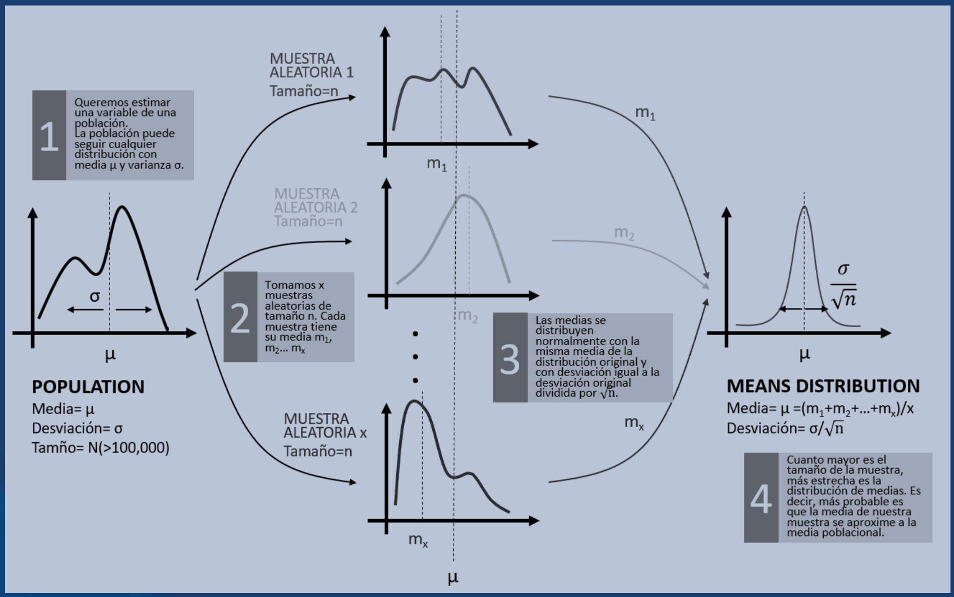
DISTRIBUCIONES MUESTRALES



DE LA MEDIA Y LA PROPORCIÓN

NOMBRE DEL AUTOR
MSC. PABLO VARGAS PEREIRA

FECHA
DICIEMBRE, 2020

Contenido

PREGUNTA DISPARADORA

La probabilidad es un concepto que la mayor parte de las personas comprende intuitivamente, por ejemplo, casi todas las personas saben que la probabilidad de ganar o perder una apuesta con el lanzamiento de una moneda es de 50% para cada una de las opciones, entonces, ¿Existe algún método probabilistico para determinar el resultado de un evento aleatorio?, por tanto, en base a la pregunta determinada podrá el o la estudiante determinar teóricamente y desde el punto de vista práctico los procesos sistemáticos de la estadística inferencial.

ABSTRACT O RESUMEN

Como se comentó en el módulo pasado, una población es el conjunto de todos los elementos o unidades que son de interés para un estudio determinado y una muestra es un subconjunto de los elementos de la población, la principal característica que debe tener una muestra estadística es que sea representativa de la población de la cual se extrae, porque el principal propósito de la obtención de muestras consiste en hacer inferencias sobre la población correspondiente, la técnica de muestreo se utiliza porque resulta más fácil, más rápido y económico en un estudio determinado. En muchas ocasiones esta técnica es mayormente sencilla y económica, esto da como resultado conclusiones que son igualmente útiles que las que se obtendrían de estudiarse la población completa, en este módulo se revisan diversos conceptos importantes relacionados con el muestreo y se revisa también con detalle el tema de las distribuciones muestrales, en especial dos de ellas, la primera es la distribución muestreal de la media y seguidamente la distribución muestral de la proporción, estas dos distribuciones son la base de donde se desprenden los mecanismos con que se hacen estimaciones de parámetros y pruebas de hipótesis, estos últimos dos conceptos se abordarán en el módulo tres del presente curso.

PALABRAS CLAVES

Muestreo, probabilidad, muestra, distribución de probabilidades, distribución normal, media, varianza, desviación estándar, error estándar, etapas de un estudio científico, proporción, pruebas de hipótesis, parámetros, estadísticos, estimación de parámetros, distribución muestral.

INTRODUCCIÓN

En el presente folleto de trabajo se desarrollará el proceso de las dos distribuciones muestrales más importantes en el campo de la estadística inferencial, la primera de ellas es la distribución muestral de la media, en este tema se abordará el desarrollo de dicha distribución, sus conclusiones en base al desprendimiento del teorema central del límite el error estándar de la media y su factor de corrección por población finita y las aplicaciones de análisis de esta distribución, como segundo tema se desarrollará la distribución muestral de la proporción, las conclusiones más importantes de las proporciones, la fórmula de error estándar y su corrección por población finita y sus consideraciones generales.

CONTENIDO

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Distribución muestral de la media: conjunto de las medias de todas las muestras de tamaño *n* que es posible obtener de una población de tamaño *N*.

Aunque el análisis que sigue de las distribuciones muestrales es principalmente teórico, es de vital importancia su interés al desarrollo que se plantea, ya que, las conclusiones que de él se derivan con el fundamento de las dos técnicas básicas de la inferencia estadística: la estimación de parámetros y las pruebas de hipótesis.
Tal como se mencionó en la sección anterior, la distribución muestral de la media es el conjunto de las medias de todas las muestras de tamaño *n* que es posible obtener de una población de tamaño *N*. En esta sección se revisa el comportamiento de las distribuciones muestrales de la media, como ya se explicó de este comportamiento se desprenden los procedimientos para aplicar los dos principales mecanismos de la inferencia estadística, la estimación de parámetros y las pruebas de hipótesis sobre medias, que a su vez representan una porción importante de las aplicaciones que se revisarán en buena parte de los capítulos restantes de este curso de estadística inferencial.

I. DESARROLLO

Antes de pasar a un ejemplo para ilustrar esta distribución muestral y considerando que se trata de un ejemplo un tanto extenso conviene visualizar antes lo que se realizará, a continuación, el procedimiento de trabajo para este tipo de distribución:

- Se supondrá una población con N elementos.
- Se calcularán la media aritmética y la desviación estándar de la población.
- Se determinará el número total de muestras distintas de tamaño n que es posible extraer y se concluirá cuáles son; es decir, se enumerará la distribución muestral (el conjunto de todas las muestras de ese tamaño que es posible obtener de esa población).
- Se resolverá la media de cada una de las muestras identificadas en el punto anterior; con esto ya se tiene la distribución muestral de la media (el conjunto de las medias de todas las muestras de ese tamaño que es posible obtener de la población).
- Se calcularán la media y la desviación estándar de la distribución muestral de la media.
- Se obtendrán las conclusiones importantes para el muestreo al comparar los parámetros poblacionales con los valores obtenidos para la distribución muestral de la media.
- Se observará también la forma de la distribución muestral de la media para obtener otra importante conclusión: la que permite utilizar la distribución normal para hacer los juicios sobre la probabilidad de estar en lo correcto cuando se hacen inferencias.

El ejemplo que se presenta en seguida supone una población con sólo 5 elementos (N = 5) y una muestra de tamaño (n = 2), esto es una situación extremadamente simplificada que tiene como propósito facilitar la explicación de los conceptos importantes del tema, estos conceptos son igualmente aplicables a poblaciones muy numerosas y muestras aún mayores que se manejan en la práctica y cotidiano dentro del entorno de las investigaciones.

EJEMPLO 1

Suponga que se tiene una población de 5 familias (N = 5) y la variable que se estudia es el número de hijos de cada familia. Los datos correspondientes aparecen en las 2 primeras columnas de la tabla 1.

Procedimiento 1:
Se supondrá una población con N elementos.

Familia	Hijos x	$X-\mu$	$(X-\mu)^2$
Pérez	2	-4	16
Gómez	4	-2	4
Durán	6	0	0
Hidalgo	8	2	4
Juárez	10	4	16
Total	**30**	**0**	**40**

Tabla 1. *Tabla de distribución y variabilidad del número de hijos por familia con N = 5*

Procedimiento 2:
Se calcularán la media aritmética y la desviación estándar de la población.

La media poblacional:

$$\mu = \frac{\sum x}{N} = \frac{30}{5} = 6$$

El promedio de hijos que poseen las cinco familias en estudio es de 6 hijos.

La desviación estándar poblacional:

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} = \sqrt{\frac{40}{5}} = 2{,}8284271$$

El valor de la desviación estándar de la población corresponde a 2,83 aproximadamente.

Procedimiento 3:
Se determinará el número total de muestras distintas de tamaño n.

$$C_n^N = \frac{N!}{n! \cdot (N-n)!}$$

Para determiner el tamaño de las muestras se trabajará con la formula de combinación, esta formula se desarrollo en el módulo anterior.

$$C_n^N = \frac{N!}{n! \cdot (N-n)!} = \frac{5!}{2!\,(5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1(3 \cdot 2 \cdot 1)} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

El número total de muestras de tamaño dos que es posible obtener de una población con 5 elementos es de 10, en la tabla 2 aparece el listado de todas estas muestras, junto con sus correspondientes valores para la variable de número de hijos la cual constituye la distribución muestral para n = 2.

Familia	**Hijos** (x)
Pérez, Gómez	2 , 4
Pérez, Durán	2 , 6
Pérez, Hidalgo	2 , 8
Pérez, Juárez	2 , 10
Gómez, Durán	4 , 6
Gómez, Hidalgo	4 , 8
Gómez, Juárez	4 , 10
Durán, Hidalgo	6 , 8
Durán, Juárez	6 , 10
Hidalgo, Juárez	8 , 10

Tabla 2. *Distribución muestral para n = 2.*

Procedimiento 4:
Se resolverá la media de cada una de las muestras identificadas en el punto anterior; con esto ya se tiene la distribución muestral de la media.

Familia	Hijos (x)	$\bar{X}$
Pérez, Gómez	2 , 4	3
Pérez, Durán	2 , 6	4
Pérez, Hidalgo	2 , 8	5
Pérez, Juárez	2 , 10	6
Gómez, Durán	4 , 6	5
Gómez, Hidalgo	4 , 8	6
Gómez, Juárez	4 , 10	7
Durán, Hidalgo	6 , 8	7
Durán, Juárez	6 , 10	8
Hidalgo, Juárez	8 , 10	9
Total		**60**

Tabla 3. *Distribución muestral de la media.*

Procedimiento 5:
Se calcularán la media y la desviación estándar de la distribución muestral de la media.

Procedimiento 6:
Se obtendrán las conclusiones importantes para el muestreo al comparar los parámetros poblacionales con los valores obtenidos para la distribución muestral de la media.

La siguiente fórmula corresponde a la media muestral

$$\mu_{\bar{x}} = \frac{\sum \bar{X}}{N_1}$$

Donde:

$\mu_{\bar{x}}$: Media muestral

$\sum \bar{X}$: Sumatoria de las medias de cada muestra

N_1: Cantidad de muestras producto de la combinación de los elementos de la variable

$$\mu_{\bar{x}} = \frac{\sum \bar{X}}{N_1} = \frac{60}{10} = 6$$

Conclusión:
La media de la distribución muestral de las medias es igual a la media de la población, en

otras palabras, el valor esperado de la media es igual a la media de la población, esto mismo expresado en símbolos corresponde a:

$$\frac{\sum x}{N} = \frac{\sum \bar{X}}{N_1} \Leftrightarrow \mu = \mu_{\bar{x}}$$

La siguiente fórmula corresponde a la desviación estándar de todas las muestras.

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{X} - \mu_{\bar{x}})^2}{N_1}}$$

Familia	**Hijos** (x)	$\bar{X}$	$\bar{X} - \mu_{\bar{x}}$	$(\bar{X} - \mu_{\bar{x}})^2$
Pérez, Gómez	2 , 4	3	-3	9
Pérez, Durán	2 , 6	4	-2	4
Pérez, Hidalgo	2 , 8	5	-1	1
Pérez, Juárez	2 , 10	6	0	0
Gómez, Durán	4 , 6	5	-1	1
Gómez, Hidalgo	4 , 8	6	0	0
Gómez, Juárez	4 , 10	7	1	1
Durán, Hidalgo	6 , 8	7	1	1
Durán, Juárez	6 , 10	8	2	4
Hidalgo, Juárez	8 , 10	9	3	9
Total		**60**	**0**	**30**

Tabla 4. *Tabla de dispersion de la distribución muestral de la media.*

$$\sigma_{\bar{x}} = \sqrt{\frac{30}{10}}$$

$$\sigma_{\bar{x}} = \sqrt{3}$$

$$\sigma_{\bar{x}} = 1{,}73$$

El valor de la desviación muestral corresponde a 1,73 aproximadamente.

A este valor se le conoce como error estándar de la media y tiene una relación con la desviación estándar de la población que se expresa mediante la siguiente ecuación:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{2{,}83}{\sqrt{2}} \sqrt{\frac{5-2}{5-1}}$$

$$\sigma_{\bar{x}} = 1{,}73$$

Entonces, se concluye que ambas fórmulas demuestran el mismo resultado, por tanto, para determinar el error estándar de la media se puede hacer uso cualquiera de las dos fórmulas.

$$\sqrt{\frac{\sum(\bar{X}-\mu_{\bar{x}})^2}{N_1}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Procedimiento 7

Construcción de la tabla de frecuencias y análisis de la forma de la distribución muestral de la media.

En la tabla 5 se presenta la distribución muestral de las medias agrupadas según las frecuencias observadas.

Medias muestrales	Frecuencia observada
3	1
4	1
5	2
6	2
7	2
8	1
9	1
Total	**10**

Tabla 5. *Frecuencias observadas de la distribución muestral de la media*

En la figura 1 se representa la gráfica de los datos de la distribución de frecuencias de la tabla 5, donde se representa la cantidad de veces que se repite cada media muestral.

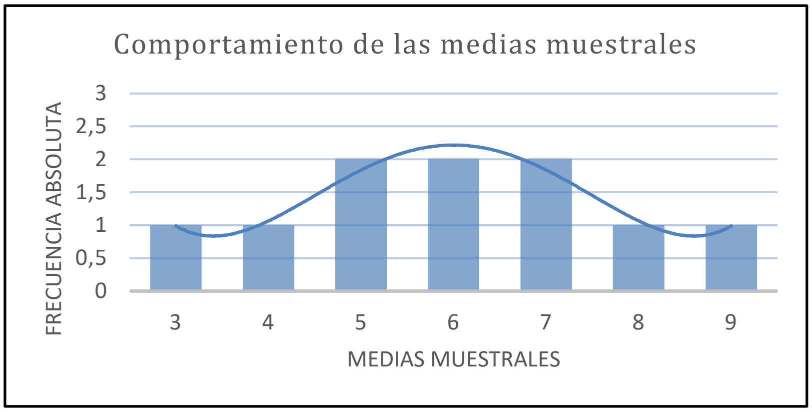


Figura 1. *Gráfica de barras del comportamiento de las frecuencias observadas de la distribución muestral de la media*

En la gráfica anterior se aprecia que las medias de las muestras tienden a agruparse alrededor (cerca) del valor de la media de la población, lo cual es una característica de la distribución normal, aunque en este ejemplo el tamaño de la muestra es exageradamente pequeño, aun así se nota que la distribución de las medias muestrales tiende a ser de forma normal. La afirmación anterior, expresada en el siguiente párrafo de manera más formal, es la última conclusión a la que se quería llegar y es lo que, junto con las dos primeras conclusiones revisadas antes, se conoce como el teorema del límite central.

Si X es una variable aleatoria para la que se conocen su media (μ) y su varianza (σ^2), la distribución muestral de la media tiende a ser normal con su media muestral ($\mu_{\bar{x}}$) y error estándar $\left(\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$ conforme aumenta el tamaño de la muestra, es importante observar que el teorema del límite central no menciona la forma de la distribución de la población en la que se analiza la variable, esto es de gran importancia porque, en términos prácticos significa que la distribución muestral de la media tiende a ser normal aun cuando la variable no se distribuya de manera normal en la población, siempre y cuando la muestra sea de tamaño grande (que tienda a infinito), en la práctica se considera que una muestra es grande cuando es mayor o igual a 30 ($n \geq 30$); como se verá más adelante el tamaño de la muestra junto con otras consideraciones tiene efecto sobre el procedimiento específico que se debe seguir para estimar parámetros o para llevar a cabo pruebas de hipótesis.

A manera de recapitulación y dada su importancia se reproducen en seguida tres conclusiones.

II. CONCLUSIONES QUE DESPRENDE DE LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA: EL TEOREMA DEL LÍMITE CENTRAL

1. La media de la distribución muestral de las medias es igual a la media de la población o, dicho en otras palabras, el valor esperado de la media es igual a la media de la población.
2. Existe una relación entre la desviación estándar de la población y la desviación estándar de la distribución muestral de las medias a la que se conoce como error estándar.
3. **Teorema del límite central:** Si *X* es una variable aleatoria para la que se conocen su media μ y su varianza σ^2, la distribución muestral de la media tiende a ser normal con media muestral y error estándar conforme aumenta el tamaño de la muestra.

III. FÓRMULA DEL ERROR ESTÁNDAR DE LA MEDIA Y FACTOR DE CORRECCIÓN POR POBLACIÓN FINITA

Para este análisis conviene observar el comportamiento de la parte $\sqrt{\frac{N-n}{N-1}}$ de la fórmula del error estándar.

En primer lugar, si la población es infinita o muy numerosa, la fracción $\frac{N-n}{N-1}$ resulta ser prácticamente igual a 1 y por consiguiente su raíz cuadrada también es igual a 1, por ello, en los casos en los que se tiene una población muy cuantiosa o con una cantidad infinita de elementos, como este factor es igual a 1 y multiplicarlo por la parte restante de la fórmula no tiene caso ya que se obtendrá el mismo resultado, en símbolos se describe de la siguiente forma:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot (1)$$

$$\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Por otro lado, en los casos en los que el tamaño de la población es reducido en comparación con el tamaño de la muestra, es necesario evaluar la conveniencia de incluir este factor, el que se conoce en los cálculos como ***factor de corrección por población finita***.

EJEMPLO 2

Si se realiza un estudio en una población de tamaño 100 con una muestra de tamaño 10, el factor de corrección corresponde a:

$Datos$	$Desarrollo$	$Factor\ de\ corrección$
$N = 100$	$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{100-10}{100-1}} = 0{,}9535$	$0{,}0465$
$n = 10$		

EJEMPLO 3

Si se realiza un estudio en una población de tamaño 1000 con una muestra de tamaño 30, el factor de corrección corresponde a:

$Datos$	$Desarrollo$	$Factor\ de\ corrección$
$N = 1000$	$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{1000-30}{1000-1}} = 0{,}9854$	$0{,}0146$
$n = 30$		

Conclusión:

El factor de corrección se puede omitir (aunque esto depende en última instancia del estudio específico que se realiza) cuando la razón del tamaño de la muestra al tamaño de la población es menor o igual a 0,05. Para mejor comprensión de este análisis se puede recurrir a la siguiente fórmula:

$$\boldsymbol{\frac{n}{N} \leq 0,05}$$

A continuación se realizará los dos procesos para una mejor comprensión teórica.

EJEMPLO 4

Si se realiza un estudio en una población de tamaño 600 con una muestra de tamaño 30, el factor de corrección corresponde a:

$$Razón\ de\ corrección\ aceptada\text{:}\ \frac{n}{N} \leq 0{,}05$$

$Datos$	$Desarrollo$	$Razón\ de\ corrección$
$N = 600$	$\frac{n}{N} = \frac{30}{600} = 0{,}05$	$0{,}05$
$n = 30$		

$Datos$ | $Desarrollo$ | $Factor\ de\ corrección$

$N = 600$ $\quad$ $\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{600-30}{600-1}} = 0{,}9755$ $\quad$ $0{,}025$

$n = 30$

Conclusión:
Realizando la prueba de razón de corrección se obtiene un resultado de 0,05 mismo valor que se encuentra dentro del parámetro de aceptación (un valor menor o igual a 0,05), también, se realiza la prueba de factor de corrección y se obtiene un resultado de 0,025, de igual manera se comprueba que 2,5% se encuentra dentro del parámetro de aceptación, por tanto, el hecho de eliminar el factor de corrección produzca una sobrestimación del valor del error estándar hace que su utilización resulte más conservadora dentro de la fiabilidad de los estadísticos de prueba y posteriormente a los resultados de una investigación, ya que se trabajaría con un error estándar (dispersión de la distribución muestral dentro del intervalo de 0,00 a 0,05) más eficaz o exacto.

IV. CONSIDERACIONES ADICIONALES SOBRE LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Se revisan en seguida otros detalles sobre la distribución muestral analizada antes para ilustrar con mayor claridad por qué las distribuciones muestrales son el fundamento de la inferencia estadística, con todo lo anterior se revisan en la siguiente sección algunos ejemplos prácticos.

EJEMPLO 5

Como se muestran los datos de la tabla 5 donde representan la totalidad de las muestras posibles, constituyen una población y por ello es posible construir su correspondiente distribución de probabilidad.
La información dada se presenta en la tabla 6, junto con los demás datos.

Medias muestrales	Frecuencia observada	Probabilidad de las medias muestrales $P(\bar{X})$
3	1	0,10
4	1	0,10
5	2	0,20
6	2	0,20
7	2	0,20
8	1	0,10
9	1	0,10
Total	**10**	**1**

Tabla 6. *Frecuencias observadas de la distribución muestral de la media y su probabilidad*

En esa tabla es fácil apreciar que existe una probabilidad de 20% de que la media de esa muestra fuese de 6 hijos por familia, por otro lado, la probabilidad de sacar una muestra cuya media esté entre 5 y 7 es de 60%. Esto ilustra que es altamente probable obtener una muestra que contenga el verdadero valor de la media de la población al igual que la importancia de las 3 conclusiones a las que se llegó antes; en el mismo sentido es altamente probable que una muestra extraída al azar de una población contenga el parámetro que se desea estimar o con el cual se desean realizar pruebas de hipótesis.

V. APLICACIONES DEL ANÁLISIS DE LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

En esta parte se revisan algunos ejemplos en los que se ilustra cómo se utilizan las conclusiones extraídas del análisis de la distribución de la media para analizar situaciones reales, vale la pena resaltar que es importante tener presente que se estará hablando de características de una población, de una muestra, también de las de la distribución muestral de la media y es de gran relevancia tener esto presente en cada caso.

EJEMPLO 6

Se extrae una muestra de n = 30 elementos, de una población que se sabe que tiene un gran número de elementos y cuya media y desviación estándar son μ = 162 y σ = 20. Encuentre la probabilidad de que la media de esa muestra:

a) Sea superior a 170.
b) Esté entre 152 y 172.

Solución:
Como se trata de extraer una sola muestra de entre todas las posibles (una cantidad prácticamente infinita de ellas), se mide la dispersión de estas medias muestrales mediante el error estándar. Además, como el tamaño de la población es prácticamente infinito, se puede utilizar la fórmula simplificada del error estándar (es decir, sin incluir el factor de corrección para una población finita).

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{30}} = 3{,}65$$

Como se trata de una muestra grande (n = 30), de acuerdo con el teorema del límite central la distribución de las medias muestrales es aproximadamente normal por lo que se utiliza el procedimiento de estandarizar los valores para convertirlos a unidades de la desviación estándar y así poder utilizar la tabla de áreas bajo la curva normal para determinar la probabilidad que se busca.

Iniciando con el punto a) la probabilidad de que la media de esa muestra sea superior a 170:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{170 - 162}{3{,}65} = 2{,}19$$

Utilizando la tabla de áreas bajo la curva normal determinamos lo siguiente:

$$P(z \geq 2{,}19) = 1 - 0{,}9857$$
$$P(z \geq 2{,}19) = 0{,}0143$$

La probabilidad de que esa muestra tenga una media igual o superior a 170 es de 1,43%.

Seguidamente el punto b) la probabilidad de que la media de esa muestra esté entre 152 y 172:

$$z_1 = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{152 - 162}{3{,}65} = -2{,}74$$
$$P(z_1 \leq -2{,}74) = 0{,}0031$$

$$z_2 = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{172 - 162}{3{,}65} = 2{,}74$$
$$P(z_2 \leq 2{,}74) = 0{,}9969$$

$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = P(z_2) - P(z_1)$$
$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = 0{,}9969 - 0{,}0031$$
$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = 0{,}9938$$

Por lo que la probabilidad de que la media de esa muestra esté entre 152 y 172 es de 99.38%, es decir, está prácticamente asegurado que esté entre esos dos valores.

EJEMPLO 7

En un estudio que se realizó a una muestra de pacientes con transtornos alimenticios se determinó que el promedio de pérdida de peso por semana es de μ = 325g, con una desviación estándar σ = 20g, si se extrae de la lista total de pacientes una muestra aleatoria de 50 pacientes. Determine lo que se le solicita:

a) ¿Cuál es la probabilidad de que el promedio de la pérdida de peso en la muestra sea inferior a 320 g?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{50}} = 2{,}828$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{320 - 325}{2{,}828} = -1{,}77$$

$$P(z \leq -1{,}77) = 0{,}0384$$

La probabilidad de que la media de la muestra de 50 pacientes con pérdida de peso tenga una media inferior a 320 g es de 3,84%.

b) ¿Cuál es la probabilidad de que la media de la muestra de los pacientes con pérdida de peso esté entre 320g y 330g?

$$z_1 = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{320 - 325}{2{,}828} = -1{,}77$$
$$P(z \leq -1{,}77) = 0{,}0384$$

$$z_1 = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{330 - 325}{2{,}828} = 1{,}77$$
$$P(z \leq 1{,}77) = 0{,}9616$$

$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = P(z_2) - P(z_1)$$
$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = 0{,}9616 - 0{,}0384$$
$$P(z_1 \leq \mu_{\bar{x}} \leq z_2) = 0{,}9232$$

Si efectivamente se determina la muestra de los 50 pacientes y resulta que su media se encuentra entre 320g y 330g se concluiría que este resultado concuerda con los valores que se suponen para la población, lo cual a su vez significaría que el proceso de selección de la muestra está bajo control o se ajusta a los parámetros poblacionales; de otro modo, si la media de la muestra fuera por ejemplo de 300 g se concluiría que ese resultado muestral no concuerda con las características que se suponen para la población (la hipótesis), ya que, la probabilidad de extraer una muestra con esa media es prácticamente de 0 (nula), por ello se sospecharía que el proceso está fuera de control y que sería necesario tomar medidas correctivas para el tipo de análisis médico y/o especializado.

EJEMPLO 8

El promedio de los depósitos en cuentas de cheques de los 1 500 cuentahabientes de una sucursal bancaria es de $4 000 con una desviación estándar de $250. ¿Cuál es la probabilidad de seleccionar una muestra de 100 cuentahabientes que arroje una media muestral de menos de $4 050?

Solución:

Como se conocen los valores de n y N se utiliza el factor de corrección por población finita para calcular el error estándar:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{250}{\sqrt{100}}\sqrt{\frac{1500-100}{1500-1}}$$

$$\sigma_{\bar{x}} = 24{,}16$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{4050 - 4000}{24{,}16} = 2{,}07$$

$$P(z \leq 2{,}07) = 0{,}9808$$

La probabilidad de obtener una muestra de 100 cuentahabientes con un promedio muestral menor a $4 050 es de 98,08%.

DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

En las secciones anteriores se abordo lo sigueinte:

- La distribución muestral es el conjunto de todas las muestras de tamaño *n* que se pueden extraer de una población de tamaño *N*.
- Se puede calcular el número de elementos de una distribución muestral así definida como el número de combinaciones de *N* elementos tomados de *n* en *n* (cantidad de muestras con tamaño n elementos).
- Si se calculan las medias de todas las muestras de una distribución muestral, entonces se tiene la distribución muestral de las medias.
- Desprendido directamente de lo anterior la distribución muestral de la proporción es el conjunto de las proporciones de todas las muestras de tamaño n que se pueden sacar de una población de tamaño N.

I. DESARROLLO

En la tabla siguiente se comparan las definiciones de las distribuciones muestral de la media y de la proporción.

Distribución Muestral de la Media	Distribución Muestral de la Proporción
El conjunto de las medias de todas las muestras posibles de determinado tamaño n que es posible obtener de una determinada población de tamaño N.	El conjunto de las proporciones de todas las muestras posibles de determinado tamaño n que es posible obtener de una determinada población de tamaño N.

Tabla 6. *Tabla comparativa de las distribuciones muestrales de la media y la proporción.*

El procedimiento que se sigue para ilustrar la distribución muestral de la proporción es el mismo que se utilizó para la distribución muestral de la media:

- Se supondrá una población con N elementos.
- Se calculará la proporción y la desviación estándar de la población.
- Se determinará el número total de muestras distintas de tamaño n que es posible extraer y se encontrará cuáles son; es decir, se enumerará la distribución muestral (el conjunto de todas las muestras de ese tamaño que es posible obtener de esa población).
- Se decidirá la proporción de cada una de las muestras identificadas en el punto anterior, con esto ya se tiene la distribución muestral de la proporción.
- Se calculará la media y la desviación estándar de la distribución muestral de la

proporción.

- Se obtendrá conclusiones similares a las conseguidas para la distribución muestral de la media y que son igualmente importantes para el muestreo, al comparar los parámetros poblacionales con los valores obtenidos para la distribución muestral de la proporción.
- Se observará también la forma de la distribución muestral de la proporción para revisar la otra importante conclusión (se refiere a la normalidad de la distribución muestral cuando el tamaño de la muestra es grande).

EJEMPLO 1

Suponga una población de N = 6 artículos, de los cuales 3 están defectuosos y 3 no tienen defecto, si se utiliza el valor "1" para representar la característica de estar defectuoso y el valor "0" para representar la característica de no estarlo, los datos de la población se presentan en la tabla 7, en esta tabla se incluyen los cálculos necesarios para determinar la proporción y la desviación estándar de esta población.

Procedimiento 1:

Se supondrá una población con N elementos.

Artículo	X	$X - \hat{p}$	$(X - \hat{p})^2$
A	1	0,5	0,25
B	1	0,5	0,25
C	1	0,5	0,25
D	0	- 0,5	0,25
E	0	- 0,5	0,25
F	0	- 0,5	0,25
Total	**3**	**0**	**1,50**

Tabla 7. *Tabla de distribución y variabilidad de los artículos con N = 6*

Procedimiento 2:

Se calculará la proporción y la desviación estándar de la población.

La proporción de la poblacional:

$$\hat{p} = \frac{\sum x}{N} = \frac{3}{6} = 0{,}5$$

La proporción de artículos es de 0,5.

La desviación estándar poblacional:

$$\sigma = \sqrt{\frac{\sum(X-\hat{p})^2}{N}} = \sqrt{\frac{1{,}5}{6}} = 0{,}5$$

La cuál también se puede calcular con la sugiente fórmula:

$$\sigma = \sqrt{p \cdot q} = \sqrt{0{,}5 \cdot 0{,}5} = 0{,}5$$

El valor de la desviación estándar de la proporción corresponde a 0,5 aproximadamente.

Procedimiento 3:
Se determinará el número total de muestras distintas de tamaño n.

$$C_n^N = \frac{N!}{n! \cdot (N-n)!}$$

Para determiner el tamaño de las muestras se trabajará con la formula de combinación, esta formula se desarrollo en el módulo anterior.

$$C_n^N = \frac{N!}{n! \cdot (N-n)!} = \frac{6!}{2!\,(6-2)!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1(4 \cdot 3 \cdot 2 \cdot 1)} = \frac{6 \cdot 5}{2 \cdot 1} = 15$$

Procedimiento 4:
Se resolverá la media de cada una de las muestras identificadas en el punto anterior; con esto ya se tiene la distribución muestral de la proporción.

Muestras	$Valor\ de\ x$	$Proporción\ (\hat{p})$
A - B	1, 1	1
A - C	1, 1	1
A - D	1, 0	0,5
A - E	1, 0	0,5
A - F	1, 0	0,5
B - C	1, 1	1
B - D	1, 0	0,5
B - E	1, 0	0,5
B - F	1, 0	0,5
C - D	1, 0	0,5
C - E	1, 0	0,5
C - F	1, 0	0,5
D - E	0, 0	0
D - F	0, 0	0
E - F	0, 0	0
Total		**7,5**

Tabla 8. *Distribución muestral de la proporción.*

Procedimiento 5:
Se calcularán la media y la desviación estándar de la distribución muestral de la proporción.

Procedimiento 6:
Se obtendrá conclusiones similares a las conseguidas para la distribución muestral de la media y que son igualmente importantes para el muestreo, al comparar los parámetros poblacionales con los valores obtenidos para la distribución muestral de la proporción.

La siguiente fórmula corresponde a la media de las proporciones.

$$\mu_{\hat{p}} = \frac{\sum \hat{p}}{N_1}$$

Donde:

$\mu_{\hat{p}}$: Media muestral

$\sum \hat{p}$: Sumatoria de las medias de cada proporción

N_1: Cantidad de muestras producto de la combinación de los elementos de la variable

$$\mu_{\hat{p}} = \frac{\sum \hat{p}}{N_1} = \frac{7{,}5}{15} = 0{,}5$$

Conclusión:
La media de la distribución muestral de las proporciones es igual a la media de la proporción poblacional, en otras palabras, el valor esperado de las proporciones es igual a la media de la proporción poblacional, esto mismo expresado en símbolos corresponde a:

$$\frac{\sum x}{N} = \frac{\sum \hat{p}}{N_1} \Leftrightarrow \hat{p} = \mu_p$$

La siguiente fórmula corresponde a la desviación estándar de todas las proporciones.

$$\sigma_{\hat{p}} = \sqrt{\frac{\sum \left(\hat{p} - \mu_{\hat{p}}\right)^2}{N_1}}$$

Para determinar la desviación estándar de esta distribución muestral de proporciones en la tabla 9 se presenta la misma distribución pero agrupada de acuerdo con la frecuencia de las proporciones muestrales y se incluyen las operaciones necesarias para calcular la desviación estándar.

Proporciones muestrales de p	f_{N_1}	$\hat{p} - \mu_p$	$(\hat{p} - \mu_p)^2$	$f \cdot (\hat{p} - \mu_p)^2$
0	3	- 0,5	0,25	0,75
0,5	9	0	0	0
1	3	0,5	0,25	0,75
Total	**15**	**0**		**1,5**

Tabla 9. *Tabla de dispersion de la distribución muestral de las proporciones agrupadas de acuerdo con su frecuencia absoluta.*

$$\sigma_{\hat{p}} = \sqrt{\frac{1,5}{15}}$$

$$\sigma_{\hat{p}} = \sqrt{0,1}$$

$$\sigma_{\hat{p}} = 0,3162$$

El valor de la desviación muestral de proporciones corresponde a 0,3162.

Al igual que la distribución muestral de la media existe una relación entre esta desviación estándar de la distribución muestral de la proporción, σ_p, y la desviación estándar de la población, p:

$$\boldsymbol{\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}}$$

Para verificar que se cumple con los datos del ejemplo se sustituyen los valores hallados:

$$\sigma_{\hat{p}} = \sqrt{\frac{0,5 \cdot 0,5}{2}} \cdot \sqrt{\frac{6-2}{6-1}}$$

$$\sigma_{\hat{p}} = 0,3162$$

Se concluye que es el mismo valor encontrado a partir de la distribución muestral.

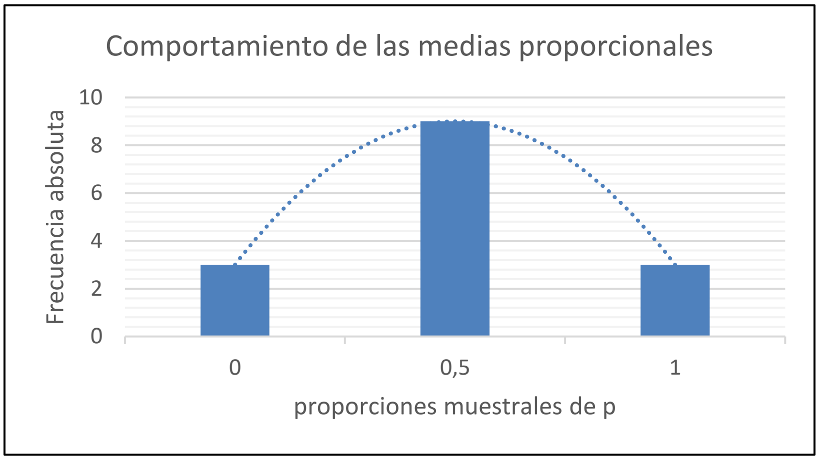


Figura 2. *Gráfica de barras del comportamiento de las frecuencias observadas de la distribución muestral de la proporción.*

En el procedimiento anterior se compararon los valores de la población con los de la distribución muestral para llegar a las conclusiones demostradas, las cuales, como se señaló son paralelas a las que se obtuvieron para la distribución muestral de la media.
Finalmente observando la distribución de frecuencias de la tabla 9 es fácil apreciar que aun para una muestra tan pequeña como n = 2 la distribución de las proporciones muestrales tiende a la normalidad (9 observaciones en el centro y 3 en cada extremo), lo cual también coincide con la tendencia observada en la distribución muestral de la media.

II. CONCLUSIONES QUE DESPRENDE DE LA DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

A continuación se indicará tres conclusiones generales de esta distribución:

1. La media de la distribución muestral de proporciones (el valor esperado de la proporción) es igual a la proporción de la población.
2. Existe una relación entre la desviación estándar de la población (binomial) y la desviación estándar de la distribución muestral de la proporción (error estándar).
3. **Teorema del límite central**: Si X es una variable aleatoria para la que se conocen su proporción p y su varianza σ^2, la distribución muestral de la proporción tiende a ser normal con media μ_p y desviación estándar (error estándar) conforme aumenta el tamaño de la muestra.

La única diferencia en la interpretación de este teorema consiste en observar que la población a la que se refiere es una población binomial y, por lo tanto, $\mu = p$, en otras palabras, quiere decir que la media de una población binomial es su proporción.

III. FÓRMULA DEL ERROR ESTÁNDAR DE LA PROPORCIÓN Y FACTOR DE CORRECCIÓN POR POBLACIÓN FINITA

Al igual que en la distribución anterior es posible eliminar el factor de corrección por población finita de la fórmula del error estándar cuando $\frac{n}{N} \leq 0{,}05$ o cuando la población tiene una cantidad muy grande o infinita de elementos, la forma simplificada del error estándar en estos casos corresponde a la siguiente:

$$\boldsymbol{\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p \cdot q}{n}}}$$

EJEMPLO 2

Una cadena de tiendas de línea blanca tiene 5000 cuentas de crédito abiertas con sus clientes, de acuerdo con la experiencia de varios años se sabe que la proporción de cuentas de crédito que se encuentran atrasadas en sus pagos (morosos) es de 10% del total; si se extrae de esta población una muestra aleatoria de 100 cuentas, ¿Cuál es la probabilidad de que más de 13 cuentas son morosas?

Datos

$n = 100$	$N = 5000$
$p = 0{,}10$	$q = 0{,}90$
$x = 13$	$P(x > 13) = ?$

Paso 1

Comprobación para el factor de corrección.

$\frac{n}{N} \leq 0{,}05$ $\quad\quad$ $\frac{100}{5000} \leq 0{,}02$ $\quad\quad$ $0{,}02 < 0{,}05$

No se utiliza el factor de corrección por población finita y se aplica la forma simplificada de la fórmula del error estándar de la proporción.

Paso 2

Determinar la desviación estándar de la población.

$$\sigma = \sqrt{p \cdot q}$$

$$\sigma = \sqrt{0{,}1 \cdot 0{,}9}$$

$$\sigma = 0{,}3$$

Paso 3
Determinar la desviación muestral de la proporción.

$$\sigma_{\hat{p}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\hat{p}} = \frac{0{,}3}{\sqrt{100}}$$

$$\sigma_{\hat{p}} = 0{,}03$$

Como se trata de una muestra grande (n > 30), se puede utilizar el teorema del límite central y considerar que la distribución del conjunto de todas las muestras implicadas tiene una forma aproximadamente normal.

Paso 4
Determinar la proporción de la muestra.

$$\hat{p} = \frac{x}{n}$$

$$\hat{p} = \frac{13}{100}$$

$$\hat{p} = 0{,}13$$

Paso 5
Determinar el valor "z" de la distribución normal.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

$$z = \frac{0{,}13 - 0{,}10}{0{,}03}$$

$$z = 1{,}00$$

De la tabla de áreas bajo la curva normal se tiene que:

$$P(z \geq 1) = 1 - 0{,}8413 = 0{,}1587$$

Respuesta
La probabilidad de que en esa muestra exista más de 13 cuentas morosas es de 15,87%

EJEMPLO 3

En el departamento de psicología de un centro hospitalario hay registrado 1 000 artículos en su inventario, de acuerdo con la experiencia que se tiene con las auditorías del inventario la proporción de registros contables por artículo que no coinciden con el inventario que realmente se tiene disponible (el inventario físico) es de 15%; si se extrae una muestra de 100 registros contables, ¿Cuál es la probabilidad de que la proporción muestral de registros que no coinciden con el inventario físico sea mayor de 20%?

Datos

$n = 100$	$N = 1000$
$p = 0{,}15$	$q = 0{,}85$
$x = 0{,}20$	$P(x > 20) = ?$

Paso 1

Comprobación para el factor de corrección.

$\frac{n}{N} \leq 0{,}05$ $\qquad$ $\frac{100}{1000} \leq 0{,}10$ $\qquad$ $0{,}10 > 0{,}05$

Como la proporción del factor de corrección no cumple la regla de ser menor al 5% de la población, sí es necesario incluir el factor de corrección por población finita en la fórmula del error estándar, pero antes de calcular este error estándar, conviene revisar alguna otra forma alternativa de la fórmula completa, en este caso usaremos la fórmula que se demostró en la página 20 de este folleto.

Paso 2

Determinar la desviación estándar de la proporción, ya que, se justifica por que el factor de corrección es mayor al 5% de la población y no es necesario en este caso determinar la desviación estándar de la población.

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0{,}15 \cdot 0{,}85}{100}} \cdot \sqrt{\frac{1000 - 100}{1000 - 1}}$$

$$\sigma_{\hat{p}} = 0{,}033891$$

La desviación muestral de la proporción es de 0,0339.

Como se trata de una muestra grande (n > 30) se puede aplicar el teorema del límite central para utilizar la distribución normal en el cálculo de la probabilidad:

Paso 3
Determinar el valor "z" de la distribución normal.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

$$z = \frac{0{,}20 - 0{,}15}{0{,}0339}$$

$$z = 1{,}47$$

De la tabla de áreas bajo la curva normal se tiene que:

$$P(z \geq 1) = 1 - 0{,}9292 = 0{,}0708$$

Respuesta
La probabilidad de que la proporción muestral de registros sea mayor de 20 activos es de 7,08%

IV. CONSIDERACIONES ADICIONALES SOBRE LA DISTRIBUCIÓN MUESTRAL DE LA PROPORCIÓN

Al igual que sucede con la media aritmética, el análisis de la distribución muestral de la proporción permite concluir que:

- Es altamente probable que una muestra extraída al azar de una población contenga el parámetro, la proporción que se desea estimar o con el cual se desean realizar pruebas de hipótesis.

- El error estándar disminuye al aumentar el tamaño de la muestra; esta conclusión se puede deducir de un análisis algebraico de la fórmula simplificada del error estándar (para encontrar este error estándar se divide la desviación estándar de la población entre la raíz cuadrada del tamaño de la muestra, al aumentar éste, aumenta el valor de dicha raíz cuadrada y por lo tanto disminuye el cociente).

EJEMPLO 4

Determine el error estándar para una situación en la que $p = 0{,}7$, cuando:

1. $n = 10$

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0{,}7 \cdot 0{,}3}{10}}$$

$$\sigma_{\hat{p}} = 0{,}1449$$

2. $n = 100$

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0{,}7 \cdot 0{,}3}{100}}$$

$$\sigma_{\hat{p}} = 0{,}0458$$

3. $n = 1000$

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0{,}7 \cdot 0{,}3}{1000}}$$

$$\sigma_{\hat{p}} = 0{,}0145$$

Observando los tres ejemplos con tamaño diferente de la muestra cada uno, se evidencia que al aumentar la muestra el error estándar disminuye

REFERENCIAS BIBLIOGRÁFICAS

Albert, J. H. y Rossman, A. (2011). Workshop Statistics. Discovery with data. A bayesian approach. Bowling Green, OH: Key College Publishing.

Díaz, A. (2013). Estadística aplicada a la administración y la economía. McGraw Hill. Méxio D.F., México.

Díaz, C. (2014). Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar. Universidad de Granada, España.

Díaz Rodríguez, M. (2019). Estadística inferencial aplicada. Universidad del Norte. https://elibro.net/es/lc/elibrocentroamerica/titulos/122378

Juan Verdoy, P. Joaquín Beltrán, M. y José Peris, M. (2016). Problemas resueltos de estadística aplicada a las ciencias sociales. Universitat Jaume I. Servei de Comunicació i Publicacions.https://elibro.net/es/lc/elibrocentroamerica/titulos/42351

Merino, J. M.; Moreno, E.; Padilla, M.; Rodríguez Miñon, P. y Villarino, A. (2012). Análisis de datos en psicología. Madrid: UNED.

Peña, D. y Romo, J. (2017) Introducción a la estadística para las ciencias sociales. Madrid: McGraw-Hill.

Puente Viedma, C. D. L. (2018). Estadística descriptiva e inferencial. Ediciones IDT. https://elibro.net/es/lc/elibrocentroamerica/titulos/59931

Serrano Angulo. J. (2013). Iniciación a la estadística bayesiana. Madrid: La Muralla.