UTN
Facultad Regional Córdoba

TUP

Tecnicatura Universitaria
en Programación

# ESTADÍSTICA

Unidad Temática N°5:
Teoría del Muestreo. Estimación

Material de estudio
1° Año – 2° Cuatrimestre

## Índice

# PROCEDIMIENTOS PARA LA OBTENCIÓN DE UNA MUESTRA

Los procedimientos para la obtención de una muestra se clasifican en:

## Procedimientos No Probabilísticos

**Muestreo de Criterio**:

Los elementos a incluir en la muestra son seleccionados de acuerdo al criterio o juicio del investigador

**Muestreo de la Muestra Disponible:**

Como su nombre lo indica, en la muestra se incluyen aquellos elementos de la población que se encuentran disponibles.

Los Procedimientos No Probabilísticos no permiten medir el Error de Muestreo, pero si puede medirse a través de los Procedimientos Probabilísticos.

## Procedimientos Probabilísticos:

**Muestreo Aleatorio Simple**:

¿Cuándo utilizarlo?

Es adecuado utilizarlo en poblaciones homogéneas (aquellas formadas por elementos de la misma categoría respecto a una determinada característica bajo estudio).

¿En qué consiste?

Consiste en extraer elementos de la población de tal manera que **todos tengan la misma probabilidad de ser incluidos en la muestra.** Por lo tanto, para seleccionar la muestra se debe utilizar alguna técnica que garantice esta condición.

En primer lugar deben enumerarse todos los elementos de la población de 1 a N (siendo N el tamaño de la población)

Posteriormente se seleccionan de la población los elementos que integrarán la muestra, para ello pueden colocarse todos los números en un bolillero y extraer tantos como sea necesario para llegar a tamaño de muestra requerido.

Es decir que se extraen “n” bolillas, siendo “n” la cantidad de elementos que deben incluirse en la muestra.

En lugar de extraer los números de un bolillero, pueden leerse de la **Tabla de Números Aleatorios**

¿En qué caso no es aconsejable su uso?

No es aconsejable el uso de este procedimiento en poblaciones heterogéneas (divididas en varias categorías respecto a cierta característica bajo estudio), ya que este procedimiento no asegura que todas las categorías estén representadas en la muestra.

A partir de una población formada por un lote de 36 artículos de los cuales se registró el precio, fueron seleccionadas 3 muestras de tamaño 5:

| Caso | Muestra_1 | Muestra_2 | Muestra_3 |
|---|---|---|---|
| 1 | 135.40 | 146.30 | 134.10 |
| 2 | 126.90 | 122.90 | 148.70 |
| 3 | 131.80 | 137.60 | 146.10 |
| 4 | 124.70 | 148.10 | 145.30 |
| 5 | 135.40 | 137.70 | 125.20 |

**Media muestral ($\bar{x}$): Media o promedio obtenido en cada muestra. La media muestral varía según de cuál muestra provenga, por lo tanto es una variable aleatoria.**

| Variable | n | Media |
|---|---|---|
| Muestra_1 | 5 | 130.84 |
| Muestra_2 | 5 | 138.52 |
| Muestra_3 | 5 | 139.88 |

**$\bar{x}_1$: 130,84 Precio promedio en la primera muestra de 5 artículos**

**$\bar{x}_2$: 138,52 Precio promedio en la segunda muestra de 5 artículos**

**$\bar{x}_3$: 139,88 Precio promedio en la tercera muestra de 5 artículos**

| Variable | n | Media |
|---|---|---|
| Precios | 36 | $\boxed{137.90}$ |

**Media poblacional o Verdadera Media: $\mu$ = 137,90 Precio promedio de los 36 artículos que forman la población.**

**La diferencia entre las medias muestrales y la media poblacional se debe a que μ es un valor constante que caracteriza a la población mientras que el estimador media muestral es una variable aleatoria cuyo valor varía según las observaciones muestrales particulares que forman parte de la muestra, es decir que varía debido a las fluctuaciones aleatorias de la media muestral.**

**Esas diferencias existentes entre las medias muestrales y la media poblacional (verdadera media) se denominan "Error de muestreo" o "Error de estimación": $|\bar{x} - \mu|$**

## MUESTREO ESTRATIFICADO

¿Cuándo utilizarlo?

**Se utiliza en poblaciones heterogéneas, es decir en poblaciones en las cuales los elementos se comportan de diferente manera respecto a la característica bajo estudio.**

Ejemplo:

Característica bajo estudio: Cantidad de cigarrillos fumados por los habitantes de cierta población que se encuentra dividida por edades.

0 -14 años

14 – 70 años

70 años o más

¿En qué consiste?

Se divide a la población en grupos denominados estratos.

Los elementos pertenecientes a un mismo estrato deben ser tan similares entre sí como sea posible; mientras que, los elementos pertenecientes a distintos estratos deben ser tan diferentes entre sí como sea posible.

Es decir, debe existir homogeneidad entre los elementos de un mismo estrato y heterogeneidad entre los elementos de diferentes estratos.

Vamos a calcular cuántos elementos de cada estrato incluiremos en la muestra

ni: cantidad de elementos que extraemos de cada estrato para incluir en la muestra.

Posteriormente se extrae de cada estrato una cantidad $n_i$ de elementos para incorporar a la muestra.

Sumando la cantidad de elementos que extraemos de cada estrato debemos obtener el tamaño de muestra requerido: ∑ni= n: tamaño de la muestra

La cantidad de elementos a extraer de cada estrato puede calcularse a través de los siguientes criterios:

**Afijación Igual**

**Afijación Proporcional**

**Afijación Optima**

**Son necesarios los siguientes datos:**

**r: cantidad de estratos**

**n: tamaño de la muestra**

**N: tamaño de la población**

**Ni: tamaño de cada estrato: Cantidad de elementos que forman cada estrato**

**$\sigma$i: Desviación de cada estrato**

## Afijación Igual

De cada estrato se extrae la misma cantidad de elementos para formar la muestra.

Este criterio es adecuado cuando todos los estratos tienen el mismo tamaño, es decir cuando éstos tienen la misma cantidad de elementos.

La cantidad de elementos a extraer de cada estrato queda determinada por la siguiente fórmula:

$$\boxed{n_i = \frac{n}{r}}$$

"**r**" es la cantidad de estratos en los cuales se encuentra dividida la población.

## Afijación Proporcional

La cantidad de elementos que se extrae de cada uno de los estratos es proporcional al tamaño de los mismos.

A través de la aplicación de este criterio logramos que en la muestra, se repita la misma proporción de elementos pertenecientes a cada estrato que hay en la población.

Este criterio es adecuado cuando los estratos tienen distinto tamaño e igual desviación estándar.

La cantidad de elementos a extraer de cada estrato queda determinada por la siguiente fórmula:

$$\boxed{ni = n\frac{Ni}{N}}$$

## Afijación Optima de Neymann

La cantidad de elementos que se extrae de cada estrato es proporcional a la desviación estándar del mismo.

Mientras mayor es la desviación del estrato, mayor proporción de elementos se extraen del mismo para integrar la muestra.

Es aconsejable cuando los estratos tienen distinta desviación estándar.

La cantidad de elementos a extraer de cada estrato queda determinada por la siguiente fórmula:

$$\boxed{ni = n\frac{Ni\,\sigma i}{\sum Ni\,\sigma i}}$$

**Ejemplo:**

Una población de estudiantes se encuentra dividida de acuerdo al nivel de conocimientos informáticos que posee de la siguiente manera:

| Estrato | Nivel de conocimientos | Ni | σi |
|---|---|---|---|
| I | Bajo | N1= 300 | 2,1 |
| II | Medio | N2=700 | 1,4 |
| III | Alto | N3= 500 | 3,3 |

r= 3 (tres estratos)

N: tamaño de la población

N= N1 + N2 + N3

= 300 + 700 + 500

N = 1500

a) Obtener una muestra del 10% de la población de acuerdo a los tres criterios de afijación conocidos.

n= 0,10*N

n= 0,10*1500

n= 150

$n_1$: cantidad de estudiantes con nivel bajo de conocimientos que deben incluirse en la muestra

$n_2$: cantidad de estudiantes con nivel medio de conocimientos que deben incluirse en la muestra

$n_3$: cantidad de estudiantes con nivel alto de conocimientos que deben incluirse en la muestra

b) ¿Cuál de los tres criterios es el más adecuado y por qué?

**Resolución**

a) N = tamaño de la población a cantidad de elementos o individuos que forman la población

   n = tamaño de la muestra

   ni = cantidad de elementos seleccionados de cada estrato

| Nivel de conocimientos | Ni | σi | Ni σi | **Afijación Igual** | **Afijación Proporcional** | **Afijación Optima** |
|---|---|---|---|---|---|---|
| Bajo | 300 | 2,1 | 300 * 2,1 = 630 | **50** | **30** | **29** |
| Medio | 700 | 1,4 | 700 * 1,4 = 980 | **50** | **70** | **45** |
| Alto | 500 | 3,3 | 500 * 3,3 = 1650 | **50** | **50** | **76** |
| | N= 1500 | | ∑Ni σi= 3260 | n = 150 | n = 150 | n = 150 |

**Afijación Igual**

$$n_1 = n_2 = n_3 = \frac{n}{r} = \frac{150}{3} = 50$$

50 + 50 + 50= 150 (n=150)

De acuerdo al criterio de **Afijación Igual** en la muestra deben incluirse: 50 estudiantes con nivel bajo de conocimientos informáticos

50 estudiantes con nivel bajo de conocimientos informáticos

50 estudiantes con nivel bajo de conocimientos informáticos

**Afijación Proporcional**

$$n_1 = n\frac{N_1}{N} = 150\frac{300}{1500} = 30$$

$$n_2 = n\frac{N_2}{N} = 150\frac{700}{1500} = 70$$

$$n_3 = n\frac{N_3}{N} = 150\frac{500}{1500} = 50$$

30 + 70 + 50= 150 (n=150)

De acuerdo al criterio de **Afijación Proporcional** en la muestra deben incluirse: 30 estudiantes con nivel bajo de conocimientos informáticos

70 estudiantes con nivel bajo de conocimientos informáticos

50 estudiantes con nivel bajo de conocimientos informáticos

**Afijación Optima**

$$n_1 = 150 \frac{300 * 2,1}{3260} = 28,99 \cong 29$$

$$n_2 = 150 \frac{700 * 1,4}{3260} = 45,09 \cong 45$$

$$n_3 = 150 \frac{500 * 3,3}{3260} = 75,92 \cong 76$$

29 + 45 + 76= 150 (n=150)

De acuerdo al criterio de **Afijación Óptima** en la muestra deben incluirse:

29 estudiantes con nivel bajo de conocimientos informáticos

45 estudiantes con nivel bajo de conocimientos informáticos

76 estudiantes con nivel bajo de conocimientos informáticos

b) ¿Cuál de los tres criterios es el más adecuado y por qué?

El criterio de Afijación Óptima es el más adecuado porque los estratos tienen distinto tamaño

## MUESTREO POR CONGLOMERADOS

¿En qué consiste?

Se divide a la población en grupos denominados conglomerados y posteriormente de cada conglomerado se extraen los elementos para formar la muestra.

Cada conglomerado debe constituir una población en miniatura.

Los elementos pertenecientes a un mismo conglomerado deben ser tan diferentes entre sí como sea posible; mientras que los elementos pertenecientes a distintos conglomerados deben ser tan similares entre sí como sea posible. Es decir, debe existir heterogeneidad entre los elementos de un mismo conglomerado y homogeneidad entre los elementos de diferentes estratos.

¿En qué caso no es aconsejable su uso?

Cuando los conglomerados en los cuales se encuentra dividida la población no cumplen con las condiciones mencionadas el Muestreo Estratificado no garantiza la obtención de una muestra representativa.

**Muestreo Aleatorio Sistemático**

Este procedimiento comprende los siguientes pasos:

1) Enumerar los elementos de la población: 1 ……… N
2) Obtener la razón de muestreo: **k = N/n**, siendo N = tamaño de la población y n=tamaño de la muestra

Ejemplo:

Población: 100 habitantes Muestra:5 habitantes

N= 100 n= 5

**k = N/n**

**= 100/5**

**k = 20**

3) Seleccionar el primer elemento para la muestra aleatoriamente de entre los k primeros de la población:

**1≤ a ≤ k** a = primer elemento seleccionado para incluir en la muestra

**1≤ a ≤ 20**

Ejemplo: El elemento número 4 de la población es el primero que se incluirá en la muestra.

En el caso planteado, habitante que tiene asignado el número 4 es el primero que se incluirá en la muestra.

**a = 4**

4) Seleccionar los restantes elementos para incluir en la muestra, cada **k** elementos de la población:

   segundo elemento seleccionado: **a + k**

   tercer elemento seleccionado: **a + k + k**

   y así sucesivamente hasta extraer los n elementos para formar la muestra

**Muestra de 5 habitantes:**

**Primer habitante incluido en la muestra: a = 4**

**Segundo habitante incluido en la muestra: a + k= 4 + 20**
**= 24**

**Tercer habitante incluido en la muestra: 24 + 20= 44**

**Cuarto habitante incluido en la muestra: 44 + 20= 64**

**Quinto habitante incluido en la muestra: 64 + 20= 84**

¿En qué caso se utiliza?

Este procedimiento se aplica a poblaciones en las cuales los elementos que se encuentran uno a continuación del otro son similares entre sí, respecto a la característica bajo estudio.

Desventaja:

Cuando los elementos de la población no están ubicados de la manera anteriormente descripta, sino que se encuentran mezclados, existe el riesgo de seleccionar aquellos en los cuales siempre se repite la misma característica, este fenómeno se denomina “periodicidad”.

# DISTRIBUCIÓN POR MUESTREO

Antes de indicar qué es una distribución por muestreo, se diferenciarán dos conceptos:

**Parámetro:**

**Es toda medida obtenida en base a datos provenientes de una población**.

Ejemplos:

| **Parámetro** | Simbología | Fórmula |
|---|---|---|
| **Media Poblacional** | **μ** | $\frac{\sum xi}{N}$ |
| **Varianza Poblacional** | **$\sigma^2$** | $\frac{\sum x_i^2}{N} - \mu^2$ |
| **Desviación Poblacional** | **σ** | $\sqrt{\frac{\sum x_i^2}{N} - \mu^2}$ |
| **Proporción Poblacional** | **P** | $\frac{X}{N}$ |

**Estimador:**

**Es toda medida obtenida en base a datos provenientes de una muestra.**

Ejemplos:

| Estimador | Simbología |
|---|---|
| Media Muestral | $\bar{x}$ |
| Varianza Muestral Corregida | $\hat{s}^2$ |
| Desviación Muestral Corregida | $\hat{s}$ |
| Proporción Muestral | $\hat{p}$ |

Cuando no es posible calcular el valor de un parámetro, entonces se recurre a una muestra y en base a dicha muestra se obtiene el valor del estimador, el cual brinda una idea respecto al verdadero valor del parámetro que se hubiera obtenido en base a la población, ese valor particular del estimador obtenido en base a una muestra se denomina **Estimación Puntual.**

Generalmente el valor del estimador no coincide con el valor del parámetro, existe una diferencia que se denomina **Error de Muestreo.**

**Error de Muestreo**:

Es la diferencia entre la **Estimación Puntual** y el **Parámetro poblacional**

El valor del estimador varía según de que muestra provenga, por lo tanto es una variable aleatoria y en consecuencia podemos obtener su distribución de probabilidades, la cual se denomina **Distribución por Muestreo.**

**Distribución por Muestreo:**

¿Qué es?

**Es la distribución de probabilidades de un estimador**.

¿Qué indica?

| |
|---|
| Indica los valores que puede asumir el estimador (por ejemplo la media muestral) y sus respectivas probabilidades. |

## Distribución por Muestreo de la Media Muestral

**Una población está compuesta por cuatro empleados que trabajan 3, 4, 5 y 6 (N=4) horas diarias. Sea x la cantidad de horas que trabaja cada empleado, se pide:**

a) Calcular el valor de los siguientes parámetros (valores poblacionales): **Media poblacional, Varianza poblacional y Desviación poblacional.**

b) Encontrar la **Distribución por muestreo del estimador** correspondiente al parámetro **Media poblacional** en base a todas las muestras de magnitud 2 que surjan por enumeración de todos los **arreglos** posibles (muestreo **con reposición**)

c) Calcular la **Esperanza Matemática del estimador**

d) Calcular la **Varianza del estimador**

e) Calcular la **Desviación Standar del estimador**

f) Comprobar la relación de la esperanza matemática y de la desviación estándar con los parámetros de la población.

**Respuesta:**

a) **Parámetros poblacionales**: Medidas que son calculadas a partir de una población.

**Media poblacional:** μ

$$\mu = \frac{\sum xi}{N} = \frac{3+4+5+6}{4}$$

**μ = 4,5**

**Varianza poblacional: $\sigma^2$**

$$\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2 = \frac{3^2+4^2+5^2+6^2}{4} - 4,5^2 = \frac{86}{4} - 20,25$$

= 21,50 – 20,25

**$\sigma^2$ = 1,25**

**Desviación poblacional:**

$$\sigma = \sqrt{\frac{\sum x_i^2}{N} - \mu^2 = \frac{3^2+4^2+5^2+6}{4} - 4,5^2} = \sqrt{1,25}$$

**σ =1,118**

Cuando no se dispone de toda la información referida a la población entonces no es posible calcular los Parámetros poblacionales.

Por ejemplo para calcular la media poblacional es necesario conocer todos los valores de la variable que se presentaron en la población, si no se dispone de dicha información entonces no será posible calcularla.

¿De qué manera se procede ante esta situación?

En tal caso se extrae una muestra y en partir de dicha muestra se obtiene la media, ese valor hallado brinda una idea del verdadero valor que se hubiese obtenido en la población.

Esa medida que se obtiene a partir de la muestra se denomina Estimador, por ejemplo la **Media Muestral ($\bar{x}$)** es un estimador de la **Media poblacional**.

El valor particular que asume un Estimador en la muestra seleccionada se denomina Estimación Puntual.

Por ejemplo, realizar una Estimación Puntual de la Media consiste en calcular la Media en la muestra seleccionada.

El valor de un Estimador varía según de cual muestra provenga.

Antes de extraer la muestra no se tiene la certeza con respecto a cuáles son los elementos que serán seleccionados, por lo tanto tampoco se tiene la certeza con respecto a que valor asumirá el estimador en la muestra, en consecuencia todo estimador es una variable aleatoria.

Si un Estimador es una variable aleatoria entonces tiene una distribución de probabilidades.

**La distribución de probabilidades de un Estimador se denomina Distribución por Muestreo e indica los posibles valores del estimador y sus respectivas probabilidades.**

Por ejemplo, la **Distribución por Muestreo de la Media Muestral (Distribución por Muestreo de $\bar{x}$)** se plantea de la siguiente manera:

| $\bar{x}$ | $P(\bar{x})$ |
|---|---|
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

Para obtener todos los posibles **valores del estimador** (que en este caso es la **media muestral**) debemos extraer todas las muestras a partir de la población dada.

En la primera columna se incluyen los posibles valores del Estimador y en la segunda columna se indican las respectivas probabilidades.

Para obtener todos los posibles valores del Estimador es necesario extraer todas las muestras posibles de tamaño "n" a partir de la población bajo estudio.

La cantidad total de muestras de tamaño n que pueden extraerse de una población de tamaño N varía según el tipo de muestreo: MCR (Muestreo Con Reposición) o MSR (Muestreo Sin Reposición)

MCR:

Si cada muestra se extrae seleccionando los elementos con reposición entonces la cantidad total de muestras se obtiene a través de la siguiente fórmula: $N^n$

Es decir que pueden extraerse $N^n$ muestras de tamaño n a partir de una población de tamaño N seleccionando los elementos **con reposición**.

MSR:

Extraer las muestras por enumeración de todas las **combinaciones** posibles

Si cada muestra se extrae seleccionando los elementos **sin reposición** entonces la cantidad total de muestras se obtiene a través de la siguiente fórmula: $C_N^n$ (combinatoria de N elementos tomados de n en n)

Es decir que pueden extraerse $C_N^n$ muestras de tamaño n a partir de una población de tamaño N seleccionando los elementos sin reposición.

A partir de la **Distribución por Muestreo de un estimador** podemos calcular los siguientes **parámetros:**

**Esperanza Matemática**

**Varianza**

**Desviación Estándar**

**Parámetros de la Distribución por Muestreo de la Media Muestral (Distribución por Muestreo de $\bar{x}$):**

**Esperanza Matemática de la media muestral (Esperanza Matemática de $\bar{x}$):**

$E(\bar{x}) = \sum \bar{x} * P(\bar{x})$

**Varianza de la media muestral (Varianza de $\bar{x}$):**

$\sigma^2(\bar{x}) = \sum \bar{x}^2 P(\bar{x}) - [E(\bar{x})]^2$

**Desviación Estándar de la media muestral (Desviación Estándar de $\bar{x}$):**

$\sigma(\bar{x}) = \sqrt{\sigma^2(\bar{x})}$

Para obtener la **Distribución por muestreo de un estimador,** previamente deben seguirse los siguientes pasos:

Paso 1- **Calcular la cantidad de muestras posibles**

Paso 2- **Obtener todas las muestras posibles**

Paso 3- **Calcular en cada muestra el valor del estimador**

Paso 1- **Calcular la cantidad de muestras posibles**

En primer lugar, se determina la cantidad de muestras que se obtendrán, para ello debe identificarse en qué condiciones son seleccionados los elementos para constituir la muestra, de la siguiente manera:

Si las muestras se obtienen por enumeración de todos los arreglos posibles, significa que los elementos a incluir en cada muestra son seleccionados con reposición, entonces la cantidad total de muestras de tamaño n que puede extraerse a partir de una población de tamaño N surge de la siguiente fórmula: $N^n$

$N^n = 4^2 = 16$

Significado:

Pueden obtenerse 16 muestras de tamaño 2 seleccionando los elementos **con reposición** (cada elemento se devuelve o repone a la población antes de extraer el siguiente) a partir de una población de tamaño 4

Paso 2- **Obtener todas las muestras posibles**

$x_1$ = cantidad de horas diarias trabajadas por el primer empleado seleccionado para formar la muestra

$x_2$ = cantidad de horas diarias trabajadas por el segundo empleado seleccionado para formar la muestra

16 muestras de 2 empleados cada una:

| $x_1$ \ $x_2$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 3 | 3 3 | 3 4 | 3 5 | 3 6 |
| 4 | 4 3 | 4 4 | 4 5 | 4 6 |
| 5 | 5 3 | 5 4 | 5 5 | 5 6 |
| 6 | 6 3 | 6 4 | 6 5 | 6 6 |

Paso 3- **Calcular en cada muestra el valor del estimador**

En este caso el estimador es la **media muestral**

**$\bar{x}$ = cantidad promedio de horas diarias por cada empleado que forma la muestra**

**El valor del estimador obtenido en una muestra seleccionada se denomina Estimación Puntual**

**Hacer una estimación puntual de la media muestral es calcular el valor de la media en la muestra seleccionada**

$$\bar{x} = \frac{x_1 + x_2}{2}$$

$x_1$: cantidad de horas diarias que trabaja el primer empleado incluido en la muestra

$x_2$: cantidad de horas diarias que trabaja el segundo empleado incluido en la muestra

| $\bar{x} = \frac{x_1 + x_2}{2}$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 3 | **3** | **3,5** | **4** | **4,5** |
| 4 | **3,5** | **4** | **4,5** | **5** |
| 5 | **4** | **4,5** | **5** | **5,5** |
| 6 | **4,5** | **5** | **5,5** | **6** |

Una vez cumplidos estos dos pasos, es posible formar la **Distribución por muestreo**, para ello:

En la primera columna de la tabla se incluye una vez cada uno de los diferentes valores que puede asumir el estimador $\bar{x}$

En la segunda columna se indican las probabilidades asociadas a cada valor, las cuales pueden calcularse mediante el cociente entre la cantidad de veces que se repite cada valor (que se obtiene contando en la tabla obtenida en el punto b.2) dividido el total de muestras

$$P(\bar{x}) = \frac{\text{cantidad de veces que se presenta cada valor de } \bar{x}}{\text{total de muestras}}$$

| $\bar{x} = \frac{x_1 + x_2}{2}$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 3 | **3** | **3,5** | **4** | **4,5** |
| 4 | **3,5** | **4** | **4,5** | **5** |
| 5 | **4** | **4,5** | **5** | **5,5** |
| 6 | **4,5** | **5** | **5,5** | **6** |

**Distribución por muestreo de la media muestral (Distribución por muestreo de $\bar{x}$): Indica los posibles valores de la media muestral y sus respectivas probabilidades.**

| $\bar{x}$ | $P(\bar{x})$ |
|---|---|
| **3** | **1/16** |
| **3,5** | **2/16** |
| **4** | **3/16** |
| **4,5** | **4/16** |
| **5** | **3/16** |
| **5,5** | **2/16** |
| **6** | **1/16** |

c) Parámetros de la Distribución de Probabilidades de la media muestral ($\bar{x}$)

**Esperanza matemática de la Media Muestral (Esperanza matemática de $\bar{x}$)**

La fórmula de cálculo es la siguiente:

**E($\bar{x}$)** = $\sum \bar{x} * P(\bar{x})$

Para calcular la esperanza matemática, se obtiene el producto de cada valor de la variable por su respectiva probabilidad y luego se suman los productos obtenidos

| $\bar{x}$ | $P(\bar{x})$ | $\bar{x}$* P ($\bar{x}$) |
|---|---|---|
| **3,0** | **1/16** | 3 * (1/16) = 3/16 |
| **3,5** | **2/16** | 3,5 * (2/16) =7/16 |
| **4,0** | **3/16** | 4 * (3/16) = 12/16 |
| **4,5** | **4/16** | 4,5 * (4/16) = 18/16 |
| **5,0** | **3/16** | 5 * (3/16) =15/16 |
| **5,5** | **2/16** | 5,5 * (2/16) = 11/16 |
| **6,0** | **1/16** | 6 * (1/16) = 6/16 |
| | | **E( $\bar{x}$ )** = 72/16 = **4,5** |

**E( $\bar{x}$ )** = $\sum \bar{x}$* P ($\bar{x}$)

= **4,5**

### d) Varianza de la Media Muestral (Varianza de $\bar{x}$ )

La fórmula de cálculo es la siguiente:

$\sigma^2(\bar{x}) = \sum \bar{x}^2 P(\bar{x}) - [E(\bar{x})]^2$

$= \mathbf{20{,}875} - (4{,}5)^2$

$= \mathbf{20{,}875} - 20{,}25$

$\sigma^2(\bar{x}) = \mathbf{0{,}625}$

Para calcular la varianza, debe calcularse el producto del cuadrado del valor de la variable por su respectiva probabilidad y a continuación se suman los productos obtenidos. A la sumatoria resultante (**20,875**) se le resta el cuadrado de la esperanza matemática (**4,5**)

| $\bar{x}$ | $P(\bar{x})$ | $\bar{x}^2 * P(\bar{x})$ |
|---|---|---|
| **3,0** | **1/16** | $3^2 * (1/16) = 9/16$ |
| **3,5** | **2/16** | $(3,5)^2 * (2/16) = 24,5/16$ |
| **4,0** | **3/16** | $(4)^2 * (3/16) = 48/16$ |
| **4,5** | **4/16** | $(4,5)^2 * (4/16) = 81/16$ |
| **5,0** | **3/16** | $(5)^2 * (3/16) = 75/16$ |
| **5,5** | **2/16** | $(5,5)^2 * (2/16) = 60,5/16$ |
| **6,0** | **1/16** | $(6)^2 * (1/16) = 36/16$ |
| | | $\sum \bar{x}^2 * P(\bar{x}) = 334/16$ <br> $=$ **20,875** |

e) **Desviación Estándar de la media muestral $\bar{x}$**

La fórmula de cálculo es la siguiente:

$$\sigma(\bar{x}) = \sqrt{\sigma 2(\bar{x})}$$

$$= \sqrt{20,875 - (4,5)^2}$$

$$= \sqrt{0,625}$$

$$= 0,79056$$

$\sigma(\bar{x}) \approx$ **0,7906**

f) Para el Muestreo Con Reposición (MCR) se verifican las siguientes relaciones entre los **Parámetros Poblacionales** y los **Parámetros de la Distribución de $\bar{x}$**:

**La Esperanza matemática de la Media Muestral es igual a la Media Poblacional:**

$E(\bar{x}) = \mu$

**4,5** = **4,5**

**La Varianza de la Media Muestral es igual al cociente entre la Varianza Poblacional y el tamaño de cada muestra:**

$\sigma^2(\overline{x}) = \dfrac{\sigma^2}{n}$

$= \dfrac{1{,}25}{2}$

**0,625** = 0,625

**La Desviación de la Media Muestral es igual al cociente entre la Desviación Poblacional y la raíz cuadrada del tamaño de cada muestra:**

$\sigma(\overline{x}) = \dfrac{\sigma}{\sqrt{n}}$

$= \dfrac{1{,}118}{\sqrt{2}}$

**0,7906** = 0,7906

## Distribución por Muestreo de la Proporción Muestral

La cantidad de hijos de un grupo de familias se indica en la siguiente tabla:

| Familias | A | B | C |
|---|---|---|---|
| Cantidad de hijos | 2 | 4 | 5 |

a) Calcular: **Media poblacional, varianza poblacional y desviación poblacional**

b) Indicar todas las muestras de 2 elementos que pueden obtenerse por enumeración de todas las **combinaciones** posibles (es decir seleccionando los elementos **sin reposición**)

c) Calcular la media en cada muestra

d) Indicar la distribución por muestreo

e) Calcular: Esperanza matemática, varianza y desviación de la media muestral

f) Comprobar las relaciones entre las medidas obtenidas en el punto anterior y los parámetros

**Resolución:**

a) **Parámetros poblacionales:**

**Media Poblacional:**

$$\mu = \frac{\Sigma Xi}{N} = \frac{2+4+5}{3} = \frac{11}{3}$$

**μ = 3,667**

**Varianza Poblacional:**

$$\sigma^2 = \frac{\Sigma X_i^2}{N} - \mu^2 = \frac{2^2+4^2+5^2}{3} - (3{,}667)^2$$

**$\sigma^2$ = 1,553**

**Desviación Poblacional:**

σ = $\sqrt{1{,}553}$

**σ = 1,246**

b) N = tamaño de la población

N = 3

n = tamaño de cada muestra

n = 2

Cantidad de muestras (**sin reposición**) = $C_N^n$

$$C_N^n = C_3^2 = \frac{N!}{n!(N-n)!} = \frac{3!}{2!(3-2)!} = \frac{3*2*1}{(2*1)*1} = 3$$

A partir de una población de tamaño 3 pueden obtenerse 3 muestras de tamaño 2 seleccionando los elementos sin reposición.

Composición de las 3 muestras:

| | 2 | 4 | 5 |
|---|---|---|---|
| 2 | | **2 4** | **2 5** |
| 4 | | | **4 5** |
| 5 | | | |

a) Se obtiene la **media** en cada muestra:

| $\bar{x}$ | 2 | 4 | 5 |
|---|---|---|---|
| 2 | | **3** | **3,5** |
| 4 | | | **4,5** |
| 5 | | | |

d) **Distribución por muestreo de la media muestral (Distribución por muestreo de $\bar{x}$): Indica los posibles valores de la media muestral y sus respectivas probabilidades.**

| $\bar{x}$ | $P(\bar{x})$ |
|---|---|
| **3** | **1/3** |
| **3,5** | **1/3** |
| **4,5** | **1/3** |

e) **Parámetros de la Distribución por muestreo de la media muestral:**

**Esperanza Matemática, Varianza y Desviación de la media muestral**

| $\bar{x}$ | P($\bar{x}$) | $\bar{x}$ P($\bar{x}$) | $\bar{x}^2$P($\bar{x}$) |
|---|---|---|---|
| **3** | **1/3** | 3*(1/3) = 3/3 | $3^2$ (1/3) = 9/3 |
| **3,5** | **1/3** | 3,5*(1/3) = 3,5/3 | $3,5^2$ (1/3) =12,25/3 |
| **4,5** | **1/3** | 4,5*(1/3) = 4,5/3 | $4,5^2$ (1/3) = 20,25/3 |
| | | E($\bar{x}$) = 11/3<br>**E($\bar{x}$) = 3,667** | $\sum \bar{x}^2$ **P($\bar{x}$)** = 41,5/3<br>= **13,833** |

**Esperanza Matemática de la media muestral:**

E($\bar{x}$) = $\sum \bar{x} * P(\bar{x})$

**E($\bar{x}$)** = **3,667**

**Varianza de la media muestral:**

$\sigma^2(\bar{x}) = \sum \bar{x}^2 P(\bar{x}) - [\, E(\bar{x}) \,]^2$
$= 13{,}833 - (3{,}667)^2$
$= 13{,}833 - 13{,}447$
**$\sigma^2(\bar{x}) = 0{,}386$**

**Desviación de la media muestral:**

$\sigma(\bar{x}) = \sqrt{0{,}386}$

**$\sigma(\bar{x}) = 0{,}621$**

f. Relaciones con los parámetros:

Para el Muestreo Sin Reposición (MSR) se verifican las siguientes relaciones entre los **Parámetros Poblacionales** y los **Parámetros de la Distribución de $\bar{x}$**

**$E(\bar{x}) = \mu$**
**$3{,}667 = 3{,}667$**

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n} * \frac{N-n}{N-1}$$

$$= \frac{1{,}553}{2} * \frac{3-2}{3-1}$$

**$0{,}386$** $\cong 0{,}388$

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$$

$$= \frac{1{,}246}{\sqrt{2}} * \sqrt{\frac{3-2}{3-1}}$$

$= 0{,}881 * 0{,}707$

**$0{,}621$** $\cong 0{,}623$

Durante el mes pasado la cantidad de inasistencias de 4 empleados de un comercio fue la siguiente:

| Empleados | A | B | C | D |
|---|---|---|---|---|
| Cantidad de inasistencias | 1 | 2 | 3 | 5 |

Se pide:

a) ¿Qué proporción de empleados tuvieron más de 3 inasistencias?
b) Obtener la **Distribución por Muestreo la Proporción Muestral** a partir todas

las muestras de tamaño 2 que surjan por enumeración de todos los arreglos posibles.

d) Calcular la **Esperanza Matemática de la Proporción Muestral**

e) Calcular la **Varianza de la Proporción Muestral**

f) Calcular la **Desviación de la Proporción Muestral**

g) Comprobar las relaciones entre los **Parámetros de la Distribución por Muestreo** y los **Parámetros Poblacionales**

**Respuesta:**

a) **Proporción Poblacional**: Es la proporción de **éxitos** que hay en la población

**P = Proporción Poblacional**

La población está dividida en dos conjuntos, uno formado por los empleados que tuvieron más de 3 inasistencias y el otro conjunto formado por los empleados que tuvieron 3 o menos inasistencias

**Éxito: empleado con más de tres inasistencias**

X: cantidad de éxitos

X= cantidad de empleados que tuvieron más de 3 inasistencias

N= tamaño de la población

**Proporción Poblacional**: $P = X/N$

$P = 1/4$

$P = 0{,}25$

b) Paso 1- **Calcular la cantidad de muestras posibles**

En primer lugar, se determina la cantidad de muestras que obtendremos, para ello debe identificarse en qué condiciones son seleccionados los elementos para constituir la muestra, de la siguiente manera:

Si las muestras se obtienen por enumeración de todos los **arreglos** posibles, significa que los elementos a incluir en cada muestra son seleccionados **con reposición**, entonces la cantidad total de muestras de tamaño n que puede obtenerse a partir de una población de tamaño N surge de la siguiente fórmula: $\mathbf{N^n}$

$\mathbf{N^n} = 4^2 = 16$

Significado:

Pueden obtenerse 16 muestras de tamaño 2 seleccionando los elementos con reposición (cada elemento no se devuelve o repone a la población antes de extraer el siguiente) a partir de una población de tamaño 4

Paso 2- **Obtener todas las muestras posibles**

Se indica la composición de las 16 muestras de dos empleados cada una.

| X2 / X1 | A | B | C | D |
|---|---|---|---|---|
| A | A A | A B | A C | A D |
| B | B A | B B | B C | B D |
| C | C A | C B | C C | C D |
| D | D A | D B | D C | D D |

Paso 3- **Calcular en cada muestra el valor del estimador** (**Proporción muestral**) esto esquivale a hacer a partir de cada muestra la estimación puntual de la proporción (En un caso real se hace la estimación puntual a partir de "una" muestra)

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

$$\hat{p} = \frac{\text{cantidad de éxitos}}{n}$$

$$\hat{p} = \frac{\text{cantidad de empleados en la muestra que tuvieron más de 3 inasistencias}}{\text{tamaño de la muestra}}$$

Ejemplo:

Si la muestra está formada por un empleado A (1inasistencia) y otro B (2 inasistencias), por lo tanto en esta muestra hay 0 empleados con más de 3 inasistencias, entonces en esta muestra la proporción se obtiene mediante el cociente 0/2 = 0

| $\hat{p}$ | A | B | C | D |
|---|---|---|---|---|
| A | 0/2 = 0 | 0/2 = 0 | 0/2 = 0 | 1/2= 0,5 |
| B | 0/2 = 0 | 0/2 = 0 | 0/2 = 0 | 1/2= 0,5 |
| C | 0/2 = 0 | 0/2 = 0 | 0/2 = 0 | 1/2= 0,5 |
| D | 1/2= 0,5 | 1/2= 0,5 | 1/2= 0,5 | 2/2= 1 |

Una vez cumplidos estos dos pasos, es posible formar la **Distribución por muestreo**, para ello:

En la primera columna de la tabla se incluye una vez cada uno de los diferentes valores que puede asumir el estimador $\hat{p}$

En la segunda columna se indican las probabilidades asociadas a cada valor, las cuales deben calcularse mediante el cociente entre la cantidad de veces que se repite cada valor (que se obtiene contando en la tabla obtenida en el punto b.2) dividido el total de muestras

$$P(\hat{p}) = \frac{\text{cantidad de veces que se presenta cada valor de } \hat{p}}{\text{total de muestras}}$$

**Distribución por muestreo de la Proporción Muestral (Distribución por muestreo de $\hat{p}$)**

| $\hat{p}$ | $P(\hat{p})$ |
|---|---|
| 0 | 9/16 |
| 0,5 | 6/16 |
| 1 | 1/16 |

d) **Esperanza matemática de la proporción muestral (Esperanza matemática de $\hat{p}$)**

**$E(\hat{p})$** $= \sum \hat{p} * P(\hat{p})$

Para calcular la esperanza matemática, debe calcularse el producto de cada valor de la variable por su respectiva probabilidad y luego se suman los productos obtenidos

| $\hat{p}$ | $P(\hat{p})$ | $\hat{p} * P(\hat{p})$ |
|---|---|---|
| 0 | 9/16 | 0 * (9/16) = 0 |
| 0,5 | 6/16 | 0,5 * (6/16) = 3/16 |
| 1 | 1/16 | 1 * (1/16) = 1/16 |
| | | $E(\hat{p})$ = 4/16 = **0,25** |

**$E(\hat{p})$** $= \sum \hat{p} * P(\hat{p}) =$ **0,25**

e) **Varianza de la proporción muestral (Varianza de $\hat{p}$)**

**$\sigma^2(\hat{p})$** $= \sum \hat{p}^2 * P(\hat{p}) - [E(\hat{p})]^2$

Para calcular la varianza, debe calcularse el producto del cuadrado del valor de la variable por su respectiva probabilidad y a continuación se suman los productos obtenidos. A la sumatoria resultante (0,15625) se le resta el cuadrado de la Esperanza Matemática (**0,25**)

| $\hat{p}$ | $P(\hat{p})$ | $\hat{p}^2 * P(\hat{p})$ |
|---|---|---|
| 0 | 9/16 | $(0)^2 * (9/16) = 0$ |
| 0,5 | 6/16 | $(0,5)^2 * (6/16) = 1,5/16$ |
| 1 | 1/16 | $(1)^2 * (1/16) = 1/16$ |
| | | $\sum \hat{p}^2 P(\hat{p}) = 2,5/16$<br>$= 0,15625$ |

$$\sigma^2(\hat{p}) = \sum \hat{p}^2 P(\hat{p}) - [E(\hat{p})]^2$$
$$= 0,15625 - (0,25)^2$$
$$= 0,15625 - 0,0625$$

**$\sigma^2(\hat{p})$ = 0,09375**

f) **Desviación de la proporción muestral (Desviación de $\hat{p}$)**

$$\sigma\,(\hat{p}) = \sqrt{\sum \hat{p}^2 * P\,(\hat{p}) \;-\; [E\,(\hat{p})]2}$$

$$= \sqrt{0{,}15625 - 0{,}0625}$$

$$= \sqrt{0{,}09375}$$

**$\sigma\,(\hat{p}) = 0{,}3062$**

g) En la distribución por muestreo de $\hat{p}$ para el muestreo con reposición se verifican las siguientes relaciones:

**La Esperanza Matemática de la proporción muestral es igual a Proporción Poblacional:**

**$E(\hat{p}) = P$**

**$0{,}25 = 0{,}25$**

La **Varianza de la proporción muestral** es igual a **Proporción de éxitos Poblacional por (Q= 1- P) que es la Proporción de fracasos en la población sobre el tamaño de cada muestra extraída**:

$$\sigma^2(\hat{p}) = \frac{P(1-P)}{n}$$

$$= \frac{0{,}25(1-0{,}25)}{2}$$

**0,09375** = 0,09375

La **Desviación de la proporción muestral** es igual a la raíz cuadrada de: **Proporción de éxitos Poblacional** por **(Q= 1- P) que es la Proporción de fracasos en la población sobre el tamaño de cada muestra extraída**:

$$\sigma(\hat{p}) = \sqrt{\frac{\mathbf{P(1-P)}}{\mathbf{n}}}$$

$$= \sqrt{\frac{0{,}25(1-0{,}25)}{2}}$$

$\sqrt{0{,}09375} = \sqrt{0{,}09375}$

**0,3062** = 0,3062

Si se emplea el Muestreo Sin Reposición (MSR), se verifican las siguientes relaciones:

**E( $\hat{p}$ ) = P** (no se agrega el factor de corrección)

$$\sigma^2(\hat{p}) = \frac{P(1-P)}{n} * \frac{N-n}{N-1}$$

$$\sigma(\hat{p}) = \sqrt{\frac{P(1-P)}{n} * \frac{N-n}{N-1}}$$

## ESTIMACIÓN

### Introducción

**Supóngase que a partir de una población formada por 36 artículos, se extrajeron 3 muestras de tamaño 5 para analizar la variable "precios"**

| Caso | Muestra_1 | Muestra_2 | Muestra_3 |
|---|---|---|---|
| 1 | 135.40 | 146.30 | 134.10 |
| 2 | 126.90 | 122.90 | 148.70 |
| 3 | 131.80 | 137.60 | 146.10 |
| 4 | 124.70 | 148.10 | 145.30 |
| 5 | 135.40 | 137.70 | 125.20 |

Se obtiene la media en cada muestra: **Media muestral ($\bar{x}$)**

| Variable | n | Media |
|---|---|---|
| Muestra_1 | 5 | 130.84 |
| Muestra_2 | 5 | 138.52 |
| Muestra_3 | 5 | 139.88 |

**130,84** es el **precio promedio en la Primera Muestra de 5 artículos**
**138,52** es el **precio promedio en la Segunda Muestra de 5 artículos**
**139,88** es el **precio promedio en la Tercera Muestra de 5 artículos**

A partir de cada una de las muestras se realizó una **Estimación Puntual de la Media**.

### Estimación Puntual

**Valor que asume el estimador en la muestra seleccionada.**

En el caso planteado, el estimador es la **Media muestral ($\bar{x}$)** mientras que el valor que asume el estimador es la **Estimación Puntual** de la media.

Calculamos la **Media poblacional o Verdadera Media**: $\mu$

| Variable | n | Media |
|---|---|---|
| Precios | 36 | 137.90 |

**Media poblacional o Verdadera Media: $\mu$ = 137,90 Precio promedio de los 36 artículos que forman la población.**

**La diferencia entre las medias muestrales y la media poblacional se debe a que μ es un valor constante que caracteriza a la población mientras que el estimador media muestral es una variable aleatoria cuyo valor varía según los valores que se obtienen en las muestras seleccionadas.**

**Esas diferencias existentes entre los valores de las medias muestrales y la media poblacional (verdadera media) se denominan "Error de muestreo" o "Error de estimación"**

$|\bar{x} - \mu|$ **Error de muestreo para la Media**

## Determinación de Error, Riesgo y Tamaño de la muestra

Cuando no es posible calcular el valor de un parámetro, entonces se recurre a una muestra y en base a dicha muestra se obtiene el valor del estimador, el cual nos da una idea respecto al verdadero valor del parámetro que se hubiera obtenido en base a la población, ese valor particular del estimador obtenido en base a una muestra se denomina **Estimación Puntual**

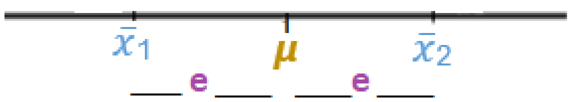
Generalmente el valor del estimador no coincide con el valor del parámetro, existe una diferencia que se denomina **Error de Muestreo**

| |
|---|
| **Error de Muestreo**: <br><br> Es la diferencia entre la **Estimación Puntual** y el **Parámetro Poblacional**. |

Para el caso de la media, el **Error de Muestreo** queda simbolizado de la siguiente manera: $|\bar{x} - \mu|$

Si se está dispuesto a aceptar que esta diferencia sea a lo sumo igual a un valor **e**, entonces **e** es el **error máximo tolerado.**

Para que la diferencia $|\bar{x} - \mu|$ sea menor o igual que **e**, $\bar{x}$ debe asumir valores comprendidos entre $\bar{x}_1$ y $\bar{x}_2$:




Si $\bar{x} < \bar{x}_1$ o si $\bar{x} > \bar{x}_1$

entonces

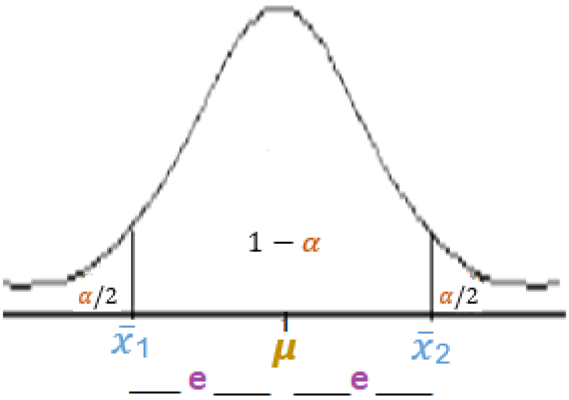
$|\bar{x} - \mu|$ será mayor que e

**No se tiene la certeza de que $|\bar{x} - \mu|$ sea menor que e**

**La probabilidad de que la diferencia $|\bar{x} - \mu|$ supere a e se denomina riesgo**

Se supondrá que la **media muestral ( $\bar{x}$ )** tiene Distribución Normal.

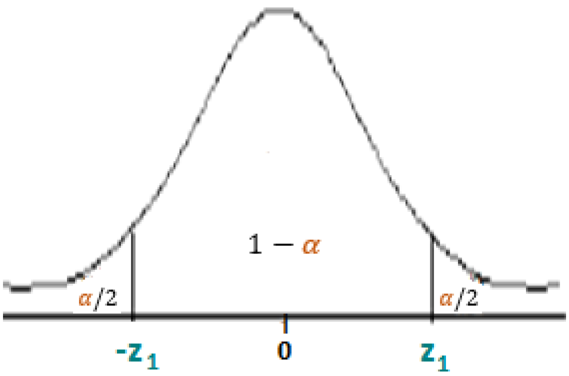
Gráficamente el riesgo queda distribuido en los dos extremos de la campana:




El **riesgo** se simboliza con la letra $\alpha$

$\alpha$ = Pr $(|\bar{x} - \mu| > e)$

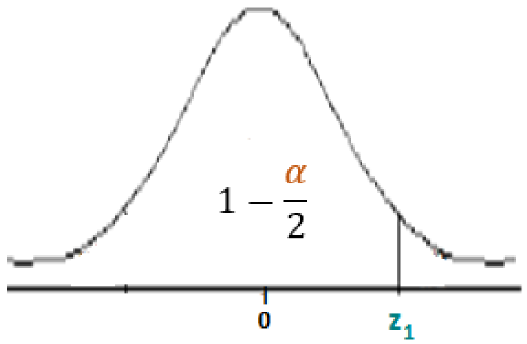
Si se estandariza $\bar{x}_1$ y si $\bar{x}_2$ se obtendría **$-z_1$** y **$z_1$**:




En la Tabla Normal puede encontrarse la siguiente probabilidad:

P(z < $z_1$)

Gráficamente esta probabilidad es el área a la izquierda de la ordenada trazada a partir de $z_1$:




P(z < $z_1$) = $1-\frac{\alpha}{2}$

Fórmula para calcular el **riesgo**:

Se despeja $\alpha$ a partir de la siguiente expresión:

P(z < $z_1$) = $1-\frac{\alpha}{2}$

$\frac{\alpha}{2}$ = 1- P(z < $z_1$)

$$\alpha = 2[1-P(z<z_1)]$$

$z_1$ se obtiene a partir de la fórmula para estandarizar $\bar{x}$:

$$z_1 = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Si se está dispuesto a aceptar que la diferencia sea a lo sumo igual a **e**, entonces se reemplaza la diferencia | $\bar{x}$ - $\mu$ | por **e**

$$z_1 = \frac{e}{\frac{\sigma}{\sqrt{n}}}$$

Si se resuelve el cociente, se obtiene la siguiente expresión:

$$z_1 = \frac{e\sqrt{n}}{\sigma}$$

$$\alpha = 2\left[1 - P\left(z < \frac{e\sqrt{n}}{\sigma}\right)\right]$$

También puede calcularse cuál es el tamaño de muestra asociado a determinado error máximo tolerado con una probabilidad $\alpha$ de que dicho error sea superado y suponiendo determinada desviación estándar.

Fórmula para calcular el tamaño de la muestra:

Se despeja la fórmula para obtener "n" a partir de la siguiente expresión:

$$z = \frac{e\sqrt{n}}{\sigma}$$

Se despeja la incógnita que es "n":

$$e\sqrt{n} = z.\,\boldsymbol{\sigma}$$

$$\sqrt{n} = \frac{z.\sigma}{e}$$

$$n = \frac{z^2\sigma^2}{e^2}$$

También puede calcularse cuál es el **error** asociado a cierto tamaño de muestra.

Fórmula para calcular el **error**:

Ser despeja **e** a partir de la siguiente expresión:

$$z = \frac{e\sqrt{n}}{\sigma}$$

$$e = \frac{(z)\sigma}{\sqrt{n}}$$

Ejemplo

Se desea estimar la vida útil media de las pilas que produce una fábrica “A” con un error máximo de 1,5 hs., una desviación estándar de 3 hs. y una confianza del 99%

a) ¿Cuál es el tamaño de muestra necesario para que se cumplan las condiciones mencionadas? Calcular e interpretar.

b) ¿Cuál es el error máximo tolerado si se toma una muestra de tamaño 49 si la desviación y la confianza son los indicados originalmente? Calcular e interpretar.

c) ¿Cuál es el riesgo si se toma una muestra de tamaño 40? Suponer que la desviación y el error son los indicados originalmente. Calcular e interpretar.

**Resolución:**

a) Datos para calcular el tamaño de la muestra:

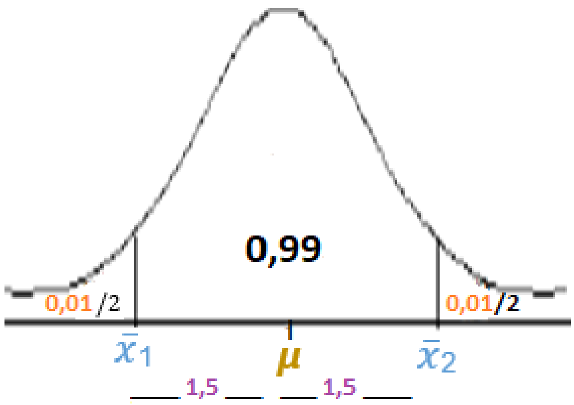
error máximo tolerado: e=1,5

confianza del 99%: 1-$\alpha$

desviación estándar: $\sigma$ =3

Significado del error máximo tolerado: e=1,5 Se a aceptar que la diferencia entre la media muestral y la media poblacional sea a lo sumo igual a 1,5.

Gráficamente:



El Nivel de Confianza es la probabilidad de que la diferencia entre la media muestral y la media poblacional sea **menor o igual** que el error máximo tolerado

Simbólicamente:

$1 - \alpha = \Pr(|\bar{x} - \mu| \leq e)$

Resolución:

En la tabla normal se busca el valor de **z** que corresponde a una probabilidad $1-\alpha/2$, previamente debe calcularse dicha probabilidad:

A partir de $1-\alpha=0,99$ despejamos $\alpha = 1- 0,99$ $\alpha = 0,01$

Calculamos $1 - \alpha/2$:

$1 - \alpha/2 = 1 - 0,01/2$
$= 1 - 0,005$

**$1 - \alpha/2 = 0,995$**

Posteriormente se busca en la Tabla Normal el valor de **z** que corresponde a una probabilidad **0,995**

| z | Probabilidad |
|---|---|
| **2,576** | **0,995** |
| | |

Entonces el valor de **z** que corresponde a una probabilidad $1 - \alpha/2$ es **2,576**

Se reemplaza el valor de **z** en la fórmula para calcular el tamaño de la muestra:

$$n = \frac{z^2 \sigma^2}{e^2}$$

$$= \frac{(2{,}576)^2(3)^2}{1{,}5^2}$$

n = 26,543

n ≥ 27 **En la muestra deben incluirse por lo menos 27 pilas**

b) Datos para calcular el error:

riesgo del 1%: α= 0,01
desviación estándar: σ=3
tamaño de la muestra: n=49

Al igual que en el punto anterior 1-α=0,99 por lo tanto en la tabla se obtendrá el mismo valor de **z** que es **2,576**

$$\mathbf{e} = \frac{(\boldsymbol{z})\sigma}{\sqrt{n}}$$

$$= \frac{(2{,}576)3}{\sqrt{49}}$$

**e=1,104**

Interpretación: Diferencia entre la **vida útil promedio en la muestra de 49 pilas** y la **verdadera vida útil promedio de todas las pilas**.

c) Datos para calcular el **riesgo**

**e=1,5 error máximo tolerado**
σ=3 desviación
n=40 tamaño de la muestra

$$\alpha = 2\left[1 - P\left(z < \frac{e\sqrt{n}}{\sigma}\right)\right] = 2\left[1 - P\left(z < \frac{1{,}5\sqrt{40}}{3}\right)\right]$$

$$\alpha = 2[1 - P(z < \mathbf{3{,}1622})]$$

$= 2[1 - P(z < \mathbf{3,16})]$

$= 2[1 - 0{,}9992]$

$= 2(0{,}0008)$

$\alpha = 0{,}0016$

**Interpretación:** **Probabilidad de que la diferencia entre la vida útil promedio en la muestra de 40 pilas y la verdadera vida útil promedio de todas las pilas, sea mayor a 1,5.**

## Estimación por intervalos de la Media

Consiste en obtener un intervalo en el cual se confía que contenga el verdadero valor de la Media.

**El nivel de confianza 1-$\alpha$ es la probabilidad de que el intervalo obtenido contenga el verdadero valor de la Media, es decir la Media Poblacional.**

Pueden presentarse los siguientes casos:

| desviación poblacional (σ) | tamaño de la muestra (n) | Distribución utilizada | $\bar{x}$- **e** ≤ μ ≤ $\bar{x}$+ **e** |
|---|---|---|---|
| σ conocida | n cualquiera | normal (z) | $\bar{x} - Z * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z * \frac{\sigma}{\sqrt{n}}$<br><br>El valor de **z** se encuentra ingresando a la |

| | | | |
|---|---|---|---|
| | | | Tabla Normal con la probabilidad 1 – α/2 |
| σ desconocida<br><br>Solo tenemos $\hat{S}$(desviación muestral) o los datos para calcular $\hat{S}$ | n ≥ 30 | normal (z) | $\bar{x} - z * \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + z * \frac{\hat{s}}{\sqrt{n}}$<br><br>El valor de **z** se encuentra ingresando a la Tabla Normal con la probabilidad 1 – α/2 |
| | n < 30 | t de Student (t) | $\bar{x} - t * \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t * \frac{\hat{s}}{\sqrt{n}}$<br><br>El valor de **t** se encuentra en la fila de n-1 y la columna de 1 – α/2 correspondiente a la Tabla t de Student. |

Ejemplo:

Se desea estimar el salario promedio que perciben los empleados de las industrias textiles ubicadas en cierta región, para ello es seleccionada una muestra de 97 empresas. Realizar una estimación de la verdadera media con una confianza del 95%.

$\boldsymbol{\sigma}$ desconocida y n=97 (n>30)por lo tanto la media muestral tiene distribución normal.

$$1-\alpha = \Pr\{\bar{x} - \mathbf{e} \leq \mu \leq \bar{x} + \mathbf{e}\}$$

$$1-\alpha = \Pr\{\bar{x} - \boldsymbol{z} * \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + \boldsymbol{z} * \frac{\hat{s}}{\sqrt{n}}\}$$

Se realiza la estimación puntual de la media, la varianza y la desviación:

$\bar{x}$ = 137,22 $\hat{s}^2$ = 74,86 $\hat{s}$ = 8,65 1-α= 0,95

$137{,}22 - 1{,}96 * \frac{8{,}65}{\sqrt{97}} \leq \mu \leq 137{,}22 + 1{,}96 * \frac{8{,}65}{\sqrt{97}}$

$137{,}22 - 1{,}96 * 8{,}878 \leq \mu \leq 137{,}22 + 1{,}96 * 8{,}878$

**137,22** – **1,7214** ≤ μ ≤ **137,22** + **1,7214**

135,499 ≤ μ ≤ 138,94

Interpretación del nivel de confianza en el caso planteado:

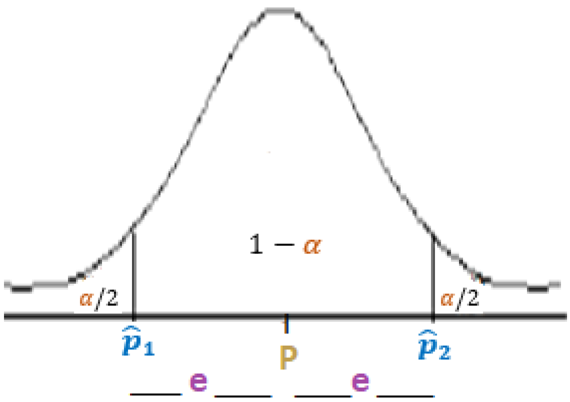
Es la probabilidad de que el intervalo obtenido contenga el verdadero salario promedio (salario promedio de todos los empleados de las industrias textiles ubicadas en cierta región)

De cada 100 intervalos obtenidos, 95 si contendrán salario promedio de todos los empleados de las industrias textiles ubicadas en cierta región, y 5 no lo contendrán.

El intervalo obtenido puede ser uno de los 95 que si contienen salario promedio de todos los empleados de las industrias textiles ubicadas en cierta región o uno de 5 que no lo contienen.

## Determinación de error, riesgo y tamaño de muestra en la estimación de la Proporción

Se parte de la fórmula para estandarizar la variable $\hat{p}$



$$z = \frac{\hat{p} - E(\hat{p})}{\sigma_{\hat{p}}}$$

$$z = \frac{\hat{p} - P}{\sqrt{\frac{P*(1-P)}{n}}}$$

$$z = \frac{e}{\sqrt{\frac{P*(1-P)}{n}}}$$

$$z = \frac{e\sqrt{n}}{\sqrt{P(1-P)}}$$

Fórmula para determinar el error de muestreo o error de estimación:

$$e = z * \sqrt{\frac{P*(1-P)}{n}}$$

Fórmula para determinar el tamaño de muestra:

$$n = \frac{(z)^2 \sqrt{P(1-P)}^2}{e^2}$$

$$n = \frac{(z)^2 P(1-P)}{e^2}$$

Fórmula para determinar el riesgo:

$$\alpha = 2\left[1 - P\left(z < \frac{e\sqrt{n}}{\sqrt{P(1-P)}}\right)\right]$$

## Estimación por intervalos de la Proporción

Consiste en obtener un intervalo en el cual se confía que contenga el verdadero valor de la Proporción.

**El nivel de confianza 1-$\alpha$ es la probabilidad de que el intervalo obtenido contenga el verdadero valor de la Proporción, es decir la Proporción Poblacional.**

Ejemplo

Un comercio que realiza ventas online fue seleccionada una muestra de 2000 clientes, de los cuales 100 son clientes habituales. Realizar una estimación por intervalos de la proporción de clientes habituales con una confianza del 99,73%

Datos:

n= 2000
x= 100 (cantidad de éxitos en la muestra)
nivel de confianza: 1- $\boldsymbol{\alpha}$ = 0,9973

$$\mathbf{1-\ \boldsymbol{\alpha}=\{\widehat{p}-\boldsymbol{e}<\boldsymbol{P}<\widehat{p}+\boldsymbol{e}\}}$$

**Paso 1: Fijar el nivel de confianza**

Nivel de confianza: **1- $\boldsymbol{\alpha}$ = 0,9973**

**Paso 2: Buscar el valor de z**

Se obtiene la probabilidad 1- α/2 a partir de 1- $\alpha$ **(Nivel de Confianza)**

1-$\boldsymbol{\alpha}$ = 0,9973

Despejamos $\boldsymbol{\alpha}$:

$\alpha = 1 - 0{,}9973$
$\alpha = 0{,}0027$
1- α/2 = 1- 0,0027/ 2
**1- α/2 = 0,9987**

El valor de **z** se encuentra en tabla Normal ingresando con 1- α/2:

TABLA VI: DISTRIBUCIÓN NORMAL
$P(Z \leq Zi)$

| z | Prob. | z | Prob. | z | Prob. | z | Prob. | z | Prob. | z | Prob. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0,00 | 0,5000 | 0,60 | 0,7257 | 1,20 | 0,8849 | 1,80 | 0,9641 | 2,40 | 0,9918 | 3,00 | 0,9987 |
| 0,01 | 0,5040 | 0,61 | 0,7291 | 1,21 | 0,8869 | 1,81 | 0,9649 | 2,41 | 0,9920 | 3,01 | 0,9987 |
| 0,02 | 0,5080 | 0,62 | 0,7324 | 1,22 | 0,8888 | 1,82 | 0,9656 | 2,42 | 0,9922 | 3,02 | 0,9987 |

1- α/2= **0,9987** → **z = 3**

## Paso 3: Calcular la proporción de éxitos en la muestra

**Calcular la proporción de éxitos en la muestra es realizar una Estimación Puntual de la Proporción:**

$$\hat{p} = \frac{x}{n} = \frac{100}{2000}$$

$$\hat{p} = 0{,}05$$

## Paso 4: Calcular el Error Típico o Error Estándar del Estimador

**En este caso debe calcularse el Error Estándar de la Proporción Muestral**

**El Error Estándar de la Proporción Muestral es la Desviación de la Proporción Muestral:** $\sigma_{\hat{p}}$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p} * (1 - \hat{p})}{n}}$$

**Paso 5: Construir el intervalo de confianza**

$$0,05 - 3 * \sqrt{\frac{0,05 * 0,95}{2000}} < P < 0,05 + 3 * \sqrt{\frac{0,05 * 0,95}{2000}}$$

$$0,05 - 0,0146 < P < 0,05 + 0,0146$$

$$0,0354 < P < 0,0646$$

**Paso 6: Conclusión**

**Con un 99,73% de confianza se puede afirmar que la proporción de clientes habituales en la población se encuentra comprendida entre 0,0354 y 0,0646**

# BIBLIOGRAFÍA

Ronald E. Walpole (2012) Probabilidad y estadística para ingeniería y ciencias, 9na. Edición Pearson Educación, México. ISBN: 978-607-32-1417-9.

Espejo, I. Fernández, F. López, M. Muñoz, M. Rodríguez, A. Sánchez, A. Valero, C. (2009) Estadística Descriptiva y Probabilidad: (Teoría y problemas). Editorial: Cádiz Universidad de Cádiz, 2009. https://libros.metabiblioteca.org/handle/001/140