer macho?*

En este caso, el éxito se define como "extraer un macho", por lo que $p = P(\text{éxito}) = 0'6$ y $X =$"Número de hembras antes de obtener el tercer macho". Entonces $X \sim BN(3, 0'6)$.

La probabilidad pedida es:

$$P(X = 5) = \binom{5 + 3 - 1}{5} 0'6^3 \cdot 0'4^5 = 0'18.$$

3. Distribución hipergeométrica

Dada una urna compuesta por N_1 bolas blancas y N_2 negras, de la que se extraen n bolas sin reemplazamiento, ¿cuál es la probabilidad de que haya r bolas blancas y $n-r$ bolas negras entre las n bolas extraídas?

Utilizando números combinatorios, esta probabilidad vale:

$$\frac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N_1+N_2}{n}}.$$

En esta situación, la distribución de la v.a. definida como número de bolas blancas (r) en una extracción de n bolas de una urna se le denomina distribución hipergeométrica. Dicha variable toma los valores $r = 0, 1, \ldots, \text{mín}\{n, N_1\}$ con probabilidad:

$$P(X = r) = \frac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N_1+N_2}{n}}.$$

Sus principales características son:

$$E[X] = \frac{nN_1}{N_1 + N_2}, \quad V[X] = n\left(\frac{N_1 + N_2 - n}{N_1 + N_2 - 1}\right)\frac{N_1 N_2}{(N_1 + N_2)^2}.$$

Se hace notar que la diferencia que existe entre un experimento con distribución binomial y uno con distribución hipergeométrica radica, en que en la primera, la extracción de las bolas se producen con reemplazamiento y en la segunda, sin reemplazamiento. Por ello, cuando N_1 y N_2 tiende a infinito y $\frac{N_1}{N_1+N_2}$ converge a un valor p, la distribución hipergeométrica se aproxima a una distribución binomial $B(n, p)$, con $p = \frac{N_1}{N_1+N_2}$; ya que en este caso la probabilidad de extraer una bola con o sin reemplazamiento es prácticamente la misma.

Ejemplo 6.5 *De una urna que contiene seis bolas negras y nueve bolas rojas se extraen cinco bolas, ¿cuál es la probabilidad de obtener tres bolas rojas?*

Sea X = "Número de bolas rojas al extraer cinco *bolas de la urna". La probabilidad pedida es:*

$$P(X=3)=\frac{\binom{6}{2}\binom{9}{3}}{\binom{15}{5}}=0'4196.$$

Ejemplo 6.6 *Si el experimento del ejemplo anterior se repite tres veces, reintegrando las bolas extraídas y añadiendo el mismo número de bolas negras y rojas obtenidas, ¿cuál es la probabilidad de obtener una secuencia de una, dos y tres bolas rojas?*

Sea X_i = "Número de bolas rojas extraídas en la extracción i-ésima". La probabilidad pedida es:

$$P(X_1=1)\cdot P(X_2=2/X_1=1)\cdot$$
$$P(X_3=3/(X_1=1,X_2=2))=$$

$$=\frac{\binom{6}{4}\binom{9}{1}}{\binom{15}{5}}\cdot\frac{\binom{10}{3}\binom{10}{2}}{\binom{20}{5}}\cdot\frac{\binom{13}{2}\binom{12}{3}}{\binom{25}{5}}=0'0051.$$

4. Proceso de Poisson

Se considera una determinada eventualidad que se produce en un **soporte continuo** (tiempo, línea, área, espacio,...), de forma **independiente** y con una cierta **estabilidad** para una unidad de soporte prefijada. Como ejemplos se pueden considerar el número de coches que pasan por un semáforo en un periodo de tiempo, el número de defectos por metro cuadrado de una pieza de tela, el número de hormigas de una cierta especie en un metro cúbico de tierra, etc. Las tres condiciones en negrita caracterizan al denominado proceso de Poisson, y su v.a. está definida por el número de eventos que se producen en una región de tamaño fijo.

4.1. Distribución de Poisson

La distribución de Poisson se obtiene como límite de la binomial cuando el número de veces que se realiza el experimento, n, tiende a infinito, la probabilidad de éxito, p, tiende a cero y el número medio de éxitos, np, se estabiliza alrededor de un número, λ, que será la media y el valor que caracterizará a la distribución. Calculando dicho límite se obtiene:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \ldots$$

La distribución, que se denota $P(\lambda)$, aparece, históricamente, al considerar el número de llamadas telefónicas realizadas por unidad de tiempo a un mismo teléfono, siendo λ el número medio de llamadas. Además, se puede observar que en la distribución Poisson la mayor parte de la masa de probabilidad queda repartida entre un número relativamente pequeño de valores, siendo posible que tome otros valores pero con una probabilidad bastante pequeña. Por ello, a la distribución Poisson se le llama distribución de los sucesos raros.

Sus principales características son:

$$E[X] = \lambda, \qquad V[X] = \lambda, \qquad M_X(t) = e^{-\lambda(1-e^t)}.$$

Todo lo anterior está referido a una unidad de soporte de tamaño 1, si se quiere generalizar a cualquier región de tamaño t la función de cuantía es:

$$P_t(X = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Ejemplo 6.7 *El número medio de vehículos por minuto que llegan a una gasolinera es igual a 2. ¿Cuál es la probabilidad de que en un minuto lleguen 5 vehículos?, ¿y la de que en 5 minutos no llegue ninguno?*

Se trata de una distribución de Poisson, para la que se conoce el número medio, que es igual al parámetro λ, por lo tanto $X \sim P(2)$ (figura 6.2). La primera probabilidad pedida es:

$$P(X = 5) = \frac{2^5}{5!} e^{-2} = 0'036.$$

Para calcular la otra probabilidad, hay que hacer un cambio de parámetro pues la longitud del intervalo cambia.

Si en un minuto el número medio de vehículos es 2, en cinco minutos será 10. Por lo tanto se tiene $Y \sim P(10)$.

$$P(Y = 0) = \frac{10^0}{0!} e^{-10} = 0'000045.$$

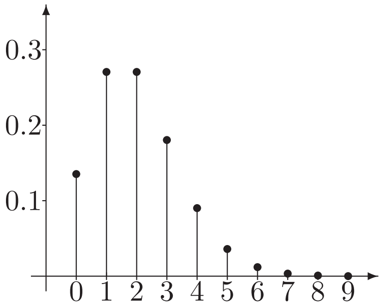


Figura 6.2: Poisson $P(2)$

4.2. Distribución exponencial

A partir de un proceso de Poisson, la distribución de una v.a. X definida como el tiempo transcurrido entre la ocurrencia de dos eventos consecutivos, se denomina exponencial y se denota por $Exp(\lambda)$. Observe que se ha dado un salto cualitativo, puesto que la variable definida es continua, tomando valores en el intervalo $[0, +\infty)$. Se verifica que:

$$P(X > t) = P(\text{cero eventos en } (0, t)) = e^{-\lambda t}.$$

y por tanto:

$$F(t) = P(X \leq t) = 1 - e^{-\lambda t}.$$

derivando se obtiene la función de densidad de X (figura 6.3):

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t}.$$

Los parámetros de la distribución son:

$$E[X] = \frac{1}{\lambda}, \quad V[X] = \frac{1}{\lambda^2}, \quad \gamma_1 = 2, \quad \gamma_2 = 9.$$

Ejemplo 6.8 *Con los datos del ejemplo 6.7 y suponiendo que acaba de llegar un vehículo, calcule la probabilidad de que transcurran más de 5 minutos hasta que aparezca el siguiente.*

El tiempo que transcurre desde que pasa un vehículo hasta el siguiente sigue una distribución exponencial de parámetro igual a 2. La probabilidad pedida es:

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) = 1 - (1 - e^{-2 \cdot 5}) \\ &= 0'000045. \end{aligned}$$

Observe que se llega al mismo resultado utilizando la distribución de Poisson que la exponencial, considerando que el suceso "transcurren más de 5 minutos en pasar algún vehículo" para la exponencial es el suceso "no pasa ningún vehículo en cinco minutos" para la Poisson.

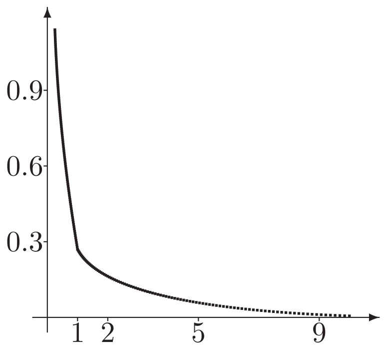


Figura 6.3: Distribución exponencial

5. Distribución uniforme continua

A continuación se estudia la distribución uniforme continua como la extensión natural de la distribución uniforme discreta, es decir, aquella que toma con igual probabilidad valores dentro de dos conjuntos cualesquiera de igual amplitud e incluidos en el intervalo de valores posibles de la variable.

La variable X sigue una distribución uniforme o rectangular en el intervalo (a,b), $U(a,b)$, cuando su función de densidad viene dada de la siguiente forma:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad a \leq x \leq b \\ 0 & \text{en el resto.} \end{cases}$$

Su representación gráfica justifica el nombre de rectangular (figura 6.4).

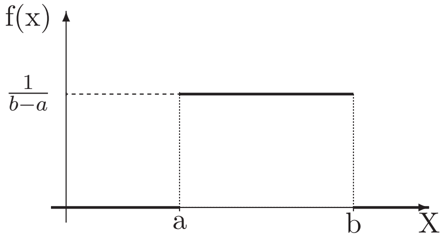


Figura 6.4: Distribución uniforme.

Sus características más importantes son:

$$E[X] = \frac{a+b}{2}, \quad V[X] = \frac{(b-a)^2}{12}, \quad M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}.$$

Ejemplo 6.9 *Un autobús pasa por cierta parada cada 15 minutos. ¿Cuál es la probabilidad de que un señor que llega en un momento dado tenga que esperar el autobús más de cinco minutos?*

Si se define X="Tiempo de espera", entonces $X \sim U(0,15)$. Se calculará en primer lugar la función de distribución

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ \frac{x}{15} & \text{si} \quad 0 \leq x \leq 15 \\ 1 & \text{si} \quad x > 15. \end{cases}$$

La probabilidad pedida viene dada por:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \frac{5}{15} = 0'67.$$

6. Distribución normal

En este epígrafe se estudia la distribución más importante del cálculo de probabilidades y de la estadística, usualmente también denominada distribución de Gauss o de Laplace. La importancia de esta distribución radica en varias razones: en primer lugar, como se verá posteriormente a través del teorema central del límite, la distribución normal es la distribución límite de una amplia gama de sucesiones de variables aleatorias independientes. En segundo lugar, la gran mayoría de las variables aleatorias que se estudian en experimentos físicos son aproximadas por una distribución normal. En tercer lugar, se ha observado que los errores aleatorios en los resultados de medida se distribuyen, de forma general, según una distribución normal. Finalmente, hay que destacar el carácter reproductivo de sus parámetros que facilita el manejo de este tipo de distribuciones.

Se dice que una variable X sigue una distribución normal si su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty \leq x \leq +\infty.$$

La distribución está caracterizada por los parámetros μ y σ, cuyo significado se trata más adelante, siendo σ necesariamente positivo. Se hace referencia a esta distribución como $N(\mu, \sigma)$. Su representación gráfica viene dada por la figura 6.5.

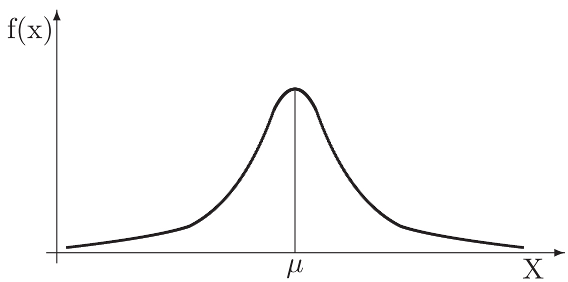


Figura 6.5: Distribución $N(\mu, \sigma)$

Propiedad 6.5 *Sean $X_i \sim N(\mu_i, \sigma_i)$ con $i = 1, \ldots, n$ variables aleatorias independientes entonces* $\displaystyle\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)$.

6.1. Distribución $N(0, 1)$

Para facilitar cálculos, se realiza el estudio de la distribución $N(0, 1)$ y mediante un cambio de variable se generalizará para la $N(\mu, \sigma)$. La función de densidad de la $N(0, 1)$ es:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \leq x \leq +\infty.$$

A continuación se realiza el estudio de dicha función.

1. Campo de existencia. Toda la recta real para x, con $f(x) > 0$.
2. Simetrías. Es simétrica respecto al eje OY, ya que $f(x) = f(-x)$.
3. Asíntotas. $y = 0$ es una asíntota horizontal.
4. Cortes con los ejes. Sólo tiene un punto de corte en $(0, \frac{1}{\sqrt{2\pi}})$.
5. Máximos y mínimos. El punto de corte anterior es el único máximo de la distribución.
6. Puntos de inflexión. Tiene dos en $(-1, \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}})$ y $(1, \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}})$.

Con todo lo anterior la curva tiene la forma de la figura 6.6.

Las características de la distribución son:

$$E[X] = 0, \quad V[X] = 1, \quad \gamma_1 = 0, \quad \gamma_2 = 3.$$

Los valores de la distribución $N(0, 1)$ están tabulados y son fácilmente calculables.

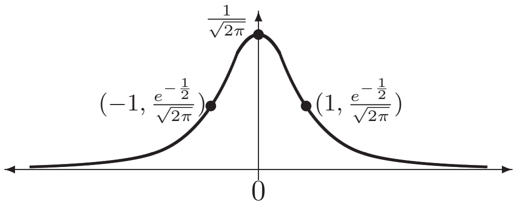


Figura 6.6: Distribución $N(0,1)$

A partir de ahora, si una variable sigue una distribución $N(0,1)$ se denotará por la letra Z. Así, si X sigue una $N(\mu,\sigma)$ se verifica que:

$$Z = \frac{X-\mu}{\sigma} \quad \text{y} \quad X = \sigma Z + \mu.$$

Con lo que se puede comprobar fácilmente que:

$$E[X] = \mu, \quad V[X] = \sigma^2.$$

Ejemplo 6.10 *El contenido de un tipo de bombonas de gas se distribuye normalmente con media 23 kg y desviación típica 0'25 kg. ¿Cuál es la probabilidad de que una bombona tenga más de 23'5 Kg?*

La probabilidad pedida es:

$$P(X > 23'5) = 1 - P(X \leq 23'5).$$

Para calcular esta probabilidad hay que tipificar la variable y hacer uso de la tabla $N(0,1)$.

$$\begin{aligned} P(X \leq 23'5) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{23'5-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{23'5-23}{0'25}\right) \\ &= P(Z \leq 2) = 0'977. \end{aligned}$$

Por lo tanto:

$$P(X > 23'5) = 1 - 0'977 = 0'023.$$

6.2. Distribuciones truncadas

A veces la variable no toma todos los valores de la recta real, imagine que sólo puede tomar valores positivos. Se supone que X no está definida fuera del intervalo $[a, b]$, entonces la función de densidad de la variable truncada en dicho intervalo que se puede denominar $f_{[a,b]}$ viene dada, en función de una variable general, por:

$$f_{[a,b]}(x) = \begin{cases} \dfrac{f(x)}{F(b) - F(a)} & \text{si} \quad x \in [a, b] \\ \\ 0 & \text{si} \quad x \notin [a, b]. \end{cases}$$

La caracterización anterior vale para cualquier tipo de distribución.

7. Relación entre binomial, Poisson y normal

A veces el cálculo para obtener la probabilidad de una variable binomial es muy dificultoso, esto suele ocurrir cuando n es muy grande. En tales casos, y bajo ciertas condiciones, es posible aproximar la distribución binomial por la normal o la Poisson, esta última también es susceptible de ser aproximada por la normal. Las siguientes propiedades resumen las relaciones existentes entre tales distribuciones.

Propiedad 6.6 *Si la probabilidad p de éxito de una variable $B(n, p)$ tiende a cero, mientras que el número de pruebas tiende a infinito, de forma que $\mu = np$ permanece constante, la distribución binomial se puede aproximar por una distribución de Poisson con media μ. En la práctica es suficiente con que n sea mayor que 100 y p menor o igual a 0'05.*

Propiedad 6.7 *Si X sigue una $B(n, p)$, siendo n grande y p no demasiado pequeño, entonces $Z = \dfrac{X - np}{\sqrt{npq}}$ se aproxima a una distribución $N(0, 1)$ a medida que n tiende a infinito. La aproximación es bastante buena siempre que:*

$$np > 5 \quad si \quad p \leq 0'5 \qquad ó \qquad nq > 5 \quad si \quad p > 0'5$$

Hay ocasiones en que es más conveniente operar con la proporción de éxitos obtenidos en n pruebas Bernouilli que con el número de éxitos. Puesto que se verifica:

$$\frac{X - np}{\sqrt{npq}} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}}.$$

Podemos afirmar lo siguiente:

Propiedad 6.8 *En las condiciones de la propiedad anterior*

$$\frac{X}{n} \sim N\left(p; \sqrt{\frac{pq}{n}}\right).$$

Propiedad 6.9 *Si X sigue una distribución de Poisson de parámetro λ suficientemente grande, entonces es posible aproximarla por una distribución $N(\lambda, \lambda^{1/2})$. En la práctica se exige a λ que sea mayor que 5.*

El gráfico de la figura 6.7 resume los resultados anteriores.

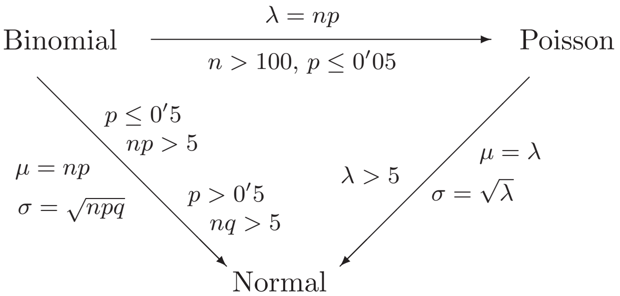


Figura 6.7: Relación entre binomial, Poisson y normal.

8. Teorema central del límite

Las aproximaciones que se han expuesto son casos particulares del denominado *teorema central del límite*. Este teorema dice que si

$X_1, X_2, \ldots, X_n$ son variables aleatorias independientes con medias μ_i, desviaciones típicas σ_i y distribuciones cualesquiera, no necesariamente la misma para todas las variables, entonces si se define $Y = X_1 + X_2 + ... + X_n$, cuando n tiende a infinito, la distribución de la variable:

$$\frac{Y - \displaystyle\sum_{i=1}^{n} \mu_i}{\sqrt{\displaystyle\sum_{i=1}^{n} {\sigma_i}^2}}$$

tiende a una distribución $N(0,1)$.

9. Distribución gamma

Para $p > 0$ se define la función $\Gamma(p)$ como

$$\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} \, \mathrm{d}\, x.$$

Esta función verifica las siguientes propiedades:

- $\Gamma(1) = 1$.
- $\Gamma(p) = (p-1)\Gamma(p-1)$ con $p > 1$.
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Se dice que una variable aleatoria, X, tiene una distribución gamma de parámetros a y p, tal que, $a > 0$ y $p > 0$, $X \sim \Gamma(a;p)$, si tiene como función de densidad

$$f(x) = \begin{cases} \dfrac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se verifica que $M_X(t) = \left(1 - \dfrac{t}{a}\right)^{-p}$, de donde se deduce que, $E[X] = \frac{p}{a}$ y que $E[X^2] = \dfrac{p(p+1)}{a^2}$, por tanto, $V[X] = \dfrac{p}{a^2}$.

Destaca como caso particular de este tipo de distribución la $\Gamma(\lambda, 1)$ que es la distribución $Exp(\lambda)$. La distribución $\Gamma(\lambda, n)$ se utiliza para calcular la distribución del tiempo transcurrido entre las ocurrencias k y $k+n$ de un proceso de Poisson de media λ.

Propiedad 6.10 *Sean $X_1, \ldots, X_n$ un conjunto de variables aleatorias independientes, tal que, $X_i \sim \Gamma(a, p_i)$ con $i = 1, \ldots, n$, se tiene que $\sum_{i=1}^{n} X_i \sim \Gamma(a, \sum_{i=1}^{n} p_i)$.*

Propiedad 6.11 *Sean $X_1, \ldots, X_n$ un conjunto de variables aleatorias independientes, tal que, $X_i \sim Exp(\lambda)$ con $i = 1, \ldots, n$, se tiene que $\sum_{i=1}^{n} X_i \sim \Gamma(\lambda, n)$.*

Propiedad 6.12 *Sea $X \sim \Gamma(a, p)$ entonces $cX \sim \Gamma\left(\frac{a}{c}, p\right)$.*

Ejemplo 6.11 Sea $X \sim \Gamma(2, 2)$, *el calculo de la función de distribución y $P(X \geq 5)$ se hace:*

$$\begin{aligned} F(x) = \int_0^x \frac{1}{\Gamma(2)}\, 2^2 e^{-2x} x \,\mathrm{d}\, x &= -2xe^{-2x}\Big|_0^x + 2\int_0^x e^{-2x} \,\mathrm{d}\, x \\ &= 1 - (1+2x)e^{-2x}. \end{aligned}$$

Con lo cual, $F(5) = 1 - 11e^{-20}$.

10. Distribución beta

Para $a > 0$ y $b > 0$ se define la función beta, como;

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} \,\mathrm{d}\, x,$$

y se verifica la siguiente relación $\beta(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Se dice que una variable aleatoria, X, sigue una distribución beta de parámetros $a > 0$ y $b > 0$, y se denota por $X \sim \beta(a,b)$, si su función de densidad viene dada por la expresión:

$$f(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)} & \text{si } 0 < x < 1 \\ 0 & \text{caso contrario.} \end{cases}$$

Se tiene que la función generatriz de momentos es

$$M_X(t) = \sum_{j=0}^{\infty} \frac{t^j}{\Gamma(j+1)} \frac{\Gamma(a+j)\Gamma(a+b)}{\Gamma(a+b+j)\Gamma(a)},$$

de donde se deduce que $E[X^k] = \dfrac{\Gamma(a+k)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+k)}$, por tanto se tiene que $E[X] = \dfrac{a}{a+b}$, $V[X] = \dfrac{ab}{(a+b)^2(a+b+1)}$.

Ejemplo 6.12 *Sea X una v.a. con distribución $\beta(4,2)$, el cálculo de $E[X]$, $V[X]$ y $P(0 < X \leq 0'7)$ es:*

$$E[X] = \frac{4}{6} = \frac{2}{3} \quad y \quad V[X] = \frac{8}{6^2 \cdot 7}.$$

Además,

$$\begin{aligned} P(0 < X \leq 0'7) &= \int_0^{0'5} \frac{1}{20}(x^3 - x^4)\,\mathrm{d}\,x \\ &= \frac{1}{20}\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big|_0^{0'5} \\ &= 0'00132. \end{aligned}$$

11. Distribución de Cauchy

Se dice que una variable aleatoria X sigue una distribución de Cauchy de parámetros μ y θ, y se denota por $X \sim C(\mu,\theta)$, si su función de densidad viene dada por la siguiente expresión:

$$f(x) = \frac{\mu}{\pi} \frac{1}{\mu^2 + (x-\theta)^2} \qquad -\infty < x < \infty, \quad \mu > 0.$$

La variable aleatoria $Y = \dfrac{X-\theta}{\mu}$ verifica que $Y \sim C(1,0)$, es decir, Y tiene como función de densidad

$$f(y) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty.$$

La distribución Cauchy es utilizada en teoría de estadísticos ordenados y como distribución a priori de leyes de probabilidad en modelos bayesianos. Modela también las duraciones de actividades sobre las que no existe suficiente información en el análisis de métodos Pert de secuenciación de actividades.

Propiedad 6.13 *Sea $X \sim C(\mu,\theta)$, se tiene que $E[X^k]$ no existe para $k \geq 1$ y que existe para $k < 1$.*

Propiedad 6.14 *Sea $X \sim C(\mu_1,\theta_1)$ e $Y \sim C(\mu_2,\theta_2)$ independientes, entonces $X+Y \sim C(\mu_1,+\mu_2,\theta_1+\theta_2)$.*

Propiedad 6.15 *Se tiene que $X \sim C(1,0)$ si y sólo si $\frac{1}{X} \sim C(1,0)$.*

Ejemplo 6.13 *Un ejercicio de orientación para una persona ciega consiste en hacerla andar en línea recta entre dos paredes paralelas que distan un kilómetro. El grado de desorientación D es la distancia entre el lugar más cercano desde el punto de partida a la segunda pared y el punto en el que la persona ciega alcanzó la segunda pared. Suponiendo que el ángulo θ que forma la primera pared y la dirección escogida por esa persona sigue una distribución uniforme* *en* $-\frac{\pi}{2}$ *y* $\frac{\pi}{2}$. *Obtenga la distribución del grado de desorientación.*

Para calcular el grado de desorientación véase que $D = \tan(\theta)$ *y que* $f_\theta(\theta) = \frac{1}{\theta}$ *para* $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$. *A partir de aquí se calcula la función de distribución de D:*

$$\begin{aligned} F_D(d) &= P(D \leq d) = P(\tan(\theta) \leq d) \\ &= P(\theta \leq \arctan(d)) \\ &= \frac{\arctan(d) + \dfrac{\pi}{2}}{\pi}. \end{aligned}$$

Con lo cual la función de densidad es:

$$f_D(d) = \frac{1}{\pi}\frac{1}{1+d^2} \qquad -\infty \leq d \leq \infty$$

Se trata de una distribución Cauchy de parámetros 1 y 0.

12. Distribuciones derivadas de la normal

12.1. Distribución lognormal

Dada una variable aleatoria $X \sim N(\mu, \sigma)$, se dice que la variable aleatoria $Y = e^X$ sigue una distribución lognormal. La función de densidad de dicha distribución es:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\mathrm{Ln}(y)-\mu)^2}, \quad y > 0.$$

Se tiene que $E[Y] = e^{\mu+\frac{\sigma^2}{2}}$ y $V[Y] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

La distribución lognormal es el resultado de un número elevado de causas independientes con efectos positivos que se componen de manera multiplicativa y donde cada una de estas causas tienen un efecto despreciable frente al global. Esto se debe a que la aditividad de los efectos conduce a una ley normal, en el caso de la ley lognormal, lo hace la proporcionalidad de los efectos.

En el campo industrial, la ley lognormal, puede recibir justificaciones teóricas como las características de un material (resistencia, dureza, etc.) que puede resultar de la combinación multiplicativa de factores elementales. También en el campo económico la ley lognormal se encuentra con frecuencia (distribución de salarios, ventas, etc.).

Ejemplo 6.14 *Se estudia la proporción de rentistas por encima de 18.000€ anuales para un sector económico cuya distribución salarial medida en miles de euros sigue un modelo logaritmo normal con parámetros $\mu = 2$ y $\sigma = 1'2$.*

Se define la variable aleatoria

$$Y = \{\text{renta en dicho sector económico}\}.$$

Puesto que Y sigue una distribución lognormal, se puede considerar una variable $X \sim N(\mu, \sigma)$, tal que, $Y = e^X$, por tanto;

$$\begin{aligned} P(Y \geq 18) &= P(e^X \geq 18) = P(X \geq 2'89) \\ &= P\left(\tfrac{X-\mu}{\sigma} \geq \tfrac{2'89-2}{1'2}\right) \\ &= P(Z \geq 0'74) = 1 - P(Z \leq 0'74) \\ &= 1 - 0'7704 = 0'2296. \end{aligned}$$

12.2. Distribución χ^2

Se dice que una variable aleatoria X sigue una distribución χ^2 con n grados de libertad, se denota por χ^2_n, si su función de densidad es

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0.$$

Se observa que la distribución χ^2_n es una distribución $\Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$, de donde se deduce que esta distribución es reproductiva en su parámetro y que $E[X] = n$, $V[X] = 2n$ y $M_X(t) = (1-2t)^{-\frac{n}{2}}$.

Propiedad 6.16 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas según una $N(0,1)$, se tiene que* $\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$.

Ejemplo 6.15 *Sea V, la velocidad (cm/seg) de un objeto que tiene una masa de 1 Kg., una variable aleatoria con distribución $N(0,25)$. Si $K = \frac{mV^2}{2}$ representa la energía cinética del objeto y se necesita saber la probabilidad de que $K < 400$.*

Puesto que $m = 1$, se tiene que;

$$\begin{aligned} P(K < 400) &= P\left(\frac{mV^2}{2} < 400\right) \\ &= P\left(\frac{V^2}{625} < \frac{400 \cdot 2}{625}\right) \\ &= P\left(\frac{V^2}{625} < 1'28\right) \\ &= P(\chi_1^2 < 1'28) = 0'725. \end{aligned}$$

12.3. Distribución t de Student

Sean X e Y dos variables aleatorias independientes distribuidas según una distribución $N(0,1)$ y χ_n^2, respectivamente.La variable aleatoria

$$T = \frac{X}{\sqrt{\frac{Y}{n}}},$$

sigue una distribución t-Student con n grados de libertad que se denota por t_n.

La función de densidad de esta distribución es

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}, \quad -\infty < x < \infty.$$

Se verifica que $E[X] = 0$ y $V[X] = \dfrac{n}{n-2}$ para $n > 2$.

Propiedad 6.17 *La distribución t de Student es simétrica con respecto al origen.*

12.4. Distribución $\mathcal{F}$ de Snedecor

Sean X e Y dos variables aleatorias independientes distribuidas según una χ^2 con n y m grados de libertad respectivamente. La variable aleatoria

$$F = \frac{\frac{X}{n}}{\frac{Y}{m}},$$

sigue una distribución $\mathcal{F}$ de Snedecor con n y m grados de libertad que se denota por $\mathcal{F}_{n,m}$.

La función de densidad de esta distribución viene dada por

$$f(x) = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} \Gamma(\frac{n+m}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} x^{\frac{m}{2}-1}(n+mx)^{-\frac{n+m}{2}}, \quad x \geq 0.$$

Se tiene que $E[X^k] = \left(\frac{n}{m}\right)^k \frac{\Gamma(k+\frac{m}{2})\Gamma(\frac{n}{2}-k)}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}$ para $n > 2k$, por tanto, $E[X] = \frac{n}{n-2}$ con $n > 2$ y que $V[X] = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$ con $n > 4$.

Propiedad 6.18 *Si $X \sim \mathcal{F}_{n,m}$ entonces $\frac{1}{X} \sim \mathcal{F}_{m,n}$.*

Propiedad 6.19 *Si $X \sim t_n$ entonces $X^2 \sim \mathcal{F}_{1,n}$.*

Ejemplo 6.16 *Las calificaciones de los alumnos en dos asignaturas A y B se distribuyen normalmente con idéntica dispersión σ^2. Se ha observado una muestra aleatoria simple de 51 alumnos presentados al examen de la asignatura A y otra, independiente de la anterior, de 19 alumnos presentados al examen de B. ¿Cuál será la probabilidad de que la varianza observada en la primera muestra sea al menos el doble de la correspondiente a la segunda?*

Sean S_A^2 y S_B^2 las varianzas muestrales de las calificaciones correspondientes a las asignaturas A y B. Puesto que S_A^2 y S_B^2 son variables aleatorias independientes tales que

$$51\frac{S_A^2}{\sigma^2} \sim \chi^2_{50}$$

y que

$$19\frac{S_B^2}{\sigma^2} \sim \chi^2_{18}.$$

Por tanto,

$$\frac{51 \cdot 18 \cdot S_A^2}{50 \cdot 19 \cdot S_B^2} \sim \mathcal{F}_{50,18}.$$

De este modo,

$$P\left(\frac{S_A^2}{S_B^2} \geq 2\right) = P\left(\frac{51 \cdot 18 \cdot S_A^2}{50 \cdot 19 \cdot S_B^2} \geq 0'48\right)$$

$$= P(F_{50,18} \geq 0'48) = 0'9787.$$

13. Distribución de Laplace

Se dice que una variable aleatoria X sigue una distribución de Laplace de parámetros λ y μ, $La(\lambda, \mu)$, si su función de densidad es

$$f(x) = \frac{\lambda}{2}e^{-\lambda|x-\mu|}, \quad -\infty < x < \infty, \quad \lambda > 0 \text{ y } -\infty < \mu < \infty.$$

La función generatriz de momentos de esta distribución tiene la expresión

$$M_X(t) = (1 - \lambda^2 t^2)^{-1} e^{\mu t}, \qquad \text{con } t < \frac{1}{\lambda},$$

de donde se deduce que $E[X] = \mu$ y $V[X] = \dfrac{2}{\lambda^2}$.

La distribución Laplace es una alternativa a la normal para medir los errores de la media.

14. Distribución logística

Se dice que una variable aleatoria X sigue una distribución logística de parámetros a y b, $Lo(a,b)$, si su función de densidad es

$$f(x) = \frac{be^{-(a+bx)}}{(1+e^{-(a+bx)})^2}; \quad -\infty < x < \infty, \ b > 0,$$

siendo su función de distribución

$$F(x) = \frac{1}{1+e^{-(a+bx)}}.$$

Este tipo de distribuciones son usuales en los fenómenos que estudian el crecimiento temporal, como por ejemplo los de origen demográfico.

15. Distribución de Pareto

Se dice que una variable aleatoria X sigue una distribución de Pareto de parámetros α y β, $P(\alpha,\beta)$, si su función de densidad viene dada por la expresión

$$f(x) = \begin{cases} \frac{\beta\alpha^\beta}{x^{\beta+1}} & \text{si } x \geq \alpha \text{ y } \alpha,\beta \geq 0 \\ 0 & \text{en caso contrario.} \end{cases}$$

Se tiene que $E[X] = \dfrac{\alpha\beta}{\beta-1}$ y que $V[X] = \dfrac{\alpha^2\beta}{(\beta-2)(\beta-1)^2}$.

Pareto introdujo esta distribución para describir unidades económicas según la extensión (salarios, rentas, empresas según ventas, ...).

Ejemplo 6.17 *Las rentas salariales anuales en cierto sector económico (X, en miles de euros) es una magnitud aleatoria distribuida según un modelo de Pareto con salario mínimo 9 y $\beta = 2'4$.*

Se desea conocer el salario esperado en el sector y la proporción de asalariados que percibe más de 18.000€/año.

Puesto que X sigue una distribución Pareto se tiene que el salario esperado viene dado por

$$E[X] = \frac{\alpha \cdot \beta}{1-\beta} = \frac{9 \cdot 2'4}{1'4} = 15'43.$$

La proporción de asalariados que perciben más de 18.000€ es

$$\begin{aligned} P(X > 18) &= 1 - P(X \leq 18) \\ &= 1 - \int_9^{18} \frac{2'4 \cdot 9^{2'4}}{x^{3'4}} \,\mathrm{d}\,x \\ &= 1 - \left[-9^{2'4} x^{-2'4}\right]_9^{18} \\ &= 1 - \left[-9^{2'4}\left(18^{-2'4} - 9^{-2'4}\right)\right] \\ &= 1 - 0'81 = 0'19. \end{aligned}$$

16. Algunos modelos multidimensionales

16.1. Distribución multinomial

La distribución multinomial es una generalización de la binomial, en la que el experimento que la genera arroja $k > 2$ resultados distintos en cada realización –la binomial es una distribución dicotómica–, donde las probabilidades de cada uno de los resultados permanece constante a lo largo de todo el proceso. La distribución multinomial se obtiene al realizar, de forma independiente, n pruebas individuales y contar el número de veces que aparece cada resultado. Si se llama A_i al i-ésimo resultado, siendo $P(A_i) = p_i$ y X_i al número de veces que se obtiene A_i en las n pruebas, X_i será la componente i-ésima de la variable k-dimensional X. La distribución de probabilidades de la variable multinomial es:

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k},$$

donde $n_1 + n_2 + \ldots + n_k = n$.

Ejemplo 6.18 *Se lanza un dado 5 veces. Determine la probabilidad de obtener un uno, dos doses y dos cuatros.*

Sea $A_i = \{Obtener\ i\ puntos\ en\ el\ dado\}$.

La probabilidad que se pide es:

$$P(X_1 = 1, X_2 = 2, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 0)$$
$$= \frac{5!}{1!\ 2!\ 0!\ 2!\ 0!\ 0!} \cdot \frac{1}{6} \cdot \frac{1}{6^2} \cdot \frac{1}{6^2} = 0'0038.$$

16.2. Distribución uniforme bidimensional

También se puede generalizar la distribución uniforme al caso de más de una dimensión, se trata el caso $n = 2$, pudiendo comprobar el lector que la generalización para cualquier otra dimensión no entraña la menor dificultad. Se dice que una variable bidimensional $X = (X_1, X_2)$ sigue una distribución uniforme si su función de densidad viene dada por la expresión:

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{(b-a)(d-c)} & \text{si} \quad x_1 \in (a, b), x_2 \in (c, d) \\ 0 & \text{en el resto.} \end{cases}$$

Ejemplo 6.19 *Dos amigos se citan entre las nueve y las diez de la noche, acordando no esperarse más de 10 minutos, ¿cuál es la probabilidad de que se encuentren?*

Se sabe que el tiempo de espera para cada uno se distribuye según una $U(0, 60)$, por lo tanto la distribución conjunta viene dado por:

$$f(x, y) = \begin{cases} \dfrac{1}{60 \cdot 60} & \text{si} \quad 0 \leq x \leq 60 \text{ y } 0 \leq y \leq 60 \\ 0 & \text{en otro caso.} \end{cases}$$

Para que se encuentren los amigos debe ocurrir que $|X - Y| \leq 10$. Por lo tanto:

$$\begin{aligned} P(|X - Y| \leq 10) &= P(-10 \leq X - Y \leq 10) \\ &= P(Y - 10 \leq X \leq Y + 10) \\ &= \int_0^{10} \int_0^{y+10} \frac{1}{3600} \,\mathrm{d}\,x \,\mathrm{d}\,y \\ &+ \int_{10}^{50} \int_{y-10}^{y+10} \frac{1}{3600} \,\mathrm{d}\,x \,\mathrm{d}\,y \\ &+ \int_{50}^{60} \int_{y-10}^{60} \frac{1}{3600} \,\mathrm{d}\,x \,\mathrm{d}\,y = \frac{11}{36}. \end{aligned}$$

16.3. Distribución normal multidimensional

Se dice que una variable n-dimensional $X = (X_1, X_2, \ldots, X_n)$ sigue una distribución normal si su función de densidad es:

$$f(X) = \frac{1}{|M|^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(X-\bar{\mu})M^{-1}(X-\bar{\mu})^t}.$$

Donde M es la matriz de covarianzas y $\bar{\mu}$ el vector de medias.

La función de densidad de la distribución para el caso de una variable bidimensional, que es la única que se puede representar, tiene forma de hongo con asíntotas en cualquier dirección del plano $x_1 x_2$. Además, al cortar la superficie de densidad con cualquier plano perpendicular al $x_1 x_2$ se obtienen curvas con forma de campana. Más formalmente se verifican las siguientes propiedades:

1. Para una variable normal n-dimensional, cualquier subconjunto de $r < n$ variables tiene conjuntamente distribución normal.
2. Si las variables $X_1, X_2, \cdots, X_n$ están incorreladas, entonces son independientes.
3. Cualquier combinación lineal de variables normales es normal. Así, si $Y = XA$, donde A es una matriz de m filas y n columnas, entonces Y es normal con media $\bar{\mu}_X A$ y matriz de covarianzas $A^t M_X A$.

Ejemplo 6.20 *Sea (X, Y) una v.a. normal, cuya función de densidad viene dada por:*

$$f(x,y) = \frac{1}{2\pi} e^{\frac{-(y^2+2x^2-2xy)}{2}} \quad \forall (x,y) \in \mathbb{R}^2.$$

Calcule el vector de medias y la matriz de covarianzas.

Igualando términos con la distribución teórica, se tiene:

- $|M| = 1$
- $-\frac{(X-\bar{\mu})M^{-1}(X-\bar{\mu})^t}{2} = -\frac{y^2+2x^2-2xy}{2}$

Por lo tanto:

$$\bar{\mu} = (0,0) \qquad M^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Con lo que:

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

17. Ejercicios

17.1. Ejercicio resuelto

6.1 La resistencia de una muestra de un determinado material viene dado por una variable aleatoria, X, con función de densidad

$$f(x) = \begin{cases} x & \text{si } 0 \le x < 1 \\ \frac{2x+1}{8} & \text{si } 1 \le x \le 2 \\ 0 & \text{en el resto.} \end{cases}$$

a) Calcule su función de distribución.
b) Calcule $P(0'5 \le X \le 1'5)$.
c) Una muestra de material se encuentra en estado ideal de resistencia si ésta se encuentra entre $0'5$ y $1'5$. Si se consideran 10 muestras de materiales, ¿cuál es la probabilidad de que al menos el 70 % de ellos tenga resistencia ideal?
d) ¿Cuál será el número medio de materiales con resistencia no ideal que se tendrá que escoger hasta encontrar uno con resistencia ideal?
e) ¿Cuál es la probabilidad de que se necesiten 10 muestras para obtener tres de resistencia ideal?

Solución:

a) Usando la definición de función de distribución se obtiene que

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(x)\,\mathrm{d}\,x = 0 \text{ si } x < 0, \\
F(x) &= \int_{-\infty}^{x} f(x)\,\mathrm{d}\,x = \int_{0}^{x} x\,\mathrm{d}\,x = \frac{x^2}{2} \quad \text{si } 0 \leq x < 1, \\
F(x) &= \int_{-\infty}^{x} f(x)\,\mathrm{d}\,x = \int_{0}^{1} x\,\mathrm{d}\,x + \int_{1}^{x} \frac{2x+1}{8} = \frac{1}{2} + \left.\frac{x^2+x}{8}\right|_{1}^{x} \\
&= \frac{1}{2} + \frac{x^2+x-2}{8} \qquad \text{si } 1 \leq x < 2, \\
F(x) &= 1 \quad \text{si } x \geq 2,
\end{aligned}
$$

Es decir, la función de distribución viene dada por

$$
F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{2} & \text{si } 0 \leq x < 1 \\ \frac{1}{2} + \frac{x^2+x-2}{8} & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2. \end{cases}
$$

b) Usando el apartado anterior se obtiene

$$
P(0'5 \leq X \leq 1'5) = F(1'5) - F(0'5) = \frac{1}{2} + \frac{(1'5)^2 + 1'5 - 2}{8} - \frac{0'5^2}{2} = 0'59.
$$

c) Se tiene que la resistencia de una muestra de material es ideal con probabilidad $0'59$ y no ideal $0'41$. Por tanto, el experimento de observar si una muestra es ideal o no, es un experimento Bernouilli con probabilidad de éxito (ser ideal) $0'59$. Con lo cual, el experimento de observar el número de muestras ideales (éxitos) en diez muestras de materiales extraídas aleatoriamente es una distribución binomial de parámetro $n = 10$ y $p = 0'59$. Si se define la variable aleatoria

$$
\begin{aligned}
Y &= \{\text{número de muestras con resistencia ideal entre 10 muestras} \\
&\quad \text{extraídas aleatoriamente}\},
\end{aligned}
$$

se tiene que $Y \sim B(10, 0'59)$. Por tanto, la probabilidad pedida es

$$
P(Y \geq 7) = P(Y = 7) + P(Y = 8) + P(Y = 9) + P(Y = 10)
$$

$$\begin{aligned} &= \binom{10}{7} \cdot 0'59^7 \cdot 0'41^3 + \binom{10}{8} \cdot 0'59^8 \cdot 0'41^2 \\ &\quad + \binom{10}{9} \cdot 0'59^9 \cdot 0'41^1 + \binom{10}{10} \cdot 0'59^10 \cdot 0'41^0 \\ &= 0'3575. \end{aligned}$$

d) Observe que el experimento descrito en este apartado consiste en ir comprobando la resistencia de los materiales hasta encontrar una que sea ideal. Se sabe, por el apartado anterior, que el experimento de probar si una muestra tiene resistencia ideal es un experimento Bernouilli, además puesto que las muestras son independientes, en este apartado, se está contando el número de fracasos (muestras no ideales) que hay que observar hasta conseguir un éxito (muestra ideal) en experimentos Bernouilli independientes. Por tanto, si se define la variable aleatoria

$$\begin{aligned} W &= \{ \text{ número de muestras no ideales (fracasos) antes de} \\ &\quad \text{encontrar una muestra ideal (éxito)}\}, \end{aligned}$$

se tiene que $W \sim Ge(0'59)$.

En este apartado se pide el número medio de fracasos hasta conseguir un éxito, es decir, $E[W] = \dfrac{0'41}{0'59} = 0'695$.

e) Siguiendo con el razonamiento anterior, en este apartado se está pidiendo el número de fracasos (muestras no ideales) antes de obtener el tercer éxito (muestra ideal). Por tanto, si se define la variable aleatoria $T = \{$número de fracasos antes del tercer éxito$\}$, se tiene que, $T \sim BN(3, 0'59)$.

Por tanto, se está pidiendo la probabilidad de que haya 7 fracasos, es decir,

$$P(T = 7) = \binom{9}{7} \cdot 0'59^3 \cdot 0'41^7 = 0'0144.$$

17.2. Ejercicios propuestos

6.1. Dada la distribución $B(10, 0'4)$ calcule las siguientes probabilidades:

a) $P(X \leq 8)$,
b) $P(2 < X \leq 5)$,
c) $P(X \geq 7)$.

6.2. Un conocido fumador gorrón ha explotado tanto a sus compañeros que por término medio cada uno de ellos le da un cigarrillo de cada diez veces que éste les pide.

a) ¿Cuál es la probabilidad de que consiga un cigarrillo en menos de cinco intentos?

b) Si pretende hacer acopio de cigarrillos para el fin de semana, ¿cuántas veces, en promedio, tendrá que pedir tabaco para conseguir 20 unidades?

6.3. A un establecimiento de apuestas deportivas llega un cliente cada tres minutos por término medio.

a) ¿Cuál es la probabilidad de que en un periodo de cinco minutos lleguen más de cinco clientes?

b) ¿Cuál es el número más probable de llegadas en media hora?

6.4. Las compañías aéreas acostumbran a reservar más plazas de las existentes en sus vuelos, dado el porcentaje de anulaciones que se produce. Si el porcentaje medio de anulaciones es del 5 %, ¿cuántas reservas deberá hacer una compañía para un vuelo con 200 plazas, si quiere garantizar al 97 % que todos sus clientes tendrán cabida en dicho vuelo?

6.5. El servicio de reclamaciones de una asociación de consumidores recibe por término medio tres quejas a la hora.

a) Calcule la probabilidad de que en una hora no reciba ninguna reclamación.

b) Calcule la probabilidad de que en dos horas reciba entre dos y seis reclamaciones.

6.6. En una pecera hay diez peces machos y ocho hembras, si se extraen aleatoriamente cinco peces, calcule la probabilidad de que tres sean machos y dos hembras.

6.7. Un jugador apuesta 5€ por tirada a un número de los 37 que componen la ruleta, si acierta, gana 180€. Calcule sus beneficios esperados al cabo de 100 jugadas.

6.8. Por una estación pasa un tren de cercanías cada treinta minutos. Si una persona, que desconoce los horarios, llega a la estación para tomar dicho tren, ¿cuál es la probabilidad de que tenga que esperar menos de cinco minutos?

6.9. ¿Cuál es la probabilidad de que de 10 personas elegidas al azar al menos 2 cumplan años en el mes de Enero?

6.10. Durante la Segunda Guerra Mundial los alemanes bombardearon repetidas veces Londres. Los expertos demostraron que se trataba de bombardeos indiscriminados y que caían en cada acción y por término medio dos bombas por cada cuadrícula de cien metros de lado. En vista a lo anterior, calcule la probabilidad de que en una cierta cuadrícula de cincuenta metros de lado no haya caído ninguna bomba durante un bombardeo.

6.11. Dada una distribución normal de media 3 y varianza 9, calcule las siguientes probabilidades:

a) $P(2 \leq X \leq 5)$,
b) $P(X \geq 3)$,
c) $P(X \leq -2)$.

6.12. Calcule en los siguientes casos el valor de a, sabiendo que $X \sim N(1,5)$.

a) $P(0 \leq X \leq a) = 0'28$,
b) $P(1 - a \leq X < 1 + a) = 0'65$.

6.13. Se sabe que la alarma de un reloj saltará en cualquier momento entre las siete y las ocho de la mañana. Si el propietario del reloj se despierta al oír dicha alarma y necesita, como mínimo, veinticinco minutos para arreglarse y llegar al trabajo,

a) ¿cuál es la probabilidad de que llegue antes de las ocho?

b) Si el dueño del reloj sigue programando el reloj de la misma manera durante 10 días, calcule el número más probable de días en que llegará después de las ocho.

6.14. De una tribu indígena se sabe que los hombres tienen una estatura que se distribuye según una ley normal con media 1'70 y desviación típica σ. Si a través de estudios realizados se conoce que la probabilidad de que su estatura sea mayor a 1'80 es 0'12, calcule la probabilidad de que un individuo elegido al azar mida entre 1'65 y 1'75.

6.15. Calcule la probabilidad de obtener más de 200 seises en 1200 lanzamientos de un dado honrado.

6.16. En una gasolinera se ofrecen tres servicios: suministrar gasolina, lavar coches y comprobar la presión de neumáticos. Estos servicios son demandados con una probabilidad de 0'9, 0'05 y 0'05 respectivamente. ¿Cuál es la probabilidad de que de diez coches que lleguen, seis vayan a repostar, tres vayan al lavado y uno a comprobar la presión?

6.17. Una máquina produce piezas que según su peso pueden clasificarse en "pesadas", "normales" y "ligeras". Por experiencia se ha estimado que el 30 % son pesadas y el 60 % normales. De cinco piezas extraídas, ¿cuál es la probabilidad de que dos sean pesadas y dos normales?

6.18. Se eligen al azar dos números en el intervalo (0,2). ¿Cómo se distribuyen esos números? ¿Cuál es la probabilidad de que la suma de dichos números sea menor que dos?

6.19. El peso de dos tipos de caracoles sigue una distribución normal bivariante con parámetros:

$$\bar{\mu} = (3, 4), \quad M = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}.$$

Calcule $P(2 \leq X \leq 3'5)$ y $P(3X - Y \leq 2)$.

6.20. Sea X una variable aleatoria que mide la cantidad de pH en una determinada sustancia y que tiene como función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < \frac{1}{4} \\ \frac{1}{2} & \text{si } \frac{1}{4} \leq x < \frac{1}{2} \\ k(x - \frac{x^2}{2}) + \frac{1}{8} & \text{si } \frac{1}{2} \leq x < 1 \\ \frac{5}{8} & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2. \end{cases}$$

a) Calcule k para que sea función de distribución.

b) Calcule $P(0 \leq X \leq 1)$.

c) Se considera que esa sustancia es altamente contaminante si su pH está comprendido entre $\frac{1}{2}$ y 1. ¿Determine la probabilidad de que entre 10 muestras de dicha sustancia, tres resulten altamente contaminantes?

6.21. Se sabe que en un colegio de primaria el 50 % de los alumnos son menores de 9 años, el 30 % tienen una edad comprendida entre 9 y 11 años y el 20 % tienen una edad superior a 11 años. Se escogen 20 alumnos al azar, se pide:

a) La probabilidad de que al menos haya uno de edad superior a 11 años.

b) El número esperado de alumnos con edad entre 9 y 11 años.

c) La varianza del número de alumnos con edad superior o igual a 9 años.

6.22. En la central del 061 se sabe que el 15 % de las llamadas corresponden a urgencias debidas a un accidente de tráfico. ¿Cuál es la probabilidad de recibir tres llamadas antes del primer aviso telefónico por accidente de tráfico? ¿Cuál es el número medio de llamadas que se reciben hasta que una de ellas avisa de un accidente de tráfico?

6.23. El número de personas que llegan a la ventanilla de un banco sigue una ley Poisson. Se sabe que la varianza del número de llegadas en un minuto es 4.

a) Calcule la probabilidad de que en un minuto no lleguen más de dos personas.

b) Si la persona que atiende la ventanilla se bloquea cuando llegan más de dos personas, ¿cuál es la probabilidad de que entre diez minutos escogidos al azar durante un día se haya bloqueado en dos?

c) ¿Cuál es la probabilidad de que después de 10 minutos se bloquee por segunda vez?

6.24. El número de llamadas recibidas en una centralita en 15 minutos sigue una Poisson de media 1.

a) Calcule la probabilidad de que si a las 5:30 se ha recibido una llamada, la siguiente se produzca después de las 6:10.

b) ¿Cuál es la probabilidad de que en 20 minutos se reciban 3 llamadas?

6.25. El número de personas que llegan a un semáforo para cruzar la calle sigue una ley Poisson. Se sabe que por término medio llegan dos personas cada cinco minutos.

a) Calcule la probabilidad de que en 7 minutos lleguen 3 personas.

b) Calcule la probabilidad de que el tiempo transcurrido entre dos llegadas sea superior a 3 minutos.

6.26. El cañón de luz de una prisión gira sobre si mismo de forma que tarda en alumbrar una misma zona 40 segundos. Un preso organiza una fuga de la prisión necesitando 27 segundos para llegar y escalar el muro. Si el preso emprende una fuga eligiendo el momento de forma aleatoria, ¿cuál es la probabilidad de que no sea visto?

6.27. En un hospital de maternidad se sabe que el peso de un niño recién nacido sigue una ley normal, que 2 de cada 10 niños pesa más de 4’5 Kg. y que 7 de cada 10 niños pesan menos de 3’5 Kg.

a) Calcule la probabilidad de que un niño pese entre 2’75 y 3’75 Kg.

b) Si se considera que un niño tiene un peso ideal si su peso esta comprendido entre 2’75 y 3’75. ¿Cuál es la probabilidad de que al considerar bebés aleatoriamente obtengamos que el segundo bebé con peso ideal sea el quinto considerado?

6.28. Un fabricante de papel de aluminio vende rollos de 8 metros. Para ello, sabe que la longitud de los rollos cortados sigue una distribución normal de media 7’5 metros. Además se sabe que el 20 % de los rollos miden más de 8 metros.

a) Calcule la probabilidad de que un rollo mida más de 9 metros.

b) Si el empresario almacena sus rollos y quiere encontrar algún rollo con más de 8 metros, ¿cuántos rollos deberá medir en promedio?

6.29. Un avión tiene la misión de bombardear un edificio cuya vista aérea es un rectángulo de 100 metros de largo y 50 metros de ancho. Se sabe que el edificio quedará seriamente dañado si la bomba cae en el círculo central de radio 2 metros o en los triángulos formados por las esquinas y que tienen 2 metros como longitud de sus catetos. Calcule la probabilidad de que el edificio resulte seriamente dañado.

6.30. Probar que si X e Y son $N(0, 1)$ e independientes, entonces $\frac{X}{Y}$ sigue una distribución de Cauchy.

6.31. En una carretera se han observado los intervalos entre el paso de dos vehículos sucesivos (X, en segundos), esta magnitud sigue un modelo gamma con a=1 y p=20’5. Calcule la probabilidad de que el tiempo transcurrido entre el paso de dos vehículos sea mayor de 28'9 segundos.

6.32. El tiempo de funcionamiento de un sistema de radar se modela como una distribución gamma con a = 1'5 y p = 2. Determine la probabilidad de que el sistema funcione al menos 1 año antes del fallo. ¿Con qué probabilidad el fallo se produce durante el segundo año desde su puesta en funcionamiento?

6.33. Se ha determinado que el tiempo de reparación en un taller de mantenimiento sigue una distribución lognormal de media 7 horas y desviación típica 4 horas.

a) ¿Con qué probabilidad una reparación superará las 10 horas?

b) ¿Qué proporción de reparaciones tienen un tiempo de ejecución entre 8 y 16 horas?

6.34. Se sabe que el tiempo de funcionamiento en años de un sistema se distribuye según una χ^2 con 7 grados de libertad. ¿Con qué probabilidad el sistema funcionará al menos tres años?

Apéndice A

Combinatoria

1. Introducción

La Combinatoria estudia las diferentes formas en que se puede llevar a cabo una cierta tarea de ordenación o agrupación de unos cuantos objetos siguiendo unas reglas prefijadas.

Es una herramienta muy importante en el cálculo de probabilidades, puesto que permite contar los casos favorables y los posibles y, por tanto, calcular probabilidades en aquellas situaciones en que todos los sucesos sean equiprobables.

2. Variaciones con repetición

Sean m elementos distintos $a_1, a_2, a_3, \ldots, a_m$; se pretende ocupar n lugares con ellos de modo que cada elemento pueda ocupar más de un lugar. Las distintas disposiciones se llaman variaciones con repetición de m elementos tomados n a n; el número total de éstas se nota por $VR_{m,n}$ y es igual a m^n.

Ejemplo A.1 *El número de quinielas de fútbol que hay que hacer para acertar 15 con seguridad es: $VR_{3,15} = 3^{15}$. Esto es así, puesto que los resultados posibles son tres: el 1, la x y el 2, en cada uno de los quince* *partidos que conforman la quiniela.*

3. Variaciones

Sean m elementos $a_1, a_2, \ldots, a_m$. Se pretende ocupar n lugares con ellos de modo que cada elemento sólo ocupe un lugar. (En este caso ha de ser $n < m$). Las distintas disposiciones se llaman variaciones de m elementos tomados n a n, formalmente $V_{m,n}$ y su número es: $V_{m,n} = m(m-1)(m-2)\ldots(m-n+1)$. O lo que es lo mismo:

$$V_{m,n} = \frac{m!}{(m-n)!}.$$

Ejemplo A.2 *En una quiniela hípica hay que acertar los tres primeros caballos que llegan a meta en una carrera en la que hay diez competidores. El número de quinielas que hay que hacer para asegurar el acierto es:* $V_{10,3} = 10 \cdot 9 \cdot 8 = 720$.

4. Permutaciones

Las permutaciones sin repetición de n elementos dan el número de ordenaciones distintas que se pueden realizar con los n elementos. El número total de éstas se nota por $P_n = n(n-1)(n-2)\ldots 2 \cdot 1$.

A este número se le llama n factorial o factorial de n y se representa por $n!$. Por otra parte es evidente que las permutaciones de n elementos coinciden con las variaciones sin repetición de n elementos tomados n a n. Es decir, $P_n = V_{n,n} = n!$.

Ejemplo A.3 *Si se tienen que colocar siete libros en una librería se puede hacer de* $P_7 = 7! = 5{,}040$ *formas distintas.*

5. Permutaciones con repetición

Dados los elementos $a_1, a_2, a_3, \ldots, a_n$, se llaman permutaciones con repetición de orden $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ a cada uno de los grupos de $\alpha_1 + \alpha_2 + \ldots + \alpha_n$ elementos que se pueden formar con la condición de que haya α_1 elementos iguales a a_1, α_2 elementos iguales a a_2; ..., α_n elementos iguales a a_n. Si $\alpha_1 + \alpha_2 + \ldots + \alpha_n = m$, el número total se nota por $P_m^{\alpha_1,\alpha_2,\ldots,\alpha_n} = PR(\alpha_1, \alpha_2, \ldots, \alpha_n)$ y vale:

$$\frac{m!}{\alpha_1!\alpha_2!\ldots\alpha_n!}$$

Ejemplo A.4 *Si se desea repartir 3 relojes, 2 bicicletas y 4 pelotas entre 9 niños, de modo que cada uno de ellos reciba un regalo, se tienen,* $PR_9^{3,2,4} = 9!/3!2!4! = 1260$ *formas de hacerlo.*

6. Combinaciones sin repetición

Se consideran m elementos distintos. Se pretenden seleccionar n, con $n < m$, de ellos sin importarnos el lugar que ocupen, sino tan solo su pertenencia al grupo. Las distintas selecciones se llaman combinaciones de m elementos tomados n a n. El número total de ellas se representa por $C_{m,n}$, siendo su valor igual a:

$$C_{m,n} = \frac{V_{m,n}}{P_n} = \frac{m!}{(m-n)!n!}.$$

Ejemplo A.5 *En una carrera donde compiten 10 corredores y se clasifican los tres primeros para la fase siguiente, puede haber tantas combinaciones de clasificados como* $C_{10,3} = 120$.

6.1. Propiedades de los números combinatorios

A los valores de $C_{m,n}$ se les llama números combinatorios y se les designa por:

$$C_{m,n} = \binom{m}{n}.$$

Los números combinatorios verifican las siguientes propiedades:

1. $\binom{m}{0} = \binom{m}{m} = 1$

2. $\binom{m}{n} = \binom{m}{m-n}$

3. $\binom{m-1}{n-1} + \binom{m-1}{n} = \binom{m}{n}$

7. Combinaciones con repetición

Se llaman combinaciones con repetición de n elementos tomados r a r a cada uno de los grupos que pueden formarse con r elementos elegidos de entre n posibles, sin importar el que se repitan. Se nota por $CR_{n,r}$ y vale:

$$\binom{n+r-1}{r}$$

Ejemplo A.6 *Si se dispone de 3 bolas iguales a las que hay que distribuir en 5 cajas distinguibles, se pueden hacer tantas combinaciones como* $\binom{7}{3} = 35$.

Observe que los elementos son las cajas y que el que una bola esté dentro de una caja sólo significa que esa caja es una de las tres seleccionadas. Si las tres bolas estuvieran en la misma caja, se seleccionaría dicha caja tres veces.

8. Ejercicios

8.1. Ejercicio resuelto

A.1 En un instituto los alumnos de 2º de Bachillerato deciden realizar un sorteo para el viaje de fin de curso. Para numerar las papeletas deciden utilizar únicamente los dígitos 1, 2, 3, 4, 5. Cuántas papeletas distintas de cuatro dígitos podrán vender si:

a) Los cuatro dígitos son distintas.
b) Pueden aparecer dígitos repetidos.
c) Aparecen 3 unos y 1 cinco.
d) Sólo se utilizan los dígitos 2, 3, 4 y 5, sin repetir ninguna.

e) Sólo se utilizan los dígitos 2, 3 ,4 y 5, pero se pueden repetir.

f) No se tiene en cuenta el orden, pero los dígitos son distintos.

g) No se tiene en cuenta el orden, pero los dígitos pueden ser repetidos.

Solución:

a) En esta situación al influir el orden y no aparecer dígitos repetidos en un mismo número, se trata de variaciones ordinarias de 5 elementos tomados de 4 en 4, por lo tanto se tendrán tantas papeletas distintas para el sorteo como $V_{5,4} = 5 \cdot 4 \cdot 3 \cdot 2 = 120$.

b) Se sigue con variaciones pero ahora serán con repetición, siendo su número de $VR_{5,4} = 5^4 = 625$.

c) Al tener 4 dígitos para generar el número, intervendrán todos en cada número, por lo que se trata de permutaciones con repetición, $PR_4 3, 1 = \frac{4!}{3!1!} = 4$.

d) Se trata de permutaciones ordinarias de 4 elementos tomados de cuatro en cuatro, $P_4 = 4! = 24$

e) Para obtener su número hay que tener en cuenta que influye el orden y que se pueden repetir los dígitos, con lo que se trata de variaciones con repetición de 4 elementos tomados de cuatro en cuatro, $VR_{4,4} = 4^4$.

f) Al no influir el orden y no repetirse ningún valor, su número se obtiene como combinaciones sin repetición de 5 elementos tomados de cuatro en cuatro, $C_{5,4} = \binom{5}{4} = 5$

g) Igual que en el caso anterior no influye el orden, pero se pueden repetir los valores, por lo tanto se trata de combinaciones con repetición de 5 elementos tomados de 4 en 4, $CR_{5,4} = \binom{5+4-1}{4} = 70$

8.2. Ejercicios propuestos

A.1. Con las cifras l, 2, 3, 4, 5, ¿cuántos números de dos cifras distintas pueden formarse?

A.2. ¿De cuántas formas distintas puede colocarse un equipo de fútbol para hacerse una foto, sabiendo que seis jugadores permanecen de pie y el resto en cuclillas delante de los primeros y que el portero siempre se sitúa de pie?

A.3. Suponga que le hacen el encargo de diseñar la bandera de un nuevo país, para lo que dispone de cinco colores; si la bandera debe tener tres bandas horizontales de igual anchura, ¿cuántas banderas diferentes podrá diseñar?

A.4. Un automóvil de cinco plazas está ocupado por dos conductores y tres no conductores. Sabiendo que los dos conductores no pueden ocupar simultáneamente las dos plazas delanteras, ¿de cuántas formas distintas pueden acomodarse los ocupantes del coche?

A.5. En un congreso de Estadística al que asisten 40 personas se han habilitado tres salas para defender, simultáneamente, las ponencias. ¿De cuántas formas distintas pueden distribuirse los asistentes entre las salas? Suponga que la capacidades de las salas son de 16, 14 y 10 personas.

A.6. El jefe de cocina de un comedor universitario dispone de cinco primeros platos, ocho segundos y cuatro postres, para combinarlos y formar menús en el mes de Noviembre. ¿Cuántos menús diferentes puede confeccionar?

A.7. ¿De cuántas formas distintas pueden acomodarse 170 pasajeros en un avión de 200 plazas?

A.8. En una competición de tenis hay 32 inscritos, ¿de cuántas formas distintas se pueden emparejar los jugadores para disputar la primera ronda?

A.9. En una compañía de baile hay diez hombres y diez mujeres. ¿Cuántas parejas distintas puede formar su director?

A.10. ¿Cuántas quinielas distintas pueden formarse con cinco x, siete 1 y tres 2?

A.11. Obtenga el número de permutaciones que se pueden formar con las letras de la palabra UNIVERSIDAD sin que haya dos consonantes seguidas.

A.12. Durante un debate 8 personas se sientan en una mesa redonda, ¿de cuántas formas distintas se pueden colocar? Conteste a la anterior cuestión si la mesa tiene forma de herradura.

A.13. Una marca de vehículos a motor dispone de 15 probadores de automóviles, 12 de motocicletas y 6 de camiones. Por cuestiones operativas, se forman equipos con 5 probadores de coche, 3 de motocicletas y 2 de camiones. ¿Cuántos equipos se pueden crear?

Apéndice B

Tablas Estadísticas

Tabla B.1: Distribución Binomial

n	k	0'01	0'05	0'10	0'15	0'20	0'25	0'30	1/3	0'35	0'40	0'45	0'50
2	0	0'9801	0'9025	0'8100	0'7225	0'6400	0'5625	0'4900	0'4444	0'4225	0'3600	0'3025	0'2500
	1	0'9999	0'9975	0'9900	0'9775	0'9600	0'9375	0'9100	0'8889	0'8775	0'8400	0'7975	0'7500
	2	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
3	0	0'9703	0'8574	0'7290	0'6141	0'5120	0'4219	0'3430	0'2963	0'2746	0'2160	0'1664	0'1250
	1	0'9997	0'9928	0'9720	0'9392	0'8960	0'8438	0'7840	0'7407	0'7182	0'6480	0'5748	0'5000
	2	1'0000	0'9999	0'9990	0'9966	0'9920	0'9844	0'9730	0'9630	0'9571	0'9360	0'9089	0'8750
	3		1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
4	0	0'9606	0'8145	0'6561	0'5220	0'4096	0'3164	0'2401	0'1975	0'1785	0'1296	0'0915	0'0625
	1	0'9994	0'9860	0'9477	0'8905	0'8192	0'7383	0'6517	0'5926	0'5630	0'4752	0'3910	0'3125
	2	1'0000	0'9995	0'9963	0'9880	0'9728	0'9492	0'9163	0'8889	0'8735	0'8208	0'7585	0'6875
	3		1'0000	0'9999	0'9995	0'9984	0'9961	0'9919	0'9876	0'9850	0'9744	0'9590	0'9375
	4			1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
5	0	0'9510	0'7738	0'5905	0'4437	0'3277	0'2373	0'1681	0'1317	0'1160	0'0778	0'0503	0'0312
	1	0'9990	0'9774	0'9185	0'8352	0'7373	0'6328	0'5282	0'4609	0'4284	0'3370	0'2562	0'1875
	2	1'0000	0'9988	0'9914	0'9734	0'9421	0'8965	0'8369	0'7901	0'7648	0'6826	0'5931	0'5000
	3		1'0000	0'9995	0'9978	0'9933	0'9844	0'9692	0'9547	0'9460	0'9130	0'8688	0'8125
	4			1'0000	0'9999	0'9997	0'9990	0'9976	0'9959	0'9947	0'9898	0'9815	0'9688
	5				1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
6	0	0'9415	0'7351	0'5314	0'3771	0'2621	0'1780	0'1176	0'0878	0'0754	0'0467	0'0277	0'0156
	1	0'9986	0'9672	0'8857	0'7765	0'6553	0'5339	0'4202	0'3512	0'3191	0'2333	0'1636	0'1094
	2	1'0000	0'9978	0'9842	0'9527	0'9011	0'8306	0'7443	0'6804	0'6471	0'5443	0'4415	0'3438
	3		0'9999	0'9987	0'9941	0'9830	0'9624	0'9295	0'8999	0'8826	0'8208	0'7447	0'6562
	4		1'0000	0'9999	0'9996	0'9984	0'9954	0'9891	0'9822	0'9777	0'9590	0'9308	0'8906
	5			1'0000	1'0000	0'9999	0'9998	0'9993	0'9986	0'9982	0'9959	0'9917	0'9844
	6					1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
7	0	0'9321	0'6983	0'4783	0'32116	0'2097	0'1335	0'0824	0'0585	0'0490	0'0280	0'0152	0'0078
	1	0'9980	0'9556	0'8503	0'7166	0'5767	0'4449	0'3294	0'2634	0'2338	0'1586	0'1024	0'0625
	2	1'0000	0'9962	0'9743	0'9262	0'8520	0'7564	0'6471	0'5706	0'5323	0'4199	0'3164	0'2266
	3		0'9998	0'9973	0'9879	0'9667	0'9294	0'8740	0'8267	0'8011	0'7102	0'6083	0'5000
	4		1'0000	0'9998	0'9988	0'9953	0'9871	0'9712	0'9547	0'9444	0'9037	0'8471	0'7734
	5			1'0000	0'9999	0'9996	0'9987	0'9962	0'9931	0'9910	0'9812	0'9643	0'9375
	6				1'0000	1'0000	0'9999	0'9998	0'9995	0'9994	0'9984	0'9963	0'9922
	7						1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
8	0	0'9227	0'6634	0'4305	0'2725	0'1678	0'1001	0'0576	0'0390	0'0319	0'0168	0'0084	0'0039
	1	0'9973	0'9428	0'8131	0'6572	0'5033	0'3671	0'2553	0'1951	0'1691	0'1064	0'0632	0'0352
	2	0'9999	0'9942	0'9619	0'8948	0'7969	0'6785	0'5518	0'4682	0'4278	0'3154	0'2201	0'1445
	3	1'0000	0'9996	0'9950	0'9786	0'9437	0'8862	0'8059	0'7413	0'7064	0'5941	0'4770	0'3633
	4		1'0000	0'9996	0'9971	0'9896	0'9727	0'9420	0'9121	0'8939	0'8263	0'7396	0'6367
	5			1'0000	0'9998	0'9988	0'9958	0'9887	0'9803	0'9747	0'9502	0'9115	0'8555
	6				1'0000	0'9999	0'9996	0'9987	0'9974	0'9964	0'9915	0'9819	0'9648
	7					1'0000	1'0000	0'9999	0.9998	0'9998	0'9993	0'9983	0'9961
	8							1'0000	1'0000	1'0000	1'0000	1'0000	1'0000
9	0	0'9135	0'6302	0'3874	0'2316	0'1342	0'0751	0'0404	0'0260	0'0207	0'0101	0'0046	0'0020
	1	0'9965	0'9288	0'7748	0'5995	0'4362	0'3003	0'1960	0'1431	0'1211	0'0705	0'0385	0'0195
	2	0'9999	0'9916	0'9470	0'8591	0'7382	0'6007	0'4628	0'3772	0'3373	0'2318	0'1495	0'0898
	3	1'0000	0'9994	0'9917	0'9661	0'9144	0'8343	0'7297	0'6503	0'6089	0'4826	0'3614	0'2539
	4		1'0000	0'9991	0'9944	0'9804	0'9511	0'9012	0'8552	0'8283	0'7334	0'6214	0'5000
	5			0'9999	0'9994	0'9969	0'9900	0'9747	0'9576	0'9464	0'9006	0'8342	0'7461
	6			1'0000	1'0000	0'9997	0'9987	0'9957	0'9917	0'9888	0'9750	0'9502	0'9102
	7					1'0000	0'9999	0'9996	0'9990	0'9986	0'9962	0'9909	0'9805
	8						1'0000	1'0000	0'9999	0'9999	0'9997	0'9992	0'9980
	9								1'0000	1'0000	1'0000	1'0000	1'0000
10	0	0'9044	0'5987	0'3487	0'1969	0'1074	0'0563	0'0282	0'0173	0'0135	0'0060	0'0025	0'0010
	1	0'9958	0'9139	0'7361	0'5443	0'3758	0'2440	0'1493	0'1040	0'0860	0'0464	0'0233	0'0107
	2	1'0000	0'9885	0'9298	0'8202	0'6778	0'5256	0'3828	0'2991	0'2616	0'1673	0'0996	0'0547
	3		0'9990	0'9872	0'9500	0'8791	0'7759	0'6496	0'5593	0'5138	0'3823	0'2660	0'1719
	4		0'9999	0'9984	0'9901	0'9672	0'9219	0'8497	0'7869	0'7515	0'6331	0'5044	0'3770
	5		1'0000	0'9999	0'9986	0'9936	0'9803	0'9527	0'9234	0'9051	0'8338	0'7384	0'6230
	6			1'0000	0'9999	0'9991	0'9965	0'9894	0'9803	0'9740	0'9452	0'8980	0'8281
	7				1'0000	0'9999	0'9996	0'9984	0'9966	0'9952	0'9877	0'9726	0'9453
	8					1'0000	1'0000	0'9999	0'9996	0'9995	0'9983	0'9955	0'9893
	9							1'0000	1'0000	1'0000	0'9999	0'9997	0'9990
	10										1'0000	1'0000	1'0000

Tabla B.2: Distribución de Poisson

k	0'1	0'2	0'3	0'4	0'5	0'6	0'7	0'8	0'9	1'0	1'1	1'2	1'3	1'4
0	0'905	0'819	0'741	0'670	0'611	0'549	0'497	0'449	0'407	0'368	0'333	0'301	0'273	0'247
1	0'995	0'982	0'963	0'938	0'910	0'878	0'844	0'809	0'772	0'736	0'699	0'663	0'627	0'592
2	1'000	0'999	0'996	0'992	0'986	0'977	0'966	0'953	0'937	0'920	0'900	0'879	0'857	0'833
3		1'000	1'000	0'999	0'998	0'997	0'994	0'991	0'987	0'981	0'974	0'966	0'957	0'946
4				1'000	1'000	1'000	0'999	0'999	0'998	0'996	0'995	0'992	0'989	0'986
5							1'000	1'000	1'000	0'999	0'999	0'998	0'998	0'997
6										1'000	1'000	1'000	1'000	0'999
7														1'000

k	1'5	1'6	1'7	1'8	1'9	2'0	2'2	2'4	2'6	2'8	3'0	3'2	3'4	3'6
0	0'223	0'202	0'183	0'165	0'150	0'135	0'111	0'091	0'074	0'061	0'050	0'041	0'033	0'027
1	0'558	0'525	0'493	0'463	0'434	0'406	0'355	0'308	0'267	0'231	0'199	0'171	0'147	0'126
2	0'809	0'783	0'757	0'731	0'704	0'677	0'623	0'570	0'518	0'469	0'423	0'380	0'340	0'303
3	0'934	0'921	0'907	0'891	0'875	0'857	0'819	0'779	0'736	0'692	0'647	0'603	0'558	0'515
4	0'981	0'976	0'970	0'964	0'956	0'947	0'928	0'904	0'877	0'848	0'815	0'781	0'744	0'706
5	0'996	0'994	0'992	0'990	0'987	0'983	0'975	0'964	0'951	0'935	0'916	0'895	0'871	0'844
6	0'999	0'999	0'998	0'997	0'997	0'995	0'993	0'988	0'983	0'976	0'966	0'955	0'942	0'927
7	1'000	1'000	1'000	0'999	0'999	0'999	0'998	0'997	0'995	0'992	0'988	0'983	0'977	0'969
8				1'000	1'000	1'000	1'000	0'999	0'999	0'998	0'996	0'994	0'992	0'988
9								1'000	1'000	0'999	0'999	0'998	0'997	0'996
10										1'000	1'000	1'000	0'999	0'999
11													1'000	1'000

k	3'8	4'0	4'2	4'4	4'6	4'8	5'0	5'2	5'4	5'6	5'8	6'0	6'5	7
0	0'022	0'018	0'015	0'012	0'010	0'008	0'007	0'006	0'005	0'004	0'003	0'002	0'002	0'001
1	0'107	0'092	0'078	0'066	0'056	0'048	0'040	0'034	0'029	0'024	0'021	0'017	0'011	0'007
2	0'269	0'238	0'210	0'185	0'163	0'143	0'125	0'109	0'095	0'082	0'072	0'062	0'043	0'030
3	0'473	0'433	0'395	0'359	0'326	0'294	0'265	0'238	0'213	0'191	0'170	0'151	0'112	0'082
4	0'668	0'629	0'590	0'551	0'513	0'476	0'440	0'406	0'373	0'342	0'313	0'285	0'223	0'173
5	0'816	0'785	0'753	0'720	0'686	0'651	0'616	0'581	0'546	0'512	0'478	0'446	0'369	0'301
6	0'909	0'889	0'867	0'844	0'818	0'791	0'762	0'732	0'702	0'670	0'638	0'606	0'527	0'450
7	0'960	0'949	0'936	0'921	0'905	0'887	0'867	0'845	0'822	0'797	0'771	0'744	0'673	0'599
8	0'984	0'979	0'972	0'964	0'955	0'944	0'932	0'918	0'903	0'886	0'867	0'847	0'792	0'730
9	0'994	0'992	0'989	0'985	0'980	0'975	0'968	0'960	0'951	0'941	0'929	0'916	0'877	0'830
10	0'998	0'997	0'996	0'994	0'992	0'990	0'986	0'982	0'977	0'972	0'965	0'957	0'933	0'901
11	0'999	0'999	0'999	0'998	0'997	0'996	0'995	0'993	0'990	0'988	0'984	0'980	0'966	0'947
12	1'000	1'000	1'000	0'999	0'999	0'999	0'998	0'997	0'996	0'995	0'993	0'991	0'984	0'973
13				1'000	1'000	1'000	0'999	0'999	0'999	0'998	0'997	0'996	0'993	0'987
14							1'000	1'000	0'999	0'999	0'999	0'999	0'997	0'994
15									1'000	1'000	1'000	0'999	0'999	0'998
16												1'000	1'000	0'999

Tabla B.3: Distribución Normal

	0	0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09
0	0'5000	0'5040	0'5080	0'5120	0'5160	0'5199	0'5239	0'5279	0'5319	0'5359
0'1	0'5398	0'5438	0'5478	0'5517	0'5557	0'5596	0'5636	0'5675	0'5714	0'5753
0'2	0'5793	0'5832	0'5871	0'5910	0'5948	0'5987	0'6026	0'6064	0'6103	0'6141
0'3	0'6179	0'6217	0'6255	0'6293	0'6331	0'6368	0'6406	0'6443	0'6480	0'6517
0'4	0'6554	0'6591	0'6628	0'6664	0'6700	0'6736	0'6772	0'6808	0'6844	0'6879
0'5	0'6915	0'6950	0'6985	0'7019	0'7054	0'7088	0'7123	0'7157	0'7190	0'7224
0'6	0'7257	0'7291	0'7324	0'7357	0'7389	0'7422	0'7454	0'7486	0'7517	0'7549
0'7	0'7580	0'7611	0'7642	0'7673	0'7704	0'7734	0'7764	0'7794	0'7823	0'7852
0'8	0'7881	0'7910	0'7939	0'7967	0'7995	0'8023	0'8051	0'8078	0'8106	0'8133
0'9	0'8159	0'8186	0'8212	0'8238	0'8264	0'8289	0'8315	0'8340	0'8365	0'8389
1	0'8413	0'8438	0'8461	0'8485	0'8508	0'8531	0'8554	0'8577	0'8599	0'8621
1'1	0'8643	0'8665	0'8686	0'8708	0'8729	0'8749	0'8770	0'8790	0'8810	0'8830
1'2	0'8849	0'8869	0'8888	0'8907	0'8925	0'8944	0'8962	0'8980	0'8997	0'9015
1'3	0'9032	0'9049	0'9066	0'9082	0'9099	0'9115	0'9131	0'9147	0'9162	0'9177
1'4	0'9192	0'9207	0'9222	0'9236	0'9251	0'9265	0'9279	0'9292	0'9306	0'9319
1'5	0'9332	0'9345	0'9357	0'9370	0'9382	0'9394	0'9406	0'9418	0'9429	0'9441
1'6	0'9452	0'9463	0'9474	0'9484	0'9495	0'9505	0'9515	0'9525	0'9535	0'9545
1'7	0'9554	0'9564	0'9573	0'9582	0'9591	0'9599	0'9608	0'9616	0'9625	0'9633
1'8	0'9641	0'9649	0'9656	0'9664	0'9671	0'9678	0'9686	0'9693	0'9699	0'9706
1'9	0'9713	0'9719	0'9726	0'9732	0'9738	0'9744	0'9750	0'9756	0'9761	0'9767
2	0'9772	0'9778	0'9783	0'9788	0'9793	0'9798	0'9803	0'9808	0'9812	0'9817
2'1	0'9821	0'9826	0'9830	0'9834	0'9838	0'9842	0'9846	0'9850	0'9854	0'9857
2'2	0'9861	0'9864	0'9868	0'9871	0'9875	0'9878	0'9881	0'9884	0'9887	0'9890
2'3	0'9893	0'9896	0'9898	0'9901	0'9904	0'9906	0'9909	0'9911	0'9913	0'9916
2'4	0'9918	0'9920	0'9922	0'9925	0'9927	0'9929	0'9931	0'9932	0'9934	0'9936
2'5	0'9938	0'9940	0'9941	0'9943	0'9945	0'9946	0'9948	0'9949	0'9951	0'9952
2'6	0'9953	0'9955	0'9956	0'9957	0'9959	0'9960	0'9961	0'9962	0'9963	0'9964
2'7	0'9965	0'9966	0'9967	0'9968	0'9969	0'9970	0'9971	0'9972	0'9973	0'9974
2'8	0'9974	0'9975	0'9976	0'9977	0'9977	0'9978	0'9979	0'9979	0'9980	0'9981
2'9	0'9981	0'9982	0'9982	0'9983	0'9984	0'9984	0'9985	0'9985	0'9986	0'9986
3	0'9987	0'9987	0'9987	0'9988	0'9988	0'9989	0'9989	0'9989	0'9990	0'9990
3'1	0'9990	0'9991	0'9991	0'9991	0'9992	0'9992	0'9992	0'9992	0'9993	0'9993
3'2	0'9993	0'9993	0'9994	0'9994	0'9994	0'9994	0'9994	0'9995	0'9995	0'9995
3'3	0'9995	0'9995	0'9995	0'9996	0'9996	0'9996	0'9996	0'9996	0'9996	0'9997
3'4	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9997	0'9998
3'5	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998	0'9998
3'6	0'9998	0'9998	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999
3'7	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999
3'8	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999	0'9999
3'9	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000	1'0000

Tabla B.4: Puntos Críticos: Distribución t de Student

	0’9995	0’995	0’9875	0’975	0’95	0’875	0’85	0’8	0’75	0’7	0’65	0’6	0’55
1	636’58	63’656	25’452	12’706	6’3137	2’4142	1’9626	1’3764	1’0000	0’7265	0’5095	0’3249	0’1584
2	31’600	9’9250	6’2054	4’3027	2’9200	1’6036	1’3862	1’0607	0’8165	0’6172	0’4447	0’2887	0’1421
3	12’924	5’8408	4’1765	3’1824	2’3534	1’4226	1’2498	0’9785	0’7649	0’5844	0’4242	0’2767	0’1366
4	8’6101	4’6041	3’4954	2’7765	2’1318	1’3444	1’1896	0’9410	0’7407	0’5686	0’4142	0’2707	0’1338
5	6’8685	4’0321	3’1634	2’5706	2’0150	1’3009	1’1558	0’9195	0’7267	0’5594	0’4082	0’2672	0’1322
6	5’9587	3’7074	2’9687	2’4469	1’9432	1’2733	1’1342	0’9057	0’7176	0’5534	0’4043	0’2648	0’1311
7	5’4081	3’4995	2’8412	2’3646	1’8946	1’2543	1’1192	0’8960	0’7111	0’5491	0’4015	0’2632	0’1303
8	5’0414	3’3554	2’7515	2’3060	1’8595	1’2403	1’1081	0’8889	0’7064	0’5459	0’3995	0’2619	0’1297
9	4’7809	3’2498	2’6850	2’2622	1’8331	1’2297	1’0997	0’8834	0’7027	0’5435	0’3979	0’2610	0’1293
10	4’5868	3’1693	2’6338	2’2281	1’8125	1’2213	1’0931	0’8791	0’6998	0’5415	0’3966	0’2602	0’1289
11	4’4369	3’1058	2’5931	2’2010	1’7959	1’2145	1’0877	0’8755	0’6974	0’5399	0’3956	0’2596	0’1286
12	4’3178	3’0545	2’5600	2’1788	1’7823	1’2089	1’0832	0’8726	0’6955	0’5386	0’3947	0’2590	0’1283
13	4’2209	3’0123	2’5326	2’1604	1’7709	1’2041	1’0795	0’8702	0’6938	0’5375	0’3940	0’2586	0’1281
14	4’1403	2’9768	2’5096	2’1448	1’7613	1’2001	1’0763	0’8681	0’6924	0’5366	0’3933	0’2582	0’1280
15	4’0728	2’9467	2’4899	2’1315	1’7531	1’1967	1’0735	0’8662	0’6912	0’5357	0’3928	0’2579	0’1278
16	4’0149	2’9208	2’4729	2’1199	1’7459	1’1937	1’0711	0’8647	0’6901	0’5350	0’3923	0’2576	0’1277
17	3’9651	2’8982	2’4581	2’1098	1’7396	1’1910	1’0690	0’8633	0’6892	0’5344	0’3919	0’2573	0’1276
18	3’9217	2’8784	2’4450	2’1009	1’7341	1’1887	1’0672	0’8620	0’6884	0’5338	0’3915	0’2571	0’1274
19	3’8833	2’8609	2’4334	2’0930	1’7291	1’1866	1’0655	0’8610	0’6876	0’5333	0’3912	0’2569	0’1274
20	3’8496	2’8453	2’4231	2’0860	1’7247	1’1848	1’0640	0’8600	0’6870	0’5329	0’3909	0’2567	0’1273
21	3’8193	2’8314	2’4138	2’0796	1’7207	1’1831	1’0627	0’8591	0’6864	0’5325	0’3906	0’2566	0’1272
22	3’7922	2’8188	2’4055	2’0739	1’7171	1’1815	1’0614	0’8583	0’6858	0’5321	0’3904	0’2564	0’1271
23	3’7676	2’8073	2’3979	2’0687	1’7139	1’1802	1’0603	0’8575	0’6853	0’5317	0’3902	0’2563	0’1271
24	3’7454	2’7970	2’3910	2’0639	1’7109	1’1789	1’0593	0’8569	0’6848	0’5314	0’3900	0’2562	0’1270
25	3’7251	2’7874	2’3846	2’0595	1’7081	1’1777	1’0584	0’8562	0’6844	0’5312	0’3898	0’2561	0’1269
26	3’7067	2’7787	2’3788	2’0555	1’7056	1’1766	1’0575	0’8557	0’6840	0’5309	0’3896	0’2560	0’1269
27	3’6895	2’7707	2’3734	2’0518	1’7033	1’1756	1’0567	0’8551	0’6837	0’5306	0’3894	0’2559	0’1268
28	3’6739	2’7633	2’3685	2’0484	1’7011	1’1747	1’0560	0’8546	0’6834	0’5304	0’3893	0’2558	0’1268
29	3’6595	2’7564	2’3638	2’0452	1’6991	1’1739	1’0553	0’8542	0’6830	0’5302	0’3892	0’2557	0’1268
30	3’6460	2’7500	2’3596	2’0423	1’6973	1’1731	1’0547	0’8538	0’6828	0’5300	0’3890	0’2556	0’1267
35	3’5911	2’7238	2’3420	2’0301	1’6896	1’1698	1’0520	0’8520	0’6816	0’5292	0’3885	0’2553	0’1266
40	3’5510	2’7045	2’3289	2’0211	1’6839	1’1673	1’0500	0’8507	0’6807	0’5286	0’3881	0’2550	0’1265
50	3’4960	2’6778	2’3109	2’0086	1’6759	1’1639	1’0473	0’8489	0’6794	0’5278	0’3875	0’2547	0’1263
60	3’4602	2’6603	2’2990	2’0003	1’6706	1’1616	1’0455	0’8477	0’6786	0’5272	0’3872	0’2545	0’1262
80	3’4164	2’6387	2’2844	1’9901	1’6641	1’1588	1’0432	0’8461	0’6776	0’5265	0’3867	0’2542	0’1261
100	3’3905	2’6259	2’2757	1’9840	1’6602	1’1571	1’0418	0’8452	0’6770	0’5261	0’3864	0’2540	0’1260
120	3’3734	2’6174	2’2699	1’9799	1’6576	1’1559	1’0409	0’8446	0’6765	0’5258	0’3862	0’2539	0’1259

Tabla B.5: Puntos Críticos: Distribución χ^2

	0'9995	0'995	0'9875	0'975	0'95	0'875	0'85	0'8	0'75	0'7	0'65	0'6	0'55
1	12'115	7'8794	6'2385	5'0239	3'8415	2'3535	2'0722	1'6424	1'3233	1'0742	0'8735	0'7083	0'5707
2	15'201	10'597	8'7641	7'3778	5'9915	4'1589	3'7942	3'2189	2'7726	2'4079	2'0996	1'8326	1'5970
3	17'731	12'838	10'861	9'3484	7'8147	5'7394	5'3170	4'6416	4'1083	3'6649	3'2831	2'9462	2'6430
4	19'998	14'860	12'762	11'143	9'4877	7'2140	6'7449	5'9886	5'3853	4'8784	4'4377	4'0446	3'6871
5	22'106	16'750	14'544	12'832	11'070	8'6248	8'1152	7'2893	6'6257	6'0644	5'5731	5'1319	4'7278
6	24'102	18'548	16'244	14'449	12'592	9'9917	9'4461	8'5581	7'8408	7'2311	6'6948	6'2108	5'7652
7	26'018	20'278	17'885	16'013	14'067	11'326	10'748	9'8032	9'0371	8'3834	7'8061	7'2832	6'8000
8	27'867	21'955	19'478	17'535	15'507	12'636	12'027	11'030	10'219	9'5245	8'9094	8'3505	7'8325
9	29'667	23'589	21'034	19'023	16'919	13'926	13'288	12'242	11'389	10'656	10'006	9'4136	8'8632
10	31'419	25'188	22'558	20'483	18'307	15'198	14'534	13'442	12'549	11'781	11'097	10'473	9'8922
11	33'138	26'757	24'056	21'920	19'675	16'457	15'767	14'631	13'701	12'899	12'184	11'530	10'920
12	34'821	28'300	25'530	23'337	21'026	17'703	16'989	15'812	14'845	14'011	13'266	12'584	11'946
13	36'477	29'819	26'985	24'736	22'362	18'939	18'202	16'985	15'984	15'119	14'345	13'636	12'972
14	38'109	31'319	28'422	26'119	23'685	20'166	19'406	18'151	17'117	16'222	15'421	14'685	13'996
15	39'717	32'801	29'843	27'488	24'996	21'384	20'603	19'311	18'245	17'322	16'494	15'733	15'020
16	41'308	34'267	31'250	28'845	26'296	22'595	21'793	20'465	19'369	18'418	17'565	16'780	16'042
17	42'881	35'718	32'644	30'191	27'587	23'799	22'977	21'615	20'489	19'511	18'633	17'824	17'065
18	44'434	37'156	34'027	31'526	28'869	24'997	24'155	22'760	21'605	20'601	19'699	18'868	18'086
19	45'974	38'582	35'399	32'852	30'144	26'189	25'329	23'900	22'718	21'689	20'764	19'910	19'107
20	47'498	39'997	36'760	34'170	31'410	27'376	26'498	25'038	23'828	22'775	21'826	20'951	20'127
21	49'010	41'401	38'113	35'479	32'671	28'559	27'662	26'171	24'935	23'858	22'888	21'992	21'147
22	50'510	42'796	39'458	36'781	33'924	29'737	28'822	27'301	26'039	24'939	23'947	23'031	22'166
23	51'999	44'181	40'794	38'076	35'172	30'911	29'979	28'429	27'141	26'018	25'006	24'069	23'185
24	53'478	45'558	42'124	39'364	36'415	32'081	31'132	29'553	28'241	27'096	26'063	25'106	24'204
25	54'948	46'928	43'446	40'646	37'652	33'247	32'282	30'675	29'339	28'172	27'118	26'143	25'222
26	56'407	48'290	44'762	41'923	38'885	34'410	33'429	31'795	30'435	29'246	28'173	27'179	26'240
27	57'856	49'645	46'071	43'195	40'113	35'570	34'574	32'912	31'528	30'319	29'227	28'214	27'257
28	59'299	50'994	47'375	44'461	41'337	36'727	35'715	34'027	32'620	31'391	30'279	29'249	28'274
29	60'734	52'335	48'674	45'722	42'557	37'881	36'854	35'139	33'711	32'461	31'331	30'283	29'291
30	62'160	53'672	49'967	46'979	43'773	39'033	37'990	36'250	34'800	33'530	32'382	31'316	30'307
35	69'197	60'275	56'365	53'203	49'802	44'753	43'640	41'778	40'223	38'859	37'623	36'475	35'386
40	76'096	66'766	62'665	59'342	55'758	50'424	49'244	47'269	45'616	44'165	42'848	41'622	40'459
50	89'560	79'490	75'039	71'420	67'505	61'647	60'346	58'164	56'334	54'723	53'258	51'892	50'592
60	102'70	91'952	87'184	83'298	79'082	72'751	71'341	68'972	66'981	65'226	63'628	62'135	60'713
80	128'26	116'32	110'99	106'63	101'88	94'709	93'106	90'405	88'130	86'120	84'284	82'566	80'927
100	153'16	140'17	134'34	129'56	124'34	116'43	114'66	111'67	109'14	106'91	104'86	102'95	101'11
120	177'60	163'65	157'37	152'21	146'57	137'99	136'06	132'81	130'05	127'62	125'38	123'29	121'28

Tabla B.6: Puntos Críticos: Distribución χ^2

	0,5	0,45	0,4	0,35	0,3	0,25	0,2	0,15	0,125	0,1	0,05	0,025	0,01	0,005
1	0,4549	0,3573	0,2750	0,2059	0,1485	0,1015	0,0642	0,0358	0,0247	0,0158	0,0039	0,0010	0,0002	0,0000
2	1,3863	1,1957	1,0217	0,8616	0,7133	0,5754	0,4463	0,3250	0,2671	0,2107	0,1026	0,0506	0,0201	0,0100
3	2,3660	2,1095	1,8692	1,6416	1,4237	1,2125	1,0052	0,7978	0,6924	0,5844	0,3518	0,2158	0,1148	0,0717
4	3,3567	3,0469	2,7528	2,4701	2,1947	1,9226	1,6488	1,3665	1,2188	1,0636	0,7107	0,4844	0,2971	0,2070
5	4,3515	3,9959	3,6555	3,3251	2,9999	2,6746	2,3425	1,9938	1,8082	1,6103	1,1455	0,8312	0,5543	0,4118
6	5,3481	4,9519	4,5702	4,1973	3,8276	3,4546	3,0701	2,6613	2,4411	2,2041	1,6354	1,2373	0,8721	0,6757
7	6,3458	5,9125	5,4932	5,0816	4,6713	4,2549	3,8223	3,3583	3,1063	2,8331	2,1673	1,6899	1,2390	0,9893
8	7,3441	6,8766	6,4226	5,9753	5,5274	5,0706	4,5936	4,0782	3,7965	3,4895	2,7326	2,1797	1,6465	1,3444
9	8,3428	7,8434	7,3570	6,8763	6,3933	5,8988	5,3801	4,8165	4,5070	4,1682	3,3251	2,7004	2,0879	1,7349
10	9,3418	8,8124	8,2955	7,7832	7,2672	6,7372	6,1791	5,5701	5,2341	4,8652	3,9403	3,2470	2,5582	2,1558
11	10,341	9,7831	9,2373	8,6952	8,1479	7,5841	6,9887	6,3364	5,9754	5,5778	4,5748	3,8157	3,0535	2,6032
12	11,340	10,755	10,182	9,6115	9,0343	8,4384	7,8073	7,1138	6,7288	6,3038	5,2260	4,4038	3,5706	3,0738
13	12,340	11,729	11,129	10,532	9,9257	9,2991	8,6339	7,9008	7,4929	7,0415	5,8919	5,0087	4,1069	3,5650
14	13,339	12,703	12,078	11,455	10,821	10,165	9,4673	8,6963	8,2662	7,7895	6,5706	5,6287	4,6604	4,0747
15	14,339	13,679	13,030	12,381	11,721	11,037	10,307	9,4993	9,0479	8,5468	7,2609	6,2621	5,2294	4,6009
16	15,338	14,656	13,983	13,310	12,624	11,912	11,152	10,309	9,8370	9,3122	7,9616	6,9077	5,8122	5,1422
17	16,338	15,633	14,937	14,241	13,531	12,792	12,002	11,125	10,633	10,085	8,6718	7,5642	6,4077	5,6973
18	17,338	16,611	15,893	15,174	14,440	13,675	12,857	11,946	11,435	10,865	9,3904	8,2307	7,0149	6,2648
19	18,338	17,589	16,850	16,109	15,352	14,562	13,716	12,773	12,242	11,651	10,117	8,9065	7,6327	6,8439
20	19,337	18,569	17,809	17,046	16,266	15,452	14,578	13,604	13,055	12,443	10,851	9,5908	8,2604	7,4338
21	20,337	19,548	18,768	17,984	17,182	16,344	15,445	14,439	13,873	13,240	11,591	10,283	8,8972	8,0336
22	21,337	20,529	19,729	18,924	18,101	17,240	16,314	15,279	14,695	14,041	12,338	10,982	9,5425	8,6427
23	22,337	21,510	20,690	19,866	19,021	18,137	17,187	16,122	15,521	14,848	13,091	11,689	10,196	9,2604
24	23,337	22,491	21,652	20,808	19,943	19,037	18,062	16,969	16,351	15,659	13,848	12,401	10,856	9,8862
25	24,337	23,472	22,616	21,752	20,867	19,939	18,940	17,818	17,184	16,473	14,611	13,120	11,524	10,520
26	25,336	24,454	23,579	22,697	21,792	20,843	19,820	18,671	18,021	17,292	15,379	13,844	12,198	11,160
27	26,336	25,437	24,544	23,644	22,719	21,749	20,703	19,527	18,861	18,114	16,151	14,573	12,878	11,808
28	27,336	26,419	25,509	24,591	23,647	22,657	21,588	20,386	19,704	18,939	16,928	15,308	13,565	12,461
29	28,336	27,402	26,475	25,539	24,577	23,567	22,475	21,247	20,550	19,768	17,708	16,047	14,256	13,121
30	29,336	28,386	27,442	26,488	25,508	24,478	23,364	22,110	21,399	20,599	18,493	16,791	14,953	13,787
35	34,336	33,306	32,282	31,246	30,178	29,054	27,836	26,460	25,678	24,797	22,465	20,569	18,509	17,192
40	39,335	38,233	37,134	36,021	34,872	33,660	32,345	30,856	30,008	29,051	26,509	24,433	22,164	20,707
50	49,335	48,099	46,864	45,610	44,313	42,942	41,449	39,754	38,785	37,689	34,764	32,357	29,707	27,991
60	59,335	57,978	56,620	55,239	53,809	52,294	50,641	48,759	47,680	46,459	43,188	40,482	37,485	35,534
80	79,334	77,763	76,188	74,583	72,915	71,145	69,207	66,994	65,722	64,278	60,391	57,153	53,540	51,172
100	99,334	97,574	95,808	94,005	92,129	90,133	87,945	85,441	83,999	82,358	77,929	74,222	70,065	67,328
120	119,33	117,40	115,46	113,48	111,42	109,22	106,81	104,04	102,44	100,62	95,705	91,573	86,923	83,852

Tabla B.7: Distribución $\mathcal{F}$ de Snedecor ($p = 0'5$)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1’000	1’500	1’709	1’823	1’894	1’942	1’977	2’004	2’025	2’042	2’056	2’067	2’077	2’086	2’093	2’100
2	0’667	1’000	1’135	1’207	1’252	1’282	1’305	1’321	1’334	1’345	1’354	1’361	1’367	1’372	1’377	1’381
3	0’585	0’881	1’000	1’063	1’102	1’129	1’148	1’163	1’174	1’183	1’191	1’197	1’203	1’207	1’211	1’215
4	0’549	0’828	0’941	1’000	1’037	1’062	1’080	1’093	1’104	1’113	1’120	1’126	1’131	1’135	1’139	1’142
5	0’528	0’799	0’907	0’965	1’000	1’024	1’041	1’055	1’065	1’073	1’080	1’085	1’090	1’094	1’098	1’101
6	0’515	0’780	0’886	0’942	0’977	1’000	1’017	1’030	1’040	1’048	1’054	1’060	1’065	1’069	1’072	1’075
7	0’506	0’767	0’871	0’926	0’960	0’983	1’000	1’013	1’022	1’030	1’037	1’042	1’047	1’051	1’054	1’057
8	0’499	0’757	0’860	0’915	0’948	0’971	0’988	1’000	1’010	1’018	1’024	1’029	1’034	1’038	1’041	1’044
9	0’494	0’749	0’852	0’906	0’939	0’962	0’978	0’990	1’000	1’008	1’014	1’019	1’024	1’028	1’031	1’034
10	0’490	0’743	0’845	0’899	0’932	0’954	0’971	0’983	0’992	1’000	1’006	1’012	1’016	1’020	1’023	1’026
11	0’486	0’739	0’840	0’893	0’926	0’948	0’964	0’977	0’986	0’994	1’000	1’005	1’010	1’013	1’017	1’020
12	0’484	0’735	0’835	0’888	0’921	0’943	0’959	0’972	0’981	0’989	0’995	1’000	1’004	1’008	1’012	1’014
13	0’481	0’731	0’832	0’885	0’917	0’939	0’955	0’967	0’977	0’984	0’990	0’996	1’000	1’004	1’007	1’010
14	0’479	0’729	0’828	0’881	0’914	0’936	0’952	0’964	0’973	0’981	0’987	0’992	0’996	1’000	1’003	1’006
15	0’478	0’726	0’826	0’878	0’911	0’933	0’949	0’960	0’970	0’977	0’983	0’989	0’993	0’997	1’000	1’003
16	0’476	0’724	0’823	0’876	0’908	0’930	0’946	0’958	0’967	0’975	0’981	0’986	0’990	0’994	0’997	1’000
17	0’475	0’722	0’821	0’874	0’906	0’928	0’943	0’955	0’965	0’972	0’978	0’983	0’988	0’991	0’995	0’997
18	0’474	0’721	0’819	0’872	0’904	0’926	0’941	0’953	0’962	0’970	0’976	0’981	0’985	0’989	0’992	0’995
19	0’473	0’719	0’818	0’870	0’902	0’924	0’939	0’951	0’961	0’968	0’974	0’979	0’984	0’987	0’990	0’993
20	0’472	0’718	0’816	0’868	0’900	0’922	0’938	0’950	0’959	0’966	0’972	0’977	0’982	0’985	0’989	0’992
21	0’471	0’717	0’815	0’867	0’899	0’921	0’936	0’948	0’957	0’965	0’971	0’976	0’980	0’984	0’987	0’990
22	0’470	0’715	0’814	0’866	0’898	0’919	0’935	0’947	0’956	0’963	0’969	0’974	0’979	0’982	0’986	0’988
23	0’470	0’714	0’813	0’864	0’896	0’918	0’934	0’945	0’955	0’962	0’968	0’973	0’977	0’981	0’984	0’987
24	0’469	0’714	0’812	0’863	0’895	0’917	0’932	0’944	0’953	0’961	0’967	0’972	0’976	0’980	0’983	0’986
25	0’468	0’713	0’811	0’862	0’894	0’916	0’931	0’943	0’952	0’960	0’966	0’971	0’975	0’979	0’982	0’985
26	0’468	0’712	0’810	0’861	0’893	0’915	0’930	0’942	0’951	0’959	0’965	0’970	0’974	0’978	0’981	0’984
27	0’467	0’711	0’809	0’861	0’892	0’914	0’930	0’941	0’950	0’958	0’964	0’969	0’973	0’977	0’980	0’983
28	0’467	0’711	0’808	0’860	0’892	0’913	0’929	0’940	0’950	0’957	0’963	0’968	0’972	0’976	0’979	0’982
29	0’467	0’710	0’808	0’859	0’891	0’912	0’928	0’940	0’949	0’956	0’962	0’967	0’971	0’975	0’978	0’981
30	0’466	0’709	0’807	0’858	0’890	0’912	0’927	0’939	0’948	0’955	0’961	0’966	0’971	0’974	0’978	0’980
35	0’465	0’707	0’804	0’856	0’887	0’909	0’924	0’936	0’945	0’952	0’958	0’963	0’968	0’971	0’974	0’977
40	0’463	0’705	0’802	0’854	0’885	0’907	0’922	0’934	0’943	0’950	0’956	0’961	0’965	0’969	0’972	0’975
50	0’462	0’703	0’800	0’851	0’882	0’903	0’919	0’930	0’940	0’947	0’953	0’958	0’962	0’966	0’969	0’972
60	0’460	0’701	0’798	0’849	0’880	0’901	0’917	0’928	0’937	0’945	0’951	0’956	0’960	0’964	0’967	0’969
70	0’460	0’700	0’796	0’847	0’879	0’900	0’915	0’927	0’936	0’943	0’949	0’954	0’958	0’962	0’965	0’968
80	0’459	0’699	0’795	0’846	0’878	0’899	0’914	0’926	0’935	0’942	0’948	0’953	0’957	0’961	0’964	0’967
90	0’459	0’699	0’795	0’846	0’877	0’898	0’913	0’925	0’934	0’941	0’947	0’952	0’956	0’960	0’963	0’966
100	0’458	0’698	0’794	0’845	0’876	0’897	0’913	0’924	0’933	0’940	0’946	0’951	0’956	0’959	0’962	0’965
120	0’458	0’697	0’793	0’844	0’875	0’896	0’912	0’923	0’932	0’939	0’945	0’950	0’955	0’958	0’961	0’964
∞	0’455	0’693	0’789	0’839	0’870	0’891	0’907	0’918	0’927	0’934	0’940	0’945	0’949	0’953	0’956	0’959

Tabla B.8: Distribución $\mathcal{F}$ de Snedecor ($p = 0$’5)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	2’105	2’110	2’115	2’119	2’135	2’145	2’153	2’158	2’163	2’166	2’172	2’175	2’178	2’180	2’185	2’198
2	1’385	1’388	1’391	1’393	1’403	1’410	1’414	1’418	1’421	1’423	1’426	1’428	1’430	1’432	1’434	1’442
3	1’218	1’220	1’223	1’225	1’234	1’239	1’243	1’246	1’249	1’251	1’254	1’256	1’257	1’258	1’261	1’268
4	1’145	1’147	1’150	1’152	1’160	1’165	1’169	1’172	1’174	1’176	1’178	1’180	1’182	1’183	1’185	1’191
5	1’104	1’106	1’109	1’111	1’118	1’123	1’127	1’130	1’132	1’134	1’136	1’138	1’139	1’140	1’143	1’149
6	1’078	1’080	1’083	1’084	1’092	1’097	1’100	1’103	1’105	1’107	1’109	1’111	1’112	1’114	1’116	1’122
7	1’060	1’062	1’064	1’066	1’074	1’079	1’082	1’085	1’087	1’088	1’091	1’093	1’094	1’095	1’097	1’103
8	1’047	1’049	1’051	1’053	1’060	1’065	1’069	1’071	1’073	1’075	1’077	1’079	1’080	1’081	1’083	1’089
9	1’037	1’039	1’041	1’043	1’050	1’055	1’058	1’061	1’063	1’064	1’067	1’068	1’070	1’071	1’073	1’079
10	1’029	1’031	1’033	1’035	1’042	1’047	1’050	1’053	1’055	1’056	1’059	1’060	1’062	1’062	1’064	1’070
11	1’022	1’025	1’027	1’028	1’035	1’040	1’043	1’046	1’048	1’050	1’052	1’054	1’055	1’056	1’058	1’064
12	1’017	1’019	1’021	1’023	1’030	1’035	1’038	1’041	1’042	1’044	1’046	1’048	1’049	1’050	1’052	1’058
13	1’012	1’015	1’017	1’019	1’026	1’030	1’033	1’036	1’038	1’039	1’042	1’043	1’045	1’046	1’048	1’053
14	1’009	1’011	1’013	1’015	1’022	1’026	1’030	1’032	1’034	1’036	1’038	1’040	1’041	1’042	1’044	1’049
15	1’005	1’008	1’010	1’011	1’018	1’023	1’026	1’029	1’031	1’032	1’034	1’036	1’037	1’038	1’040	1’046
16	1’003	1’005	1’007	1’009	1’015	1’020	1’023	1’026	1’028	1’029	1’032	1’033	1’034	1’035	1’037	1’043
17	1’000	1’002	1’004	1’006	1’013	1’017	1’021	1’023	1’025	1’027	1’029	1’031	1’032	1’033	1’035	1’040
18	0’998	1’000	1’002	1’004	1’011	1’015	1’018	1’021	1’023	1’024	1’027	1’028	1’030	1’030	1’032	1’038
19	0’996	0’998	1’000	1’002	1’009	1’013	1’016	1’019	1’021	1’022	1’025	1’026	1’027	1’028	1’030	1’036
20	0’994	0’996	0’998	1’000	1’007	1’011	1’015	1’017	1’019	1’020	1’023	1’024	1’026	1’027	1’029	1’034
21	0’992	0’995	0’997	0’998	1’005	1’010	1’013	1’015	1’017	1’019	1’021	1’023	1’024	1’025	1’027	1’032
22	0’991	0’993	0’995	0’997	1’004	1’008	1’011	1’014	1’016	1’017	1’020	1’021	1’022	1’023	1’025	1’031
23	0’990	0’992	0’994	0’996	1’002	1’007	1’010	1’013	1’014	1’016	1’018	1’020	1’021	1’022	1’024	1’030
24	0’988	0’991	0’993	0’994	1’001	1’006	1’009	1’011	1’013	1’015	1’017	1’019	1’020	1’021	1’023	1’028
25	0’987	0’989	0’991	0’993	1’000	1’005	1’008	1’010	1’012	1’014	1’016	1’017	1’019	1’020	1’022	1’027
26	0’986	0’988	0’990	0’992	0’999	1’003	1’007	1’009	1’011	1’013	1’015	1’016	1’018	1’019	1’020	1’026
27	0’985	0’988	0’989	0’991	0’998	1’003	1’006	1’008	1’010	1’012	1’014	1’015	1’017	1’018	1’020	1’025
28	0’984	0’987	0’989	0’990	0’997	1’002	1’005	1’007	1’009	1’011	1’013	1’015	1’016	1’017	1’019	1’024
29	0’984	0’986	0’988	0’990	0’996	1’001	1’004	1’006	1’008	1’010	1’012	1’014	1’015	1’016	1’018	1’023
30	0’983	0’985	0’987	0’989	0’996	1’000	1’003	1’006	1’008	1’009	1’011	1’013	1’014	1’015	1’017	1’022
35	0’980	0’982	0’984	0’986	0’992	0’997	1’000	1’002	1’004	1’006	1’008	1’010	1’011	1’012	1’014	1’019
40	0’977	0’980	0’981	0’983	0’990	0’994	0’998	1’000	1’002	1’003	1’006	1’007	1’008	1’009	1’011	1’017
50	0’974	0’976	0’978	0’980	0’987	0’991	0’994	0’997	0’999	1’000	1’002	1’004	1’005	1’006	1’008	1’013
60	0’972	0’974	0’976	0’978	0’984	0’989	0’992	0’994	0’996	0’998	1’000	1’002	1’003	1’004	1’006	1’011
70	0’970	0’972	0’974	0’976	0’983	0’987	0’990	0’993	0’995	0’996	0’998	1’000	1’001	1’002	1’004	1’009
80	0’969	0’971	0’973	0’975	0’982	0’986	0’989	0’992	0’993	0’995	0’997	0’999	1’000	1’001	1’003	1’008
90	0’968	0’970	0’972	0’974	0’981	0’985	0’988	0’991	0’993	0’994	0’996	0’998	0’999	1’000	1’002	1’007
100	0’968	0’970	0’972	0’973	0’980	0’984	0’988	0’990	0’992	0’993	0’996	0’997	0’998	0’999	1’001	1’007
120	0’966	0’969	0’971	0’972	0’979	0’983	0’986	0’989	0’991	0’992	0’994	0’996	0’997	0’998	1’000	1’005
∞	0’961	0’963	0’965	0’967	0’974	0’978	0’981	0’984	0’985	0’987	0’989	0’991	0’992	0’993	0’995	1’000

Tabla B.9: Distribución $\mathcal{F}$ de Snedecor ($p = 0$'75)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	5'828	7'500	8'200	8'581	8'820	8'983	9'102	9'192	9'263	9'320	9'367	9'406	9'440	9'468	9'493	9'515
2	2'571	3'000	3'153	3'232	3'280	3'312	3'335	3'353	3'366	3'377	3'386	3'393	3'400	3'405	3'410	3'414
3	2'024	2'280	2'356	2'390	2'409	2'422	2'430	2'436	2'441	2'445	2'448	2'450	2'452	2'454	2'455	2'456
4	1'807	2'000	2'047	2'064	2'072	2'077	2'079	2'080	2'081	2'082	2'082	2'083	2'083	2'083	2'083	2'083
5	1'692	1'853	1'884	1'893	1'895	1'894	1'894	1'892	1'891	1'890	1'889	1'888	1'887	1'886	1'885	1'884
6	1'621	1'762	1'784	1'787	1'785	1'782	1'779	1'776	1'773	1'771	1'769	1'767	1'765	1'764	1'762	1'761
7	1'573	1'701	1'717	1'716	1'711	1'706	1'701	1'697	1'693	1'690	1'687	1'684	1'682	1'680	1'678	1'676
8	1'538	1'657	1'668	1'664	1'658	1'651	1'645	1'640	1'635	1'631	1'627	1'624	1'622	1'619	1'617	1'615
9	1'512	1'624	1'632	1'625	1'617	1'609	1'602	1'596	1'591	1'586	1'582	1'579	1'576	1'573	1'570	1'568
10	1'491	1'598	1'603	1'595	1'585	1'576	1'569	1'562	1'556	1'551	1'547	1'543	1'540	1'537	1'534	1'531
11	1'475	1'577	1'580	1'570	1'560	1'550	1'542	1'535	1'528	1'523	1'518	1'514	1'510	1'507	1'504	1'501
12	1'461	1'560	1'561	1'550	1'539	1'529	1'520	1'512	1'505	1'500	1'495	1'490	1'486	1'483	1'480	1'477
13	1'450	1'545	1'545	1'534	1'521	1'511	1'501	1'493	1'486	1'480	1'475	1'470	1'466	1'462	1'459	1'456
14	1'440	1'533	1'532	1'519	1'507	1'495	1'485	1'477	1'470	1'463	1'458	1'453	1'449	1'445	1'441	1'438
15	1'432	1'523	1'520	1'507	1'494	1'482	1'472	1'463	1'456	1'449	1'443	1'438	1'434	1'430	1'426	1'423
16	1'425	1'514	1'510	1'497	1'483	1'471	1'460	1'451	1'443	1'437	1'431	1'426	1'421	1'417	1'413	1'410
17	1'419	1'506	1'502	1'487	1'473	1'460	1'450	1'441	1'433	1'426	1'420	1'414	1'409	1'405	1'401	1'398
18	1'413	1'499	1'494	1'479	1'464	1'452	1'441	1'431	1'423	1'416	1'410	1'404	1'399	1'395	1'391	1'388
19	1'408	1'493	1'487	1'472	1'457	1'444	1'432	1'423	1'414	1'407	1'401	1'395	1'390	1'386	1'382	1'378
20	1'404	1'487	1'481	1'465	1'450	1'437	1'425	1'415	1'407	1'399	1'393	1'387	1'382	1'378	1'374	1'370
21	1'400	1'482	1'475	1'459	1'444	1'430	1'419	1'409	1'400	1'392	1'386	1'380	1'375	1'370	1'366	1'362
22	1'396	1'477	1'470	1'454	1'438	1'424	1'413	1'402	1'394	1'386	1'379	1'374	1'368	1'364	1'359	1'355
23	1'393	1'473	1'466	1'449	1'433	1'419	1'407	1'397	1'388	1'380	1'374	1'368	1'362	1'357	1'353	1'349
24	1'390	1'470	1'462	1'445	1'428	1'414	1'402	1'392	1'383	1'375	1'368	1'362	1'357	1'352	1'347	1'343
25	1'387	1'466	1'458	1'441	1'424	1'410	1'398	1'387	1'378	1'370	1'363	1'357	1'352	1'347	1'342	1'338
26	1'384	1'463	1'454	1'437	1'420	1'406	1'393	1'383	1'374	1'366	1'359	1'352	1'347	1'342	1'337	1'333
27	1'382	1'460	1'451	1'433	1'417	1'402	1'390	1'379	1'370	1'361	1'354	1'348	1'342	1'337	1'333	1'329
28	1'380	1'457	1'448	1'430	1'413	1'399	1'386	1'375	1'366	1'358	1'350	1'344	1'338	1'333	1'329	1'325
29	1'378	1'455	1'445	1'427	1'410	1'395	1'383	1'372	1'362	1'354	1'347	1'340	1'335	1'330	1'325	1'321
30	1'376	1'452	1'443	1'424	1'407	1'392	1'380	1'369	1'359	1'351	1'343	1'337	1'331	1'326	1'321	1'317
35	1'368	1'443	1'432	1'413	1'395	1'380	1'367	1'355	1'345	1'337	1'329	1'323	1'317	1'311	1'306	1'302
40	1'363	1'435	1'424	1'404	1'386	1'371	1'357	1'345	1'335	1'327	1'319	1'312	1'306	1'300	1'295	1'291
50	1'355	1'425	1'413	1'393	1'374	1'358	1'344	1'332	1'321	1'312	1'304	1'297	1'291	1'285	1'280	1'275
60	1'349	1'419	1'405	1'385	1'366	1'349	1'335	1'323	1'312	1'303	1'294	1'287	1'280	1'274	1'269	1'264
70	1'346	1'414	1'400	1'379	1'360	1'343	1'329	1'316	1'305	1'296	1'287	1'280	1'273	1'267	1'262	1'257
80	1'343	1'411	1'396	1'375	1'355	1'338	1'324	1'311	1'300	1'291	1'282	1'275	1'268	1'262	1'256	1'251
90	1'341	1'408	1'393	1'372	1'352	1'335	1'320	1'307	1'296	1'287	1'278	1'270	1'263	1'257	1'252	1'246
100	1'339	1'406	1'391	1'369	1'349	1'332	1'317	1'304	1'293	1'283	1'275	1'267	1'260	1'254	1'248	1'243
120	1'336	1'402	1'387	1'365	1'345	1'328	1'313	1'300	1'289	1'279	1'270	1'262	1'255	1'249	1'243	1'237
∞	1'324	1'387	1'370	1'347	1'326	1'307	1'292	1'278	1'266	1'255	1'246	1'238	1'230	1'223	1'217	1'211

Tabla B.10: Distribución $\mathcal{F}$ de Snedecor ($p = 0$'75)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	9'535	9'552	9'567	9'581	9'634	9'670	9'695	9'714	9'729	9'741	9'759	9'772	9'782	9'789	9'804	9'848
2	3'418	3'421	3'424	3'426	3'436	3'443	3'448	3'451	3'454	3'456	3'459	3'462	3'464	3'465	3'468	3'476
3	2'458	2'459	2'459	2'460	2'463	2'465	2'466	2'467	2'468	2'469	2'470	2'470	2'471	2'471	2'472	2'474
4	2'083	2'083	2'083	2'083	2'083	2'082	2'082	2'082	2'082	2'082	2'082	2'082	2'081	2'081	2'081	2'081
5	1'884	1'883	1'882	1'882	1'880	1'878	1'877	1'876	1'876	1'875	1'874	1'874	1'873	1'873	1'872	1'869
6	1'760	1'759	1'758	1'757	1'753	1'751	1'749	1'748	1'747	1'746	1'744	1'743	1'742	1'742	1'741	1'737
7	1'675	1'674	1'672	1'671	1'667	1'663	1'661	1'659	1'658	1'657	1'655	1'654	1'653	1'652	1'650	1'645
8	1'613	1'612	1'610	1'609	1'603	1'600	1'597	1'595	1'593	1'591	1'589	1'588	1'586	1'586	1'584	1'578
9	1'566	1'564	1'563	1'561	1'555	1'551	1'547	1'545	1'543	1'541	1'539	1'537	1'536	1'535	1'533	1'526
10	1'529	1'527	1'525	1'523	1'517	1'512	1'508	1'506	1'503	1'502	1'499	1'497	1'495	1'494	1'492	1'484
11	1'499	1'497	1'495	1'493	1'486	1'481	1'477	1'474	1'471	1'469	1'466	1'464	1'463	1'461	1'459	1'451
12	1'474	1'472	1'470	1'468	1'460	1'454	1'450	1'447	1'445	1'443	1'439	1'437	1'435	1'434	1'431	1'422
13	1'453	1'451	1'449	1'447	1'438	1'432	1'428	1'425	1'422	1'420	1'416	1'414	1'412	1'411	1'408	1'398
14	1'435	1'433	1'431	1'428	1'420	1'414	1'409	1'405	1'403	1'400	1'397	1'394	1'392	1'391	1'387	1'377
15	1'420	1'417	1'415	1'413	1'404	1'397	1'392	1'389	1'386	1'383	1'380	1'377	1'375	1'373	1'370	1'359
16	1'407	1'404	1'401	1'399	1'390	1'383	1'378	1'374	1'371	1'369	1'365	1'362	1'360	1'358	1'354	1'343
17	1'395	1'392	1'389	1'387	1'377	1'370	1'365	1'361	1'358	1'355	1'351	1'348	1'346	1'344	1'341	1'329
18	1'384	1'381	1'379	1'376	1'366	1'359	1'354	1'350	1'346	1'344	1'340	1'336	1'334	1'332	1'328	1'317
19	1'375	1'372	1'369	1'367	1'356	1'349	1'344	1'339	1'336	1'333	1'329	1'326	1'323	1'321	1'317	1'305
20	1'367	1'363	1'361	1'358	1'348	1'340	1'335	1'330	1'327	1'324	1'319	1'316	1'313	1'311	1'307	1'295
21	1'359	1'356	1'353	1'350	1'340	1'332	1'326	1'322	1'318	1'315	1'311	1'307	1'305	1'303	1'298	1'285
22	1'352	1'349	1'346	1'343	1'332	1'324	1'319	1'314	1'310	1'307	1'303	1'299	1'296	1'294	1'290	1'276
23	1'346	1'342	1'339	1'337	1'326	1'318	1'312	1'307	1'303	1'300	1'295	1'292	1'289	1'287	1'282	1'268
24	1'340	1'337	1'333	1'331	1'319	1'311	1'305	1'300	1'297	1'293	1'289	1'285	1'282	1'280	1'275	1'261
25	1'335	1'331	1'328	1'325	1'314	1'306	1'299	1'294	1'291	1'287	1'282	1'279	1'276	1'273	1'269	1'254
26	1'330	1'326	1'323	1'320	1'309	1'300	1'294	1'289	1'285	1'282	1'277	1'273	1'270	1'268	1'263	1'248
27	1'325	1'322	1'318	1'315	1'304	1'295	1'289	1'284	1'280	1'276	1'271	1'267	1'264	1'262	1'257	1'242
28	1'321	1'317	1'314	1'311	1'299	1'291	1'284	1'279	1'275	1'271	1'266	1'262	1'259	1'257	1'252	1'236
29	1'317	1'313	1'310	1'307	1'295	1'286	1'280	1'275	1'270	1'267	1'262	1'258	1'254	1'252	1'247	1'231
30	1'313	1'310	1'306	1'303	1'291	1'282	1'276	1'270	1'266	1'263	1'257	1'253	1'250	1'247	1'242	1'226
35	1'298	1'294	1'291	1'288	1'275	1'266	1'258	1'253	1'248	1'245	1'239	1'234	1'231	1'228	1'223	1'205
40	1'286	1'283	1'279	1'276	1'263	1'253	1'245	1'240	1'235	1'231	1'225	1'220	1'217	1'214	1'208	1'189
50	1'270	1'266	1'263	1'259	1'245	1'235	1'227	1'221	1'216	1'212	1'205	1'200	1'196	1'193	1'186	1'165
60	1'260	1'255	1'252	1'248	1'234	1'223	1'215	1'208	1'203	1'198	1'191	1'186	1'182	1'178	1'172	1'148
70	1'252	1'248	1'244	1'240	1'225	1'214	1'206	1'199	1'193	1'189	1'181	1'176	1'171	1'168	1'161	1'135
80	1'246	1'242	1'238	1'234	1'219	1'208	1'199	1'192	1'186	1'181	1'174	1'168	1'163	1'160	1'152	1'125
90	1'242	1'237	1'233	1'229	1'214	1'202	1'194	1'186	1'180	1'176	1'168	1'162	1'157	1'153	1'145	1'117
100	1'238	1'234	1'229	1'226	1'210	1'198	1'189	1'182	1'176	1'171	1'163	1'157	1'152	1'148	1'140	1'110
120	1'233	1'228	1'224	1'220	1'204	1'192	1'183	1'175	1'169	1'164	1'156	1'149	1'144	1'140	1'131	1'100
∞	1'206	1'201	1'196	1'192	1'174	1'161	1'150	1'141	1'134	1'128	1'117	1'109	1'103	1'097	1'085	1'019

Tabla B.11: Distribución $\mathcal{F}$ de Snedecor ($p = 0'9$)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	39’86	49’50	53’59	55’83	57’24	58’20	58’91	59’44	59’86	60’19	60’47	60’71	60’90	61’07	61’22	61’35
2	8’526	9’000	9’162	9’243	9’293	9’326	9’349	9’367	9’381	9’392	9’401	9’408	9’415	9’420	9’425	9’429
3	5’538	5’462	5’391	5’343	5’309	5’285	5’266	5’252	5’240	5’230	5’222	5’216	5’210	5’205	5’200	5’196
4	4’545	4’325	4’191	4’107	4’051	4’010	3’979	3’955	3’936	3’920	3’907	3’896	3’886	3’878	3’870	3’864
5	4’060	3’780	3’619	3’520	3’453	3’405	3’368	3’339	3’316	3’297	3’282	3’268	3’257	3’247	3’238	3’230
6	3’776	3’463	3’289	3’181	3’108	3’055	3’014	2’983	2’958	2’937	2’920	2’905	2’892	2’881	2’871	2’863
7	3’589	3’257	3’074	2’961	2’883	2’827	2’785	2’752	2’725	2’703	2’684	2’668	2’654	2’643	2’632	2’623
8	3’458	3’113	2’924	2’806	2’726	2’668	2’624	2’589	2’561	2’538	2’519	2’502	2’488	2’475	2’464	2’454
9	3’360	3’006	2’813	2’693	2’611	2’551	2’505	2’469	2’440	2’416	2’396	2’379	2’364	2’351	2’340	2’330
10	3’285	2’924	2’728	2’605	2’522	2’461	2’414	2’377	2’347	2’323	2’302	2’284	2’269	2’255	2’244	2’233
11	3’225	2’860	2’660	2’536	2’451	2’389	2’342	2’304	2’274	2’248	2’227	2’209	2’193	2’179	2’167	2’156
12	3’177	2’807	2’606	2’480	2’394	2’331	2’283	2’245	2’214	2’188	2’166	2’147	2’131	2’117	2’105	2’094
13	3’136	2’763	2’560	2’434	2’347	2’283	2’234	2’195	2’164	2’138	2’116	2’097	2’080	2’066	2’053	2’042
14	3’102	2’726	2’522	2’395	2’307	2’243	2’193	2’154	2’122	2’095	2’073	2’054	2’037	2’022	2’010	1’998
15	3’073	2’695	2’490	2’361	2’273	2’208	2’158	2’119	2’086	2’059	2’037	2’017	2’000	1’985	1’972	1’961
16	3’048	2’668	2’462	2’333	2’244	2’178	2’128	2’088	2’055	2’028	2’005	1’985	1’968	1’953	1’940	1’928
17	3’026	2’645	2’437	2’308	2’218	2’152	2’102	2’061	2’028	2’001	1’978	1’958	1’940	1’925	1’912	1’900
18	3’007	2’624	2’416	2’286	2’196	2’130	2’079	2’038	2’005	1’977	1’954	1’933	1’916	1’900	1’887	1’875
19	2’990	2’606	2’397	2’266	2’176	2’109	2’058	2’017	1’984	1’956	1’932	1’912	1’894	1’878	1’865	1’852
20	2’975	2’589	2’380	2’249	2’158	2’091	2’040	1’999	1’965	1’937	1’913	1’892	1’875	1’859	1’845	1’833
21	2’961	2’575	2’365	2’233	2’142	2’075	2’023	1’982	1’948	1’920	1’896	1’875	1’857	1’841	1’827	1’815
22	2’949	2’561	2’351	2’219	2’128	2’060	2’008	1’967	1’933	1’904	1’880	1’859	1’841	1’825	1’811	1’798
23	2’937	2’549	2’339	2’207	2’115	2’047	1’995	1’953	1’919	1’890	1’866	1’845	1’827	1’811	1’796	1’784
24	2’927	2’538	2’327	2’195	2’103	2’035	1’983	1’941	1’906	1’877	1’853	1’832	1’814	1’797	1’783	1’770
25	2’918	2’528	2’317	2’184	2’092	2’024	1’971	1’929	1’895	1’866	1’841	1’820	1’802	1’785	1’771	1’758
26	2’909	2’519	2’307	2’174	2’082	2’014	1’961	1’919	1’884	1’855	1’830	1’809	1’790	1’774	1’760	1’747
27	2’901	2’511	2’299	2’165	2’073	2’005	1’952	1’909	1’874	1’845	1’820	1’799	1’780	1’764	1’749	1’736
28	2’894	2’503	2’291	2’157	2’064	1’996	1’943	1’900	1’865	1’836	1’811	1’790	1’771	1’754	1’740	1’726
29	2’887	2’495	2’283	2’149	2’057	1’988	1’935	1’892	1’857	1’827	1’802	1’781	1’762	1’745	1’731	1’717
30	2’881	2’489	2’276	2’142	2’049	1’980	1’927	1’884	1’849	1’819	1’794	1’773	1’754	1’737	1’722	1’709
35	2’855	2’461	2’247	2’113	2’019	1’950	1’896	1’852	1’817	1’787	1’761	1’739	1’720	1’703	1’688	1’674
40	2’835	2’440	2’226	2’091	1’997	1’927	1’873	1’829	1’793	1’763	1’737	1’715	1’695	1’678	1’662	1’649
50	2’809	2’412	2’197	2’061	1’966	1’895	1’840	1’796	1’760	1’729	1’703	1’680	1’660	1’643	1’627	1’613
60	2’791	2’393	2’177	2’041	1’946	1’875	1’819	1’775	1’738	1’707	1’680	1’657	1’637	1’619	1’603	1’589
70	2’779	2’380	2’164	2’027	1’931	1’860	1’804	1’760	1’723	1’691	1’665	1’641	1’621	1’603	1’587	1’572
80	2’769	2’370	2’154	2’016	1’921	1’849	1’793	1’748	1’711	1’680	1’653	1’629	1’609	1’590	1’574	1’559
90	2’762	2’363	2’146	2’008	1’912	1’841	1’785	1’739	1’702	1’670	1’643	1’620	1’599	1’581	1’564	1’550
100	2’756	2’356	2’139	2’002	1’906	1’834	1’778	1’732	1’695	1’663	1’636	1’612	1’592	1’573	1’557	1’542
120	2’748	2’347	2’130	1’992	1’896	1’824	1’767	1’722	1’684	1’652	1’625	1’601	1’580	1’562	1’545	1’530
∞	2’707	2’304	2’085	1’946	1’848	1’775	1’718	1’671	1’633	1’600	1’572	1’547	1’525	1’506	1’489	1’473

Tabla B.12: Distribución $\mathcal{F}$ de Snedecor ($p = 0'9$)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	61'46	61'57	61'66	61'74	62'05	62'26	62'42	62'53	62'62	62'69	62'79	62'87	62'93	62'97	63'06	63'32
2	9'433	9'436	9'439	9'441	9'451	9'458	9'463	9'466	9'469	9'471	9'475	9'477	9'479	9'480	9'483	9'491
3	5'193	5'190	5'187	5'184	5'175	5'168	5'163	5'160	5'157	5'155	5'151	5'149	5'147	5'145	5'143	5'134
4	3'858	3'853	3'848	3'844	3'828	3'817	3'810	3'804	3'799	3'795	3'790	3'786	3'782	3'780	3'775	3'761
5	3'223	3'217	3'212	3'207	3'187	3'174	3'165	3'157	3'152	3'147	3'140	3'135	3'132	3'129	3'123	3'105
6	2'855	2'848	2'842	2'836	2'815	2'800	2'789	2'781	2'775	2'770	2'762	2'756	2'752	2'749	2'742	2'723
7	2'615	2'607	2'601	2'595	2'571	2'555	2'544	2'535	2'528	2'523	2'514	2'508	2'504	2'500	2'493	2'471
8	2'446	2'438	2'431	2'425	2'400	2'383	2'371	2'361	2'354	2'348	2'339	2'333	2'328	2'324	2'316	2'293
9	2'320	2'312	2'305	2'298	2'272	2'255	2'242	2'232	2'224	2'218	2'208	2'202	2'196	2'192	2'184	2'160
10	2'224	2'215	2'208	2'201	2'174	2'155	2'142	2'132	2'124	2'117	2'107	2'100	2'095	2'090	2'082	2'056
11	2'147	2'138	2'130	2'123	2'095	2'076	2'062	2'052	2'043	2'036	2'026	2'019	2'013	2'009	2'000	1'973
12	2'084	2'075	2'067	2'060	2'031	2'011	1'997	1'986	1'977	1'970	1'960	1'952	1'946	1'942	1'932	1'904
13	2'032	2'023	2'014	2'007	1'978	1'958	1'943	1'931	1'923	1'915	1'904	1'896	1'890	1'886	1'876	1'847
14	1'988	1'978	1'970	1'962	1'933	1'912	1'897	1'885	1'876	1'869	1'857	1'849	1'843	1'838	1'828	1'798
15	1'950	1'941	1'932	1'924	1'894	1'873	1'857	1'845	1'836	1'828	1'817	1'808	1'802	1'797	1'787	1'756
16	1'917	1'908	1'899	1'891	1'860	1'839	1'823	1'811	1'801	1'793	1'782	1'773	1'766	1'761	1'751	1'719
17	1'889	1'879	1'870	1'862	1'831	1'809	1'793	1'781	1'771	1'763	1'751	1'742	1'735	1'730	1'719	1'686
18	1'864	1'854	1'845	1'837	1'805	1'783	1'766	1'754	1'744	1'736	1'723	1'714	1'707	1'702	1'691	1'658
19	1'841	1'831	1'822	1'814	1'782	1'759	1'743	1'730	1'720	1'711	1'699	1'690	1'683	1'677	1'666	1'632
20	1'821	1'811	1'802	1'794	1'761	1'738	1'721	1'708	1'698	1'690	1'677	1'667	1'660	1'655	1'643	1'608
21	1'803	1'793	1'784	1'776	1'742	1'719	1'702	1'689	1'678	1'670	1'657	1'647	1'640	1'634	1'623	1'587
22	1'787	1'777	1'768	1'759	1'726	1'702	1'685	1'671	1'661	1'652	1'639	1'629	1'622	1'616	1'604	1'568
23	1'772	1'762	1'753	1'744	1'710	1'686	1'669	1'655	1'645	1'636	1'622	1'613	1'605	1'599	1'587	1'550
24	1'759	1'748	1'739	1'730	1'696	1'672	1'654	1'641	1'630	1'621	1'607	1'597	1'590	1'584	1'571	1'534
25	1'746	1'736	1'726	1'718	1'683	1'659	1'641	1'627	1'616	1'607	1'593	1'583	1'576	1'569	1'557	1'519
26	1'735	1'724	1'715	1'706	1'671	1'647	1'629	1'615	1'604	1'594	1'581	1'570	1'562	1'556	1'544	1'505
27	1'724	1'714	1'704	1'695	1'660	1'636	1'617	1'603	1'592	1'583	1'569	1'558	1'550	1'544	1'531	1'492
28	1'715	1'704	1'694	1'685	1'650	1'625	1'607	1'592	1'581	1'572	1'558	1'547	1'539	1'533	1'520	1'479
29	1'705	1'695	1'685	1'676	1'640	1'616	1'597	1'583	1'571	1'562	1'547	1'537	1'529	1'522	1'509	1'468
30	1'697	1'686	1'676	1'667	1'632	1'606	1'588	1'573	1'562	1'552	1'538	1'527	1'519	1'512	1'499	1'457
35	1'662	1'651	1'641	1'632	1'595	1'569	1'550	1'535	1'523	1'513	1'497	1'486	1'478	1'471	1'457	1'413
40	1'636	1'625	1'615	1'605	1'568	1'541	1'521	1'506	1'493	1'483	1'467	1'455	1'447	1'439	1'425	1'378
50	1'600	1'588	1'578	1'568	1'529	1'502	1'481	1'465	1'452	1'441	1'424	1'412	1'402	1'395	1'379	1'328
60	1'576	1'564	1'553	1'543	1'504	1'476	1'454	1'437	1'424	1'413	1'395	1'382	1'372	1'364	1'348	1'293
70	1'559	1'547	1'536	1'526	1'486	1'457	1'435	1'418	1'404	1'392	1'374	1'361	1'350	1'342	1'325	1'267
80	1'546	1'534	1'523	1'513	1'472	1'443	1'420	1'403	1'388	1'377	1'358	1'344	1'334	1'325	1'307	1'246
90	1'536	1'524	1'513	1'503	1'461	1'432	1'409	1'391	1'377	1'365	1'346	1'332	1'321	1'312	1'293	1'230
100	1'528	1'516	1'505	1'494	1'453	1'423	1'400	1'382	1'367	1'355	1'336	1'321	1'310	1'301	1'282	1'216
120	1'516	1'504	1'493	1'482	1'440	1'409	1'386	1'368	1'353	1'340	1'320	1'305	1'294	1'284	1'265	1'195
∞	1'458	1'445	1'433	1'422	1'377	1'344	1'318	1'297	1'280	1'265	1'242	1'224	1'209	1'197	1'171	1'037

Tabla B.13: Distribución $\mathcal{F}$ de Snedecor ($p = 0’95$)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	161’4	199’5	215’7	224’6	230’2	234’0	236’8	238’9	240’5	241’9	243’0	243’9	244’7	245’4	245’9	246’5
2	18’51	19’00	19’16	19’25	19’30	19’33	19’35	19’37	19’38	19’40	19’40	19’41	19’42	19’42	19’43	19’43
3	10’13	9’552	9’277	9’117	9’013	8’941	8’887	8’845	8’812	8’785	8’763	8’745	8’729	8’715	8’703	8’692
4	7’709	6’944	6’591	6’388	6’256	6’163	6’094	6’041	5’999	5’964	5’936	5’912	5’891	5’873	5’858	5’844
5	6’608	5’786	5’409	5’192	5’050	4’950	4’876	4’818	4’772	4’735	4’704	4’678	4’655	4’636	4’619	4’604
6	5’987	5’143	4’757	4’534	4’387	4’284	4’207	4’147	4’099	4’060	4’027	4’000	3’976	3’956	3’938	3’922
7	5’591	4’737	4’347	4’120	3’972	3’866	3’787	3’726	3’677	3’637	3’603	3’575	3’550	3’529	3’511	3’494
8	5’318	4’459	4’066	3’838	3’688	3’581	3’500	3’438	3’388	3’347	3’313	3’284	3’259	3’237	3’218	3’202
9	5’117	4’256	3’863	3’633	3’482	3’374	3’293	3’230	3’179	3’137	3’102	3’073	3’048	3’025	3’006	2’989
10	4’965	4’103	3’708	3’478	3’326	3’217	3’135	3’072	3’020	2’978	2’943	2’913	2’887	2’865	2’845	2’828
11	4’844	3’982	3’587	3’357	3’204	3’095	3’012	2’948	2’896	2’854	2’818	2’788	2’761	2’739	2’719	2’701
12	4’747	3’885	3’490	3’259	3’106	2’996	2’913	2’849	2’796	2’753	2’717	2’687	2’660	2’637	2’617	2’599
13	4’667	3’806	3’411	3’179	3’025	2’915	2’832	2’767	2’714	2’671	2’635	2’604	2’577	2’554	2’533	2’515
14	4’600	3’739	3’344	3’112	2’958	2’848	2’764	2’699	2’646	2’602	2’565	2’534	2’507	2’484	2’463	2’445
15	4’543	3’682	3’287	3’056	2’901	2’790	2’707	2’641	2’588	2’544	2’507	2’475	2’448	2’424	2’403	2’385
16	4’494	3’634	3’239	3’007	2’852	2’741	2’657	2’591	2’538	2’494	2’456	2’425	2’397	2’373	2’352	2’333
17	4’451	3’592	3’197	2’965	2’810	2’699	2’614	2’548	2’494	2’450	2’413	2’381	2’353	2’329	2’308	2’289
18	4’414	3’555	3’160	2’928	2’773	2’661	2’577	2’510	2’456	2’412	2’374	2’342	2’314	2’290	2’269	2’250
19	4’381	3’522	3’127	2’895	2’740	2’628	2’544	2’477	2’423	2’378	2’340	2’308	2’280	2’256	2’234	2’215
20	4’351	3’493	3’098	2’866	2’711	2’599	2’514	2’447	2’393	2’348	2’310	2’278	2’250	2’225	2’203	2’184
21	4’325	3’467	3’072	2’840	2’685	2’573	2’488	2’420	2’366	2’321	2’283	2’250	2’222	2’197	2’176	2’156
22	4’301	3’443	3’049	2’817	2’661	2’549	2’464	2’397	2’342	2’297	2’259	2’226	2’198	2’173	2’151	2’131
23	4’279	3’422	3’028	2’796	2’640	2’528	2’442	2’375	2’320	2’275	2’236	2’204	2’175	2’150	2’128	2’109
24	4’260	3’403	3’009	2’776	2’621	2’508	2’423	2’355	2’300	2’255	2’216	2’183	2’155	2’130	2’108	2’088
25	4’242	3’385	2’991	2’759	2’603	2’490	2’405	2’337	2’282	2’236	2’198	2’165	2’136	2’111	2’089	2’069
26	4’225	3’369	2’975	2’743	2’587	2’474	2’388	2’321	2’265	2’220	2’181	2’148	2’119	2’094	2’072	2’052
27	4’210	3’354	2’960	2’728	2’572	2’459	2’373	2’305	2’250	2’204	2’166	2’132	2’103	2’078	2’056	2’036
28	4’196	3’340	2’947	2’714	2’558	2’445	2’359	2’291	2’236	2’190	2’151	2’118	2’089	2’064	2’041	2’021
29	4’183	3’328	2’934	2’701	2’545	2’432	2’346	2’278	2’223	2’177	2’138	2’104	2’075	2’050	2’027	2’007
30	4’171	3’316	2’922	2’690	2’534	2’421	2’334	2’266	2’211	2’165	2’126	2’092	2’063	2’037	2’015	1’995
35	4’121	3’267	2’874	2’641	2’485	2’372	2’285	2’217	2’161	2’114	2’075	2’041	2’012	1’986	1’963	1’942
40	4’085	3’232	2’839	2’606	2’449	2’336	2’249	2’180	2’124	2’077	2’038	2’003	1’974	1’948	1’924	1’904
50	4’034	3’183	2’790	2’557	2’400	2’286	2’199	2’130	2’073	2’026	1’986	1’952	1’921	1’895	1’871	1’850
60	4’001	3’150	2’758	2’525	2’368	2’254	2’167	2’097	2’040	1’993	1’952	1’917	1’887	1’860	1’836	1’815
70	3’978	3’128	2’736	2’503	2’346	2’231	2’143	2’074	2’017	1’969	1’928	1’893	1’863	1’836	1’812	1’790
80	3’960	3’111	2’719	2’486	2’329	2’214	2’126	2’056	1’999	1’951	1’910	1’875	1’845	1’817	1’793	1’772
90	3’947	3’098	2’706	2’473	2’316	2’201	2’113	2’043	1’986	1’938	1’897	1’861	1’830	1’803	1’779	1’757
100	3’936	3’087	2’696	2’463	2’305	2’191	2’103	2’032	1’975	1’927	1’886	1’850	1’819	1’792	1’768	1’746
120	3’920	3’072	2’680	2’447	2’290	2’175	2’087	2’016	1’959	1’910	1’869	1’834	1’803	1’775	1’750	1’728
∞	3’843	2’998	2’607	2’374	2’216	2’100	2’011	1’940	1’882	1’833	1’791	1’754	1’722	1’694	1’668	1’646

Tabla B.14: Distribución $\mathcal{F}$ de Snedecor ($p = 0'95$)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	246'9	247'3	247'7	248'0	249'3	250'1	250'7	251'1	251'5	251'8	252'2	252'5	252'7	252'9	253'3	254'3
2	19'44	19'44	19'44	19'45	19'46	19'46	19'47	19'47	19'47	19'48	19'48	19'48	19'48	19'48	19'49	19'50
3	8'683	8'675	8'667	8'660	8'634	8'617	8'604	8'594	8'587	8'581	8'572	8'566	8'561	8'557	8'549	8'527
4	5'832	5'821	5'811	5'803	5'769	5'746	5'729	5'717	5'707	5'699	5'688	5'679	5'673	5'668	5'658	5'629
5	4'590	4'579	4'568	4'558	4'521	4'496	4'478	4'464	4'453	4'444	4'431	4'422	4'415	4'409	4'398	4'366
6	3'908	3'896	3'884	3'874	3'835	3'808	3'789	3'774	3'763	3'754	3'740	3'730	3'722	3'716	3'705	3'670
7	3'480	3'467	3'455	3'445	3'404	3'376	3'356	3'340	3'328	3'319	3'304	3'294	3'286	3'280	3'267	3'231
8	3'187	3'173	3'161	3'150	3'108	3'079	3'059	3'043	3'030	3'020	3'005	2'994	2'986	2'980	2'967	2'929
9	2'974	2'960	2'948	2'936	2'893	2'864	2'842	2'826	2'813	2'803	2'787	2'776	2'768	2'761	2'748	2'708
10	2'812	2'798	2'785	2'774	2'730	2'700	2'678	2'661	2'648	2'637	2'621	2'609	2'601	2'594	2'580	2'539
11	2'685	2'671	2'658	2'646	2'601	2'570	2'548	2'531	2'517	2'507	2'490	2'478	2'469	2'462	2'448	2'406
12	2'583	2'568	2'555	2'544	2'498	2'466	2'443	2'426	2'412	2'401	2'384	2'372	2'363	2'356	2'341	2'297
13	2'499	2'484	2'471	2'459	2'412	2'380	2'357	2'339	2'325	2'314	2'297	2'284	2'275	2'267	2'252	2'208
14	2'428	2'413	2'400	2'388	2'341	2'308	2'284	2'266	2'252	2'241	2'223	2'210	2'201	2'193	2'178	2'132
15	2'368	2'353	2'340	2'328	2'280	2'247	2'223	2'204	2'190	2'178	2'160	2'147	2'137	2'130	2'114	2'067
16	2'317	2'302	2'288	2'276	2'227	2'194	2'169	2'151	2'136	2'124	2'106	2'093	2'083	2'075	2'059	2'011
17	2'272	2'257	2'243	2'230	2'181	2'148	2'123	2'104	2'089	2'077	2'058	2'045	2'035	2'027	2'011	1'962
18	2'233	2'217	2'203	2'191	2'141	2'107	2'082	2'063	2'048	2'035	2'017	2'003	1'993	1'985	1'968	1'918
19	2'198	2'182	2'168	2'155	2'106	2'071	2'046	2'026	2'011	1'999	1'980	1'966	1'955	1'947	1'930	1'879
20	2'167	2'151	2'137	2'124	2'074	2'039	2'013	1'994	1'978	1'966	1'946	1'932	1'922	1'913	1'896	1'844
21	2'139	2'123	2'109	2'096	2'045	2'010	1'984	1'965	1'949	1'936	1'916	1'902	1'891	1'883	1'866	1'813
22	2'114	2'098	2'084	2'071	2'020	1'984	1'958	1'938	1'922	1'909	1'889	1'875	1'864	1'856	1'838	1'784
23	2'091	2'075	2'061	2'048	1'996	1'961	1'934	1'914	1'898	1'885	1'865	1'850	1'839	1'830	1'813	1'758
24	2'070	2'054	2'040	2'027	1'975	1'939	1'912	1'892	1'876	1'863	1'842	1'828	1'816	1'808	1'790	1'734
25	2'051	2'035	2'021	2'007	1'955	1'919	1'892	1'872	1'855	1'842	1'822	1'807	1'796	1'787	1'768	1'712
26	2'034	2'018	2'003	1'990	1'938	1'901	1'874	1'853	1'837	1'823	1'803	1'788	1'776	1'767	1'749	1'692
27	2'018	2'002	1'987	1'974	1'921	1'884	1'857	1'836	1'819	1'806	1'785	1'770	1'758	1'749	1'731	1'673
28	2'003	1'987	1'972	1'959	1'906	1'869	1'841	1'820	1'803	1'790	1'769	1'754	1'742	1'733	1'714	1'656
29	1'989	1'973	1'958	1'945	1'891	1'854	1'827	1'806	1'789	1'775	1'754	1'738	1'726	1'717	1'698	1'639
30	1'976	1'960	1'945	1'932	1'878	1'841	1'813	1'792	1'775	1'761	1'740	1'724	1'712	1'703	1'683	1'624
35	1'924	1'907	1'892	1'878	1'824	1'786	1'757	1'735	1'718	1'703	1'681	1'665	1'652	1'643	1'623	1'560
40	1'885	1'868	1'853	1'839	1'783	1'744	1'715	1'693	1'675	1'660	1'637	1'621	1'608	1'597	1'577	1'511
50	1'831	1'814	1'798	1'784	1'727	1'687	1'657	1'634	1'615	1'599	1'576	1'558	1'544	1'534	1'511	1'440
60	1'796	1'778	1'763	1'748	1'690	1'649	1'618	1'594	1'575	1'559	1'534	1'516	1'502	1'491	1'467	1'391
70	1'771	1'753	1'737	1'722	1'664	1'622	1'591	1'566	1'546	1'530	1'505	1'486	1'471	1'459	1'435	1'355
80	1'752	1'734	1'718	1'703	1'644	1'602	1'570	1'545	1'525	1'508	1'482	1'463	1'448	1'436	1'411	1'327
90	1'737	1'720	1'703	1'688	1'629	1'586	1'554	1'528	1'508	1'491	1'465	1'445	1'429	1'417	1'391	1'304
100	1'726	1'708	1'691	1'676	1'616	1'573	1'541	1'515	1'494	1'477	1'450	1'430	1'415	1'402	1'376	1'286
120	1'709	1'690	1'674	1'659	1'598	1'554	1'521	1'495	1'474	1'457	1'429	1'408	1'392	1'379	1'352	1'257
∞	1'625	1'606	1'589	1'573	1'508	1'461	1'425	1'396	1'373	1'353	1'321	1'296	1'277	1'260	1'225	1'048

Tabla B.15: Distribución $\mathcal{F}$ de Snedecor ($p = 0'975$)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	647'8	799'5	864'2	899'6	921'8	937'1	948'2	956'6	963'3	968'6	973'0	976'7	979'8	982'5	984'9	986'9
2	38'51	39'00	39'17	39'25	39'30	39'33	39'36	39'37	39'39	39'40	39'41	39'41	39'42	39'43	39'43	39'44
3	17'44	16'04	15'44	15'10	14'88	14'73	14'62	14'54	14'47	14'42	14'37	14'34	14'30	14'28	14'25	14'23
4	12'22	10'65	9'979	9'604	9'364	9'197	9'074	8'980	8'905	8'844	8'794	8'751	8'715	8'684	8'657	8'633
5	10'01	8'434	7'764	7'388	7'146	6'978	6'853	6'757	6'681	6'619	6'568	6'525	6'488	6'456	6'428	6'403
6	8'813	7'260	6'599	6'227	5'988	5'820	5'695	5'600	5'523	5'461	5'410	5'366	5'329	5'297	5'269	5'244
7	8'073	6'542	5'890	5'523	5'285	5'119	4'995	4'899	4'823	4'761	4'709	4'666	4'628	4'596	4'568	4'543
8	7'571	6'059	5'416	5'053	4'817	4'652	4'529	4'433	4'357	4'295	4'243	4'200	4'162	4'130	4'101	4'076
9	7'209	5'715	5'078	4'718	4'484	4'320	4'197	4'102	4'026	3'964	3'912	3'868	3'831	3'798	3'769	3'744
10	6'937	5'456	4'826	4'468	4'236	4'072	3'950	3'855	3'779	3'717	3'665	3'621	3'583	3'550	3'522	3'496
11	6'724	5'256	4'630	4'275	4'044	3'881	3'759	3'664	3'588	3'526	3'474	3'430	3'392	3'359	3'330	3'304
12	6'554	5'096	4'474	4'121	3'891	3'728	3'607	3'512	3'436	3'374	3'321	3'277	3'239	3'206	3'177	3'152
13	6'414	4'965	4'347	3'996	3'767	3'604	3'483	3'388	3'312	3'250	3'197	3'153	3'115	3'082	3'053	3'027
14	6'298	4'857	4'242	3'892	3'663	3'501	3'380	3'285	3'209	3'147	3'095	3'050	3'012	2'979	2'949	2'923
15	6'200	4'765	4'153	3'804	3'576	3'415	3'293	3'199	3'123	3'060	3'008	2'963	2'925	2'891	2'862	2'836
16	6'115	4'687	4'077	3'729	3'502	3'341	3'219	3'125	3'049	2'986	2'934	2'889	2'851	2'817	2'788	2'761
17	6'042	4'619	4'011	3'665	3'438	3'277	3'156	3'061	2'985	2'922	2'870	2'825	2'786	2'753	2'723	2'697
18	5'978	4'560	3'954	3'608	3'382	3'221	3'100	3'005	2'929	2'866	2'814	2'769	2'730	2'696	2'667	2'640
19	5'922	4'508	3'903	3'559	3'333	3'172	3'051	2'956	2'880	2'817	2'765	2'720	2'681	2'647	2'617	2'591
20	5'871	4'461	3'859	3'515	3'289	3'128	3'007	2'913	2'837	2'774	2'721	2'676	2'637	2'603	2'573	2'547
21	5'827	4'420	3'819	3'475	3'250	3'090	2'969	2'874	2'798	2'735	2'682	2'637	2'598	2'564	2'534	2'507
22	5'786	4'383	3'783	3'440	3'215	3'055	2'934	2'839	2'763	2'700	2'647	2'602	2'563	2'528	2'498	2'472
23	5'750	4'349	3'750	3'408	3'183	3'023	2'902	2'808	2'731	2'668	2'615	2'570	2'531	2'497	2'466	2'440
24	5'717	4'319	3'721	3'379	3'155	2'995	2'874	2'779	2'703	2'640	2'586	2'541	2'502	2'468	2'437	2'411
25	5'686	4'291	3'694	3'353	3'129	2'969	2'848	2'753	2'677	2'613	2'560	2'515	2'476	2'441	2'411	2'384
26	5'659	4'265	3'670	3'329	3'105	2'945	2'824	2'729	2'653	2'590	2'536	2'491	2'452	2'417	2'387	2'360
27	5'633	4'242	3'647	3'307	3'083	2'923	2'802	2'707	2'631	2'568	2'514	2'469	2'429	2'395	2'364	2'337
28	5'610	4'221	3'626	3'286	3'063	2'903	2'782	2'687	2'611	2'547	2'494	2'448	2'409	2'374	2'344	2'317
29	5'588	4'201	3'607	3'267	3'044	2'884	2'763	2'669	2'592	2'529	2'475	2'430	2'390	2'355	2'325	2'298
30	5'568	4'182	3'589	3'250	3'026	2'867	2'746	2'651	2'575	2'511	2'458	2'412	2'372	2'338	2'307	2'280
35	5'485	4'106	3'517	3'179	2'956	2'796	2'676	2'581	2'504	2'440	2'387	2'341	2'301	2'266	2'235	2'207
40	5'424	4'051	3'463	3'126	2'904	2'744	2'624	2'529	2'452	2'388	2'334	2'288	2'248	2'213	2'182	2'154
50	5'340	3'975	3'390	3'054	2'833	2'674	2'553	2'458	2'381	2'317	2'263	2'216	2'176	2'140	2'109	2'081
60	5'286	3'925	3'343	3'008	2'786	2'627	2'507	2'412	2'334	2'270	2'216	2'169	2'129	2'093	2'061	2'033
70	5'247	3'890	3'309	2'975	2'754	2'595	2'474	2'379	2'302	2'237	2'183	2'136	2'095	2'059	2'028	1'999
80	5'218	3'864	3'284	2'950	2'730	2'571	2'450	2'355	2'277	2'213	2'158	2'111	2'071	2'035	2'003	1'974
90	5'196	3'844	3'265	2'932	2'711	2'552	2'432	2'336	2'259	2'194	2'140	2'092	2'051	2'015	1'983	1'955
100	5'179	3'828	3'250	2'917	2'696	2'537	2'417	2'321	2'244	2'179	2'124	2'077	2'036	2'000	1'968	1'939
120	5'152	3'805	3'227	2'894	2'674	2'515	2'395	2'299	2'222	2'157	2'102	2'055	2'014	1'977	1'945	1'916
∞	5'027	3'692	3'119	2'788	2'569	2'411	2'290	2'194	2'116	2'051	1'995	1'947	1'905	1'868	1'835	1'806

Tabla B.16: Distribución $\mathcal{F}$ de Snedecor ($p = 0'975$)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	988'7	990'3	991'8	993'1	998'1	1001	1004	1006	1007	1008	1010	1011	1012	1013	1014	1018
2	39'44	39'44	39'45	39'45	39'46	39'46	39'47	39'47	39'48	39'48	39'48	39'48	39'49	39'49	39'49	39'50
3	14'21	14'20	14'18	14'17	14'12	14'08	14'06	14'04	14'02	14'01	13'99	13'98	13'97	13'96	13'95	13'90
4	8'611	8'592	8'575	8'560	8'501	8'461	8'433	8'411	8'394	8'381	8'360	8'346	8'335	8'326	8'309	8'259
5	6'381	6'362	6'344	6'329	6'268	6'227	6'197	6'175	6'158	6'144	6'123	6'107	6'096	6'087	6'069	6'017
6	5'222	5'202	5'184	5'168	5'107	5'065	5'035	5'012	4'995	4'980	4'959	4'943	4'932	4'923	4'904	4'850
7	4'521	4'501	4'483	4'467	4'405	4'362	4'332	4'309	4'291	4'276	4'254	4'239	4'227	4'218	4'199	4'144
8	4'054	4'034	4'016	3'999	3'937	3'894	3'863	3'840	3'821	3'807	3'784	3'768	3'756	3'747	3'728	3'672
9	3'722	3'701	3'683	3'667	3'604	3'560	3'529	3'505	3'487	3'472	3'449	3'433	3'421	3'411	3'392	3'334
10	3'474	3'453	3'435	3'419	3'355	3'311	3'279	3'255	3'237	3'221	3'198	3'182	3'169	3'160	3'140	3'081
11	3'282	3'261	3'243	3'226	3'162	3'118	3'086	3'061	3'042	3'027	3'004	2'987	2'974	2'964	2'944	2'884
12	3'129	3'108	3'090	3'073	3'008	2'963	2'931	2'906	2'887	2'871	2'848	2'831	2'818	2'808	2'787	2'726
13	3'004	2'983	2'965	2'948	2'882	2'837	2'805	2'780	2'760	2'744	2'720	2'703	2'690	2'680	2'659	2'597
14	2'900	2'879	2'861	2'844	2'778	2'732	2'699	2'674	2'654	2'638	2'614	2'597	2'583	2'573	2'552	2'489
15	2'813	2'792	2'773	2'756	2'689	2'644	2'610	2'585	2'565	2'549	2'524	2'506	2'493	2'482	2'461	2'397
16	2'738	2'717	2'698	2'681	2'614	2'568	2'534	2'509	2'488	2'472	2'447	2'429	2'415	2'405	2'383	2'318
17	2'673	2'652	2'633	2'616	2'548	2'502	2'468	2'442	2'422	2'405	2'380	2'362	2'348	2'337	2'315	2'249
18	2'617	2'596	2'576	2'559	2'491	2'445	2'410	2'384	2'364	2'347	2'321	2'303	2'289	2'278	2'256	2'189
19	2'567	2'546	2'526	2'509	2'441	2'394	2'359	2'333	2'312	2'295	2'270	2'251	2'237	2'226	2'203	2'135
20	2'523	2'501	2'482	2'464	2'396	2'349	2'314	2'287	2'266	2'249	2'223	2'205	2'190	2'179	2'156	2'087
21	2'483	2'462	2'442	2'425	2'356	2'308	2'273	2'246	2'225	2'208	2'182	2'163	2'148	2'137	2'114	2'044
22	2'448	2'426	2'407	2'389	2'320	2'272	2'237	2'210	2'188	2'171	2'145	2'125	2'111	2'099	2'076	2'005
23	2'416	2'394	2'374	2'357	2'287	2'239	2'204	2'176	2'155	2'137	2'111	2'091	2'077	2'065	2'041	1'970
24	2'386	2'365	2'345	2'327	2'257	2'209	2'173	2'146	2'124	2'107	2'080	2'060	2'045	2'034	2'010	1'937
25	2'360	2'338	2'318	2'300	2'230	2'182	2'146	2'118	2'096	2'079	2'052	2'032	2'017	2'005	1'981	1'907
26	2'335	2'314	2'294	2'276	2'205	2'157	2'120	2'093	2'071	2'053	2'026	2'006	1'991	1'979	1'954	1'880
27	2'313	2'291	2'271	2'253	2'183	2'133	2'097	2'069	2'047	2'029	2'002	1'982	1'966	1'954	1'930	1'855
28	2'292	2'270	2'251	2'232	2'161	2'112	2'076	2'048	2'025	2'007	1'980	1'959	1'944	1'932	1'907	1'831
29	2'273	2'251	2'231	2'213	2'142	2'092	2'056	2'028	2'005	1'987	1'959	1'939	1'923	1'911	1'886	1'809
30	2'255	2'233	2'213	2'195	2'124	2'074	2'037	2'009	1'986	1'968	1'940	1'920	1'904	1'892	1'866	1'789
35	2'183	2'160	2'140	2'122	2'049	1'999	1'961	1'932	1'909	1'890	1'861	1'840	1'824	1'811	1'785	1'704
40	2'129	2'107	2'086	2'068	1'994	1'943	1'905	1'875	1'852	1'832	1'803	1'781	1'764	1'751	1'724	1'639
50	2'056	2'033	2'012	1'993	1'919	1'866	1'827	1'796	1'772	1'752	1'721	1'698	1'681	1'667	1'639	1'548
60	2'008	1'985	1'964	1'944	1'869	1'815	1'775	1'744	1'719	1'699	1'667	1'643	1'625	1'611	1'581	1'485
70	1'974	1'950	1'929	1'910	1'833	1'779	1'739	1'707	1'681	1'660	1'628	1'604	1'585	1'570	1'539	1'438
80	1'948	1'925	1'904	1'884	1'807	1'752	1'711	1'679	1'653	1'632	1'599	1'574	1'555	1'540	1'508	1'403
90	1'929	1'905	1'884	1'864	1'787	1'731	1'690	1'657	1'631	1'610	1'576	1'551	1'531	1'516	1'483	1'374
100	1'913	1'890	1'868	1'849	1'770	1'715	1'673	1'640	1'614	1'592	1'558	1'532	1'512	1'496	1'463	1'351
120	1'890	1'866	1'845	1'825	1'746	1'690	1'647	1'614	1'587	1'565	1'530	1'504	1'483	1'467	1'433	1'314
∞	1'779	1'754	1'732	1'711	1'629	1'569	1'523	1'487	1'457	1'432	1'392	1'361	1'337	1'317	1'273	1'057

Tabla B.17: Distribución $\mathcal{F}$ de Snedecor ($p = 0$'99)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107	6126	6143	6157	6170
2	98'50	99'00	99'16	99'25	99'30	99'33	99'36	99'38	99'39	99'40	99'41	99'42	99'42	99'43	99'43	99'44
3	34'12	30'82	29'46	28'71	28'24	27'91	27'67	27'49	27'34	27'23	27'13	27'05	26'98	26'92	26'87	26'83
4	21'20	18'00	16'69	15'98	15'52	15'21	14'98	14'80	14'66	14'55	14'45	14'37	14'31	14'25	14'20	14'15
5	16'26	13'27	12'06	11'39	10'97	10'67	10'46	10'29	10'16	10'05	9'963	9'888	9'825	9'770	9'722	9'680
6	13'75	10'92	9'780	9'148	8'746	8'466	8'260	8'102	7'976	7'874	7'790	7'718	7'657	7'605	7'559	7'519
7	12'25	9'547	8'451	7'847	7'460	7'191	6'993	6'840	6'719	6'620	6'538	6'469	6'410	6'359	6'314	6'275
8	11'26	8'649	7'591	7'006	6'632	6'371	6'178	6'029	5'911	5'814	5'734	5'667	5'609	5'559	5'515	5'477
9	10'56	8'022	6'992	6'422	6'057	5'802	5'613	5'467	5'351	5'257	5'178	5'111	5'055	5'005	4'962	4'924
10	10'04	7'559	6'552	5'994	5'636	5'386	5'200	5'057	4'942	4'849	4'772	4'706	4'650	4'601	4'558	4'520
11	9'646	7'206	6'217	5'668	5'316	5'069	4'886	4'744	4'632	4'539	4'462	4'397	4'342	4'293	4'251	4'213
12	9'330	6'927	5'953	5'412	5'064	4'821	4'640	4'499	4'388	4'296	4'220	4'155	4'100	4'052	4'010	3'972
13	9'074	6'701	5'739	5'205	4'862	4'620	4'441	4'302	4'191	4'100	4'025	3'960	3'905	3'857	3'815	3'778
14	8'862	6'515	5'564	5'035	4'695	4'456	4'278	4'140	4'030	3'939	3'864	3'800	3'745	3'698	3'656	3'619
15	8'683	6'359	5'417	4'893	4'556	4'318	4'142	4'004	3'895	3'805	3'730	3'666	3'612	3'564	3'522	3'485
16	8'531	6'226	5'292	4'773	4'437	4'202	4'026	3'890	3'780	3'691	3'616	3'553	3'498	3'451	3'409	3'372
17	8'400	6'112	5'185	4'669	4'336	4'101	3'927	3'791	3'682	3'593	3'518	3'455	3'401	3'353	3'312	3'275
18	8'285	6'013	5'092	4'579	4'248	4'015	3'841	3'705	3'597	3'508	3'434	3'371	3'316	3'269	3'227	3'190
19	8'185	5'926	5'010	4'500	4'171	3'939	3'765	3'631	3'523	3'434	3'360	3'297	3'242	3'195	3'153	3'116
20	8'096	5'849	4'938	4'431	4'103	3'871	3'699	3'564	3'457	3'368	3'294	3'231	3'177	3'130	3'088	3'051
21	8'017	5'780	4'874	4'369	4'042	3'812	3'640	3'506	3'398	3'310	3'236	3'173	3'119	3'072	3'030	2'993
22	7'945	5'719	4'817	4'313	3'988	3'758	3'587	3'453	3'346	3'258	3'184	3'121	3'067	3'019	2'978	2'941
23	7'881	5'664	4'765	4'264	3'939	3'710	3'539	3'406	3'299	3'211	3'137	3'074	3'020	2'973	2'931	2'894
24	7'823	5'614	4'718	4'218	3'895	3'667	3'496	3'363	3'256	3'168	3'094	3'032	2'977	2'930	2'889	2'852
25	7'770	5'568	4'675	4'177	3'855	3'627	3'457	3'324	3'217	3'129	3'056	2'993	2'939	2'892	2'850	2'813
26	7'721	5'526	4'637	4'140	3'818	3'591	3'421	3'288	3'182	3'094	3'021	2'958	2'904	2'857	2'815	2'778
27	7'677	5'488	4'601	4'106	3'785	3'558	3'388	3'256	3'149	3'062	2'988	2'926	2'872	2'824	2'783	2'746
28	7'636	5'453	4'568	4'074	3'754	3'528	3'358	3'226	3'120	3'032	2'959	2'896	2'842	2'795	2'753	2'716
29	7'598	5'420	4'538	4'045	3'725	3'499	3'330	3'198	3'092	3'005	2'931	2'868	2'814	2'767	2'726	2'689
30	7'562	5'390	4'510	4'018	3'699	3'473	3'305	3'173	3'067	2'979	2'906	2'843	2'789	2'742	2'700	2'663
35	7'419	5'268	4'396	3'908	3'592	3'368	3'200	3'069	2'963	2'876	2'803	2'740	2'686	2'639	2'597	2'560
40	7'314	5'178	4'313	3'828	3'514	3'291	3'124	2'993	2'888	2'801	2'727	2'665	2'611	2'563	2'522	2'484
50	7'171	5'057	4'199	3'720	3'408	3'186	3'020	2'890	2'785	2'698	2'625	2'563	2'508	2'461	2'419	2'382
60	7'077	4'977	4'126	3'649	3'339	3'119	2'953	2'823	2'718	2'632	2'559	2'496	2'442	2'394	2'352	2'315
70	7'011	4'922	4'074	3'600	3'291	3'071	2'906	2'777	2'672	2'585	2'512	2'450	2'395	2'348	2'306	2'268
80	6'963	4'881	4'036	3'563	3'255	3'036	2'871	2'742	2'637	2'551	2'478	2'415	2'361	2'313	2'271	2'233
90	6'925	4'849	4'007	3'535	3'228	3'009	2'845	2'715	2'611	2'524	2'451	2'389	2'334	2'286	2'244	2'206
100	6'895	4'824	3'984	3'513	3'206	2'988	2'823	2'694	2'590	2'503	2'430	2'368	2'313	2'265	2'223	2'185
120	6'851	4'787	3'949	3'480	3'174	2'956	2'792	2'663	2'559	2'472	2'399	2'336	2'282	2'234	2'191	2'154
∞	6'640	4'609	3'786	3'323	3'021	2'806	2'643	2'515	2'411	2'324	2'251	2'188	2'133	2'085	2'042	2'004

Tabla B.18: Distribución $\mathcal{F}$ de Snedecor ($p = 0'99$)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	6181	6191	6201	6209	6240	6260	6275	6286	6296	6302	6313	6321	6326	6331	6340	6366
2	99'44	99'44	99'45	99'45	99'46	99'47	99'47	99'48	99'48	99'48	99'48	99'48	99'48	99'49	99'49	99'50
3	26'79	26'75	26'72	26'69	26'58	26'50	26'45	26'41	26'38	26'35	26'32	26'29	26'27	26'25	26'22	26'13
4	14'11	14'08	14'05	14'02	13'91	13'84	13'79	13'75	13'71	13'69	13'65	13'63	13'61	13'59	13'56	13'47
5	9'643	9'609	9'580	9'553	9'449	9'379	9'329	9'291	9'262	9'238	9'202	9'176	9'157	9'142	9'112	9'023
6	7'483	7'451	7'422	7'396	7'296	7'229	7'180	7'143	7'115	7'091	7'057	7'032	7'013	6'998	6'969	6'882
7	6'240	6'209	6'181	6'155	6'058	5'992	5'944	5'908	5'880	5'858	5'824	5'799	5'781	5'766	5'737	5'652
8	5'442	5'412	5'384	5'359	5'263	5'198	5'151	5'116	5'088	5'065	5'032	5'007	4'989	4'975	4'946	4'861
9	4'890	4'860	4'833	4'808	4'713	4'649	4'602	4'567	4'539	4'517	4'483	4'459	4'441	4'426	4'398	4'313
10	4'487	4'457	4'430	4'405	4'311	4'247	4'201	4'165	4'138	4'115	4'082	4'058	4'039	4'025	3'996	3'911
11	4'180	4'150	4'123	4'099	4'005	3'941	3'895	3'860	3'832	3'810	3'776	3'752	3'734	3'719	3'690	3'605
12	3'939	3'910	3'883	3'858	3'765	3'701	3'654	3'619	3'592	3'569	3'535	3'511	3'493	3'478	3'449	3'363
13	3'745	3'716	3'689	3'665	3'571	3'507	3'461	3'425	3'398	3'375	3'341	3'317	3'298	3'284	3'255	3'168
14	3'586	3'556	3'529	3'505	3'412	3'348	3'301	3'266	3'238	3'215	3'181	3'157	3'138	3'124	3'094	3'006
15	3'452	3'423	3'396	3'372	3'278	3'214	3'167	3'132	3'104	3'081	3'047	3'022	3'004	2'989	2'959	2'871
16	3'339	3'310	3'283	3'259	3'165	3'101	3'054	3'018	2'990	2'967	2'933	2'908	2'889	2'875	2'845	2'755
17	3'242	3'212	3'186	3'162	3'068	3'003	2'956	2'920	2'892	2'869	2'835	2'810	2'791	2'776	2'746	2'655
18	3'158	3'128	3'101	3'077	2'983	2'919	2'871	2'835	2'807	2'784	2'749	2'724	2'705	2'690	2'660	2'568
19	3'084	3'054	3'027	3'003	2'909	2'844	2'797	2'761	2'732	2'709	2'674	2'649	2'630	2'614	2'584	2'492
20	3'018	2'989	2'962	2'938	2'843	2'778	2'731	2'695	2'666	2'643	2'608	2'582	2'563	2'548	2'517	2'424
21	2'960	2'931	2'904	2'880	2'785	2'720	2'672	2'636	2'607	2'584	2'548	2'523	2'503	2'488	2'457	2'363
22	2'908	2'879	2'852	2'827	2'733	2'667	2'620	2'583	2'554	2'531	2'495	2'469	2'450	2'434	2'403	2'308
23	2'861	2'832	2'805	2'780	2'686	2'620	2'572	2'536	2'506	2'483	2'447	2'421	2'401	2'386	2'354	2'258
24	2'819	2'789	2'762	2'738	2'643	2'577	2'529	2'492	2'463	2'440	2'403	2'377	2'357	2'342	2'310	2'213
25	2'780	2'751	2'724	2'699	2'604	2'538	2'490	2'453	2'424	2'400	2'364	2'337	2'317	2'302	2'270	2'172
26	2'745	2'715	2'688	2'664	2'569	2'503	2'454	2'417	2'388	2'364	2'327	2'301	2'281	2'265	2'233	2'134
27	2'713	2'683	2'656	2'632	2'536	2'470	2'421	2'384	2'354	2'330	2'294	2'267	2'247	2'231	2'198	2'099
28	2'683	2'653	2'626	2'602	2'506	2'440	2'391	2'354	2'324	2'300	2'263	2'236	2'216	2'200	2'167	2'067
29	2'656	2'626	2'599	2'574	2'478	2'412	2'363	2'325	2'296	2'271	2'234	2'207	2'187	2'171	2'138	2'037
30	2'630	2'600	2'573	2'549	2'453	2'386	2'337	2'299	2'269	2'245	2'208	2'181	2'160	2'144	2'111	2'009
35	2'527	2'497	2'470	2'445	2'348	2'281	2'231	2'193	2'162	2'137	2'099	2'072	2'050	2'034	2'000	1'894
40	2'451	2'421	2'394	2'369	2'271	2'203	2'153	2'114	2'083	2'058	2'019	1'991	1'969	1'952	1'917	1'808
50	2'348	2'318	2'290	2'265	2'167	2'098	2'046	2'007	1'975	1'949	1'909	1'880	1'857	1'839	1'803	1'686
60	2'281	2'251	2'223	2'198	2'098	2'028	1'976	1'936	1'904	1'877	1'836	1'806	1'783	1'764	1'726	1'604
70	2'234	2'204	2'176	2'150	2'050	1'980	1'927	1'886	1'853	1'826	1'785	1'754	1'730	1'711	1'672	1'544
80	2'199	2'169	2'141	2'115	2'015	1'944	1'890	1'849	1'816	1'788	1'746	1'714	1'690	1'671	1'630	1'498
90	2'172	2'142	2'114	2'088	1'987	1'916	1'862	1'820	1'787	1'759	1'716	1'684	1'659	1'639	1'598	1'461
100	2'151	2'120	2'092	2'067	1'965	1'893	1'839	1'797	1'763	1'735	1'692	1'659	1'634	1'614	1'572	1'431
120	2'119	2'089	2'060	2'035	1'932	1'860	1'806	1'763	1'728	1'700	1'656	1'623	1'597	1'576	1'533	1'385
∞	1'969	1'937	1'908	1'882	1'776	1'700	1'642	1'596	1'559	1'527	1'477	1'439	1'409	1'384	1'330	1'068

Tabla B.19: Distribución $\mathcal{F}$ de Snedecor ($p = 0'995$)

	n_1															
n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	16212	19997	21614	22501	23056	23440	23715	23924	24091	24222	24334	24427	24505	24572	24632	24684
2	198'5	199'0	199'2	199'2	199'3	199'3	199'4	199'4	199'4	199'4	199'4	199'4	199'4	199'4	199'4	199'4
3	55'55	49'80	47'47	46'20	45'39	44'84	44'43	44'13	43'88	43'68	43'52	43'39	43'27	43'17	43'08	43'01
4	31'33	26'28	24'26	23'15	22'46	21'98	21'62	21'35	21'14	20'97	20'82	20'70	20'60	20'51	20'44	20'37
5	22'78	18'31	16'53	15'56	14'94	14'51	14'20	13'96	13'77	13'62	13'49	13'38	13'29	13'21	13'15	13'09
6	18'63	14'54	12'92	12'03	11'46	11'07	10'79	10'57	10'39	10'25	10'13	10'03	9'950	9'878	9'814	9'758
7	16'24	12'40	10'88	10'05	9'522	9'155	8'885	8'678	8'514	8'380	8'270	8'176	8'097	8'028	7'968	7'915
8	14'69	11'04	9'597	8'805	8'302	7'952	7'694	7'496	7'339	7'211	7'105	7'015	6'938	6'872	6'814	6'763
9	13'61	10'11	8'717	7'956	7'471	7'134	6'885	6'693	6'541	6'417	6'314	6'227	6'153	6'089	6'032	5'983
10	12'83	9'427	8'081	7'343	6'872	6'545	6'303	6'116	5'968	5'847	5'746	5'661	5'589	5'526	5'471	5'422
11	12'23	8'912	7'600	6'881	6'422	6'102	5'865	5'682	5'537	5'418	5'320	5'236	5'165	5'103	5'049	5'001
12	11'75	8'510	7'226	6'521	6'071	5'757	5'524	5'345	5'202	5'085	4'988	4'906	4'836	4'775	4'721	4'674
13	11'37	8'186	6'926	6'233	5'791	5'482	5'253	5'076	4'935	4'820	4'724	4'643	4'573	4'513	4'460	4'413
14	11'06	7'922	6'680	5'998	5'562	5'257	5'031	4'857	4'717	4'603	4'508	4'428	4'359	4'299	4'247	4'201
15	10'80	7'701	6'476	5'803	5'372	5'071	4'847	4'674	4'536	4'424	4'329	4'250	4'181	4'122	4'070	4'024
16	10'58	7'514	6'303	5'638	5'212	4'913	4'692	4'521	4'384	4'272	4'179	4'099	4'031	3'972	3'920	3'875
17	10'38	7'354	6'156	5'497	5'075	4'779	4'559	4'389	4'254	4'142	4'050	3'971	3'903	3'844	3'793	3'747
18	10'22	7'215	6'028	5'375	4'956	4'663	4'445	4'276	4'141	4'030	3'938	3'860	3'793	3'734	3'683	3'637
19	10'07	7'093	5'916	5'268	4'853	4'561	4'345	4'177	4'043	3'933	3'841	3'763	3'696	3'638	3'587	3'541
20	9'944	6'987	5'818	5'174	4'762	4'472	4'257	4'090	3'956	3'847	3'756	3'678	3'611	3'553	3'502	3'457
21	9'829	6'891	5'730	5'091	4'681	4'393	4'179	4'013	3'880	3'771	3'680	3'602	3'536	3'478	3'427	3'382
22	9'727	6'806	5'652	5'017	4'609	4'322	4'109	3'944	3'812	3'703	3'612	3'535	3'469	3'411	3'360	3'315
23	9'635	6'730	5'582	4'950	4'544	4'259	4'047	3'882	3'750	3'642	3'551	3'474	3'408	3'351	3'300	3'255
24	9'551	6'661	5'519	4'890	4'486	4'202	3'991	3'826	3'695	3'587	3'497	3'420	3'354	3'296	3'246	3'201
25	9'475	6'598	5'462	4'835	4'433	4'150	3'939	3'776	3'645	3'537	3'447	3'370	3'304	3'247	3'196	3'152
26	9'406	6'541	5'409	4'785	4'384	4'103	3'893	3'730	3'599	3'492	3'402	3'325	3'259	3'202	3'151	3'107
27	9'342	6'489	5'361	4'740	4'340	4'059	3'850	3'687	3'557	3'450	3'360	3'284	3'218	3'161	3'110	3'066
28	9'284	6'440	5'317	4'698	4'300	4'020	3'811	3'649	3'519	3'412	3'322	3'246	3'180	3'123	3'073	3'028
29	9'230	6'396	5'276	4'659	4'262	3'983	3'775	3'613	3'483	3'376	3'287	3'211	3'145	3'088	3'038	2'993
30	9'180	6'355	5'239	4'623	4'228	3'949	3'742	3'580	3'451	3'344	3'255	3'179	3'113	3'056	3'006	2'961
35	8'976	6'188	5'086	4'479	4'088	3'812	3'607	3'447	3'318	3'212	3'124	3'048	2'983	2'926	2'876	2'831
40	8'828	6'066	4'976	4'374	3'986	3'713	3'509	3'350	3'222	3'117	3'028	2'953	2'888	2'831	2'781	2'737
50	8'626	5'902	4'826	4'232	3'849	3'579	3'376	3'219	3'092	2'988	2'900	2'825	2'760	2'703	2'653	2'609
60	8'495	5'795	4'729	4'140	3'760	3'492	3'291	3'134	3'008	2'904	2'817	2'742	2'677	2'620	2'570	2'526
70	8'403	5'720	4'661	4'076	3'698	3'431	3'232	3'076	2'950	2'846	2'759	2'684	2'619	2'563	2'513	2'468
80	8'335	5'665	4'611	4'028	3'652	3'387	3'188	3'032	2'907	2'803	2'716	2'641	2'577	2'520	2'470	2'425
90	8'282	5'623	4'573	3'992	3'617	3'352	3'154	2'999	2'873	2'770	2'683	2'608	2'544	2'487	2'437	2'393
100	8'241	5'589	4'542	3'963	3'589	3'325	3'127	2'972	2'847	2'744	2'657	2'583	2'518	2'461	2'411	2'367
120	8'179	5'539	4'497	3'921	3'548	3'285	3'087	2'933	2'808	2'705	2'618	2'544	2'479	2'423	2'373	2'328
∞	7'886	5'304	4'284	3'720	3'355	3'096	2'901	2'749	2'625	2'523	2'437	2'363	2'298	2'241	2'191	2'146

Tabla B.20: Distribución $\mathcal{F}$ de Snedecor ($p = 0'995$)

	n_1															
n_2	17	18	19	20	25	30	35	40	45	50	60	70	80	90	120	∞
1	24728	24766	24803	24837	24959	25041	25101	25146	25183	25213	25254	25284	25306	25325	25358	25462
2	199'4	199'4	199'4	199'4	199'4	199'5	199'5	199'5	199'5	199'5	199'5	199'5	199'5	199'5	199'5	199'5
3	42'94	42'88	42'83	42'78	42'59	42'47	42'38	42'31	42'26	42'21	42'15	42'10	42'07	42'04	41'99	41'83
4	20'31	20'26	20'21	20'17	20'00	19'89	19'81	19'75	19'71	19'67	19'61	19'57	19'54	19'52	19'47	19'33
5	13'03	12'98	12'94	12'90	12'76	12'66	12'58	12'53	12'49	12'45	12'40	12'37	12'34	12'32	12'27	12'15
6	9'709	9'664	9'625	9'589	9'451	9'358	9'291	9'241	9'201	9'170	9'122	9'088	9'062	9'042	9'001	8'882
7	7'868	7'826	7'788	7'754	7'623	7'534	7'471	7'422	7'385	7'354	7'309	7'276	7'251	7'232	7'193	7'079
8	6'718	6'678	6'641	6'608	6'482	6'396	6'334	6'288	6'251	6'222	6'177	6'145	6'121	6'102	6'065	5'953
9	5'939	5'899	5'864	5'832	5'708	5'625	5'564	5'519	5'483	5'454	5'410	5'379	5'356	5'337	5'300	5'190
10	5'379	5'340	5'306	5'274	5'153	5'071	5'011	4'966	4'931	4'902	4'859	4'828	4'805	4'787	4'750	4'641
11	4'959	4'921	4'886	4'855	4'736	4'654	4'595	4'551	4'516	4'488	4'445	4'414	4'391	4'373	4'337	4'228
12	4'632	4'595	4'561	4'530	4'412	4'331	4'272	4'228	4'193	4'165	4'123	4'092	4'069	4'051	4'015	3'907
13	4'372	4'334	4'301	4'270	4'153	4'073	4'015	3'970	3'936	3'908	3'866	3'835	3'812	3'794	3'758	3'649
14	4'159	4'122	4'089	4'059	3'942	3'862	3'804	3'760	3'725	3'697	3'655	3'625	3'602	3'584	3'547	3'439
15	3'983	3'946	3'913	3'883	3'766	3'687	3'629	3'585	3'550	3'523	3'480	3'450	3'427	3'409	3'372	3'263
16	3'834	3'797	3'764	3'734	3'618	3'539	3'481	3'437	3'403	3'375	3'332	3'302	3'279	3'261	3'224	3'114
17	3'707	3'670	3'637	3'607	3'492	3'412	3'355	3'311	3'276	3'248	3'206	3'175	3'152	3'134	3'097	2'987
18	3'597	3'560	3'527	3'498	3'382	3'303	3'245	3'201	3'167	3'139	3'096	3'065	3'042	3'024	2'987	2'876
19	3'501	3'464	3'432	3'402	3'287	3'208	3'150	3'106	3'071	3'043	3'000	2'970	2'946	2'928	2'891	2'779
20	3'416	3'380	3'348	3'318	3'203	3'123	3'066	3'022	2'987	2'959	2'916	2'885	2'861	2'843	2'806	2'693
21	3'342	3'305	3'273	3'243	3'128	3'049	2'991	2'947	2'912	2'884	2'841	2'810	2'786	2'768	2'730	2'617
22	3'275	3'239	3'206	3'176	3'061	2'982	2'924	2'880	2'845	2'817	2'774	2'742	2'719	2'700	2'663	2'548
23	3'215	3'179	3'146	3'116	3'001	2'922	2'864	2'820	2'785	2'756	2'713	2'682	2'658	2'639	2'602	2'487
24	3'161	3'125	3'092	3'062	2'947	2'868	2'810	2'765	2'730	2'702	2'658	2'627	2'603	2'584	2'546	2'431
25	3'111	3'075	3'043	3'013	2'898	2'819	2'761	2'716	2'681	2'652	2'609	2'577	2'553	2'534	2'496	2'379
26	3'067	3'031	2'998	2'968	2'853	2'774	2'716	2'671	2'636	2'607	2'563	2'532	2'508	2'489	2'450	2'333
27	3'026	2'990	2'957	2'927	2'812	2'733	2'674	2'630	2'594	2'565	2'522	2'490	2'466	2'447	2'408	2'290
28	2'988	2'952	2'919	2'890	2'775	2'695	2'636	2'592	2'556	2'527	2'483	2'451	2'427	2'408	2'369	2'250
29	2'953	2'917	2'885	2'855	2'740	2'660	2'601	2'557	2'521	2'492	2'448	2'416	2'391	2'372	2'333	2'213
30	2'921	2'885	2'853	2'823	2'708	2'628	2'569	2'524	2'488	2'459	2'415	2'383	2'358	2'339	2'300	2'179
35	2'791	2'755	2'723	2'693	2'577	2'497	2'438	2'392	2'356	2'327	2'282	2'249	2'224	2'204	2'164	2'039
40	2'697	2'661	2'628	2'598	2'482	2'401	2'342	2'296	2'259	2'230	2'184	2'150	2'125	2'105	2'064	1'935
50	2'569	2'533	2'500	2'470	2'353	2'272	2'211	2'164	2'127	2'097	2'050	2'015	1'989	1'968	1'925	1'790
60	2'486	2'450	2'417	2'387	2'270	2'187	2'126	2'079	2'041	2'010	1'962	1'927	1'900	1'878	1'834	1'692
70	2'428	2'392	2'359	2'329	2'211	2'128	2'067	2'019	1'980	1'949	1'900	1'864	1'837	1'815	1'769	1'622
80	2'385	2'349	2'316	2'286	2'168	2'084	2'022	1'974	1'935	1'903	1'854	1'817	1'789	1'767	1'720	1'568
90	2'353	2'316	2'283	2'253	2'134	2'051	1'988	1'939	1'900	1'868	1'818	1'781	1'752	1'730	1'682	1'525
100	2'326	2'290	2'257	2'227	2'108	2'024	1'961	1'912	1'873	1'840	1'790	1'752	1'723	1'700	1'652	1'490
120	2'288	2'251	2'218	2'188	2'069	1'984	1'921	1'871	1'831	1'798	1'747	1'709	1'679	1'655	1'606	1'436
∞	2'105	2'069	2'035	2'004	1'882	1'794	1'727	1'674	1'631	1'595	1'538	1'494	1'460	1'431	1'370	1'076

Apéndice C

Bibliografía

Bibliografía

[1] G. Arnaiz. *Introducción a la Estadística Teórica.* Lex Nova, 1986.

[2] S. J. Baró Llina. *Estadística Descriptiva.* Parramón, 1985.

[3] J. Baró Llinás. *Cálculo de Probabilidades.* Parramón, 1987.

[4] G. Calot. *Curso de Estadística Descriptiva.* Paraninfo, 1970.

[5] F. Calvo. *Estadística Aplicada.* Deusto, 1989.

[6] G. C. Canavos. *Probabilidad y Estadística: Aplicaciones y Métodos.* McGraw Hill, 1992.

[7] E. Casa Aruta. *200 Problemas de Estadística Descriptiva.* Vicens Vives, 1979.

[8] H. Cramer. *Teoría de Probabilidades y Aplicaciones.* Aguilar, 1968.

[9] C. M. Cuadras. *Problemas de Probabilidades y Estadística, Vol. I: Probabilidades.* EUB, 1995.

[10] M. H. DeGrout. *Probabilidad y Estadística.* Addison–Wesley, 1968.

[11] A. I. Durand and S. L. Ipiña. *Introducción a la Teoría de la Probabilidad y la Inferencia Estadística.* Rueda, 1994.

[12] R. Escuder Vallés. *Estadística Económica y Empresarial.* Tebar Flores, 1982.

[13] W. Feller. *An Introduction to Probability Theory and its Applications.* Wiley, 1978.

[14] N. L. Johnson and S. Kotz. *Discrete Distributions.* Wiley, 1969.

[15] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions.* Wiley, 1970.

[16] J. G. Kalbfleisch. *Probabilidad e Inferencia Estadística.* AC, 1984.

[17] J. M. Keynes. *A Treatise on Probability.* Macmillan, 1921.

[18] P. S. Laplace. *Théorie Analytique des Probabilities.* Gauthier Villars, 1812.

[19] P. S. Laplace. *Ensayo Filosófico sobre las Probabilidades.* Alianza, 1985.

[20] J. Lóbez Urquía and E. Casa Aruta. *Estadística Intermedia.* Vicens Vives, 1975.

[21] M. Loeve. *Teoría de la Probabilidad.* Tecnos, 1976.

[22] A. Martín Andrés and J. D. Luna del Castillo. *Bioestadística para Las Ciencias de la Salud.* Norma, 1994.

[23] P. Martín-Guzmán, F. J. Martín Pliego, and otros. *Curso Básico de Estadística Económica.* AC, 1989.

[24] F. J. Martín Pliego. *Introducción a la Estadística Económica y Empresarial.* AC, 1994.

[25] F. J. Martín Pliego and L Ruiz-Maya. *Estadística I: Probabilidad.* AC, 1995.

[26] J. Montero, L. Pardo, D. Morales, and V. Quesada. *Ejercicios Y Problemas de Cálculo de Probabilidades.* Díaz de Santos, 1988.

[27] D. Montgomery. *Diseño y Análisis de Experimentos.* Grupo Editorial Iberoamérica, 1991.

[28] A. Mood and F. Graybill. *Introducción a la Teoría de la Estadística.* Aguilar, 1978.

[29] P. A. P. Morán. *An Introduction to Probability Theory.* Oxford University Press, 1968.

[30] E. Parzen. *Teoría Moderna de Probabilidades y sus Aplicaciones.* Limusa, 1979.

[31] D. Peña. *Estadística. Modelos y Métodos.* Alianza Universidad, 1991.

[32] D. Peña Sánchez de Rivera. *Estadística, Modelos y Métodos. Vol. I, Fundamentos.* AUT, 1992.

[33] R. Pérez Suárez, A. López, C. Caso, M. J. Río, M. Alvargonzález, N. Muñoz, and J. Baudilio García. *Análisis de Datos Económicos I, Métodos Descriptivos.* Pirámide, 1993.

[34] V. Quesada and A. García. *Curso Básico de Cálculo de Probabilidades.* ICE, 1985.

[35] V. Quesada and A. García. *Lecciones de Cálculo de Probabilidades.* Díaz de Santos, 1988.

[36] A. Rényi. *Cálculo de Probabilidades.* Reverté, 1976.

[37] S. Ríos. *Métodos Estadísticos.* Castillo, 1967.

[38] S. Ríos. *Iniciación Estadística.* ICE, 1977.

[39] V. K. Rohatgi. *An Introduction to Probability Theory and Mathematical Statistics.* Wiley, 1977.

[40] L. Ruiz Maya. *Problemas de Estadística.* AC, 1986.

[41] A. Sarabia and C. Maté. *Problemas de Probabilidad y Estadística.* Clagsa, 1993.

[42] M. R. Spiegel. *Estadística.* McGraw–Hill, 1970.

[43] H. G. Tucker. *Introducción a la Teoría Matemática de las Probabilidades y la Estadística.* Vicens Vives, 1972.

[44] E. Uriel and M. Muñiz. *Estadística Económica y Empresarial.* AC, 1988.

Glosario