

Estadística Aplicada a las Ciencias y la Ingeniería

Emilio L. Cano

2021-08-31

Índice general

Prefacio	5
Estándares y software	5
Sobre el autor	7
Agradecimientos	8
A. Símbolos	9
A.1. Letras griegas	9
A.2. Símbolos	9
B. Tablas estadísticas	11
B.1. Distribución normal	11
B.2. Resumen modelos de distribución de probabilidad	13
C. Repaso	15
C.1. Logaritmos y exponenciales	15
C.2. Combinatoria	15
D. Ampliación	19
D.1. Función característica	20
D.2. Cambio de variable	20
D.3. Variables aleatorias unidimensionales mixtas	20
D.4. Variables aleatorias bidimensionales mixtas	20
D.5. Algunos modelos de distribución continuos más	20
D.6. Modelos de distribución de probabilidad multivariantes	20
D.7. Modelos de distribución de probabilidad relacionadas con la normal	20
D.8. Simulación de variables aleatorias	20
E. Demostraciones	21
E.1. Variable aleatoria discreta	21
F. Créditos	23

Prefacio

Este libro incluye los contenidos habitualmente presentes en el currículo de asignaturas de **Estadística** de los grados Ciencias e Ingenierías de universidades españolas. 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*. 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 Google**⁴ y también en **Microsoft EXCEL**⁵ que, aunque no es software libre, su uso

¹A la espera de encontrar editorial.

²<http://maxima.sourceforge.net/es/>

³<https://es.libreoffice.org>

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

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

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).

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 `#>`.

```
sessionInfo()
#> R version 4.1.1 (2021-08-10)
#> Platform: x86_64-apple-darwin17.0 (64-bit)
#> Running under: macOS Big Sur 10.16
#>
#> Matrix products: default
#> BLAS: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
#>
#> locale:
#> [1] es_ES.UTF-8/es_ES.UTF-8/es_ES.UTF-8/C/es_ES.UTF-8/es_ES.UTF-8
#>
#> attached base packages:
#> [1] stats graphics grDevices utils datasets
#> [6] methods base
```

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>

```
#>
#> loaded via a namespace (and not attached):
#> [1] compiler_4.1.1 magrittr_2.0.1 fastmap_1.1.0
#> [4] bookdown_0.23.4 htmltools_0.5.2 tools_4.1.1
#> [7] yaml_2.2.1      stringi_1.7.4    rmarkdown_2.10
#> [10] knitr_1.33      stringr_1.4.0    digest_0.6.27
#> [13] xfun_0.25       rlang_0.4.11     evaluate_0.14
```

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.

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.

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.

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. Sus intereses de investigación incluyen Estadística Aplicada, Aprendizaje Estadístico y Metodologías para la Calidad. afiliaciones. Previamente ha sido profesor en la Universidad de Castilla-La Mancha, donde sigue 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: <https://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, Felipe Ortega, Mariano Prieto, Miguel Ángel Tarancón, Víctor M. Casero, 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 especial al R Core Team y al equipo de RStudio.

Apéndice A

Símbolos

A.1. Letras griegas

Letra	Se lee
α	alfa
λ	lambda
η	eta
μ	mu
ω	omega
Ω	Omega*
σ	sigma
Σ	Sigma*
ρ	ro
θ	teta
ξ	xi

* Mayúsculas

A.2. 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)

Símbolo	Se lee
$E[\cdot]$	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

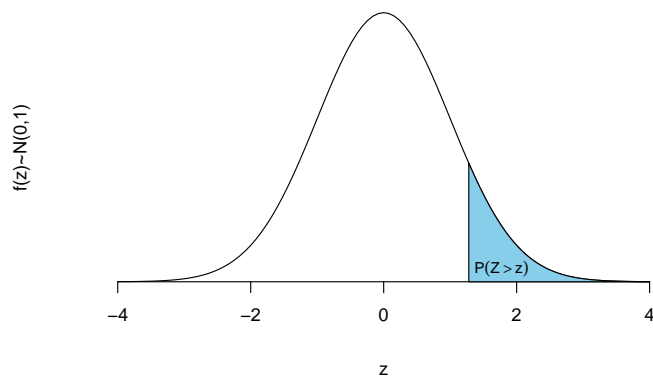
²En este libro se puede utilizar para tomar el entero superior o inferior según el contexto

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
$\text{Bernoulli}(\mathit{Ber}(p))$	$X = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p \end{cases}$

Apéndice C

Repaso

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:

$$\begin{aligned} \binom{m}{m} &= \binom{m}{0} = 1. \\ \binom{m}{1} &= \binom{m}{m-1} = m. \\ \binom{m}{n} + \binom{m}{n+1} &= \binom{m+1}{n+1} \end{aligned}$$

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 económica y empresarial, y que no se han incluido en el cuerpo de los capítulos para mantener el nivel de un primer curso 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.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.

- 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.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Xie, Y. (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. <https://github.com/rstudio/bookdown>, <https://pkgs.rstudio.com/bookdown/>.