

Estadística Aplicada a las Ciencias y la Ingeniería

Emilio L. Cano

2021-09-23

Índice general

Bienvenida	7
Estándares y software	7
Estructura del libro	8
Sobre el autor	10
Agradecimientos	10
 I Estadística descriptiva	 11
1. Introducción	13
1.1. Estadística y análisis de datos	13
1.2. Los dos grandes bloques de la Estadística	14
1.3. La esencia de la Estadística	15
1.4. Tipos de datos	16
1.5. La esencia de la Estadística	20
1.6. Organización de los datos - <i>Tidy data</i>	20
1.7. Ejemplo: datos bien organizados	21
1.8. Escalas	23
1.9. Conversión	23
1.10. 2. La Estadística y el método científico	23
1.11. Bioestadística	24
1.12. La Estadística aplicada a la Biología	24
1.13. 3. Plantear una hipótesis	25
1.14. 6. Comunicar resultados	27
1.15. 3. Estadística, Calidad y Sostenibilidad	27
1.16. Inspección por muestreo	31
1.17. Objetivos de Desarrollo Sostenible (ODS)	36
1.18. Bioestadística y sostenibilidad	36
 2. Análisis exploratorio univariante	 37
 3. Análisis exploratorio bivalente	 39

II Probabilidad	41
4. Introducción a la Probabilidad	43
5. Variable aleatoria univariante	45
6. Variable aleatoria bivariante	47
7. Modelos de distribución de probabilidad	49
III Inferencia estadística	51
8. Muestreo y estimación	53
9. Comparación de grupos	55
10. Modelos de regresión	57
11. Diseño de experimentos	59
IV Control estadístico de la calidad	61
12. Introducción	63
13. Control Estadístico de Procesos	65
14. Inspección por muestreo	67
A. Símbolos, abreviaturas y acrónimos	69
A.1. Acrónimos	69
A.2. Letras griegas	69
A.3. Símbolos	70
B. Tablas estadísticas	71
B.1. Distribución normal	71
B.2. Resumen modelos de distribución de probabilidad	73
C. Repaso	75
C.1. Logaritmos y exponenciales	75
C.2. Combinatoria	75
D. Ampliación	79
D.1. Función característica	79
D.2. Cambio de variable	79
D.3. Variables aleatorias unidimensionales mixtas	79
D.4. Variables aleatorias bidimensionales mixtas	79

<i>ÍNDICE GENERAL</i>	5
D.5. Algunos modelos de distribución continuos más	79
D.6. Modelos de distribución de probabilidad multivariantes	83
D.7. Modelos de distribución de probabilidad relacionadas con la normal	83
D.8. Simulación de variables aleatorias	83
E. Demostraciones	85
E.1. Variable aleatoria discreta	85
F. Créditos	87

Bienvenida

Este libro incluye los contenidos habitualmente presentes en el currículo de asignaturas de **Estadística** de los grados Ciencias e Ingenierías de universidades españolas. Aunque no aparezca en el título, el manual incluye también los contenidos de **Probabilidad** necesarios. Si bien existe abundante material bibliográfico que cubre los contenidos de estas asignaturas, quería elaborar un material propio que no fuera solamente para mis clases sino algo más *global*. En los últimos años ya lo hice para asignaturas de grado y Máster en ADE (López Cano, 2018, 2019). Por otra parte, me motiva cubrir el hueco de los materiales de acceso gratuito con la opción de comprar una edición impresa¹ y con el enfoque que se menciona en el siguiente apartado. Por otra parte, los libros publicados originalmente en inglés y traducidos al español a menudo me resultan lejanos a nuestro idioma (por muy buenas que sean las traducciones, los ejemplos en *acres* no son muy intuitivos para un lector español). Espero que también sirva para lectores de otros países de habla hispana.

Estándares y software

Los contenidos de este libro se basan en dos paradigmas que están presentes en los intereses de investigación y docencia del autor: los **estándares** y el **software libre**. En lo que se refiere a estándares, la notación utilizada, definiciones y fórmulas se ajustarán el máximo posible a la utilizada en normas nacionales e internacionales sobre metodología estadística. Estas normas se citarán pertinentemente a lo largo del texto. En cuanto al software libre, se proporcionarán instrucciones para resolver los ejemplos que ilustran la teoría utilizando software libre. No obstante, el uso del software es auxiliar al texto y se puede seguir sin necesidad de utilizar los programas. Según lo que proceda en cada caso, se utilizará software de hoja de cálculo, el software estadístico y lenguaje de programación **R** (R Core Team, 2021), y el software de álgebra computacional **Máxima**². Respecto al software de hoja de cálculo, las fórmulas utilizadas se han probado en el software libre **LibreOffice**³, en **Hojas de Cálculo de Goo-**

¹A la espera de encontrar editorial.

²<http://maxima.sourceforge.net/es/>

³<https://es.libreoffice.org>

gle⁴ y también en **Microsoft EXCEL**⁵ que, aunque no es software libre, su uso está más que generalizado y normalmente los estudiantes disponen de licencia de uso a través de su universidad. En caso de que el nombre de la función sea distinta en EXCEL, se indicará en el propio ejemplo.




Las normas son clave para el desarrollo económico de un país. Estudios en diversos países, incluido España, han demostrado que la aportación de la normalización a su economía es del 1 % del PIB⁶. La Asociación Española de Normalización (UNE) es el organismo legalmente responsable del desarrollo y difusión de las normas técnicas en España. Además, representa a España en los organismos internacionales de normalización como ISO⁷ y CEN⁸.

Las normas sobre estadística que surgen de ISO las elabora el *Technical Committee* ISO TC 69⁹ *Statistical Methods*. Por su parte, el subcomité técnico de normalización CTN 66/SC 3¹⁰, Métodos Estadísticos, participa como miembro nacional en ese comité ISO. Las normas que son de interés en España, se ratifican en inglés o se traducen al español como normas UNE. Para una descripción más completa de la elaboración de normas, véase Cano et al. (2015).

Estructura del libro

Este libro se ha elaborado utilizando el lenguaje *Markdown* con el propio software **R** y el paquete **bookdown** (Xie, 2021). Se incluyen una gran cantidad de ejemplos resueltos tanto de forma analítica como mediante software. En algunos casos se proporciona el uso de funciones en hojas de cálculo (y el resultado obtenido con un recuadro). En otros, código de R, que aparecen en el texto sombreados y con la sintaxis coloreada, como el fragmento a continuación donde se puede comprobar la sesión de R en la que ha sido generado este material. Obsérvese que los resultados se muestran precedidos de los símbolos `#>{r}` `sessionInfo()`

Normalmente, la descripción o enunciado de los ejemplos se incluyen en bloques con el siguiente aspecto:

Esto es un ejemplo. A continuación puede mostrarse código o no. Los ejemplos pueden ir precedidos por un icono para identificar su campo de aplicación, por ejemplo  Biología,  Ciencia y tecnología de Alimentos, o  Ciencia e Ingeniería Ambiental.

⁴<https://www.google.es/intl/es/sheets/about/>

⁵<https://products.office.com/es-es/excel>

⁶<http://www.aenor.es/DescargasWeb/normas/como-beneficia-es.pdf>


⁷<https://www.iso.org/>

⁸<https://www.cen.eu/>

⁹<https://www.iso.org/committee/49742/x/catalogue/>

¹⁰<https://www.une.org/encuentra-tu-norma/comites-tecnicos-de-normalizacion/comite/?c=CTN%2066/SC%203>


Cuando el ejemplo incluya explicaciones sobre cómo resolverlo con software, estas explicaciones aparecerán en bloques con el siguiente aspecto:




HOJA DE CÁLCULO
La función **FACT** obtiene el factorial de un número x ($x!$):
`=FACT(5)` 120

También se incluirán con el formato anterior indicaciones para usar la calculadora científica, cuando esto sea posible.

El texto incluye otros bloques con información de distinto tipo, como los siguientes:



Este contenido se considera avanzado. El lector principiante puede saltarse estos apartados y volver sobre ellos en una segunda lectura.



Estos bloques están pensados para incluir información curiosa o complementaria para poner en contexto las explicaciones.

Este volumen cubre los contenidos de asignaturas básicas de Estadística en un amplio rango de grados. Puede servir también como repaso para alumnos de posgrado o incluso egresados que necesiten refrescar conocimientos o aprender a aplicarlos con software moderno. Un segundo volumen cubrirá en el futuro métodos y modelos avanzados para entornos más exigentes.

El libro está dividido en 4 partes. La primera parte está dedicada a la Estadística Descriptiva, y consta de un capítulo introductorio seguido de sendos capítulos para el análisis exploratorio univariante y bivalente. La segunda parte trata la Probabilidad en 4 capítulos, uno introductorio, dos dedicados a las variables aleatorias univariantes y bivariantes respectivamente, y finalmente un capítulo que trata los modelos de distribución de probabilidad. En la tercera parte se aborda la inferencia estadística, con una introducción al muestreo y la estimación puntual, seguida de capítulos dedicados a los contrastes de comparación de grupos, análisis de regresión y diseño de experimentos. La última parte está dedicada al control estadístico de la calidad, en la que, tras un capítulo introductorio, se tratan las dos herramientas más importantes en este campo: el control estadístico de procesos (SPC, *Statistical Process Control*, por sus siglas en inglés) y los muestreos de aceptación o, dicho de otra forma, la inspección por muestreo. Finalmente, una serie de apéndices con diverso material complementan el libro en su conjunto.

Sobre el autor

Actualmente soy Profesor Ayudante Doctor en la Escuela Técnica Superior de Ingeniería Informática e investigador en el Data Science Laboratory de la Universidad Rey Juan Carlos. Mis intereses de investigación incluyen Estadística Aplicada, Aprendizaje Estadístico y Metodologías para la Calidad. Previamente he sido profesor e investigador en la Universidad de Castilla-La Mancha, donde sigo colaborando en docencia e investigación, y Estadístico en empresas del sector privado de diversos sectores.

Presidente del subcomité técnico de normalización UNE (miembro de ISO) CTN 66/SC 3 (Métodos Estadísticos). Profesor en la Asociación Española para la Calidad (AEC). Presidente de la asociación Comunidad R Hispano.

Más sobre mí, información actualizada y publicaciones: <http://emilio.lcano.com>.
Contacto: emilio@lcano.com

El material se proporciona bajo licencia CC-BY-NC-ND. Todos los logotipos y marcas comerciales que puedan aparecer en este texto son propiedad de sus respectivos dueños y se incluyen en este texto únicamente con fines formativos. Se ha puesto especial cuidado en la adecuada atribución del material no elaborado por el autor, véase el Apéndice F. Aún así, si detecta algún uso indebido de material protegido póngase en contacto con el autor y será retirado. Igualmente, contacte con el autor **si desea utilizar este material con fines comerciales**.



Este obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

Agradecimientos

Este libro es el resultado de años de trabajo en la docencia, investigación y transferencia de conocimiento en el campo de la Estadística. Está construido a partir de las contribuciones a lo largo de los años de compañeros y amigos como Javier M. Moguerza, Andrés Redchuk, David Ríos, Felipe Ortega, Mariano Prieto, Miguel Ángel Tarancón, Víctor M. Casero, Virgilio Gómez-Rubio, Matías Gámez, y muchos otros (perdón a l@s omitid@s por no ser más exhaustivo).

Especial agradecimiento a toda la comunidad del software libre y lenguaje de programación R, y en particular al *R Core Team* y al equipo de RStudio.

Parte I

Estadística descriptiva

Capítulo 1

Introducción

1.1. Estadística y análisis de datos

Antes de introducirnos en el estudio de la Estadística y sus métodos, vamos a intentar tener una visión de todo lo que abarca. Así pues, ¿qué es la Estadística? La primera fuente que podemos consultar es la definición de la Real Academia Española, y encontramos estas acepciones:

estadístico, ca

La forma f., del al. Statistik, y este der. del it. statista ‘hombre de Estado’.

1. adj. Perteneciente o relativo a la estadística.
2. m. y f. Especialista en estadística.
3. f. **Estudio de los datos** cuantitativos de la población, de los recursos naturales e industriales, del tráfico o de cualquier otra manifestación de las sociedades humanas.
4. f. Conjunto de **datos** estadísticos.
5. f. Rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener **inferencias** basadas en el **cálculo de probabilidades**.

RAE

Las acepciones que nos interesan son sobre todo la tercera y la cuarta, en las que aparecen conceptos que veremos en este capítulo introductorio y en los que profundizaremos en el resto del libro. La tercera acepción, “Conjunto de **datos** estadísticos”, es lo que muchas personas entienden cuando oyen la palabra Estadística: La estadística del paro, la estadística de los precios, etc. Pero la Estadística es mucho más amplia. En primer lugar, esos “datos estadísticos”

han tenido que ser recopilados y tratados de alguna forma antes de llegar a su publicación. Además, los datos estadísticos así entendidos son el resultado de un estudio pormenorizado (acepción 3) y normalmente de la aplicación de técnicas de **inferencia** (acepción 5). Algunas de estas técnicas forma parte de lo que vulgarmente se conoce como “la cocina” de las estadísticas.

Podemos hablar entonces de la Estadística, de forma muy resumida, como la ciencia de analizar datos. Encontramos a menudo¹ una definición de la Estadística como “la ciencia que establece los métodos necesarios para la recolección, organización, presentación y análisis de datos relativos a un conjunto de elementos o individuos”. Pero esta definición se centra solo en los métodos. Una definición más completa sería la siguiente:

[...] la estadística es la parte de la matemática que estudia la **variabilidad** y el proceso aleatorio que la genera siguiendo leyes de **probabilidad**.

Esta variabilidad puede ser debida al azar, o bien estar producida por causas ajenas a él, correspondiendo al **razonamiento estadístico** diferenciar entre la variabilidad casual y la variabilidad causal.

Ocaña-Riola (2017)

Aquí vemos uno de los conceptos clave que guiará todo el estudio y aplicación de la Estadística: la variabilidad es la clave de todo. Entender el concepto de variabilidad ayudará enormemente a entender los métodos por complejos que sean.

Variation is the reason for being of statistics

Cano et al. (2012)

La Estadística ha sido siempre importante en los estudios de Ciencias e Ingeniería. No obstante, en los últimos tiempos la alta disponibilidad tanto de datos como de tecnología para tratarlos, hace imprescindible un dominio de las técnicas estadísticas y su aplicación en el dominio específico.

1.2. Los dos grandes bloques de la Estadística

La Estadística se divide en dos grandes bloques de estudio, que son la **Estadística Descriptiva** y la **Inferencia Estadística**. A la Estadística Descriptiva también se la conoce como *Análisis Exploratorio de Datos* (EDA, *Exploratory Data Analysis*, por sus siglas en inglés). Esta disciplina tuvo un gran desarrollo gracias al trabajo de Tukey (Tukey et al., 1977), que todavía hoy es una referencia. Pero en los últimos años ha cobrado si cabe más importancia por la alta disponibilidad de datos y la necesidad de analizarlos.

¹Por ejemplo en el Curso de Estadística Práctica Aplicada a la Calidad de la AEC.

La **Estadística Descriptiva** se aplica sobre un conjunto de datos concretos, del que obtenemos resúmenes numéricos y visualización de datos a través de los gráficos apropiados. Con la Estadística Descriptiva se identifican **relaciones** y **patrones**, guiando el trabajo posterior de la Inferencia Estadística.

La **Estadística Inferencial** utiliza los datos y su análisis anterior para, a través de las Leyes de la **Probabilidad**, obtener conclusiones de diverso tipo, como explicación de fenómenos, confirmación de relaciones de causa-efecto, realizar predicciones o comparar grupos. En definitiva, tomar decisiones por medio de modelos estadísticos y basadas en los datos.

1.3. La esencia de la Estadística

La figura 1.1 representa la esencia de la Estadística y sus métodos. Estudiamos alguna **característica** observable en una serie de **elementos** (sujetos, individuos, ...) identificables y únicos. Los datos que analizamos, provienen de una determinada **población** que es objeto de estudio. Pero estos datos, no son más que una **muestra**, es decir, un subconjunto representativo de la población. Incluso cuando “creemos” que tenemos todos los datos, debemos tener presente que trabajamos con muestras, ya que generalmente tomaremos decisiones o llegaremos a conclusiones sobre el futuro, y esos datos seguro que no los tenemos. Por eso es importante considerar siempre este paradigma población-muestra, donde la población es desconocida y sus propiedades teóricas. La **Estadística Descriptiva** se ocupa del análisis exploratorio de datos en sentido amplio, que aplicaremos sobre los datos concretos de la muestra en esta unidad y la siguiente. La **Inferencia Estadística** hace referencia a los métodos mediante los cuales, a través de los datos de la muestra, tomaremos decisiones, explicaremos relaciones, o haremos predicciones sobre la población. Para ello, haremos uso de la **Probabilidad**, que veremos más adelante, aplicando el método más adecuado. En estos métodos será muy importante considerar el método de obtención de la muestra que, en términos generales, debe ser representativa de la población para que las conclusiones sean válidas.

🌀 En un ensayo clínico, se eligen una serie de participantes en el estudio a los que se le suministran distintos tratamientos según el diseño del ensayo. Los participantes en el estudio son sujetos que constituyen la **muestra**. A través de los resultados de esta muestra, obtendremos conclusiones para toda la **población**, que estará definida en el propio ensayo clínico. Por ejemplo, en el estudio del efecto de un determinado tratamiento para la diabetes, la población serían todos los enfermos de diabetes.



Otro concepto clave inherente a la Estadística, es que casi siempre estaremos investigando sobre esta fórmula:

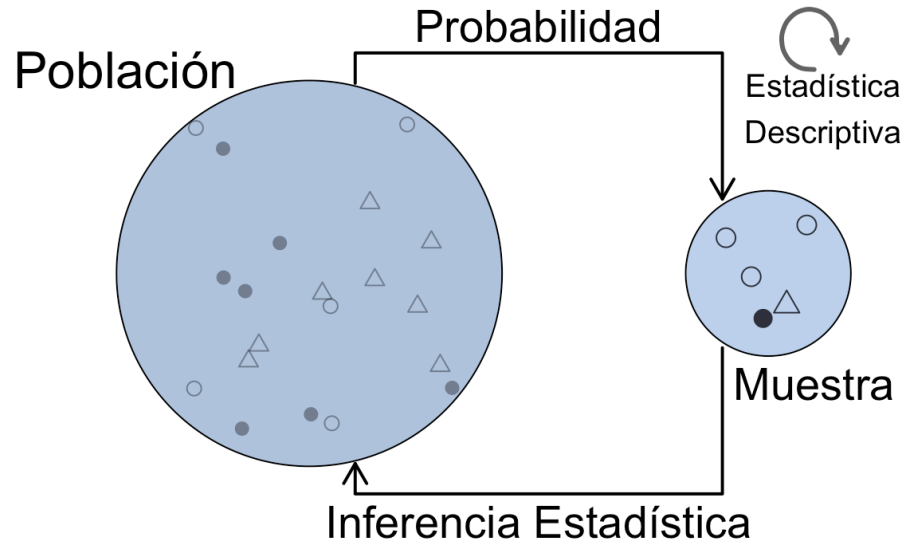


Figura 1.1: La esencia de los métodos estadísticos

$$Y = f(X)$$

Es decir, buscamos encontrar la relación entre una variable respuesta Y y una o varias variables explicativas X . Casi toda la Ciencia de Datos consiste en encontrar esa f . Es fundamental interiorizar este concepto para después aplicar el método adecuado, ya que según sean la/s Y , la/s X y el objetivo de nuestro estudio, los caminos pueden ser muy diferentes.

1.4. Tipos de datos

Las **características** que observamos en los **elementos** de la muestra (o que estudiamos en una población) pueden ser distintos tipos. Nos referiremos genéricamente a estas características como **variables**, aunque en algunos ámbitos como el Control Estadístico de Procesos (SPC, *Statistical Process Control* por sus siglas en inglés) este término se refiere solo a las variables continuas que ahora definiremos.

Denotaremos las variables con letras mayúsculas del alfabeto latino (X, Y, A, \dots). Cuando observamos la característica, la variable toma un **valor**. Estos valores pueden ser agrupados en **clases**, de forma que cada posible valor pertenezca a una y solo una clase. En ocasiones los datos con los que trabajamos están ya clasificados en clases.



Cuando se recogen datos utilizando cuestionarios, a menudo en las preguntas para recoger características cuantitativas se ofrece elegir un intervalo en vez de preguntar el **valor** exacto. Por ejemplo, al preguntar la edad de una persona, se pueden dar las opciones: 1) menos de 20 años; 2) entre 20 y 40 años; 3) entre 40 y 60 años; 4) Más de 60 años. Así, si una persona tiene 30 años, el **valor** de la variable es 30 (en el caso de la encuesta no lo conoceremos exactamente) que pertenece a la **clase** “entre 20 y 40 años”.

Datos univariantes, bivariantes, multivariantes

La importancia de la variabilidad

Calidad y Estadística

Inferencia Estadística y sus técnicas

Ciencia de Datos, Big Data y otras chuches

ODS: algo

- Pueden tomar cualquier valor en su **dominio** (conjunto de **posibles** valores que puede tomar la variable)

—
 .media.centrado[Siguen una distribución de **probabilidad**]

```
class:
lar-
ge
##
Pa-
rá-
me-
tros
y Es-
ta-
dísti-
cos
.pull-
left[
##
Pa-
rá-
me-
tros
```

* Se

defi-

nen

so-

bre

la

po-

bla-

ción

*

[Ca-

si]

siem-

pre

des-

co-

noci-

dos

*

Va-

lores

teó-

ricos

* So-

bre

los

que

hare-

mos

infe-

ren-

cia

* Le-

tras

grie-

gas

]

.pull-

right[

##

Es-

ta-

dísti-

cos

*
 Fun-
 ción
 defi-
 nida
 so-
 bre
 los
 da-
 tos
 de
 una
**mues-
 tra**
 (va-
 lores
 de
 una
 o
 más
 va-
 ria-
 bles)
 * En
 cada
 mues-
 tra
 se-
 rán
 dis-
 tin-
 tos
 (va-
 ria-
 bili-
 dad)

* Si-
guen
una
**dis-
tri-
bu-
ción**
en el
mues-
treo
* Le-
tras
lati-
nas
]

class: large background-image: url(/images/dogma2.png) background-position:
50 % 50 % background-size: 75 %

1.5. La esencia de la Estadística

class: large background-image: url(images/tidy_data.png) background-position:
50 % 85 % background-size: 60 %

1.6. Organización de los datos - *Tidy data*

- Datos rectangulares
- Una columna para cada variable (mismo tipo de datos)
- Una fila para cada observación (elemento, individuo)

—



```
.idea[
Analista y software deben entender lo mismo
]
```

```
class: large background-image: url(images/rectangular.png) background-
position: 95 % 50 % background-size: 45 %
```

1.7. Ejemplo: datos bien organizados

```
.pull-left[ - Separar la capa de datos de las capas de presentación y lógica -
Datos para humanos vs datos para máquinas - El análisis posterior se simplifica
si se preparan los datos para la máquina - Importancia de los .red[metadatos]
(diccionarios de datos)]
```

```
class:
lar-
ge
```

 Ti-
 pos
 de
 va-
 ria-
 bles
 *
 Cuan-
 titati-
 vas
 o
 Nu-
 mé-
 ricas
 +
 Con-
 ti-
 nuas
 +
 Dis-
 cre-
 tas
 *
 Cua-
 lita-
 titivas
 o
 Ca-
 tegó-
 ricas
 +
 Mul-
 tini-
 vel
 +
 Di-
 cotó-
 mi-
 cas
 +
 Or-
 dina-
 les

*
Mar-
cas
de
tiem-
po e
iden-
tifi-
ca-
do-
res

class: large

1.8. Escalas

- Nominal: atributos, factores, etiquetas
- Ordinal: atributos con un orden lógico
- Métrica: Permiten medir diferencias entre individuos

1.9. Conversión

- Fechas a categóricas (por ejemplo, mes, día de la semana, ...)
- Cualitativas a discretas (clases)
- Ordinales como numérica: cuidado, sobre todo si hay pocos datos (<100).
Mejor combinar en índices
- Variables calculadas con otras (por ejemplo, IMC)

class: large, inverse, toc, middle

1.10. 2. La Estadística y el método científico

class: large

1.11. Bioestadística

1.12. La Estadística aplicada a la Biología

- Cualquier análisis de datos, como cualquier disciplina.
- Énfasis en:
 - Diseños experimentales
 - Ensayos clínicos
 - Análisis genómico (vínculo con Bioinformática)

```

class:
lar-
ge
background-
image:
url(images/mc.jpg)
background-
position:
95 %
50 %
background-
size:
35 %

```

```

class:
lar-
ge
##
2.
In-
ves-
tiga-
ción
de
base
*

```

```

Aná-
lisis
ex-
plo-
rato-
rio
de
da-
tos

```

```

class:
lar-
ge
background-
image:
url(images/mc.jpg)
background-
position:
95 %
50 %
background-
size:
35 %

```

```

*

```

Iden-
tifi-
car
rela-
cio-
nes
*

Posi-
ble-
men-
te,
cam-
biar
la
pre-
gun-
ta
del
pri-
mer
paso

class: large

1.13. 3. Plantear una hipótesis

- Formalizarla en términos de Hipótesis nula, H_0 , e hipótesis alternativa, H_1
- El planteamiento de la hipótesis determina el método estadístico a utilizar, y el diseño del experimento (en sentido amplio)

class:
lar-
ge

class:
lar-
ge

5.

Ana-
lizar
re-
sul-
ta-
dos
y ex-
traer
con-
clu-
sio-
nes
*

Aná-
lisis
ex-
plo-
rato-
rio
*

Con-
tras-
tes
de
hi-
pó-
tesis
*

Vali-
da-
ción
de
los
mo-
de-
los

class: large

1.14. 6. Comunicar resultados

- Informes reproducibles (RMarkdown)
- Gráficos efectivos
- Resultados clave
- Resultados negativos

class: large, inverse, toc, middle

1.15. 3. Estadística, Calidad y Sostenibilidad

class:

lar-

ge

##

Con-

trol

Es-

ta-

dísti-

co

de

la

Cali-

dad

class:

lar-

ge

##

Cali-

dad

y va-

ria-

bili-

dad

class:
lar-
ge

Con-
trol
Es-
ta-
dísti-
co
de
la
Cali-
dad

.media[
>
**Ca-
li-
dad:**
Gra-
do
en el
que
un
con-
jun-
to
de
.red[características]
inhe-
ren-
tes
de
un
obje-
to >
cum-
ple
con
los
.red[requisitos]
> >
ISO
9001:2015
3.6.2

class:
lar-
ge

Con-
trol
Es-
ta-
dísti-
co
de
la
Cali-
dad

Los
re-
qui-
sitos
son
.red[**especificaciones**]
de
la
ca-
rac-
terís-
tica,
que
pue-
den
ser
bila-
tera-
les o
uni-
late-
ra-
les.]

class: large background-image: url(images/histos-1.png) background-position:
50 % 70 % background-size: 80 % ## La media y la variabilidad

—

Misma media, distinta capacidad de proceso

???

class: large background-image: url(images/taguchi-1.png) background-position:
50 % 90 % background-size: 50 % ## Función de pérdida de Taguchi

—

La Calidad se mide como la pérdida total que un producto causa a la sociedad

Genichi Taguchi

???

class:
lar-
ge
background-
image:
url(images/spc.png)
background-
position:
90 %
50 %
background-
size:
40 %

SPC:
Sta-
tisti-
cal
Pro-
cess
Con-
trol
.pull-
left[
*
Grá-
ficos
de-
con-
trol

*
 Aná-
 lisis
 de
 la
 ca-
 paci-
 dad
 del
 pro-
 ceso
 *
 Com-
 bina-
 dos
 con
 otras
 téc-
 ni-
 cas
 esta-
 dísti-
 cas
]

class: large background-image: url(images/qcpass.png) background-position:
 80 % 80 % background-size: 40 %

1.16. Inspección por muestreo

- AKA Muestreos de aceptación
- Aceptación: dentro de los límites de especificación
- Por atributos y por variables
- La base: probabilidad de aceptar/rechazar un lote defectuoso/correcto
- Muestreos por lotes
- Planes simples
- Planes dobles y múltiples
- Planes secuenciales
- MIL-STD -> ISO 2859

```

class:
lar-
ge
background-
image:
url(images/micro.jpg)
background-
position:
90 %
90 %
background-
size:
35 %
##
En-
sa-
yos
inter-
laboratorios

```

```

class:
lar-
ge
##
Me-
to-
dolo-
gías
y es-
tán-
da-
res
*
ISO
TC69
+
UNE
CT66/SC3

```


```
class:
lar-
ge
background-
image:
url(images/micro.jpg)
background-
position:
90 %
90 %
background-
size:
35 %
##
En-
sa-
yos
inter-
laboratorios
```

```
*
Me-
to-
dolo-
gía
Seis
Sig-
ma
y el
ciclo
DMAIC
*
Lean
Six
Sig-
ma
```

```
class:
lar-
ge
background-
image:
url(images/micro.jpg)
background-
position:
90 %
90 %
background-
size:
35 %
##
En-
sa-
yos
inter-
laboratorios
```

*
ISO
9000
+
UNE-
ISO
TR
1017
(Orien-
ta-
ción
so-
bre
las
téc-
ni-
cas
esta-
dísti-
cas
para
la
Nor-
ma
ISO
9001:2020)

```
class:
lar-
ge
background-
image:
url(images/micro.jpg)
background-
position:
90 %
90 %
background-
size:
35 %
##
En-
sa-
yos
inter-
laboratorios
```

```
.idea[ 
En
la
bi-
bli-
oteca
de
la
URJC
te-
néis
dis-
po-
ni-
ble
la
co-
lecc-
ción
de
nor-
mas
UNE]
```

class: huge

1.17. Objetivos de Desarrollo Sostenible (ODS)

- Iniciativa de la ONU: *Sustainable Development Goals* (SDG)
- 17 objetivos
- 169 metas

El 25 de septiembre de 2015, los líderes mundiales adoptaron un conjunto de **objetivos globales** para erradicar la pobreza, proteger el planeta y asegurar la prosperidad para todos como parte de una nueva agenda de desarrollo sostenible. Cada objetivo tiene **metas específicas** que deben alcanzarse en los próximos 15 años.

Naciones Unidas

class: large background-image: url(./images/nature.jpg) background-position: 90 % 90 % background-size: 25 %

1.18. Bioestadística y sostenibilidad

- Analizar datos relacionados con los ODS (Investigar)
- Ser sostenible en los análisis
- Relacionar con ODS e intentar contribuir sea cual sea el objetivo de la investigación
- ¿Cómo puede contribuir este trabajo/estudio/investigación/... a conseguir los Objetivos de Desarrollo Sostenible?

Capítulo 2

Análisis exploratorio univariante

Resúmenes numéricos

Resúmenes gráficos

Valores atípicos

Valores perdidos

En preparación.

Capítulo 3

Análisis exploratorio bivariante

Representación gráfica

Correlación

Regresión

Intro multivariante

En preparación.

Parte II

Probabilidad

Capítulo 4

Introducción a la Probabilidad

En preparación.

Definiciones

Propiedades

Probabilidad total y Bayes

Capítulo 5

Variable aleatoria univariante

En preparación.

Definición

Función de distribución

VA discreta

VA continua

Capítulo 6

Variable aleatoria bivalente

En preparación.

Distribución conjunta

Correlación y regresión

Capítulo 7

Modelos de distribución de probabilidad

En preparación.

Introducción

Modelos discretos

Modelos continuos

Modelos multivariantes*

Parte III

Inferencia estadística

Capítulo 8

Muestreo y estimación

En preparación.

Muestreo estadístico

Estimación y contrastes

Estadísticos

Estimadores puntuales (medias, proporciones, varianzas)

Estimación por intervalos

Estimación no paramétrica

Inferencia Bayesiana*

Capítulo 9

Comparación de grupos

En preparación.

Comparación de atributos

Comparación de dos grupos

Comparación de más de dos grupos

Capítulo 10

Modelos de regresión

En preparación.

Regresión lineal simple

Regresión no lineal

Regresión lineal múltiple

Otros modelos*
(GLM, GAM, ...)

Capítulo 11

Diseño de experimentos

En preparación.

Intro

Diseños factoriales

Diseños 2^k

Diseños fraccionales

Parte IV

Control estadístico de la calidad

Capítulo 12

Introducción

En preparación.

Historia de la calidad

Estadística y calidad

Gestión de la calidad

Mejora de procesos vs control de calidad

Metodologías

Intro Six Sigma*

Capítulo 13

Control Estadístico de Procesos

En preparación.

Intro SPC

Gráficos de control

Capacidad y rendimiento

Capítulo 14

Inspección por muestreo

En preparación.

Intro

Planes para atributos

Planes para variables

Apéndice A

Símbolos, abreviaturas y acrónimos

A.1. Acrónimos

Acrónimo	Descripción
SPC	Statistical Process Control

A.2. Letras griegas

Letra	Se lee
α	alfa
β	beta
γ	gamma
Γ	Gamma*
λ	lambda
η	eta
μ	mu
ω	omega
Ω	Omega*
σ	sigma
Σ	Sigma*
ρ	ro
θ	zeta (<i>theta</i> , teta)
ξ	xi
χ	chi (o <i>ji</i>)

Letra	Se lee
π	pi
ε	épsilon

* Mayúsculas

A.3. Símbolos

Símbolo	Se lee
\emptyset	Conjunto vacío o suceso imposible
\aleph	Aleph
\wp	Probabilidad (como función)
:	Tal que
$P(\cdot)$	Probabilidad de \cdot (sucesos)
$P[\cdot]$	Probabilidad de \cdot (variables aleatorias)
$E[\cdot]$	Esperanza de \cdot
\cdot	<i>lo que sea</i> (representa cualquier objeto matemático)
	Condicionado a
\sum	Sumatorio
$\sum_{i=1}^n$	Sumatorio desde i igual a uno hasta n
\prod	Producto
$\prod_{i=1}^n$	Producto desde i igual a uno hasta n
\forall	Para todo
\in	Pertenece/perteneciente
\exists	Existe
\Rightarrow	Implica/entonces
∂	Derivada parcial
\simeq	Aproximadamente igual ¹
\approx	Aproximadamente ²
\equiv	Equivalente
\mathbb{R}	Conjunto de los números reales
\cup	Unión
\cap	Intersección
\subset	Incluido
\subseteq	Incluido o igual

¹En este libro se usa sobre todo para indicar que se ha redondeado un número decimal

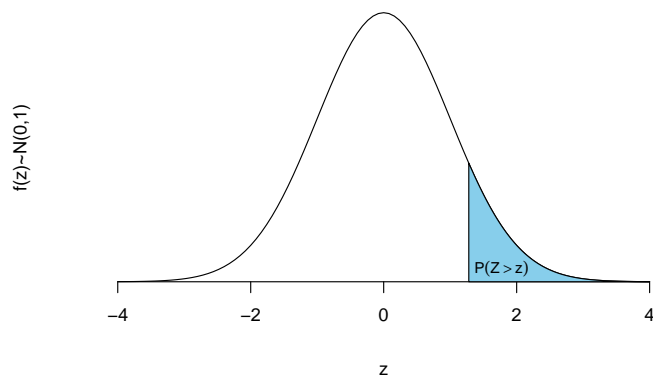
²En este libro se puede utilizar para tomar el entero superior o inferior según el contexto

Apéndice B

Tablas estadísticas

B.1. Distribución normal

La siguiente tabla contiene la probabilidad de la cola superior de la distribución normal estándar $Z \sim N(0; 1)$, es decir $1 - F(z) = P[Z > z]$.



[illegible]

B.2. Resumen modelos de distribución de probabilidad

Distribución	Probabilidad/Densidad/Distribución
$\text{Bernoulli} \mathrel{\mathop{\rule{0pt}{0.4pt}}{\rightarrow}} \mathit{Ber}(p)$	$X = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p \end{cases}$

Apéndice C

Repaso

Este apéndice cubre algunas cuestiones matemáticas básicas que el lector de este libro con seguridad habrá aprendido con anterioridad. Se incluyen como referencia para facilitar el repaso a aquellos que lo necesiten.

C.1. Logaritmos y exponenciales

C.2. Combinatoria

Una de las definiciones de probabilidad implica **contar** el número de veces que puede ocurrir un suceso determinado. Por tanto, en muchas ocasiones el cálculo de probabilidades empieza contando las posibilidades de que ocurra un suceso. La Combinatoria es la parte de la Matemática discreta que nos ayuda en esta tarea. Incluimos un breve resumen con ejemplos de las fórmulas más habituales y su cálculo con R.

C.2.1. Ejemplo ilustrativo

Habitualmente se utilizan ejemplos de juegos de azar para introducir el cálculo de probabilidades, como lanzando monedas y dados, o combinaciones de cartas en barajas de naipes. Para darle un enfoque práctico, utilizaremos a lo largo del módulo un ejemplo ilustrativo que, aunque totalmente inventado, se puede encontrar el lector en el futuro con ligeras variaciones según su ámbito de actuación. Utilizaremos en lo posible las cifras usadas en los problemas de azar para ver la utilidad de aquéllos ejemplos en casos más prácticos.

Datos básicos:

- 52 posibles usuarios de un servicio
- La mitad son mujeres

- 4 directivos, 12 mandos, resto operarios
- 13 jóvenes, 26 adultos, 13 mayores (5, 18 y 3 mujeres en cada grupo respectivamente)
- 1 de cada seis hombres contratará el servicio (el doble si es mujer)

Nótese cómo podemos *traducir* el concepto de servicio a cualquier ámbito: usuarios de salud o educación, enfermos de una determinada patología, equipos de una infraestructura, etc. Asimismo las categorías pueden ser cualesquiera aplicables a los elementos de los conjuntos.

C.2.2. Principio básico de conteo

Definición: Realizamos k experimentos sucesivamente, cada uno de ellos con n_i posibles resultados ($i = 1, \dots, k$). Entonces el número total de resultados posibles es:

$$n_1 \cdot n_2 \cdot \dots \cdot n_k$$

Ejemplo: Resultados posibles si tomamos al azar un individuo y observamos su grupo de edad y si contratará o no el servicio.

Código

```
3*2
#> [1] 6
```

C.2.3. Permutaciones

Definición: De cuántas formas posibles podemos ordenar un conjunto de n elementos sin repetirlos.

$$P_n = n! = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1$$

Ejemplo: De cuántas formas podemos ordenar un conjunto de tres individuos, uno de cada categoría laboral.

Código

```
factorial(3)
#> [1] 6
```

C.2.4. Variaciones (muestreo sin reemplazamiento)

Definición: De cuántas formas posibles podemos seleccionar una muestra de n elementos de un conjunto total de m , sin que se repitan. Una ordenación distinta, es una posibilidad distinta.

$$V_{m,n} = m \cdot (m-1) \cdot (m-2) \cdot \dots \cdot (m-n+1) = \frac{m!}{(m-n)!}$$

Ejemplo: De cuántas formas podemos seleccionar una muestra de 5 individuos en nuestro conjunto de 52 sin que se repitan (por ejemplo para asignar un ranking)

Código

```
factorial(52)/factorial(52-5)
#> [1] 311875200
```

C.2.5. Variaciones con repetición (muestreo con reemplazamiento)

Definición: De cuántas formas posibles podemos seleccionar una muestra de n elementos de un conjunto total de m , pudiéndose repetir. Una ordenación distinta, es una posibilidad distinta.

$$VR_{m,n} = m^n$$

Ejemplo: De cuántas formas podemos seleccionar una muestra de 5 individuos en nuestro conjunto de 52 pudiéndose repetir (por ejemplo para asignar premios consecutivamente)

Código

```
52^5
#> [1] 380204032
```

C.2.6. Combinaciones (muestras equivalentes)

Definición: De cuántas formas posibles podemos seleccionar una muestra de n elementos de un conjunto total de m , sin importar el orden.

$$C_{m,n} = \binom{m}{n} = \frac{m!}{n!(m-n)!}$$

$\binom{m}{n}$ se lee *m sobre n*, y se le conoce como *número combinatorio*. Algunas propiedades importantes de los números combinatorios:

$$\binom{m}{m} = \binom{m}{0} = 1.$$

$$\binom{m}{1} = \binom{m}{m-1} = m.$$

$$\binom{m}{n} + \binom{m}{n+1} = \binom{m+1}{n+1}$$

Por otra parte, por convenio se tiene que:

$$0! = 1,$$

$$\text{si } a < b \implies \binom{a}{b} = 0.$$

Ejemplo: De cuántas formas podemos seleccionar una muestra de 5 individuos en nuestro conjunto de 52 sin importar el orden (por ejemplo para asignar premios de una sola vez)

Código

```
choose(52, 5)
#> [1] 2598960
```

C.2.7. Combinaciones y permutaciones con repetición

Las combinaciones y permutaciones también se pueden dar con repetición, siendo las fórmulas para calcularlas las siguientes:

$$CR_{m,n} = C_{m+n-1,n} = \frac{(m+n-1)!}{n! \cdot (m-1)!}$$

$$PR = \frac{n!}{a! \cdot b! \cdot \dots \cdot z!}$$

La primera situación es aquella en la que los elementos se pueden repetir, pero no nos importa el orden en que lo hagan. La segunda aparece cuando el elemento A del conjunto total de elementos aparece a veces, y así sucesivamente.

Apéndice D

Ampliación

En este apéndice se incluyen temas avanzados que pueden ser útiles al lector más allá de un curso básico de estadística para ciencias o ingeniería, y que no se han incluido en el cuerpo de los capítulos para mantener el nivel de una asignatura de grado.

D.1. Función característica

D.2. Cambio de variable

D.3. Variables aleatorias unidimensionales mixtas

D.4. Variables aleatorias bidimensionales mixtas

D.5. Algunos modelos de distribución continuos más

D.5.1. Distribución Beta

La distribución Beta se utiliza en problemas de inferencia relativos a proporciones, especialmente en inferencia bayesiana.

$$X \sim Be(\alpha, \beta)$$

Función de densidad

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{si } 0 < x < 1 \\ 0 & \text{resto} \end{cases}$$

En matemáticas, la función Gamma (Γ) es una integral indefinida que tiene entre otras las siguientes propiedades:

- $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$
- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$
- $n \in \mathbb{N} - \{0\} \implies \Gamma(n) = (n-1)!$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

**** Características****

- Esperanza: $E[X] = \frac{\alpha}{\alpha+\beta}$
- Varianza: $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Caso particular: $Be(1, 1) = U(0, 1)$.

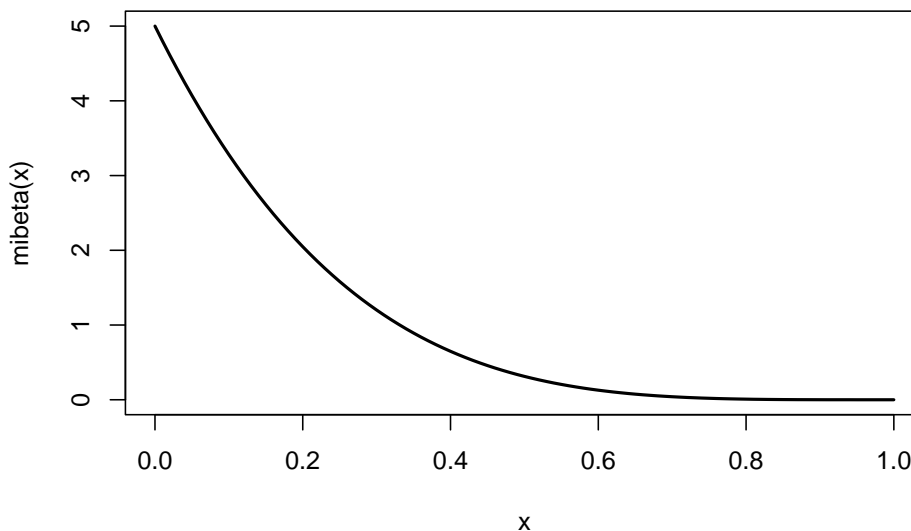
Ejemplo

X : Proporción de clientes que contratarán el servicio

$$X \sim Be(1, 5)$$

Código

```
mibeta <- function(x) dbeta(x, 1, 5)
curve(mibeta, lwd = 2)
```



D.5.2. Distribución Gamma

La distribución Gamma se utiliza, entre otros, para modelizar tiempos de espera hasta que suceden α eventos en un proceso de Poisson. De hecho, en inferencia bayesiana gamma es la distribución a priori de la distribución de Poisson.

$$X \sim Ga(a, b)$$

Función de densidad

$$f(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & \text{si } 0 < x < \infty \\ 0 & \text{resto} \end{cases}$$

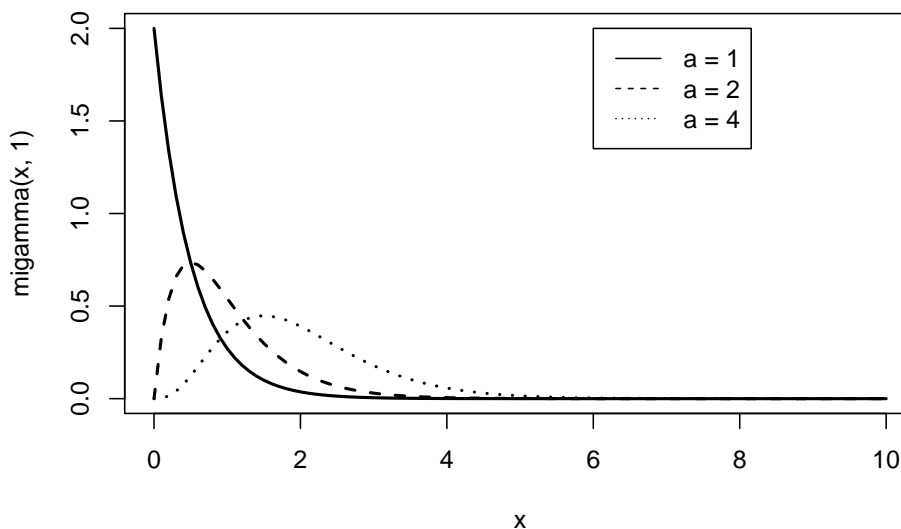
Características

- Esperanza: $E[X] = \frac{a}{b}$
- Varianza: $Var[X] = \frac{a}{b^2}$
- $\int_0^\infty x^{a-1} e^{-bx} dx = \frac{\Gamma(a)}{b^a}$
- La exponencial es un caso particular

Código

```
migamma <- function(x, a) dgamma(x, a, 2)
curve(migamma(x, 1), lwd = 2, xlim = c(0,10),
      main = "Distribución Gamma b = 2")
curve(migamma(x, 2), lwd = 2, add = TRUE, lty = 2)
curve(migamma(x, 4), lwd = 2, add = TRUE, lty = 3)
legend(x = 6, y = 2, c("a = 1", "a = 2", "a = 4"), lty = 1:3)
```

Distribución Gamma b = 2



D.5.3. Distribución de Weibull

La distribución Gamma presenta algunos inconvenientes al modelizar tiempos de vida, y por eso algunas veces se prefiere la distribución de Weibull, que básicamente sirve para lo mismo. Véase Ugarte et al. (2015) para los detalles.

$$X \sim We(a, b)$$

Función de densidad

$$f(x) = \begin{cases} \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} e^{-(x/b)^a} & \text{si } x > 0 \\ 0 & \text{resto} \end{cases}$$

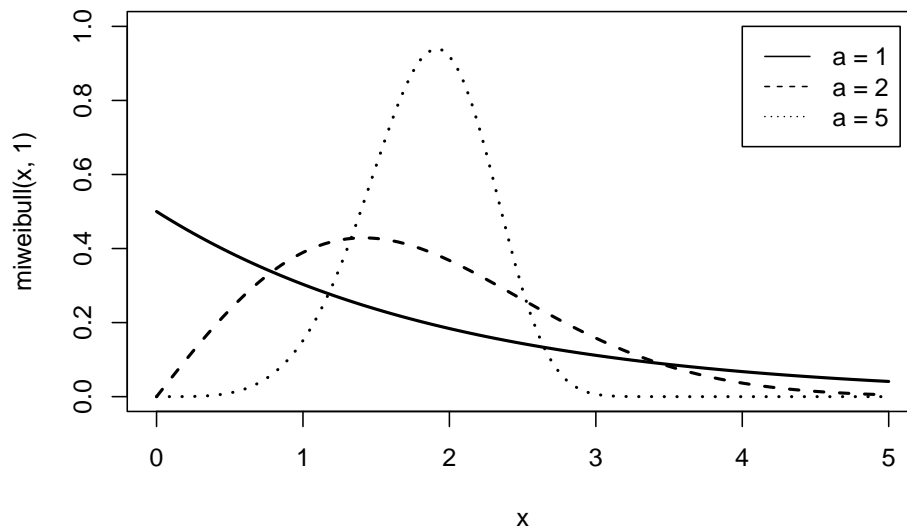
Características

- Esperanza: $E[X] = b\Gamma\left(1 + \frac{1}{a}\right)$
- Varianza: $Var[X] = b^2 \left(\Gamma\left(1 + \frac{2}{a}\right) - \left(\Gamma\left(1 + \frac{1}{a}\right)\right)^2 \right)$

Código

```
miweibull <- function(x, a) dweibull(x, a, 2)
curve(miweibull(x, 1), lwd = 2, xlim = c(0,5),
      ylim = c(0, 1),
      main = "Distribución Weibull b = 2")
curve(miweibull(x, 2), lwd = 2, add = TRUE, lty = 2)
curve(miweibull(x, 5), lwd = 2, add = TRUE, lty = 3)
legend(x = 4, y = 1, c("a = 1", "a = 2", "a = 5"), lty = 1:3)
```

Distribución Weibull $b = 2$



D.6. Modelos de distribución de probabilidad multivariantes

D.7. Modelos de distribución de probabilidad relacionadas con la normal

D.8. Simulación de variables aleatorias

$U(0; 1)$: Generador de probabilidades aleatorias. Dada cualquier función de distribución F , se pueden generar valores de esa VA obteniendo $F^{-1}(U(0; 1))$

Apéndice E

Demostraciones

Em este apéndice se incluyen aquellas demostraciones de teoremas y propiedades no incluidas en los capítulos para mantener el carácter práctico del mismo.

E.1. Variable aleatoria discreta

E.1.1. Función de probabilidad

E.1.2. Esperanza

E.1.3. Varianza

Apéndice F

Créditos

Los gráficos y diagramas generados son creación y propiedad del autor, salvo que se indique lo contrario. Su licencia de uso es la misma que la del resto de la obra, véase el Prefacio.

La imagen de la portada es de dominio público, obtenida en pixabay.com, gracias al usuario Manuchi.

Las imágenes de tipo *clipart* usadas en esta obra y las fotografías no atribuidas pertenecen al dominio público gracias a openclipart.org, unplash.com o pixabay.com.

The R logo is (c) 2016 The R Foundation.

Bibliografía

- Cano, E. L., Moguerza, J. M., and Corcoba, M. P. (2015). *Quality Control with R. An ISO Standards Approach*. Use R! Springer.
- Cano, E. L., Moguerza, J. M., and Redchuk, A. (2012). *Six Sigma with R. Statistical Engineering for Process Improvement*, volume 36 of *Use R!* Springer, New York.
- López Cano, E. (2018). Estadística económica y empresarial. Libro de apuntes con licencia Creative Commons.
- López Cano, E. (2019). Análisis de datos con r aplicado a la economía, la empresa y la industria. Libro de apuntes con licencia Creative Commons.
- Ocaña-Riola, R. (2017). La necesidad de convertir la estadística en profesión regulada. *Estadística Española*, 59(194):193–212.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Tukey, J. W. et al. (1977). *Exploratory data analysis*, volume 2. Reading, Mass.
- Ugarte, M., Militino, A., and Arnholt, A. (2015). *Probability and Statistics with R, Second Edition*. CRC Press.
- Xie, Y. (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.24.