UAX_Texto(01).dot

Estadística Empresarial

# 4. Teoría de Muestras

DOCUMENTO1

# ÍNDICE

# MOTIVACIÓN

Cuando se está analizando una población y no se dispone de toda la información sobre la misma, resulta necesario tomar una muestra que es un subconjunto de elementos de la población. El estudio de la Teoría de Muestras es muy importante, ya que a partir de la muestra se van a extraer conclusiones de tipo general sobre la población como veremos en las siguientes Unidades Didácticas.

En esta Unidad Didáctica nos basaremos en la Estadística Descriptiva (Unidad Didáctica 1), y en la Teoría de la Probabilidad (Unidades Didácticas 2 y 3).

# PROPÓSITOS

Los principales propósitos de esta Unidad Didáctica son:

- Distinguir entre población y muestra.
- Conocer las características del muestreo aleatorio simple.
- Entender el concepto de estadístico.
- Comprender las distribuciones de probabilidad de los estadísticos.
- Aprender a calcular probabilidades referentes a estadísticos muestrales.

# PREPARACIÓN PARA LA UNIDAD

Esta unidad didáctica comienza estableciendo la distinción entre población y muestra. A continuación, se define el concepto de muestreo y se exponen las características del muestreo aleatorio simple. Seguidamente, se analiza el concepto de estadístico y se presentan los principales estadísticos. Finalmente, se estudian las distribuciones de probabilidad de los estadísticos en poblaciones normales y no normales.

En esta Unidad Didáctica aprenderás a distinguir entre población y muestra, calcular los principales estadísticos a partir de una muestra e interpretar los resultados obtenidos, y saber calcular e interpretar probabilidades asociadas a estadísticos muestrales.

# 1. POBLACIÓN Y MUESTRA

Una **población** es el conjunto de personas o unidades experimentales sobre las que se va a medir una serie de características medibles o no medibles. Se llama tamaño de una población al número de elementos que la forman. Las poblaciones pueden ser infinitas o finitas. En este capitulo consideraremos poblaciones infinitas.

Los parámetros poblacionales son medidas que caracterizan a la población.

- **Esperanza poblacional**: es la media calculada con todos los elementos de la población. Se representa por $\mu$.
- **Varianza poblacional**: es la varianza calculada con todos los elementos de la población. Se representa por $\sigma^2$.
- **Desviación típica poblacional**: es la desviación típica calculada con todos los elementos de la población. Se representará por $\sigma$.

Una **muestra** es un subconjunto de la población. El tamaño de la muestra es el número de elementos que componen la muestra. Se representará por la letra $n$.

El conjunto de todas las posibles muestras lo representamos mediante un vector aleatorio n dimensional: $X = (X_1, X_2, ..., X_n)$ siendo $X_i \quad i = 1, 2, ..., n$ la variable aleatoria que representa todos los posibles valores que puede tomar el i-ésimo elemento muestral.

Una muestra concreta la representamos mediante un vector numérico $x = (x_1, x_2, ..., x_n)$ siendo $x_i \quad i = 1, 2, ..., n$ el valor numérico del i-ésimo elemento muestral.

A continuación, se presenta un ejemplo para clarificar los conceptos que se acaban de exponer.

**Ejemplo 1.** Se está realizando un estudio sobre las empresas del sector de las telecomunicaciones en España. Se van a medir una serie de características de las mismas: ingresos por ventas, beneficios, número de trabajadores, salarios etc. Consideramos, por ejemplo, la variable número de trabajadores.

a) Identificar la esperanza y la varianza poblacionales.
b) Distinguir entre la esperanza poblacional y la media muestral y entre la varianza poblacional y la varianza muestral.

a) La esperanza poblacional es el número medio de trabajadores calculado teniendo en cuenta todas las empresas del sector de las telecomunicaciones en España. Igualmente la varianza poblacional es la varianza calculada con todas las empresas del sector.

b) Supongamos que no se puede tener acceso a todas las empresas y se toma, por ejemplo, una muestra de 15 empresas. Si a partir de esa información se calculan estas medidas, la media y la varianza obtenidas serían muestrales, puesto que se han calculado a partir de un subconjunto de la población. Obsérvese que si se tomase otra muestra diferente de 15 empresas, se obtendrían otra media y otra varianza muestral que podrían ser diferentes.

Los parámetros poblacionales son únicos, porque están calculados con todos los elementos de la población. Sin embargo, la media y la varianza muestrales no son únicas, puesto que están calculadas a partir de un subconjunto de elementos de la población. Cada muestra tiene una media y una varianza asociada.

Se denomina **muestreo** al proceso de obtención de una muestra. Se distinguen distintos tipos de muestreo.

En esta unidad didáctica se estudiará el **muestreo aleatorio simple**.

Este muestreo se caracteriza porque los elementos muestrales se seleccionan:

1. Al azar.
2. Con reemplazamiento.

Estas características tienen dos implicaciones:

1. Cada elemento muestral se distribuye como la población. Por tanto, la esperanza y la varianza de cada elemento muestral coinciden con los de la población.

$$E(X_i) = \mu \qquad i = 1, 2, \ldots, n$$

$$V(X_i) = \sigma^2 \qquad i = 1, 2, \ldots, n$$

2. Las observaciones muestrales son independientes entre sí.

# 2. PRINCIPALES ESTADÍSTICOS

Un estadístico es una función de los elementos muestrales que se presentará por:

$$T(X_1, X_2, \ldots, X_n)$$

Para cada muestra que se tome, el estadístico tendrá un valor numérico que podría ser diferente. Son ejemplos de estadísticos, la media muestral, la varianza muestral, etc.

Cada estadístico tendrá una media y una varianza asociada. Estos dos conceptos son también muy importantes, ya que se aplicarán en las Unidades Didácticas 5 y 6.

¿Qué es la esperanza de un estadístico?

Si se pudiesen tomar todas las muestras posibles de un tamaño n fijo, se obtendría un valor del estadístico con cada muestra.

Si se calculase la media de todos los valores del estadístico que se han obtenido, el resultado obtenido sería la esperanza del estadístico.

¿Qué es la varianza de un estadístico?

Si se pudiesen tomar todas las muestras posibles de un tamaño n fijo, se obtendría un valor del estadístico con cada muestra.

Si se calculase la varianza de todos los valores del estadístico que se han obtenido, el resultado obtenido sería la varianza del estadístico.

A continuación, se presentan los estadísticos más utilizados.

## MEDIA MUESTRAL

La media muestral, que se representará por $\bar{x}$, es la suma de todas las observaciones muestrales dividida entre el tamaño muestral.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

El operador $\sum$ nos permite expresar una suma de forma abreviada. $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.

**Ejemplo 2.** Se ha observado el precio en euros de las acciones de una empresa A, durante dos semanas. Los precios han sido los siguientes:

12; 10; 11; 14; 12; 11; 10; 12; 12; 14

Determinar la media muestral.

El precio medio de la acción es:

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{12 + 10 + \ldots + 14}{10} = 11{,}80 \text{ €}$$

Obsérvese que si se hubiese tomado otra muestra diferente, se podría haber obtenido otro valor de la media muestral.

Como se acaba de exponer en el ejemplo anterior, la media muestral tomará valores diferentes para las distintas muestras. Por tanto, tendrá una esperanza y una varianza asociadas. A continuación se determinará la esperanza y la varianza de la media muestral.

La esperanza de la media muestral es igual a la esperanza poblacional.

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

Demostración

$$E(\bar{X}) = E(\frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] = \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right] =$$

$$= \frac{n\mu}{n} = \mu$$

Para realizar esta demostración se ha tenido en cuenta que cada elemento muestral se distribuye como la población y por tanto, la esperanza de cada elemento muestral coincide con la poblacional.

Por tanto, la esperanza de la media muestral es:

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

La varianza de la media muestral es igual al cociente entre la varianza poblacional y el tamaño de la muestra.

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Demostración

$$V(\overline{X}) = V(\frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{1}{n^2}\left[V(X_1) + V(X_2) + \cdots + V(X_n)\right] = \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \cdots + \sigma^2\right] =$$
$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Para realizar esta demostración se ha tenido en cuenta que:

1. Cada elemento muestral se distribuye como la población y por tanto, la varianza de cada elemento muestral coincide con la poblacional.
2. Las observaciones muestrales son independientes entre si. Por tanto, la varianza de la suma es la suma de las varianzas.

Por tanto, la varianza de la media muestral es:

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

**VARIANZA MUESTRAL**

La varianza muestral, que se representará por $s^2$, es la suma de los cuadrados de las desviaciones de los valores de la muestra respecto a la media muestral dividida entre el tamaño de la muestra.

$$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

Obsérvese que:

1. La varianza siempre es mayor o igual que cero, pues-to que se trata de una suma de cuadrados.
2. La varianza será nula cuando todas las observaciones muestrales sean iguales y por tanto, iguales a la me-dia.

La varianza mide la dispersión, es decir, la proximidad o alejamiento de las observaciones muestrales con respecto a la media muestral.

Si se desarrolla el cuadrado de la última expresión, se puede obtener otra expresión para la varianza que facilita el cálculo de la misma.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$$

Ejemplo 3. A partir de los precios de las acciones de la empresa A que se presentan en el ejemplo 2, calcular la varianza del precio de las acciones.

$$s^2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - \left( \frac{\sum_{i=1}^{10} x_i}{10} \right)^2 = \frac{12^2 + 10^2 + \ldots + 14^2}{10} - (11{,}8)^2 = 1{,}76\,€^2$$

Obsérvese que si se hubiese tomado otra muestra diferente, se podría haber obtenido otro valor de la varianza muestral.

Como se acaba de exponer en el ejemplo anterior, la varianza muestral tomará valores diferentes para las distintas muestras. Por tanto, tendrá una esperanza asociada.

La esperanza de la varianza muestral es:

$$E(s^2) = \frac{(n-1)\sigma^2}{n}$$

## DESVIACIÓN TÍPICA MUESTRAL

La varianza presenta el inconveniente de que aparece expresada en las unidades de la variable al cuadrado, lo que hace difícil su interpretación. Para resolver este problema se define la desviación típica. La desviación típica, que se representará por $s$ , es la raíz cuadrada positiva de la varianza.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2}$$

**Ejemplo 4.** A partir de la varianza calculada en el ejemplo 3, determinar la desviación típica del precio de las acciones de la empresa A.

La desviación típica es la raíz cuadrada positiva de la varianza. Por tanto, se tiene que:

$$s = \sqrt{1{,}76\text{€}^2} = 1{,}326\text{€}$$

Obsérvese que la desviación típica aparece expresada en las mismas unidades que la variable.

## CUASIVARIANZA MUESTRAL

La cuasivarianza muestral, que se representa por $S^2$, se define de la siguiente forma:

$$S^2 = \sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{n-1}$$

Obsérvese que:

1. La cuasivarianza siempre es mayor o igual que cero, puesto que se trata de una suma de cuadrados.
2. La cuasivarianza será nula cuando todas las observaciones muestrales sean iguales y por tanto, iguales a la media.

La cuasivarianza al igual que la varianza mide la dispersión, es decir, la proximidad o alejamiento de las observaciones muestrales con respecto a la media.

A partir de las expresiones de la varianza y la cuasivarianza muestrales se deduce que:

$$S^2 = \frac{n}{n-1} s^2$$

Esta expresión nos permite calcular la cuasivarianza muestral a partir del tamaño muestral y de la varianza muestral.

**Ejemplo 5.** A partir de la varianza calculada en el ejemplo 3, determinar la cuasivarianza del precio de las acciones.

Para calcular la cuasivarianza muestral, se tendrá en cuenta la relación que existe entre la cuasivarianza y la varianza muestral.

$$S^2 = \frac{n}{n-1} s^2 = \frac{10}{9} \times 1{,}76 = 1{,}95€^2$$

Obsérvese que si se hubiese tomado otra muestra diferente, se podría haber obtenido otro valor de la cuasivarianza muestral.

Como se acaba de exponer en el ejemplo anterior, la cuasivarianza muestral tomará valores diferentes para las distintas muestras. Por tanto, tendrá una esperanza asociada.

La esperanza de la cuasivarianza muestral es igual a la varianza poblacional.

$$E(S^2) = \sigma^2$$

<u>Demostración</u>

$$E(S^2) = E(\frac{n}{n-1}s^2) = \frac{n}{n-1}E(s^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Para realizar esta demostración se ha tenido en cuenta que:

1. La relación entre la cuasivarianza y la varianza muestral $S^2 = \frac{n}{n-1}s^2$.
2. La esperanza de la varianza muestral $E(s^2) = \frac{(n-1)\sigma^2}{n}$.

Por tanto, la esperanza de la cuasivarianza muestral es:

$$E(S^2) = \sigma^2$$

### CUASIDESVIACIÓN TÍPICA MUESTRAL

La cuasivarianza presenta el inconveniente de que aparece expresada en las unidades de la variable al cuadrado, lo que hace difícil su interpretación. Para resolver este problema se define la cuasidesviación típica.

La cuasidesviación típica, que se representará por $S$, es la raíz cuadrada positiva de la cuasivarianza.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

**Ejemplo 6.** A partir de la cuasivarianza calculada en el ejemplo 5, obtener la cuasidesviación típica del precio de las acciones.

La cuasidesviación típica es la raíz cuadrada positiva de la cuasvarianza. Por tanto, se tiene que:

$$S = \sqrt{1{,}95\,€^2} = 1{,}39\,€$$

Obsérvese que la cuasidesviación típica aparece expresada en las mismas unidades que la variable.

# 3. DISTRIBUCIONES EN EL MUESTREO DE POBLACIONES NORMALES

En esta sección se presentan las distribuciones de probabilidad de los estadísticos cuando la población sigue una distribución normal.

**DISTRIBUCIÓN DE LA MEDIA DE UNA MUESTRA PROCEDENTE DE UNA POBLACION NORMAL CON VARIANZA CONOCIDA**

Se considera una población que sigue una distribución normal con varianza conocida. Se toma una muestra aleatoria simple de tamaño n. En este caso el estadístico media muestral sigue una distribución normal.

$$\overline{X} \cong N(\mu, \frac{\sigma}{\sqrt{n}})$$

**Ejemplo 7.** Una población sigue una distribución normal con esperanza 12 y varianza 16. Se toma una muestra aleatoria simple de tamaño 9. Determinar la probabilidad de que la media de la muestra tenga un valor superior a 14.

La variable sigue una distribución normal y la varianza poblacional es conocida. Por tanto, la distribución de probabilidad del estadístico media muestral viene dada por la siguiente expresión:

$$\overline{X} \cong N(\mu, \frac{\sigma}{\sqrt{n}})$$

Particularizando para los datos que proporciona el enunciado se tiene que:

$$\overline{X} \cong N(12, \frac{4}{\sqrt{9}})$$

Se pide la probabilidad de que la media muestral $\overline{X}$ tome valores mayores que 14.

$$P(\overline{X} > 14) = P\left(\frac{\overline{X} - 12}{4/3} > \frac{14-12}{4/3}\right) = P(\overline{X}^* > 1{,}5) = 0{,}0668$$

En primer lugar, se ha transformado la variable en una distribución normal con esperanza 0 y varianza 1, ya que esta es la distribución cuyas probabilidades están tabuladas.

Para ello, a cada lado de la desigualdad se ha restado la media de la variable y se ha dividido entre la desviación típica.

En segundo lugar, se ha determina la probabilidad a partir de la tabla de la distribución normal que se presenta en el anexo de la Unidad Didáctica 3.

La tabla proporciona áreas a la derecha de valores positivos. Se busca 1,5 en la primera columna y 0,00 en la primera fila. El valor correspondiente a esa fila y a esa columna que aparece dentro de la tabla es la probabilidad. Dicho valor es 0,0668.

## DISTRIBUCIÓN DE LA MEDIA DE UNA MUESTRA PROCEDENTE DE UNA POBLACION NORMAL CON VARIANZA DESCONOCIDA

Se considera una población que sigue una distribución normal con varianza desconocida. Se toma una muestra aleatoria simple de tamaño n.

Se distinguen dos casos:

a) Si el tamaño muestral es menor o igual a 30 ($n \leq 30$), la distribución de probabilidad del estadístico media muestral, viene dada por la siguiente expresión:

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong t_{n-1}$$

siendo $t_{n-1}$ una distribución t de Student con n-1 grados de libertad.

**Ejemplo 8.** De una población normal con esperanza 5 se extrae una muestra aleatoria simple de tamaño 21 obteniéndose a partir de dicha muestra una cuasidesviación típica igual a 8. Determinar la probabilidad de que la media muestral sea mayor que 5,875.

La población sigue una distribución normal, la varianza poblacional es desconocida y el tamaño de la muestra es menor que 30. Por tanto, la distribución de probabilidad del estadístico media muestral viene dada por la siguiente expresión:

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong t_{n-1}$$

Particularizando para los valores que indica el enunciado se tiene que:

$$\frac{\overline{X}-5}{\frac{8}{\sqrt{21}}} = \frac{\overline{X}-5}{1{,}75} \cong t_{20}$$

Se pide la probabilidad de que la media muestral $\overline{X}$ tome valores mayores que 5,875.

$$P(\overline{X} > 5{,}875) = P\left(\frac{\overline{X}-5}{1{,}75} > \frac{5{,}875-5}{1{,}75}\right) = P(t_{20} > 0{,}5) \cong 0{,}3$$

En primer lugar, se ha restado 5 y se ha dividido entre 1,75 a cada lado de la desigualdad, con el fin de tener la expresión correspondiente a la distribución t de Student.

En segundo lugar, se ha determinado la probabilidad a partir de la tabla de la distribución t de Student que se presenta en el anexo de la unidad didáctica 3.

Se busca 20 en la primera columna (grados de libertad de la distribución). En la fila correspondiente a 20, se busca 0,5. El valor más próximo es 0,533. La probabilidad aproximada es el valor que aparece en la primera fija y que corresponde a la columna en la que esta situado 0,533. Dicho valor es 0,3.

b) Si el tamaño muestral es mayor que 30 ($n > 30$), la distribución de probabilidad del estadístico media muestral viene dada por la siguiente expresión:

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \cong N(0,1)$$

El estadístico sigue una distribución normal estándar.

**DISTRIBUCIÓN DE LA VARIANZA DE UNA MUESTRA PROCEDENTE DE UNA POBLACIÓN NORMAL**

Se considera una población que sigue una distribución normal con varianza conocida. Se toma una muestra aleatoria simple de tamaño n. La distribución de probabilidad del estadístico varianza muestral viene dada por la siguiente expresión:

$$\frac{ns^2}{\sigma^2} \cong \chi^2{}_{n-1}$$

siendo $\chi^2{}_{n-1}$ una distribución $\chi^2$ de Pearson con n-1 grados de libertad.

**Ejemplo 9.** De una población normal con esperanza 6 y desviación típica 1,5 se extrae una muestra aleatoria simple de tamaño 10. Determinar la probabilidad de que la varianza muestral sea superior a 0,94.

La distribución de probabilidad del estadístico varianza muestral viene dada por la siguiente expresión:

$$\frac{ns^2}{\sigma^2} \cong \chi^2{}_{n-1}$$

Particularizando para los valores que indica el enunciado se tiene que:

$$\frac{10s^2}{1{,}5^2} \cong \chi_9^2$$

Se pide la probabilidad de que la varianza muestral $s^2$ tome valores mayores que 0,94.

$$P(s^2 > 0{,}94) = P\left(\frac{10s^2}{1{,}5^2} > \frac{0{,}94 \times 10}{1{,}5^2}\right) = P(\chi^2{}_9 > 4{,}177) \cong 0{,}9$$

Cada lado de la desigualdad se ha multiplicado por $\frac{10}{1{,}5^2}$ con el fin de conseguir la distribución $\chi^2$ de Pearson. A continuación, se ha determinado la probabilidad a partir de la tabla de dicha distribución que se presentan en el anexo de la Unidad Didáctica 3.

Se busca 9 en la primera columna (grados de libertad de la distribución). En la fila correspondiente a 9, se busca el valor 4,177. El valor más próximo es 4,168. La probabilidad aproximada es el valor que aparece en la primera fija y que corresponde a la columna en la que esta situado 4,168. Dicho valor es 0,9.

## DISTRIBUCIÓN DE LA CUASIVARIANZA DE UNA MUESTRA PRODEDENTE DE UNA POBLACION NORMAL

Se considera una población que sigue una distribución normal con varianza $\sigma^2$ conocida. Se toma una muestra aleatoria simple de tamaño n. La distribución de probabilidad del estadístico cuasivarianza muestral viene dada por la siguiente expresión:

$$\frac{(n-1)S^2}{\sigma^2} \cong \chi^2{}_{n-1}$$

siendo $\chi^2{}_{n-1}$ una distribución $\chi^2$ de Pearson con n-1 grados de libertad.

## DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES PROCEDENTES DE DOS POBLACIONES NORMALES CON VARIANZAS CONOCIDAS

Se consideran dos poblaciones que siguen distribuciones normales con varianzas conocidas. Sean $\mu_1$ y $\mu_2$ las esperanzas de la primera y de la segunda población respectivamente y $\sigma_1^2$ y $\sigma_2^2$ las varianzas. Se toman dos muestras aleatorias simples de tamaños $n_1$ y $n_2$ respectivamente. La distribución de probabilidad del estadístico diferencia de medias muestrales es:

$$\overline{X}_1 - \overline{X}_2 \cong N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^2}{n_2}})$$

**Ejemplo 10**. La duración de los aparatos electrónicos fabricados por una empresa A sigue una distribución normal con esperanza 2500 horas y desviación típica 500 horas, mientras que la duración de los aparatos electrónicos fabricados por una empresa B sigue una distribución normal con esperanza 2300 horas y desviación típica 800 horas. Se ha tomado una muestra aleatoria simple de 300 aparatos de la empresa A y de 200 aparatos de la empresa B.

Determinar la probabilidad de que la diferencia entre la duración media de los aparatos de la empresa A y la duración media de los obtenidos en B sea inferior a 100 horas.

$\xi_1$= " Duración de los aparatos de la empresa A"

$\mu_1 = 2500h$ $\sigma_1 = 500h$ $n_1 = 300$

$\xi_2$ = " Duración de los aparatos de la empresa B"

$\mu_2 = 2300h$ $\sigma_1 = 800h$ $n_2 = 200$

Se consideran dos poblaciones normales con varianzas conocidas. La distribución de probabilidad del estadístico diferencia de medias muestrales es:

$$\overline{X}_1 - \overline{X}_2 \cong N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^2}{n_2}})$$

Particularizando para los valores que indica el enunciado se tiene que:

$$\overline{X}_1 - \overline{X}_2 \cong N(200, \sqrt{\frac{500^2}{300} + \frac{800^2}{200}})$$

$$\overline{X}_1 - \overline{X}_2 \cong N(200,\ 63{,}58)$$

Se pide la probabilidad de que la diferencia de medias muestrales sea inferior a 100.

$$P\left[(\overline{X}_1 - \overline{X}_2) < 100\right] = P\left[\frac{(\overline{X}_1 - \overline{X}_2) - 200}{63{,}58} < \frac{100 - 200}{63{,}58}\right] = P(Z^* < -1{,}57) =$$
$$= (Z^* > 1{,}57) = 0{,}0582$$

En primer lugar, se ha transformado la variable en una distribución normal con esperanza 0 y varianza 1, ya que esta es la distribución cuyas probabilidades están tabuladas.

Para ello, a cada lado de la desigualdad se ha restado la media de la variable y se ha dividido entre la desviación típica.

En segundo lugar, se tiene que determinar la probabilidad a partir de la tabla de la distribución normal que se presenta en el anexo de la unidad didáctica 3.

En la tabla de la distribución normal se presentan áreas a la derecha de valores positivos.

Como la función de densidad de la distribución normal tiene un eje de simetría en 0, se cumple que el área a la izquierda de -1,57 es igual al área a la derecha de 1,57.

A continuación, se busca 1,5 en la primera columna y 0,07 en la primera fila. El valor correspondiente a esa fila y a esa columna que aparece dentro de la tabla es la probabilidad. Dicho valor es 0,0582.

# 4. DISTRIBUCIONES EN EL MUESTREO DE POBLACIONES NO NORMALES

En esta sección se presentan las distribuciones de probabilidad de los estadísticos cuando la población no sigue una distribución normal. Dichas distribuciones están basadas en el Teorema Central del Límite que se enuncia a continuación.

### TEOREMA CENTRAL DEL LIMITE

Sean $\xi_1, \xi_2,...,\xi_n$ variables aleatorias independientes entre sí, y con esperanzas $\mu_1, \mu_2,...,\mu_n$ y varianzas $\sigma_1^2, \sigma_2^2,...,\sigma_n^2$ finitas. Se define una nueva variable $\eta = \xi_1 + \xi_2 + ... + \xi_n$. Si $n > 30$, la variable aleatoria $\eta$ sigue una distribución normal aproximada. $\eta \approx N(\sum_i^n \mu_i, \sqrt{\sum_{i=1}^n \sigma_i^2})$.

### DISTRIBUCIÓN APROXIMADA DE LA MEDIA DE UNA MUESTRA PROCEDENTE DE UNA POBLACION NO NORMAL CON VARIANZA CONOCIDA

Se considera una población con esperanza y varianza finitas. Se toma una muestra aleatoria simple de tamaño n>30.

Como el muestreo es aleatorio simple:

1. Cada observación muestral $X_i$ se distribuye como la población. Por tanto, la esperanza y la varianza de cada observación muestral coinciden con las de la población.

$$E(X_i) = \mu \qquad i = 1, 2, \ldots, n$$

$$V(X_i) = \sigma^2 \qquad i = 1, 2, \ldots, n$$

2. Las observaciones muestrales son independientes entre si.

La media muestral viene dada por la siguiente expresión:

$$\overline{X} = \frac{X_1 + X_2 + \ldots\ldots\ldots\ldots + X_n}{n}$$

Por tanto, es una suma de variables aleatorias con igual distribución, independientes entre si y con esperanza y varianza finitas.

Se cumplen los supuestos del Teorema Central del Límite. La media muestral sigue una distribución normal aproximada.

$$\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$$

**Ejemplo 11.** De una población con esperanza 62,5 y varianza 468,75 se toma una muestra aleatoria simple de tamaño 100. Determinar la probabilidad de que la media muestral tome valores mayores que 65.

Se considera una población con varianza conocida y el tamaño de la muestra es mayor que 30. Por tanto, la distribución de probabilidad del estadístico media muestral es:

$$\overline{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$$

Particularizando para los valores que indica el enunciado se tiene que:

$$\overline{X} \approx N(62{,}5, \frac{\sqrt{468{,}75}}{\sqrt{100}})$$

Se pide la probabilidad de que la media muestral $\overline{X}$ tome valores mayores que 65.

$$P(\overline{X} > 65) = P\left(\frac{\overline{X} - 62{,}5}{\sqrt{4{,}6875}} > \frac{65 - 62{,}5}{\sqrt{4{,}6875}}\right) = P(\overline{X}^* > 1{,}15) = 0{,}1251$$

En primer lugar, se ha transformado la variable en una distribución normal con esperanza 0 y varianza 1, ya que esta es la distribución cuyas probabilidades están tabuladas.

Para ello, a cada lado de la desigualdad se ha restado la media de la variable y se ha dividido entre la desviación típica.

En segundo lugar, se determina la probabilidad a partir de la tabla de la distribución normal que se presenta en el anexo de la unidad didáctica 3.

La tabla proporciona áreas a la derecha de valores positivos. Se busca 1,1 en la primera columna y 0,05 en la primera fila. El valor correspondiente a esa fila y a esa columna que aparece dentro de la tabla es la probabilidad. Dicho valor es 0,1251.

## DISTRIBUCIÓN APROXIMADA DE LA MEDIA DE UNA MUESTRA PROCEDENTE DE UNA POBLACION NO NORMAL CON VARIANZA DESCONOCIDA

Se considera una población con varianza desconocida. Se toma una muestra aleatoria simple de tamaño n mayor que 30. La distribución aproximada del estadístico media muestral viene dada por la siguiente expresión:

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \approx N(0{,}1)$$

El estadístico sigue una distribución normal estándar aproximada.

## DISTRIBUCIÓN APROXIMADA DE LA DIFERENCIA DE MEDIAS MUESTRALES PROCEDENTES DE DOS POBLACIONES NO NORMALES CON VARIANZA CONOCIDAS

Se consideran dos poblaciones con varianzas conocidas. Sean $\mu_1$ y $\mu_2$ las esperanzas de la primera y de la segunda población respectivamente, y $\sigma_1^2$ y $\sigma_2^2$ las varianzas. Se toman dos muestras aleatorias simples de tamaños $n_1$ y $n_2$ respectivamente ambas de tamaños mayores que 30. El estadístico diferencia de medias muestrales sigue una distribución normal aproximada.

$$\overline{X}_1 - \overline{X}_2 \approx N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^2}{n_2}})$$

## CUADROS

### CUADRO 1. DISTRIBUCIONES DE LOS ESTADÍSTICOS EN POBLACIONES NORMALES

| | |
|---|---|
| **Distribución de la media muestral** | **Varianza conocida** $\overline{X} \cong N(\mu, \frac{\sigma}{\sqrt{n}})$ **Varianza desconocida** **n<30** $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong t_{n-1}$ **n>30** $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong N(0,1)$ |
| **Distribución de la varianza muestral** | $\frac{\mathrm{ns}^2}{\sigma^2} \cong \chi^2{}_{\mathrm{n}-1}$ |
| **Distribución de la cuasivarianza muestral** | $\frac{(n-1)S^2}{\sigma^2} \cong \chi^2{}_{n-1}$ |
| **Distribución de la diferencia de medias muestrales** | $\overline{X}_1 - \overline{X}_2 \cong N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2^2}{n_2}})$ |

**CUADRO 2. DISTRIBUCIONES APROXIMADAS DE LOS ESTADÍSTICOS EN POBLACIONES NO NORMALES n > 30**

| | |
|---|---|
| **Distribución de la media muestral** | **Varianza conocida** $\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$ **Varianza desconocida** $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$ |
| **Distribución de la diferencia de medias muestrales** | $\bar{X}_1 - \bar{X}_2 \approx N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$ |

# CONCLUSIONES

Cuando se está analizando una población y no se dispone de toda la información sobre la misma, resulta necesario tomar una muestra que es un subconjunto de elementos de la población.

A partir de la muestra se van a extraer conclusiones de tipo general sobre la población aplicando métodos estadísticos que se estudiarán en las unidades didácticas siguientes.

Por tanto, se debe tener especial cuidado con el proceso de obtención de la muestra. La muestra debe ser representativa de la población objeto de estudio para que los resultados obtenidos sean fiables.

# RECAPITULACIÓN

Una **población** es el conjunto de personas o unidades experimentales sobre las que se va a medir una serie de características medibles o no medibles.

Una **muestra** es un subconjunto de la población. Se denomina muestreo al proceso de obtención de una muestra. Se distinguen distintos tipos de muestreo.

El **muestreo aleatorio simple** se caracteriza porque los elementos muestrales se seleccionan al azar y con reemplazamiento. Estas características tienen dos implicaciones:

1. Cada elemento muestral se distribuye como la población.
2. Las observaciones muestrales son independientes entre sí.

Un **estadístico** es una función de los elementos muestrales. Para cada muestra tomará un valor concreto.

# AUTOCOMPROBACIÓN

1. La esperanza poblacional es:

    a) La media calculada con todos los elementos de una población.

    b) La media calculada con un subconjunto de elementos de una población.

    c) No es única.

    d) Ninguna respuesta es correcta.

2. La varianza poblacional es:

    a) La varianza calculada con un subconjunto de elementos de una población.

    b) La varianza calculada con todos los elementos de una población.

    c) No es única.

    d) Ninguna respuesta es correcta.

3. Se denomina muestreo a:

    a) El proceso de análisis de una población.

    b) El proceso de asignación de valores a los parámetros desconocidos de una población.

    c) El proceso de obtención de un subconjunto de elementos de una población.

    d) Ninguna respuesta es correcta.

4. El muestreo aleatorio simple se caracteriza por:

   a) Los elementos muestrales no se seleccionan al azar.

   b) Los elementos muestrales se seleccionan sin reemplazamiento.

   c) Los elementos maestrales se seleccionan al azar y con reemplazamiento.

   d) Ninguna respuesta es correcta.

5. Si se realiza un muestreo aleatorio simple:

   a) No se conoce la distribución de probabilidad de cada elemento muestral.

   b) Cada elemento muestral se distribuye como la población y los elementos muestrales son independientes entre si.

   c) Los elementos muestrales son dependientes.

   d) Ninguna respuesta es correcta.

6. Un estadístico es:

   a) Una función de las observaciones muestrales.

   b) Un parámetro desconocido.

   c) Una afirmación sobre un parámetro de una población.

   d) Ninguna respuesta es correcta.

7. La media muestral es:

   a) La media calculada con todos los elementos de una población.

   b) La media calculada con un subconjunto de elementos de una población.

   c) Es única.

   d) Ninguna respuesta es correcta.

8. La varianza muestral es:

   a) La varianza calculada con un subconjunto de elementos de una población.

   b) La varianza calculada con todos los elementos de una población.

   c) Es única.

   d) Ninguna respuesta es correcta.

9. La varianza muestral:

   a) Mide la dispersión, es decir, la proximidad o alejamiento de las observaciones muestrales con respecto a la media muestral.
   b) Puede tomar valores negativos.
   c) Está expresada en las mismas unidades que la muestra.
   d) Ninguna respuesta es correcta.

10. La diferencia entre la varianza y la desviación típica de una muestra es:

    a) La varianza es una medida de dispersión y la desviación típica no.
    b) La desviación típica es una medida de dispersión y la varianza no.
    c) La varianza está expresada en las unidades al cuadrado y la desviación típica en las mismas unidades que la muestra.
    d) Ninguna respuesta es correcta.

# SOLUCIONARIO

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1. | a | 2. | b | 3. | c | 4. | c | 5. | b |
| 6. | a | 7. | b | 8. | a | 9. | a | 10. | c |

# PROPUESTAS DE AMPLIACIÓN

Si estás interesado en ampliar los conocimientos sobre teoría de muestras, puedes consultar el siguiente libro:

Azorín F. y Sánchez- Crespo J. L. (1995). Métodos y aplicaciones del muestreo. Editorial Alianza.

# BIBLIOGRAFÍA

Martín Pliego F. J. y Ruiz Maya L. (2005). Fundamentos de Inferencia Estadística. Editorial Thomson.

López de la Manzanara Barbero J. (1996). Problemas de Estadística. Editorial Pirámide.