7.- Si Z es la distribución normal tipificada, hállese:
a) $P(Z>0,63)$ b) $P(-0'65<Z<0,65)$ c) $P(1<Z1,32)$

Sol: a) 0,2643; b) 0,5156; c) 0,0653

8.- Si Z es la normal tipificada, hállese k en los siguientes casos:
a) $P(Z<k) = 0,7823$; b) $P(Z>k) = 0,0838$; c) $P(Z<K) = 0,9948$; d) $P(Z<K) = 0,2206$.

Sol: a) k=0,78; b) k=1,38; c) k=2,56; d) k=-0,77

9.- En una finca agrícola dedicada a la producción de manzanas se ha comprobado que el peso de las manzanas sigue una distribución normal con media 100 g y desviación 10. A la hora de comercializarlas se toman para la clase A las comprendidas entre 80 y 120 gr. Hallar la probabilidad de que escogida una manzana al azar:

a) corresponda a la clase A
b) pese menos de 70 g
c) pese más de 120 g

Sol: a) 0'9544; b) 0'0013; c) 0'0228

10.- Una patrulla de tráfico realiza un control de alcoholemia en una carretera y llegan a la conclusión de que el nivel de alcohol en sangre de los conductores sigue una distribución N(0'25,0'1). Si el nivel de alcoholemia permitido es de 0'5 para los conductores expertos y 0'3 para los conductores novatos.

a) ¿Cuál es la probabilidad de que un conductor al azar tenga más de 0'5?
b) ¿Cuál es la probabilidad de que tenga más de 0'3?
c) ¿Cuál es la probabilidad de que un conductor diese positivo en la de 0'3 pero no en la de 0'5?

Sol: a) 0'0062; b) 0'3085; c) 0'3023

11.- Una empresa lleva a cabo una prueba para seleccionar nuevos empleados. Por la experiencia de pruebas anteriores, se sabe que las puntuaciones siguen una distribución normal de media 80 y desviación típica 25.
¿Qué porcentaje de candidatos obtendrá entre 75 y 100 puntos?

(Sol. 36,74%)

12.- Un profesor de matemáticas ha observado que las notas obtenidas por sus alumnos en los exámenes de Estadística siguen una distribución *N(6; 2,5).*
Se han presentado al último examen 32 alumnos, ¿cuántos sacaron al menos un 7?.

(Sol. 11)

13.- La nota media de un examen tipo test fue de 40'3 y su desviación típica 3'3. Si las calificaciones siguen una distribución normal y se considera aprobado a los que superen 37 puntos.

a) ¿Cuál es el porcentaje de aprobados?

b) Si a ese examen se presentaron 400 personas ¿cuántas aprobaron?

c) Si quisiésemos que sólo aprobasen 50 personas ¿cuál tendría que ser la nota de corte?

Sol: a) 84'13%; b) 336; c) 44'1

14.- Las alturas de 200 estudiantes se distribuyen normalmente, con una media de 175 cm y una desviación típica de 10 cm. ¿Cuántos de estos estudiantes tienen altura:

a) mayor de 180 cm. b) menor de 165 cm.

b) c) entre 160 cm y 180 cm. d) igual a 182 cm.

Sol: a) 61,7; b) 31,74; c) 124,94; d) 0

15.- La cantidad de azúcar depositada en cada bolsa por una máquina envasadora automática sigue una distribución normal con media μ=1050 grs y desviación típica σ=50 grs.

a) Calcula el porcentaje de bolsas con un peso mayor a 1 Kg

b) Calcula el tanto por ciento de paquetes con un contenido que tiene un peso comprendido entre 900 y 1000 grs.

Sol: a) 84'13%; b) 15'74%

16.- Se ha elegido una muestra de 500 tornillos fabricados por una máquina. La media de los diámetros de dichos tornillos es de 2,9 mm y la desviación típica, de 1 mm. Un cliente considera que un tornillo es inservible si su diámetro es inferior a 2,85 mm o superior a 3,1 mm.

a) Sabiendo que los diámetros se distribuyen normalmente, hállese qué porcentaje de tornillos son defectuosos.

b) Si para otro cliente son válidos desde 2,75 hasta 3,15, ¿qué porcentaje son inservibles?

Sol: a) 90%; b) 84,17%

17.- El coeficiente de inteligencia de una población es una v.a. cuya distribución sigue una ley normal del tipo N(100,10). Calcúlese, según esos datos, qué porcentaje de personas cabe esperar que tengan coeficiente de inteligencia:

a) superior a 120 b) entre 90 y 120 c) inferior a 80;

d) Si se escogen 5000 personas al azar, ¿cuántas tendrán un coeficiente de inteligencia mayor de 125?

Sol: a) 2,28%; b) 81,85%; c) 2,28%; d) 31

18.- Los gastos diarios de una familia siguen una distribución normal, con media de 50 euros y desviación típica de 5 euros.

a) Calcular el porcentaje de días en los que los gastos son inferiores a 55 euros.

b) Calcular el porcentaje de días en que los gastos superan los 60 euros.

c) Calcular el porcentaje de días en los gastos son superiores a 50 euros e inferiores a 60 euros.

Sol: a) 84,13%; b) 2,28%; c) 47,72%

19.- Una empresa de transporte de viajeros afirma que el tiempo de retraso de sus viajes sigue una distribución normal, con un retraso medio de 5 minutos y desviación típica de 2 minutos. Calcular:

a) Probabilidad de que un viaje no tenga retraso

b) Probabilidad de que el próximo llegue con más de 5 minutos de retraso.

c) Probabilidad de que el próximo llegue con más de 10 minutos de retraso.

Sol: a) 0,0062; b) 0,5; c) 0,0062.

20.- Una de las pruebas de acceso a la Universidad para mayores de 25 años consiste en un test con 100 preguntas, cada una de las cuales tiene 4 posibles respuestas y sólo una correcta. Para superar esta prueba deben obtenerse, al menos, 30 respuestas correctas.

Si una persona contesta al azar, ¿Qué probabilidad tendrá de superar la prueba?

(Sol. Utilizando la aproximación a través de la normal: p= 0,1492)

TEMA 3.- TEORÍA DE MUESTRAS

1.- POBLACIÓN Y MUESTRA

Entendemos por ***población*** un conjunto de elementos que poseen una característica o propiedad común, y que constituyen la totalidad de los individuos de interés para nuestro estudio.

En particular, nos interesa obtener información acerca de algún valor que caracteriza a la población, como una media, una varianza, una mediana, etc. A estos valores que se refieren a la totalidad de la población se les denomina ***parámetros poblacionales*** y en su notación es común utilizar el alfabeto griego: μ (media poblacional), σ^2 (varianza poblacional), σ (desviación típica poblacional), etc.

Como las poblaciones en las que se pretende estudiar una determinada variable aleatoria son grandes, es muy caro o imposible estudiar a todos sus individuos; lo que se hace, es estudiar una ***muestra*** (una parte) de la población:

Una ***muestra*** es cualquier subconjunto de la población sobre el que se realizan estudios para obtener conclusiones acerca de las características de la población. Cualquier valor obtenido a partir de los datos de la muestra se denomina ***estadístico muestral***. Ejemplos de estadísticos muestrales son: $\bar{x}$ (media muestral), S^2 (varianza muestral), S (desviación típica muestral), etc.

La ***teoría del muestreo*** tiene por objetivo, el estudio de las relaciones existentes entre la distribución de un carácter en dicha población y las distribuciones de dicho carácter en todas sus muestras.

Las ventajas de estudiar una población a partir de sus muestras son principalmente:

- **Coste reducido:**

  Si los datos que buscamos los podemos obtener a partir de una pequeña parte del total de la población, los gastos de recogida y tratamiento de los datos serán menores. Por ejemplo, cuando se realizan encuestas previas a un referéndum, es más barato preguntar a 4.000 personas su intención de voto, que a 30.000.000;

- **Mayor rapidez:**

  Estamos acostumbrados a ver cómo con los resultados del escrutinio de las primeras mesas electorales, se obtiene una aproximación bastante buena del resultado final de unas elecciones, muchas horas antes de que el recuento final de votos haya finalizado;

- **Más posibilidades:**

  Para hacer cierto tipo de estudios, por ejemplo el de duración de cierto tipo de bombillas, no es posible en la práctica destruirlas todas para conocer su vida media, ya que no quedaría nada que vender. Es mejor destruir sólo una pequeña parte de ellas y sacar conclusiones sobre las demás.

De este modo se ve que al hacer estadística inferencial debemos enfrentarnos con dos problemas:

- Elección de la muestra *(muestreo)*, que es a lo que nos dedicaremos en este capítulo.
- Extrapolación de las conclusiones obtenidas sobre la muestra, al resto de la población *(inferencia).*

Es decir, el principal objetivo de la mayoría de los estudios, análisis o investigaciones, es hacer generalizaciones *"acertadas"* con base en muestras de poblaciones de las que se derivan tales muestras. Obsérvese la palabra *"acertadas"* porque no es fácil responder cuándo y en que condiciones las muestras permiten tales generalizaciones.

Por ejemplo si queremos calcular la cantidad de dinero que se gasta una persona en vacaciones, ¿tomaríamos como muestra lo que gastan los viajeros que lo hacen en primera clase? Es obvio que no, pero saber a que tipo de personas debemos incluir en nuestra muestra no es algo intuitivo ni evidente.

Sobre las muestras hay dos aspectos que resultan fundamentales: el tamaño (nº de elementos de la muestra) y la forma en que se realiza la selección de los individuos que la forman.

La elección de la muestra influirá de manera determinante en los resultados, por lo que hay que evitar muestras *"sesgadas"* (parciales y subjetivas) que produzcan errores incontrolables (aparte de los errores propios de sustituir el estudio de una población por el de una muestra).

El tipo de muestreo más importante es el ***muestreo aleatorio***, en el que todos los elementos de la población tienen la misma probabilidad de ser extraídos; Aunque dependiendo del problema y con el objetivo de reducir los costes o aumentar la precisión, otros tipos de muestreo pueden ser considerados como veremos a continuación.

2.- TIPOS DE MUESTREO

1.- Muestreo Aleatorio Simple

Consiste en seleccionar *n* elementos sin reemplazamiento de entre los *N* que componen la población, de tal modo que todas las muestras tengan la misma probabilidad de ser elegidas.

En la práctica, se enumeran los individuos y se sortean cuáles de ellos se elegirán. Si los individuos son, por ejemplo, tornillos que se encuentran en una caja, se eligen al azar por simple extracción.

2.- Muestreo Aleatorio Sistemático

Se elige un individuo al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra.

Por ejemplo si tenemos una población formada por 100 elementos y queremos extraer una muestra de 25 elementos, en primer lugar debemos establecer el intervalo de selección que será igual a 100/25=4. A continuación elegimos el elemento de arranque, tomando aleatoriamente un número entre el 1 y el 4, y a partir de él obtenemos los restantes elementos de la muestra:

2 , 6 , 10 , 14 , … , 98

3.- Muestreo Aleatorio Estratificado

Si conocemos que la población puede dividirse en partes o estratos, en relación con variables que pueden ser de interés en nuestro estudio, de modo que en cada uno de los estratos los elementos posean una gran homogeneidad respecto al carácter que se estudia (sexo, grupos de edad, nivel de estudios,…), se puede aumentar la precisión si muestreamos los estratos por separado. La forma de repartir los elementos de la muestra, determinando cuantos deben corresponder a cada estrato se denomina ***afijación***, y puede ser de varios tipos:

- ***Afijación uniforme***: Si se toma el mismo número de elementos en cada estrato.
- ***Afijación Proporcional***: si el número de elementos que se toma en cada estrato es proporcional al tamaño del estrato. Se utiliza cuando las varianzas de los estratos no difieren mucho entre si.

Los elementos de cada estrato se toman mediante muestreo aleatorio simple.

Ejemplo:

En una ciudad se quiere hacer un estudio para conocer qué tipo de actividades se realizan en el tiempo dedicado al ocio. Para ello van a ser encuestadas 300 personas elegidas al azar mediante muestreo aleatorio estratificado con afijación proporcional. Teniendo en cuenta que de un total de 15.000 habitantes, 7.500 son adultos, 3.000 ancianos y 4.500 niños, definir los estratos y el tamaño muestral correspondiente a cada estrato.

Solución:

Basta con hacer una simple regla de 3 para determinar el tamaño muestral de cada estrato:

Niños: 15.000 -------- 4.500

300 -------- x

$$x = \frac{4.500 * 300}{15.000} = 90 \text{ niños}$$

Adultos: $$x = \frac{7.500 * 300}{15.000} = 150 \text{ adultos}$$

Ancianos: $$x = \frac{3.000 * 300}{15.000} = 60 \text{ ancianos}$$

4.- Muestreo Por Conglomerados

En este tipo de muestreo, llamado ***muestreo por conglomerados***, se divide la población total en un número determinado de subdivisiones relativamente pequeñas y se seleccionan al azar algunas de estas subdivisiones o conglomerados, para incluirlos en la muestra total. Si estos conglomerados coinciden con áreas geográficas, este muestreo se llama también ***muestreo por áreas***.

Por ejemplo, supongamos que una gran empresa quiere estudiar los patrones variables de los gastos familiares de una ciudad como Granada. Al intentar elaborar los programas de gastos de una muestra de 1200 familias, nos encontramos con la dificultad de realizar un muestreo aleatorio simple, (es complicado tener una lista actualizada de todos los habitantes de una ciudad). Una manera de tomar una muestra en esta situación es dividir el área total (Granada en este caso) en áreas más pequeñas que no se solapen (Por ejemplo Distritos postales, manzanas etc...) En este caso seleccionaríamos algunas áreas al azar y todas las familias (o muestras de éstas) que residen en estos distritos postales o manzanas, constituirían la muestra definitiva.

Aunque las estimaciones basadas en el muestreo por conglomerados, por lo general no son tan fiables como las obtenidas por muestreos aleatorios simples del mismo tamaño, son más baratas. Volviendo al ejemplo anterior, es mucho más económico visitar a familias que viven en el mismo vecindario, que ir visitando a familias que viven en un área muy extensa.

En la práctica se pueden combinar el uso de varios de los métodos de muestreo que hemos analizados para un mismo estudio.

Ejercicios:

1.- Consideremos la población formada por 5 bolas contenidas en una urna y numeradas del 1 al 5. Obtener todas las muestras de tamaño 2 extraídas mediante muestreo aleatorio simple.

2.- En un instituto de enseñanza secundaria en que se ofertan los siguientes tipos de enseñanza:

- Ciclos de grado superior: 110 alumnos.
- Bachillerato: 162 alumnos.
- Ciclos de grado medio : 210 alumnos
- 2º ciclo de enseñanza secundaria obligatoria: 338 alumnos.

Se pretende valorar las faltas de ortografía que cometen los alumnos del centro mediante una prueba-dictado de un texto de 20 líneas; la prueba se pasará a una muestra de 50 alumnos, para minimizar el costo en tiempo y medios.

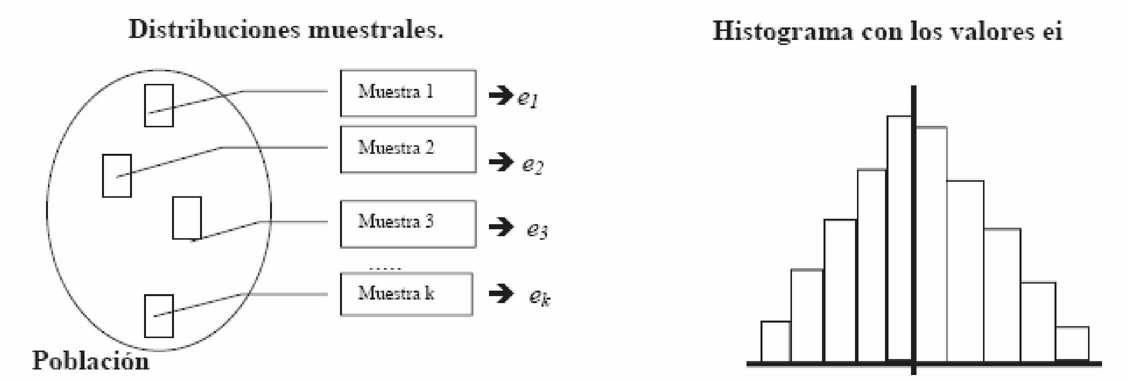
¿Cómo se obtiene una muestra adecuada a esta población?

3.- DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES

La idea de ***distribución muestral*** es la siguiente:

Partimos de una población de tamaño *N*. Obtenemos *k* muestras (todas las posibles de tamaño *n*) y de cada una de ellas se calcula una medida (media, mediana, varianza, desviación típica,..) obteniendo *k* valores: $e_1, e_2, e_3, ...e_k$.

Estos valores pueden representarse mediante un histograma para observar su distribución:



Se puede observar que este histograma va adquiriendo forma de campana de Gauss, y a medida que *k* aumenta, se va pareciendo cada vez más a la forma de una distribución Normal.

Lo vemos con un ***ejemplo***:

Elaboraremos la distribución muestral de la media de una muestra aleatoria de tamaño n = 2 tomada sin reemplazo de la población finita de tamaño N = 5, cuyos elementos son: 3,5,7,9,11.

La media de esta población es: $\mu = \dfrac{3+5+7+9+11}{5} = 7$

y su desviación típica es:

$$\sigma^2 = \frac{3^2+5^2+7^2+9^2+11^2}{5} - 7^2 = 8 \quad \Rightarrow \sigma = \sqrt{8}$$

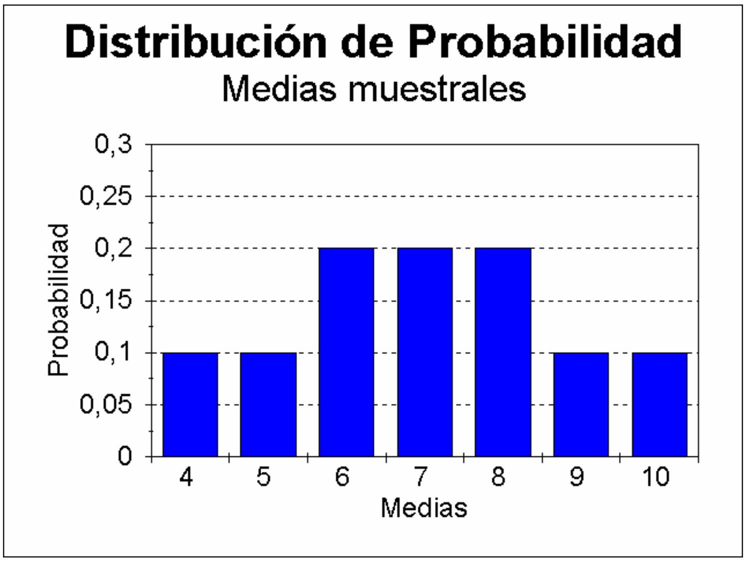
Ahora si tomamos una muestra aleatoria de tamaño n = 2 de esta población hay 20 posibilidades:

nº muestra	Muestras		$\overline{X}$ (media de la muestra)
1	3	5	4
2	3	7	5
3	3	9	6
4	3	11	7
5	5	3	4
6	5	7	6
7	5	9	7
8	5	11	8
9	7	3	5
10	7	5	6
11	7	9	8
12	7	11	9
13	9	3	4
14	9	5	7
15	9	7	8
16	9	11	10
17	11	3	7
18	11	5	8
19	11	7	9
20	11	9	10

Si hacemos la función de probabilidad de la variable $\overline{X}$ de las medias muestrales:

Media	Probabilidad
4	2/20
5	2/20
6	4/20
7	4/20
8	4/20
9	2/20
10	2/20

Cuyo histograma sería:



Si calculamos la media y la desviación típica de la distribución de las medias obtenemos que: $\mu_{\overline{X}} = 7$ y $\sigma_{\overline{X}} = \sqrt{3}$, luego la media μ_x coincide con la media de la población y la desviación típica ha disminuido.

Este ejemplo puede generalizarse para cualquier distribución según el siguiente teorema:

Teorema Central del Límite

Si una población tiene media μ y desviación típica σ, y tomamos muestras de tamaño n (n>30, ó cualquier tamaño si la población es "normal"), las medias de estas muestras siguen aproximadamente la distribución:

$$\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Destacar que si la población de la que se obtienen las muestras es *normal*, las medias muestrales también se distribuyen según una distribución normal, independientemente del tamaño de la muestra.

Al saber cómo funcionan las medias muestrales, podemos obtener probabilidades relacionadas con ellas.

Ejemplo 1:

Se sabe que las bolsas de azúcar producidas por una máquina tiene una media de 500 g. y una desviación típica de 35 g. Dichas bolsas se empaquetan en cajas de 100 unidades.

a) ¿Cómo se distribuyen las medias de los pesos de las bolsas de cada caja?

b) Calcular la probabilidad de que la media de los pesos de las bolsas de un paquete sea inferior a 495 g.

Solución:

a) Tenemos $\mu = 500$, $\sigma = 35$, $n = 100$

Por el Teorema Central del Límite (aunque la variable peso no sea normal, tenemos n > 30), la variable $\overline{X}$ = “media de los pesos de la muestra” se distribuye según una normal:

$$\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(500, \frac{35}{\sqrt{100}}\right) = N(500, 3'5)$$

b) $$P(\overline{X} < 495) = tipificando = P\left(\frac{\overline{X} - 500}{3'5} < \frac{495 - 500}{3'5}\right) = P(Z < -1'43) =$$
$$= P(Z > 1'43) = 1 - P(Z < 1'43) = 0'0764$$

Lo que vendría a significar que en el 7’6 % de las cajas el peso medio de las bolsas es inferior a 495 g.

Ejemplo 2:

Las estaturas, en centímetros, de un grupo de soldados se distribuyen normalmente con media 173 y desviación típica 6.

a) Elegido un soldado al azar, ¿cuál es la probabilidad de que mida menos de 175 cm.?

b) Si se toma una muestra de 12 soldados, ¿cuál es la probabilidad de que su estatura media supere el 1'76?

Solución:

a) Cuidado, en este apartado no hay muestra, ni por tanto medias muestrales, sino un simple ejercicio de la normal.
Si X = "altura", sabemos que sigue una N(173,6)
Luego

$$P(X < 175) = tipificando = P\left(\frac{X-173}{6} < \frac{175-173}{6}\right) =$$
$$= P(Z < 0'33) = 0'6293$$

b) Aquí si tenemos una muestra con *n = 12*
Aunque sea n < 30, como la población de partida es normal, podemos también aplicar el Teorema Central del Límite:

$$\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(173, \frac{6}{\sqrt{12}}\right) = N(173{,}1'73)$$

Luego:

$$P(\overline{X} > 176) = tipificando) P\left(\frac{\overline{X}-173}{1'73} > \frac{176-173}{1'73}\right) = P(Z > 1'73) =$$
$$= 1 - P(Z < 1'73) = 1 - 0'9582 = 0'0418$$

Ejercicios:

1.- Se supone que la distribución de la temperatura del cuerpo humano en la población tiene de media 37º y de desviación típica 0'85º. Se elige una muestra de 105 personas. Hallar las probabilidades de que:

a) La media sea menor que 36'9º

b) La media esté comprendida entre 36'5º y 37'2º

2.- Las notas de cierto examen se distribuyen según una normal de media 5,8 y desviación típica 2,4. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7

3.- Consideremos la población formada por los cuatro elementos 0, 3, 4, 6. Hallar:

a) Todas las muestras posibles de tamaño 2 extraídas mediante muestreo aleatorio simple

b) La media y la desviación típica poblacionales

c) La media y la desviación típica de las medias muestrales

4.- DISTRIBUCIÓN DE LAS PROPORCIONES MUESTRALES

En lugar de calcular medias de las muestras, ahora vamos a trabajar con proporciones.

En una población, la proporción de individuos que poseen una determinada característica es p. (Llamaremos $q = 1 - p$)
Extraemos todas las posibles muestras de tamaño n que podemos extraer de esa población. La proporción de individuos de cada una de esas muestras con esa característica ser $\hat{p}$.
Llamaremos $\hat{P}$ a la variable aleatoria que toma los distintos valores de esas proporciones muestrales.

Si n es lo suficientemente grande ($n > 30$), se puede demostrar que la variable $\hat{P}$ sigue una distribución normal de parámetros:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right)$$

(Esta fórmula proviene de la aproximación de una binomial por una normal)

Ejemplo 1:

Una nueva droga ha curado al 85% de los enfermos a los que se les ha aplicado. Si se toman muestras de 30 personas, ¿Cuál es la distribución de las proporciones muestrales? ¿Y si las muestras son de 100 personas? ¿Y si son de 1000?

Solución:
Tenemos $p = 0'85$, por lo que $q = 0'15$

Si $n = 30$:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right) \Rightarrow \hat{P} \to N\left(0'85, \sqrt{\frac{0'85 \cdot 0'15}{30}}\right) \Rightarrow \hat{P} \to N(0'85, 0'065)$$

Si $n = 100$:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right) \Rightarrow \hat{P} \to N\left(0'85, \sqrt{\frac{0'85 \cdot 0'15}{100}}\right) \Rightarrow \hat{P} \to N(0'85, 0'036)$$

Si $n = 1000$:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right) \Rightarrow \hat{P} \to N\left(0'85, \sqrt{\frac{0'85 \cdot 0'15}{1000}}\right) \Rightarrow \hat{P} \to N(0'85, 0'011)$$

Como ya vimos con las medias muestrales, al aumentar el tamaño de las muestras disminuye la varianza de las distribuciones muestrales.

Ejemplo 2:

Se sabe que el 15 % de los jóvenes entre 18 y 25 años son miopes.

a) ¿Cómo se distribuye la proporción de jóvenes miopes en muestras de 40 individuos?

b) ¿Cuál es la probabilidad de que en dicha muestra la proporción de miopes esté entre el 8 y el 22 %?

Solución:

a) Tenemos *p = 0'15*, por lo que *q = 0'85*. Además *n = 40*. Por tanto:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right) \Rightarrow \hat{P} \to N\left(0'15, \sqrt{\frac{0'15 \cdot 0'85}{40}}\right) \Rightarrow \hat{P} \to N(0'15, 0'0565)$$

b) Usando el apartado anterior:

$$P\left(0'08 < \hat{P} < 0'22\right) = tipificando = P\left(\frac{0'08 - 0'15}{0'0565} < \hat{P} < \frac{0'22 - 0'15}{0'0565}\right) =$$
$$= P\left(-1'24 < \hat{P} < 1'24\right) = P\left(\hat{P} < 1'24\right) - P\left(\hat{P} < -1'24\right) = (*)$$

Calculamos

$$P\left(\hat{P} < -1'24\right) = P\left(\hat{P} > 1'24\right) = 1 - P\left(\hat{P} < 1'24\right) = 1 - 0'8925 = 0'1075$$

Y por tanto:
(*) = 0'8925 – 0'1075 = 0'785

Esto significaría que en el 78'5 % de las muestras que extrajésemos de tamaño 40, la proporción de individuos miopes estaría entre el 8 y el 22%.

Ejercicios:

1.- Una máquina fabrica piezas de precisión y en su producción habitual tiene un 3% de piezas defectuosas. Se empaquetan en cajas de 200, ¿cuál es la probabilidad de encontrar entre 5 y 7 piezas defectuosas en una caja?

2.- Si tiramos una moneda no trucada 100 veces, ¿cuál es la probabilidad de que obtengamos más de 55 caras?

3.- Se ha realizado una encuesta a 20.000 universitarios sobre su actitud ante el botellón. De ellos, 13.200 son partidarios y el resto no. Se toman 100 muestras de 30 universitarios cada una. Hallar:

a) La distribución de la proporción muestral de partidarios del botellón

b) La probabilidad de que en una de estas muestras se manifiesten favorables al botellón más de 21 alumnos

c) ¿En cuántas muestras cabe esperar que más de 15 y menos de 19 estudiantes se muestren partidarios del botellón?

EJERCICIOS

1.- En el instituto de cierta localidad se imparten 4 niveles diferentes: 1º, 2º, 3º y 4º ESO, estando matriculados un total de 800 alumnos. En 1º hay 160 alumnos, en 2º hay 240 y en 3º hay 208. Explica cómo obtener una muestra de 50 alumnos mediante:

a) Muestreo aleatorio sistemático
b) Muestreo aleatorio estratificado con afijación proporcional

2.- En un barrio hay 4000 habitantes, distribuidos en cuatro urbanizaciones: el 12% viven en A, el 20% en B, el 36% en C y el 32% en D. Además, los porcentajes de mujeres en cada urbanización son: 50, 60, 66 y 75 % respectivamente. Se desea obtener una muestra de 50 habitantes que sea representativa en cuanto al sexo y a las urbanizaciones mediante muestreo aleatorio estratificado proporcional. ¿Cuántas personas (hombres y mujeres) de cada urbanización habrá que seleccionar?

3.- Dada la población {2, 4, 6, 8}:

a) Escribir todas las muestras posibles de tamaño 2 mediante muestreo aleatorio simple
b) Calcular la media y la varianza de las medias muestrales
c) Hacer lo mismo con todas las muestras posibles de tamaño 3

4.- La media de edad de los lectores de una determinada revista es de 17’2 años y la desviación típica de 2’3 años. Si elegimos muestras de 100 lectores:

a) ¿Cuál es la distribución de las medias muestrales?
b) ¿Cuál es la probabilidad de que la media de edad de la muestra esté comprendida entre 16’7 y 17’5 años?

5.- En una determinada población, los pesos se distribuyen según una normal de media 65 kg. y varianza 49. Si extraemos muestras de tamaño 64, ¿cuál es la probabilidad de que la media de los pesos de una de esas muestras sea mayor que 66’5 kg?

6.- La edad de los miembros de una determinada asociación se distribuye normalmente con media 52 años y desviación típica 3 años.

a) ¿Cuál es la probabilidad de que un miembro elegido al azar sea mayor de 60 años?
b) Si se toman muestras de tamaño 36, ¿Cuál es la probabilidad de que la edad media de dichas muestras sea superior a los 60 años?

7.- La edad de los alumnos de 2º Bachillerato de cierto instituto sigue una distribución *N(17’6 , 0’5)*. Los agrupamos al azar de 10 en 10 para una competición. ¿Cuál es la probabilidad de que en uno de esos grupos la edad media esté comprendida en el intervalo (17,18)?

8.- La duración (en horas) de un determinado tipo de pilas sigue una distribución normal con $\mu = 50$ y $\sigma = 5$. Empaquetamos las pilas en cajas de 16. ¿Cuál es la probabilidad de que la duración media de las pilas de una de las cajas sea inferior a 48 horas?

9.- En un instituto A las pruebas de acceso a la Universidad han obtenido una media de 5'8 con una desviación típica de 1'25, mientras que en otro instituto B la media ha sido de 5'6 con una desviación típica de 1'5. Si se toman al azar muestras de 100 personas de cada instituto, ¿en cuál de ellos es más probable que la nota media sea superior al 5'7?

10.- El peso de las truchas de una piscifactoría sigue una ley *N(200,50)*. Se toman muestras de 60 truchas y se calcula su peso medio. Hallar las probabilidades de que la media muestral:

a) Sea mayor que 210g
b) Sea menor que 185g
c) Esté entre 210 y 225g

11.- Una máquina produce el 5% de tornillos defectuosos. Si se empaquetan en cajas de 400 tornillos:

a) ¿Cuál es la distribución de la proporción de tornillos defectuosos en las cajas?
b) ¿Cuál es la probabilidad de que dicha proporción supere el 3%?
c) ¿Y de que esté entre el 3 y el 6%?

12.- En una población el 35% son fumadores. Se toman muestras de 60, 100 y 200 personas. ¿Cuál será la distribución de la proporción de fumadores en cada una de esas muestras?

13.- En el 77% de los hogares españoles hay teléfono móvil. Si se extrae una muestra de 500 viviendas, ¿Cuál es la probabilidad de que en más del 80% haya teléfono móvil? ¿Y si la muestra se toma de 1000 viviendas?

14.- Tras unas elecciones, se sabe que el candidato elegido obtuvo el 42% de los votos. Hallar la probabilidad de que de 1000 individuos elegidos al azar de entre los votantes hubiese obtenido dicho candidato más de 450 votos

15.- El 17% de familias de una determinada población tienen 3 o más hijos. Encuestadas 150 familias, ¿Cuál es la probabilidad de que más de 25 de ellas tengan 3 o más hijos?

16.- La duración de las baterías de determinados coches eléctricos se distribuye normalmente con media 225 y desviación típica 24 minutos. Si se eligen al azar 10 baterías, ¿Cuál es la probabilidad de que su duración media esté entre 220 y 235 minutos?

17.- Se ha determinado que 60% de los estudiantes de una universidad grande fuman cigarrillos. Se toma una muestra aleatoria de 800 estudiantes. Calcule la probabilidad de que la proporción de la muestra de la gente que fuma cigarrillos sea menor que 0.55.

18.- Se sabe que la verdadera proporción de los componentes defectuosos fabricados por una firma es de 4%, y encuentre la probabilidad de que una muestra aleatoria de tamaño 60 tenga:

a) Menos del 3% de los componentes defectuosos.

b) Más del 1% pero menos del 5% de partes defectuosas.

19.- Las estaturas de 1000 estudiantes están distribuidas aproximadamente en forma normal con una media de 174.5 centímetros y una desviación estándar de 6.9 centímetros. Si se extraen 200 muestras aleatorias de tamaño 25 sin reemplazo de esta población, determine:

a) El número de medias muestrales que caen entre 172’5 y 175’8 cm.
b) El número de medias muestrales que caen por debajo de 172 cm.

20.- Un medicamento para malestar estomacal tiene la advertencia de que algunos usuarios pueden presentar una reacción adversa a él, más aún, se piensa que alrededor del 3% de los usuarios tienen tal reacción. Si una muestra aleatoria de 150 personas con malestar estomacal usa el medicamento, encuentre la probabilidad de que la proporción de la muestra de los usuarios que realmente presentan una reacción adversa exceda el 4%.

TEMA 4.- INTERVALOS DE CONFIANZA

1.- INTRODUCCIÓN

En una población cuya distribución es conocida pero desconocemos algún parámetro, podemos estimar dicho parámetro a partir de una muestra representativa.

Un ***estimador*** es un valor que puede calcularse a partir de los datos muestrales y que proporciona información sobre el valor del parámetro. Por ejemplo la media muestral es un estimador de la media poblacional, la proporción observada en la muestra es un estimador de la proporción en la población,…

Una estimación es ***puntual*** cuando se obtiene un sólo valor para el parámetro. Los estimadores más probables en este caso son los estadísticos obtenidos en la muestra, aunque es necesario cuantificar el riesgo que se asume al considerarlos. Por ejemplo, es obviamente inútil concluir que si el sueldo medio de una muestra de una ciudad es de 1.240 € entonces el sueldo medio de los habitantes de dicha ciudad también será ése. Lógicamente la posibilidad de equivocarnos es demasiado grande.

Más útil es la ***estimación mediante intervalos de confianza***, que consiste en determinar un posible rango de valores o intervalo, en los que pueda precisarse, con una determinada probabilidad, que el valor de un parámetro de la población se encuentra dentro de esos límites. Este parámetro será habitualmente una proporción en el caso de variables dicotómicas, y la media para distribuciones normales.

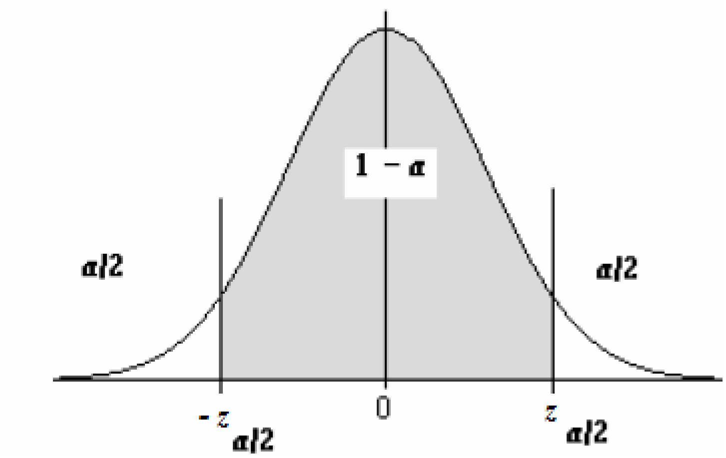
Evidentemente esta técnica no tiene porqué dar un resultado correcto. A la probabilidad de que hayamos acertado al decir que el parámetro estaba contenido en dicho intervalo se la denomina ***nivel de confianza:***

> ***Nivel de confianza*** es la "probabilidad" de que el intervalo calculado contenga al verdadero valor del parámetro. Se indica por $1-\alpha$ y habitualmente se da en porcentaje $(1-\alpha)100\%$ (Hablaremos de un nivel de confianza del 90%, del 95%, del 99%,…). Hablamos de nivel de confianza y no de probabilidad ya que una vez extraída la muestra, el intervalo de confianza contendrá al verdadero valor del parámetro o no, lo que sabemos es que si repitiésemos el proceso con muchas muestras podríamos afirmar que el $(1-\alpha)\%$ de los intervalos así construidos contendría al verdadero valor del parámetro.
>
> A la probabilidad de equivocarnos se le denomina **nivel de significación**, y lo representamos por α.
>
> Lógicamente, cuanto más pequeño sea α (es decir, cuanto más grande sea el nivel de confianza), la probabilidad de equivocarnos será menor, pero el intervalo que calcularemos será más grande y por tanto la precisión de la estimación será menor.

Se trata pues de encontrar un equilibrio entre que la probabilidad de equivocarnos no sea muy grande y que el intervalo tampoco para obtener mayor precisión. Se suelen para ello prefijar niveles de confianza superiores al 90%.

Dado un nivel de confianza, $1-\alpha$, se llama ***valor crítico*** $z_{\alpha/2}$ al valor que en una N(0,1) cumple que: $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$

Es decir:



Para calcular el valor crítico tenemos en cuenta que si $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$, entonces $P\left(Z \geq z_{\alpha/2}\right) = \dfrac{\alpha}{2}$ (Ver dibujo) y por tanto $P\left(Z \leq z_{\alpha/2}\right) = 1-\dfrac{\alpha}{2}$ y eso lo podemos buscar en la tabla de la N(0,1).

Vemos un ejemplo práctico de cómo calcular el valor crítico:

Ejemplo: Calcular el valor crítico correspondiente a un nivel de confianza del 99%.

Como $1-\alpha = 0'99 \;\Rightarrow\; \alpha = 0'01 \;\Rightarrow\; \dfrac{\alpha}{2} = 0'005 \;\Rightarrow\; 1-\dfrac{\alpha}{2} = 0'995$
Y buscamos ahora en la tabla el valor de z que deja a la izquierda una probabilidad de 0’995, obteniendo:

$$z_{\alpha/2} = 2'575$$

Ejercicio: Calcular los valores críticos para niveles de confianza del 90, 92, 95, 98 y 99’5%.

2.- INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL

Supongamos que la población de partida es $N(\mu,\sigma)$, y queremos estimar mediante un intervalo la media de la población, μ, que es desconocida. Para ello escogemos una muestra aleatoria de tamaño n y calculamos la media muestral, $\bar{x}$.

Como vimos en el tema anterior, la media muestral tiene una distribución conocida:

$$\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Y por tanto, tipificando: $$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$$

Fijado un nivel de confianza, $1-\alpha$, queremos dos valores tales que la probabilidad de que la media de la población, μ, se encuentre entre ellos sea precisamente $1-\alpha$.

Si nos fijamos en la definición de valor crítico:

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1-\alpha$$

De donde $$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

Despejando: $$P\left(-z_{\alpha/2} \cdot \sigma/\sqrt{n} \le \bar{X}-\mu \le z_{\alpha/2} \cdot \sigma/\sqrt{n}\right) = 1-\alpha$$

Y por tanto: $$P\left(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} \le \mu \le \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}\right) = 1-\alpha$$

Es decir:

El intervalo de confianza para el parámetro μ de una población $N(\mu,\sigma)$ al nivel de confianza $1-\alpha$ viene dado por:

$$\boxed{\left(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}\ ,\ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}\right)}$$

Si σ es desconocida, se sustituye por la desviación típica de la muestra, s.

Nota: tenemos que tener en cuenta que, o bien $n \geq 30$, o bien la distribución de partida es normal, pues sólo así conocemos la distribución de las medias muestrales que es en lo que nos basamos para calcular el intervalo de confianza.

Ejemplo 1:

Se sabe que la desviación típica de las tallas de los alumnos de una universidad es de 5cm. Se desea estimar la talla media de dichos alumnos, para lo que se escoge una muestra de 100 estudiantes y se obtiene que la media muestral es de 172cm. Hallar el intervalo de confianza para la talla media de la universidad para los niveles de confianza del 90 y del 95%.

Solución:
Tenemos $\sigma = 5$, n = 100 y $\bar{x} = 172$

Calculamos el valor crítico para el nivel de confianza del 90%:

$$1-\alpha = 0'90 \quad \Rightarrow \quad \alpha = 0'1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'05 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 0'95$$

Y buscamos ahora en la tabla el valor de z que deja a la izquierda una probabilidad de 0’95, obteniendo (aprox.):

$$z_{\alpha/2} = 1'64$$

Sustituyendo en la fórmula del intervalo de confianza:

$$\left(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} \ , \ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \right)$$

$$172 - 1'64 \cdot \frac{5}{\sqrt{100}} = 172 - 0'82 = 171'18$$

$$172 + 1'64 \cdot \frac{5}{\sqrt{100}} = 172 + 0'82 = 172'82$$

Luego el intervalo de confianza para el 90% será: $IC_{90} = (171'18, 172'82)$

Hacemos lo mismo para el nivel de confianza del 90%:

$$1-\alpha = 0'99 \quad \Rightarrow \quad \alpha = 0'01 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'005 \quad \Rightarrow \quad 1-\frac{\alpha}{2} = 0'995$$

Y buscamos ahora en la tabla el valor de z que deja a la izquierda una probabilidad de 0’95, obteniendo (aprox.):

$$z_{\alpha/2} = 2'58$$

Sustituyendo en la fórmula del intervalo de confianza:

$$\left(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} \ , \ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \right)$$

$$172 - 2'58 \cdot \frac{5}{\sqrt{100}} = 172 - 1'29 = 170'71$$

$$172 + 2'58 \cdot \frac{5}{\sqrt{100}} = 172 + 1'29 = 173'29$$

Luego el intervalo de confianza para el 99% será: $IC_{99} = (170'71, 173'29)$

Obsérvese que a mayor nivel de confianza mayor es el intervalo, con lo que la precisión de la estimación es menor.

Ejemplo 2:

Un psicólogo escolar quiere estimar la media de tiempo de reacción a un determinado estímulo de los alumnos de 1º de Primaria. Para ello ha elegido una muestra de 35 niños obteniendo un tiempo medio de 1'12 minutos y una desviación típica de 0'21 minutos. Hallar el intervalo de confianza para el tiempo medio de reacción con un nivel de significación del 8%.

Solución:
Como *n* es grande, podemos usar la fórmula del intervalo de confianza usando la desviación típica muestral pues desconocemos la de la población:

$$\left(\overline{X} - z_{\alpha/2} \cdot s/\sqrt{n} \,,\, \overline{X} + z_{\alpha/2} \cdot s/\sqrt{n} \right)$$

Tenemos $s = 0'21$, n = 35 y $\bar{x} = 1'12$

Calculamos el valor crítico para el nivel de confianza del 92% (pues $\alpha = 0'08$):

$$1 - \alpha = 0'92 \quad \Rightarrow \quad \alpha = 0'08 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'04 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 0'96$$

Y buscamos ahora en la tabla el valor de *z* que deja a la izquierda una probabilidad de 0'96, obteniendo (aprox.):

$$z_{\alpha/2} = 1'75$$

Luego:

$$1'12 - 1'75 \cdot \frac{0'21}{\sqrt{35}} = 1'12 - 0'06 = 1'06$$

$$1'12 + 1'75 \cdot \frac{0'21}{\sqrt{35}} = 1'12 + 0'06 = 1'18$$

Y por tanto el intervalo de confianza será: $IC_{92} = (1'06, 1'18)$

Error Máximo Admisible y Tamaño de la Muestra

Si observamos la fórmula obtenida para el intervalo de confianza:

$$\left(\overline{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} \ , \ \overline{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \right)$$

la media muestral siempre será el centro de dicho intervalo, mientras que su amplitud depende del valor $E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Con un nivel de confianza del $(1-\alpha)100\%$ admitimos que la diferencia entre la estimación para la media de la población a partir de la muestra y su valor real es menor que *E*, que llamaremos ***error máximo admisible***.

El ***tamaño de la muestra*** depende del nivel de confianza que se desee para los resultados y de la amplitud de intervalo de confianza, es decir, del error máximo que se esté dispuesto a admitir.

Fijados $1-\alpha$ y *E*, podemos calcular el tamaño mínimo de la muestra que emplearemos despejando de la expresión de *E*:

$$E = z_{\alpha/2} \cdot \sigma/\sqrt{n} \quad \Rightarrow \quad \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \quad \Rightarrow \quad n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Notas:

- *A mayor tamaño de la muestra, menor es el error*
- *A mayor nivel de confianza, mayor es el error*
- *A mayor nivel de confianza, mayor tamaño de la muestra (con un error fijo)*

Ejemplo:

Se quiere estimar el peso medio de las truchas de una piscifactoría. Por estudios previos, se sabe que la desviación típica del peso de las truchas es de 45 gramos. Se quiere construir un intervalo de confianza al 99% sin que el error de la estimación supere los 4’1 gramos. ¿Cómo deberá ser de grande la muestra?

Solución:

Los datos son:

E = 4’1 , $\sigma = 45$, $1-\alpha = 0'99$

Calculando el valor crítico correspondiente (hacerlo): $z_{\alpha/2} = 2'58$

Sustituyendo en la fórmula del error (no merece la pena aprenderse la fórmula del tamaño ya despejado):

$$E = z_{\alpha/2} \cdot \sigma/\sqrt{n} \quad \Rightarrow \quad \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \quad \Rightarrow \quad \sqrt{n} = \frac{2'58 \cdot 45}{4'1} = 28'317 \quad \Rightarrow \quad n = 801'85$$

Por lo que debemos tomar una muestra de *802* truchas (si tomásemos 801, el error sería superior al permitido. Esta aproximación siempre se hace por encima, aunque hubiera dado 801’01).

Ejercicios:

1.- La media de las estaturas de una muestra aleatoria de 400 personas de una ciudad es 1,75 m. Se sabe que la estatura de las personas de esa ciudad es una variable aleatoria que sigue una distribución normal con varianza $\sigma^2 = 0{,}16\ m^2$.

a) Construye un intervalo, de un 95% de confianza, para la media de las estaturas de la población.

b) ¿Cuál sería el mínimo tamaño muestral necesario para que pueda decirse que la verdadera media de las estaturas está a menos de 2 cm de la media muestral, con un nivel de confianza del 90%?

2.- Un sociólogo está estudiando la duración del noviazgo en una extensa área rural. Se tomó una muestra aleatoria formada por 56 familias y se obtuvo que la duración media de su noviazgo fue 3'4 años, con una desviación típica de 1'2 años.

a) Hallar un intervalo de confianza para la duración media del noviazgo para la población de familias en dicha área rural al nivel de confianza del 85%.

b) Repetir el apartado anterior para niveles del 95 y del 99% y comparar los intervalos obtenidos.

c) ¿Cuál debería ser el tamaño de la muestra para estar seguro al nivel del 90% de que error máximo cometido es de 0'05?

3.- Las ventas mensuales de una tienda de electrodomésticos se distribuyen según una ley normal, con desviación típica 900 €. En un estudio estadístico de las ventas realizadas en los últimos nueve meses, se ha encontrado un intervalo de confianza para la media mensual de las ventas, cuyos extremos son 4.663 € y 5.839 €.

a) ¿Cuál ha sido la media de las ventas en estos nueve meses?

b) ¿Cuál es el nivel de confianza para este intervalo?

3.- INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

Deseamos ahora estimar la proporción *p* con la que una determinada característica se da en una población. Para ello extraemos una muestra de tamaño *n* y obtenemos la proporción muestral, es decir,

$$\hat{p} = \frac{\textit{número de individuos que cumplen la característica}}{\textit{tamaño de la muestra}}$$

Como vimos en el tema anterior, la distribución de las proporciones muestrales es:

$$\hat{P} \to N\left(p, \sqrt{\frac{pq}{n}}\right)$$

donde *q=1 – p*

Dado un nivel de confianza, $1-\alpha$, y haciendo lo mismo que en el caso de la media, se obtiene el siguiente intervalo de confianza para la proporción de la población:

$$\boxed{\left(\hat{p}-z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p}+z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)}$$

Ejemplo:

Tomando al azar una muestra de 300 personas mayores de 15 años en una gran ciudad, se encuentra que 104 de ellas leían el periódico habitualmente. Hallar, con un nivel de confianza del 90%, un intervalo para estimar la proporción de lectores de periódico entre los habitantes de esa ciudad mayores de 15 años.

Solución:

La proporción muestral es: $\hat{p}=\frac{104}{300}=0'347$

El valor crítico para un nivel de confianza del 90%:

$1-\alpha=0'90 \quad\Rightarrow\quad \alpha=0'1 \quad\Rightarrow\quad \frac{\alpha}{2}=0'05 \quad\Rightarrow\quad 1-\frac{\alpha}{2}=0'95$

Luego $z_{\alpha/2}=1'64$

Luego sustituyendo en la fórmula: $\left(\hat{p}-z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\ ,\ \hat{p}+z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

$$0'347-1'64\cdot\sqrt{\frac{0'347\cdot 0'653}{300}}=0'347-0'045=0'302$$

$$0'347+1'64\cdot\sqrt{\frac{0'347\cdot 0'653}{300}}=0'347+0'045=0'392$$

Luego el intervalo pedido es $IC_{90}=(0'302\,,0'392)$

<u>Error Máximo Admisible y Tamaño de la Muestra</u>

Los conceptos y notas ha tener en cuenta son los mismos que en los intervalos de confianza para la media, con los cambios obvios en las fórmulas correspondientes.

En cuanto al error:

$$E=z_{\alpha/2}\cdot\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Y en cuanto al tamaño de la muestra

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \Rightarrow \quad \sqrt{\frac{\hat{p}\hat{q}}{n}} = \frac{E}{z_{\alpha/2}} \quad \Rightarrow \quad \frac{\hat{p}\hat{q}}{n} = \left(\frac{E}{z_{\alpha/2}}\right)^2 \quad \Rightarrow$$

$$n = \frac{\hat{p}\hat{q}}{\left(\frac{E}{z_{\alpha/2}}\right)^2} \quad \Rightarrow \quad n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Ejemplo:

Una empresa dedicada a la venta de palomitas compra el maíz directamente a los agricultores. Antes de efectuar la compra, un agente de la compañía quiere estimar la probabilidad de que el grano de maíz se abra al freírlo. Ha realizado un estudio sobre una pequeña muestra de 60 granos, de los que 48 sea abrían. ¿Cuántos granos deberá examinar para estar seguro al nivel del 90% de que el error que cometa no superará el 1%?

Solución:

Sabemos que $\hat{p} = \frac{48}{60} = 0'8$

Además $z_{\alpha/2} = 1'64$ (hecho en el ejemplo anterior)

De la fórmula del error:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \Rightarrow \quad 0'01 = 1'64 \cdot \sqrt{\frac{0'8 \cdot 0'2}{n}} \quad \Rightarrow \quad 0'0061 = \sqrt{\frac{0'16}{n}}$$

$$0'000037 = \frac{0'16}{n} \quad \Rightarrow \quad n = \frac{0'16}{0'000037} = 4324'32$$

Luego para que el error no supere el 1% debería tomar una muestra de 4.325 granos de maíz (siempre por exceso).

Ejercicio:

¿De qué tamaño habría que elegir una muestra para estimar la proporción de alumnos del instituto que le gusta el fútbol con un nivel de confianza del 95% y un error inferior a 0.05, si en una muestra de 10 alumnos, 6 de ellos respondieron que les gustaba el fútbol?

EJERCICIOS

1.- En una población una variable aleatoria sigue una ley normal de media desconocida y desviación típica 2.

a) Observada una muestra de tamaño 400, tomada al azar, se ha obtenido una media muestra al igual a 50. Calcule un intervalo, con el 97 % de confianza, para la media de la población.

b) Con el mismo nivel de confianza, ¿qué tamaño mínimo debe tener la muestra para qué la amplitud del intervalo que se obtenga sea, como máximo, 1?

2.- El número de horas que duermen los estudiantes de bachillerato de la comunidad autónoma de Andalucía sigue una distribución normal con varianza 9.

a) A partir de una muestra de tamaño 30, se ha obtenido una media muestral igual a 7 horas. Halla un intervalo de confianza al 96% para la media de horas de sueño de dicha población.

b) ¿Qué tamaño de muestra sería necesario si el error de estimación deseamos que sea menor de 30 minutos con un nivel de significación de 0.05?

3.- Tomada al azar una muestra de 120 estudiantes de una Universidad, se encontró que 54 de ellos hablaban inglés. Halle, con un nivel de confianza del 90%, un intervalo para estimar la proporción de estudiantes de esa Universidad que hablan inglés.

4.- Se sabe que el contenido de fructosa de cierto alimento sigue una distribución normal, cuya varianza es conocida, teniendo un valor de 0,25. Se desea estimar el valor de la media poblacional mediante el valor de la media de una muestra, admitiéndose un error máximo de 0,2 con una confianza del 95%. ¿Cuál ha de ser el tamaño de la muestra?

5.- Los pesos en una determinada población siguen una distribución normal de media desconocida y desviación típica igual a 5 kg. Pesando a l0 individuos de dicha población, se obtuvieron los siguientes resultados medidos en kilogramos:

62	65	63	58	64	60	57	62	60	58

Halla un intervalo de confianza al 90% para el peso medio de la población

6.- En un determinado barrio se seleccionó al azar una muestra de 100 personas cuya media de ingresos mensuales resultaba igual a 1.060 € con una desviación típica de 200 €.

a) Si se toma un nivel de confianza del 95%, ¿cuál es el intervalo de confianza para la media de los ingresos mensuales de toda la población?

b) Si se toma un nivel de significación igual a 0,01, ¿cuál es el tamaño muestral necesario para estimar la media de ingresos mensuales con un error menor de 30 €?

7.- Para estimar la estatura media de los jóvenes entre 15 y 25 años de una localidad se ha medido a 40 de ellos, obteniendo los siguientes resultados:

ESTATURA (cm)	[148, 153)	[153, 158)	[158, 163)	[163, 168)	[168, 173)	[173, 178)
Nº JÓVENES	2	4	11	14	5	4

Estima, con un nivel de confianza del 99%, el valor de la estatura media de los jóvenes entre 15 y 25 años de dicha localidad

8.- Se desea estimar la proporción p de individuos daltónicos de una población a través del porcentaje observado en una muestra aleatoria de individuos de tamaño n.

a) Si el porcentaje de individuos daltónicos en la muestra es igual al 30%, calcula el valor de n para que, con un nivel de confianza de 0,95, el error cometido en la estimación sea inferior al 3,1%.

b) Si el tamaño de la muestra es de 64 individuos y el porcentaje de individuos daltónicos en la muestra es del 35%, determina, usando un nivel de significación del 1%, el correspondiente intervalo de confianza para la proporción de daltónicos de la población. ¿Qué error máximo cometeremos con esta estimación?

9.- A partir de la información suministrada por una muestra aleatoria de 100 familias de cierta ciudad se ha determinado el intervalo de confianza al 99% para el gasto medio mensual por familia (en euros) en electricidad, obteniéndose el intervalo (42,58). Determinar, justificando las respuestas:

a) La media de gasto en la muestra y en error cometido en la estimación realizada.

b) ¿Qué número de familias tendríamos que seleccionar al azar como mínimo para garantizarnos, con una confianza del 99%, una estimación de dicho gasto medio con un error máximo no superior a 3 euros?

10.- Una muestra de 100 votantes, elegidos al azar entre todos los de un distrito, indicó que el 55% de ellos estaba a favor de un candidato determinado. Halla el intervalo de confianza al 99'73% para la proporción de todos los votantes del distrito que estaban a favor del candidato.

11.- La edad de los alumnos que se presentan a las pruebas de acceso a la universidad sigue una distribución normal con varianza 0'36. Deseamos estimar la edad media de dichos estudiantes con un error menor de 0'2 años y con una confianza del 99'5%. ¿De qué tamaño, como mínimo, debemos seleccionar la muestra? ¿Y si no queremos que el error supere los 0'1 años? ¿Cuál sería el error cometido si la muestra tuviera 40 individuos?

12.- Sea X una población normal. Se sabe que, a partir de una muestra de 40 unidades, se saca un media muestral igual a 100 y una varianza muestral igual a 4. Calcular el intervalo de confianza para la media al 95% y 99%, sabiendo que la varianza de la población es igual a 36.
¿Que pasaría en el cálculo precedente si no se supiera nada cerca de la distribución de X?

13.- En una muestra de 300 universitarios 240 de ellos respondieron que asisten semanalmente al cine. Entre qué valores se encuentra, con un nivel de confianza del 95%, la proporción de universitarios que acude semanalmente al cine.
¿Cuáles son esos valores si el nivel de confianza es del 98%?

14.- El tiempo, en minutos, que esperan los clientes de un determinado banco hasta que son atendidos sigue una distribución normal de media desconocida y varianza 9. Los tiempos que esperaron 10 clientes elegidos al azar fueron:
1’5 , 2 , 2’5 , 3 , 1 , 5 , 5’5 , 4’5 , 3 , 3
Determinar el intervalo de confianza al 95% para el tiempo medio de espera.

15.- En una población una variable aleatoria sigue una ley normal de media desconocida y desviación típica 2. Observada una muestra de 400 tomada al azar, el intervalo de confianza obtenido fue *(49’78 , 50’22)*
Obtenga cuál fue la media muestral y el nivel de confianza al que se realizó el intervalo.

16.- De 1500 personas encuestadas en un sondeo preelectoral, 800 manifiestan su intención de votar. ¿Entre qué valores puede estimarse, con un 95% de confianza, que se encontrará el nivel de abstención del conjunto del censo?

17.- La duración de un determinado proceso industrial es una variable aleatoria con distribución desconocida. Examinando dicho proceso en 200 ocasiones elegidas al azar, se observó una duración media muestral de 1’25 horas y una desviación típica muestral de 1’3 horas. Determinar el intervalo de confianza al 92% para la duración media de dicho proceso. ¿Cuál es el error máximo cometido? ¿Cuál tendría que ser el tamaño de la muestra para reducir ese error a la mitad?

18.- A través de una encuesta realizada a 800 personas sobre la elección de alcalde de una ciudad, se estimó que la proporción de votantes al candidato A estaba entre el 54% y el 59%. ¿Con qué nivel de confianza se realizó la estimación?

19.- Se pasa un test de aritmética a una muestra de 250 mujeres entre 21 y 25 años, obteniendo una puntuación media de 275 puntos. Se sabe que la desviación típica de la puntuación del test es de 60 puntos.

a) Calcula los intervalos de confianza al 90, 95 y 99% para la media de los resultados en el test de toda la población de mujeres entre 21 y 25 años. ¿Cómo afecta el nivel de confianza al error de estimación del intervalo?

b) Supongamos que ahora la muestra es de 4000 mujeres, con la misma puntuación media. Construya un intervalo al 95% de nivel de confianza. ¿Cómo afecta el tamaño muestral al error de estimación del intervalo?

20.- Mediante una muestra aleatoria de tamaño 400 se estima la proporción de residentes en Sevilla que tienen intención de asistir a un partido de fútbol entre el Betis y el C.F. Sevilla. Si para un nivel de confianza del 95% resulta un error máximo en la estimación del 3%, ¿Cuál ha sido la proporción muestral de encuestados que tienen intención de asistir al partido?

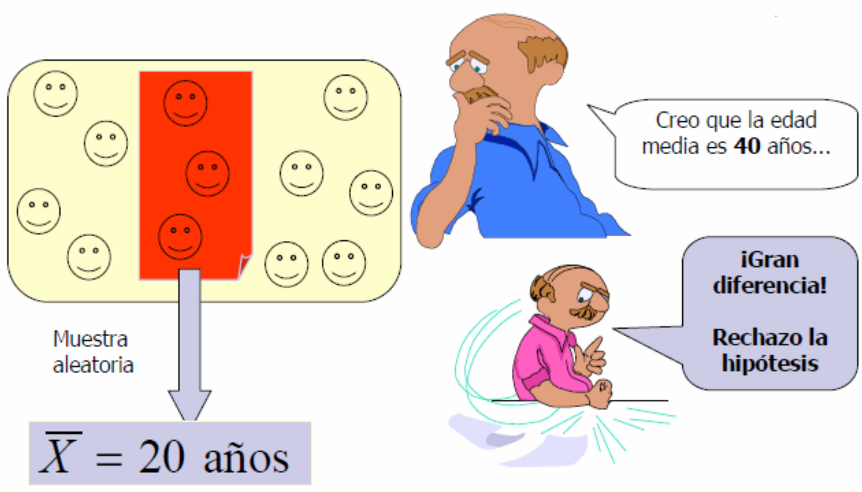
TEMA 5.- TEST DE HIPÓTESIS

1.- INTRODUCCIÓN

En el tema anterior vimos como, a partir de los datos de una muestra, podíamos estimar un parámetro de la población (media o proporción) mediante un intervalo.

En este tema abordaremos el importante aspecto de la toma de decisiones, es decir, plantearemos determinadas hipótesis sobre los parámetros de una población y a partir de los datos de una muestra decidiremos si podemos o no aceptar la hipótesis inicial.

Las hipótesis en estadística inferencial son afirmaciones que involucran al total de la población. Su verdad o falsedad podría establecerse con exactitud si dispusiéramos de la oportunidad de evaluar a todos los individuos que la componen. Como esto no es posible o no se lleva a cabo, el criterio para aceptar o rechazar una hipótesis estadística se basa en un razonamiento de tipo probabilístico: a través del estudio de una o varias muestras se determina la probabilidad de que los resultados obtenidos sean compatibles con la hipótesis establecida. Si es altamente improbable que, de ser cierta la hipótesis, se hayan producido dichos resultados la rechazaremos. Si no es así, lo más que podemos decir es que no existen razones para pensar que tal hipótesis no sea cierta.



Ejemplos:

a) Hace algunos años, la media de estatura de los españoles adultos varones era de 170 cm y su desviación típica 9 cm. Pasado el tiempo, un muestreo realizado a 36 adultos da una medida de 172 cm. ¿Puede afirmarse que esa diferencia de 2 cm es debida al azar o realmente la estatura media ha aumentado?.

b) Supongamos que, respecto a una determinada ley, el 52 % de los ciudadanos está en contra. Pasado el tiempo, una encuesta realizada a 400 personas indica que los ciudadanos en contra han descendido hasta el 49 %.¿Ha cambiado realmente la opinión pública o tal resultado es debido al azar?.

2.- ELEMENTOS DE UN TEST DE HIPÓTESIS

2.1 Hipótesis

Trataremos de utilizar los datos obtenidos en una muestra para tomar decisiones sobre la población. Para ello, debemos realizar ciertos supuestos o conjeturas sobre las poblaciones. Estos supuestos, que pueden ser o no ciertos, se llaman ***hipótesis estadísticas***.

Podemos, entonces, definir el ***test de hipótesis o contraste de hipótesis*** como el procedimiento estadístico mediante el cuál se investiga la verdad o falsedad de una hipótesis acerca de una población o poblaciones.

Dichas hipótesis se formularán normalmente sobre la media poblacional μ o la proporción poblacional p.

Llamaremos ***hipótesis nula, H_0***, a la hipótesis que se formula y por tanto que se quiere contrastar o rechazar, es decir, la que mantendremos salvo que los datos muestres de forma evidente su falsedad.

Llamaremos ***hipótesis alternativa, H_1***, a cualquier otra hipótesis que sea diferente de la ya formulada y que sea contraria a H_0, de forma que la aceptación de la hipótesis nula H_0 implica el rechazo de la alternativa H_1 y viceversa, el rechazo de H_0 implica la aceptación de H_1.

Ejemplos:

1.- Decidir la inocencia o culpabilidad de una persona en un país en el que se sigue el principio de presunción de inocencia:

Como se quiere evitar condenar a una persona inocente, sólo se hará cuando haya una fuerte evidencia de su culpabilidad, cuando esté demostrada ésta. En caso de duda, se primará la inocencia frente a la culpabilidad. Por tanto, en la terminología propuesta sería:

$$\begin{cases} H_0 : Inocente \\ H_1 : Culpable \end{cases}$$

2.- Decidir si un alumno sabe o no la asignatura de Estadística, y por tanto aprueba o suspende la asignatura:

Desde el punto de vista del profesorado, un estudiante no sabe la asignatura mientras no demuestre lo contrario; es decir, el examen ha de presentar pruebas suficientes de que conoce la asignatura. En general, en caso de duda o de falta de datos, se primará el suspenso frente al aprobado. Por tanto, en la terminología propuesta sería:

$$\begin{cases} H_0 : El\ estudiante\ NO\ sabe\ la\ asignatura\ (suspenso) \\ H_1 : El\ estudiante\ SI\ sabe\ la\ asignatura\ (aprobado) \end{cases}$$

Observaciones:

Sobre la metodología de los test de hipótesis hay que tener en cuenta que:

1. No sirve para demostrar H_0
2. Sirve para decidir que, a partir de los datos de la muestra, o no puede rechazarse H_0, o es aceptable suponer que H_0 es cierta
3. Sirve para demostrar H_1 en el sentido de que, a partir de los datos de la muestra, hay una fuerte evidencia de que H_1 es cierta en comparación con H_0

2.2 Errores

Cuando trabajamos con el método del contraste de hipótesis podemos cometer dos tipos de errores:

	Rechazar H_0	**No rechazar H_0**
H_0 **cierta**	**Error de tipo I**	Decisión correcta
H_1 **cierta**	Decisión correcta	**Error de tipo II**

En los ejemplos anteriores:

1.- Decidir la inocencia o culpabilidad de una persona en un estado en el que se sigue el principio de presunción de inocencia:

<table>
<tr><td rowspan="3">H_0:Inocente
H_1:Culpable</td><td></td><td>Rechazar H_0</td><td>No rechazar H_0</td></tr>
<tr><td>H_0 cierta</td><td>Error de tipo I: Se condena a una persona que es inocente</td><td>Se absuelve a una persona que es inocente</td></tr>
<tr><td>H_1 cierta</td><td>Se condena a una persona que es culpable</td><td>Error de tipo II: Se absuelve a una persona que es culpable</td></tr>
</table>

2.- Decidir si un alumno sabe o no la asignatura de Estadística, y por tanto aprueba o suspende la asignatura:

		Rechazar H_0	**No rechazar H_0**
H_0 :*No sabe la asignatura* H_1**:** *Sí sabe la asignatura*	***H_0* cierta**	**Error de tipo I:** Se aprueba a un estudiante que NO sabe la asignatura	Se suspende a un estudiante que no sabe la asignatura
	***H_1* cierta**	Se aprueba a un estudiante que sí sabe la asignatura	**Error de tipo II:** Se suspende a un estudiante que SÍ sabe la asignatura

2.3 Nivel de Significación y Potencia

Llamaremos ***nivel de significación,*** α**,** a la probabilidad de cometer un error de tipo I, es decir, $\alpha = P(\text{Re}\,chazar\ H_0 / H_0\ es\ cierta)$.

Llamaremos ***potencia*** del contraste al valor de $1-\beta$, siendo β la probabilidad de cometer un error de tipo II, es decir, $\beta = P(NO\ \text{Re}\,chazar\ H_0 / H_0\ es\ \ falsa)$

Lo ideal sería minimizar α y β, pero esto no puede hacerse simultáneamente pues si disminuye uno aumenta el otro y viceversa.

Así, si un examen es muy exigente se disminuye α, es decir, la probabilidad de aprobar a un estudiante que no sabe; sin embargo, se aumenta β, la probabilidad de suspender a un estudiante que si sabe. Pero si el examen es poco exigente disminuye la probabilidad de suspender a un alumno que si sabe la asignatura (β), pero aumenta la de aprobar a uno que no sabe lo suficiente (α).

La única manera de disminuir los dos tipos de errores a la vez es aumentando el tamaño de la muestra (preguntar muchas cosas, para tener más datos sobre lo que sabe o no el estudiante)

En general, se fija de antemano un ***nivel de confianza (1-α)***, que asegure un error de tipo I admisible (haciendo mínima la probabilidad de "condenar a un inocente") y de entre todos los contrastes con dicho nivel de confianza se elige el de mayor potencia. (El estudio de la potencia de un test se escapa al nivel de este curso, así que daremos por hecho que los contrastes de este tema cumplen esa condición)

2.4 Región de Aceptación y Región Crítica

Sabemos ya formular la hipótesis nula y la hipótesis alternativa. Lo que necesitamos ahora es un criterio para saber si debemos aceptar una u otra, es decir, ¿con cuál de las dos hipótesis nos quedamos?.

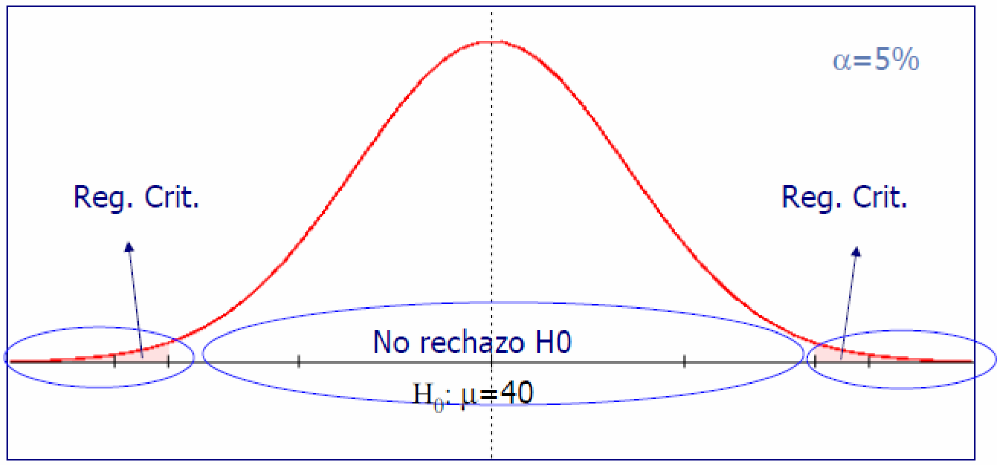
Al tener ya formulada la hipótesis nula, es necesario que las evidencias sean muy fuertes para rechazarla; es decir, puede que haya cambios debidos al azar, en cuyo caso el cambio no es significativo, y no cambiamos , pero puede que los cambios sean debidos a otras causas. En este último caso es cuando el cambio es significativo y rechazaremos.

Por lo tanto, lo primero que debemos hacer es fijar un cierto intervalo dentro del cual es normal que haya cambios, es decir, una región tal que si el parámetro (en nuestro caso media o proporción) se mantiene en dicho intervalo, nos seguimos quedando con H_0, pues esas pequeñas variaciones son debidas al azar. Ese intervalo o región se denomina ***región de aceptación***, y será mayor o menor dependiendo del nivel de confianza que precisemos, $1-\alpha$.

La región que quede fuera de la región de aceptación indica que en este caso los cambios no se pueden atribuir al azar, y por tanto hemos de rechazar H_0 y aceptar H_1. Tal región se llama ***región crítica o de rechazo***.

Así, gráficamente, si nuestro test de hipótesis viene planteado por: $\begin{cases} H_0 : \mu = 40 \\ H_1 : \mu \neq 40 \end{cases}$

y fijamos un nivel de confianza del 95 %:

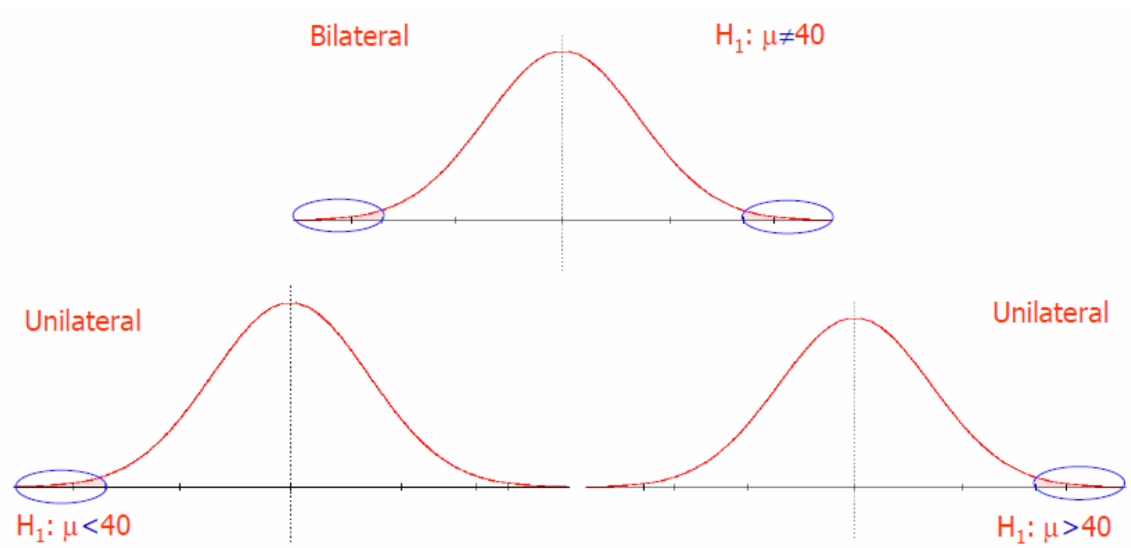


Llegados a este punto, hemos de distinguir entre dos tipos de contraste o test, que determinan la región de aceptación y la región de rechazo.

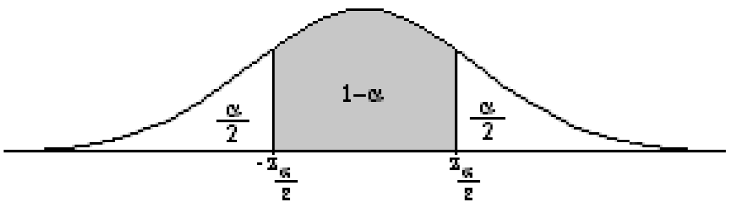
a) Contraste Bilateral (o de dos colas):
En este caso la región de rechazo o región crítica está formada por los dos extremos fuera del intervalo. Dicho caso se presenta cuando la hipótesis nula es del tipo H_0 : $\mu = k$ (o bien H_0 : $p = k$) y la hipótesis alternativa, por tanto, es del tipo $H_1 :\ \mu \neq k$ (o bien $H_1 :\ p \neq k$).

b) Contraste Unilateral (o de una cola):
En este caso la región de rechazo o región crítica está formada por sólo uno de los extremos fuera del intervalo. Dicho caso se presenta cuando la hipótesis nula es del tipo H_0 : $\mu \geq k$ (o bien H_0 : $p \geq k$) y la hipótesis alternativa, por tanto, es del tipo $H_1 :\ \mu < k$ (o bien $H_1 :\ p < k$). (El sentido de las desigualdades pueden cambiar).

Gráficamente:



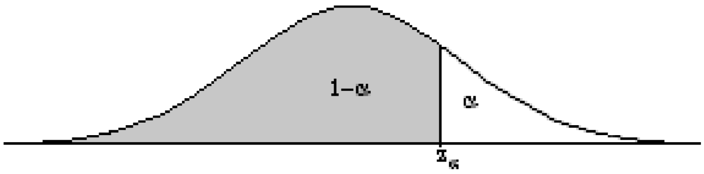
En el caso de distribuciones normales (que son las que vemos en este tema), y para un contraste bilateral, la región de aceptación será:



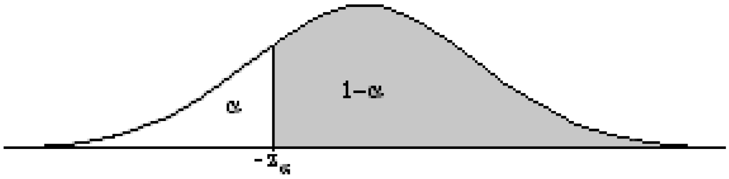
Donde $z_{\frac{\alpha}{2}}$ es el valor crítico cuyo cálculo ya se estudió en el tema anterior de intervalos de confianza, y la región de aceptación no es más que dicho intervalo.

En el caso unilateral cambia un poco:

Unilateral por la derecha: $(H_1 : \mu > k)$



Unilateral por la izquierda: $(H_1 : \mu < k)$



Donde z_α será el valor que en una *N(0,1)* deje por la izquierda una probabilidad de 1-α

3.- METODOLOGÍA GENERAL DE UN TEST DE HIPÓTESIS

Los procedimientos seguidos en las pruebas de hipótesis correspondientes a las situaciones de decisi´on estadística se encuentran totalmente prefijados y se llevan a cabo en una serie de etapas que facilitan su comprensión, y que son:

1. Enunciar la hipótesis nula H_0 y la alternativa H_1.
   Dichas hipótesis deben ser excluyentes y tener en cuenta en su formulación todo lo explicado en la introducción del tema.
   Una vez enunciadas, se analizará si el contraste es bilateral (la hipótesis alternativa es del tipo $\neq$) o si se trata de un contraste unilateral (la hipótesis alternativa es del tipo > o <).

2. Se elige un estadístico cuya distribución muestral es conocida. En nuestro caso será la media o la proporción muestral. Este estadístico se llama *estadístico de* contraste.

3. Determinar, a partir del nivel de confianza, 1-α, o del de significación, α, el valor de $z_{\frac{\alpha}{2}}$ para contrastes bilaterales o el de z_α para contrastes unilaterales, y con dichos valores se construyen las regiones de aceptación correspondientes (y por tanto también las regiones de rechazo):
   $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$ para contrastes bilaterales
   $(-\infty, z_\alpha)$ para contrastes unilaterales a la derecha
   $(-z_\alpha, +\infty)$ para contrastes unilaterales a la izquierda

4. Calcular el valor concreto del estadístico de contraste a partir de la muestra.

5. Aplicar el test, es decir, dependiendo de si el estadístico de contraste cae en la región de aceptación o de rechazo, tomar la decisión de rechazar H_0 o de indicar que no existen evidencias estadísticas significativas como para rechazarla.

Estudiaremos ahora tres de los tipos de test de hipótesis más habituales: test para la media de una población, test para la proporción de una población y test de comparación de medias de dos poblaciones:

4.- TEST DE HIPÓTESIS PARA LA MEDIA DE UNA POBLACIÓN

El planteamiento para este tipo de test será del tipo:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$ para contrastes bilaterales

o del tipo:

$$\begin{cases} H_0 : \mu < \mu_0 \\ H_1 : \mu \geq \mu_0 \end{cases}$$ (Unilateral por la derecha)

o

$$\begin{cases} H_0 : \mu > \mu_0 \\ H_1 : \mu \leq \mu_0 \end{cases}$$ (Unilateral por la izquierda)

Recordemos que si la población de partida es $N(\mu,\sigma)$, la media muestral se distribuye: $\overline{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, donde *n* es el tamaño de la muestra.

Si la hipótesis nula fuese cierta, se tendría que:

$$\overline{X} \to N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$$

Y tipificando obtendremos el estadístico de contraste:

$$\boxed{Z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}}$$ que sigue una N(0,1)

Nota: En todos los casos, si desconocemos la desviación típica de la población, σ, usaremos la desviación típica de la muestra, s, siempre y cuando el tamaño muestral sea suficientemente grande $(n \geq 30)$

Ejemplo 1:

Se cree que el tiempo medio de ocio que dedican al día los estudiantes de Bachillerato sigue una distribución normal de media 350 minutos y desviación típica 60 minutos. Para contrastar esta hipótesis, se toma una muestra aleatoria formada por 100 alumnos, y se observa que el tiempo medio de ocio es de 320 minutos. Con un nivel de significación del 10%, ¿se contradice la afirmación inicial?

Solución:

Paso 1: Formular las hipótesis

$$\begin{cases} H_0 : \mu = 350 \\ H_1 : \mu \neq 350 \end{cases}$$

En este caso es un test bilateral.

Paso 2: Elegir el estadístico de contraste

Como se trata de un contraste sobre la media muestral, el estadístico de contraste será:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$ que sigue una N(0,1)

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Como se trata de un test bilateral, la región de aceptación será:

$$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}} \right)$$

Calculamos $z_{\frac{\alpha}{2}}$

Como $\alpha = 0'1 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'05 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 0'95$

Y buscamos ahora en la tabla el valor de *z* que deja a la izquierda una probabilidad de 0’95, obteniendo:

$$z_{\alpha/2} = 1'645$$

La región de aceptación será, en este caso: (-1’645 , 1’645)

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{320 - 350}{60 / \sqrt{100}} = -5$$

Paso 5: Toma de decisión

Como Z = -5 no pertenece al intervalo (-1’645 , 1’645), **rechazaremos** la hipótesis nula, lo que indica que hay evidencias estadísticas significativas para suponer que el tiempo medio diario de ocio de los alumnos de Bachillerato no es de 350 minutos.

Ejemplo 2:

Una encuesta, realizada a 64 empleados de una fábrica, concluyó que el tiempo medio de duración de un empleo en la misma era de 6’5 años con una desviación típica de 4. ¿Sirve esta afirmación para aceptar, con un nivel de significación del 1%, que el tiempo medio de empleo en esa fábrica es menor o igual que 6?

Solución:

Paso 1: Formular las hipótesis

Como se deduce claramente del enunciado:

$$\begin{cases} H_0 : \mu \le 6 \\ H_1 : \mu > 6 \end{cases}$$

En este caso es un test unilateral por la derecha.

Paso 2: Elegir el estadístico de contraste

Como se trata de un contraste sobre la media muestral, el estadístico de contraste será:

$Z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ que sigue una N(0,1)

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Como se trata de un test unilateral por la derecha, la región de aceptación será:

$$(-\infty, z_\alpha)$$

Calculamos z_α

Como $\alpha = 0'1 \quad \Rightarrow \quad 1 - \alpha = 0'9$

Y buscamos ahora en la tabla el valor de *z* que deja a la izquierda una probabilidad de 0’9, obteniendo:

$$z_\alpha = 1'28$$

La región de aceptación será, en este caso: $(-\infty, 1'28)$

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

$$Z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{6'5 - 6}{4 / \sqrt{64}} = 1$$

Paso 5: Toma de decisión

Como $Z = 1 \in (-\infty, 1'28)$, **no podemos rechazar** la hipótesis nula, es decir, no hay evidencias estadísticas significativas que indiquen que el tiempo medio de empleo en dicha fábrica no sea inferior a 6 años.

Consideraciones a tener en cuenta:

- En la práctica, la muestra se toma después de haber formulado las hipótesis, con el fin de que el resultado de la muestra no influya en el planteamiento de éstas.

- Al disminuir el nivel de significación, α , aumenta la región de aceptación y por tanto es posible que una hipótesis que se rechace con un nivel de significación del 10% no se pueda rechazar a un nivel de significación del 5%.

- Cuanto más *"fuera"* de la región de aceptación se encuentre nuestro estadístico de contraste, con mayor *confianza* podremos rechazar la hipótesis nula y por tanto mayor seguridad tendremos en que nuestra decisión es la correcta. De la misma manera, cuanto más *"dentro"* de la región de aceptación se encuentre, mayor seguridad tendremos a la hora de no rechazar la hipótesis nula.

Ejercicios:

1.- Las estaturas de las alumnas de COU eran, en 1990, de media 167 cm. y desviación típica 7 cm. Se toma una muestra de 60 alumnas actuales de 2º de Bachillerato y se obtiene una estatura media de 168'72 cm.
¿Es posible afirmar, con un nivel de confianza del 90 %, que la estatura media de las alumnas actuales de 2º de Bachillerato es la misma? ¿Y con un nivel de confianza del 95 %?

Sol: en el primer caso se rechaza la hipótesis nula y en el segundo no

2.- Pablo y Virginia quieren contrastar si el consumo medio en teléfono móvil entre los estudiantes es, como máximo, de 10 euros mensuales.
Pablo, en una muestra de 36 estudiantes, obtuvo una media de 10'41 euros y una desviación típica de 2 euros.
Virginia, en una muestra de 49 estudiantes, , obtuvo una media de 10'39 euros y una desviación típica de 2 euros.
Si acuerdan tomar un nivel de significación del 10 %, ¿qué decisión tomará cada uno?

Sol: Pablo acepta la hipótesis nula y Virginia la rechaza

3.- Gracias a trabajos realizados por expertos, se sabe que la velocidad lectora media de los niños de 6 años es de 40 palabras por minuto, con una desviación típica de 12. Hemos tomado una muestra aleatoria de 49 niños y les hemos medido su velocidad lectora, resultando una media de 42 palabras por minuto. ¿Podemos afirmar que nuestra media es compatible con la de los expertos a un nivel de confianza del 95 %?

Sol: si

5.- TEST DE HIPÓTESIS PARA LA PROPORCIÓN

Si llamamos *p* a la proporción de individuos de una población que tienen una determinada característica, el planteamiento para este tipo de test será del tipo:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$ para contrastes bilaterales

o del tipo:

$$\begin{cases} H_0 : p < p_0 \\ H_1 : p \geq p_0 \end{cases}$$ (Unilateral por la derecha)

o

$$\begin{cases} H_0 : p > p_0 \\ H_1 : p \leq p_0 \end{cases}$$ (Unilateral por la izquierda)

Como vimos en temas anteriores, la distribución de las proporciones muestrales es:

$$\hat{P} \rightarrow N\left(p, \sqrt{\frac{pq}{n}} \right)$$

donde $q=1-p$

Si la hipótesis nula fuera cierta, se tendría que $\hat{P} \rightarrow N\left(p_0, \sqrt{\frac{p_0 q_0}{n}} \right)$

Y tipificando obtendremos el estadístico de contraste para este test:

$$\boxed{Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$ que sigue una N(0,1)

Salvo los pequeños cambios lógicos, estos test son muy parecidos a los test de hipótesis para la media, y las consideraciones realizadas anteriormente siguen siendo válidas en este caso.

Ejemplo 1:

El ayuntamiento de una ciudad afirma que el 65 % de los accidentes juveniles de los fines de semana son debidos al alcohol. Un investigador decide contrastar dicha hipótesis, para lo cual toma una muestra formada por 35 accidentes y observa que 24 de ellos han sido debidos al alcohol. Con un nivel de confianza del 99 %, ¿qué podemos decir sobre la afirmación del ayuntamiento?

Solución:

Paso 1: Formular las hipótesis

$$\begin{cases} H_0 : p = 0'65 \\ H_1 : p \neq 0'65 \end{cases}$$

En este caso es un test bilateral.

Paso 2: Elegir el estadístico de contraste

Como se trata de un contraste sobre la proporción muestral, el estadístico de contraste será:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$ que sigue una N(0,1)

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Como se trata de un test bilateral, la región de aceptación será:

$$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$$

Calculamos $z_{\alpha/2}$

Como $\alpha = 0'01 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'005 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 0'995$

Y buscamos ahora en la tabla el valor de z que deja a la izquierda una probabilidad de 0’995, obteniendo:

$$z_{\alpha/2} = 2'575$$

La región de aceptación será, en este caso: (-2’575 , 2’575)

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

Como $\hat{p} = \frac{24}{35} = 0'686$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0'686 - 0'65}{\sqrt{\frac{0'65 \cdot 0'35}{35}}} = \frac{0'036}{0'08} = 0'45$$

Paso 5: Toma de decisión

Como $Z = 0'45 \in (-2'575, 2'575)$, **no podemos rechazar** la hipótesis nula, es decir, no hay evidencias estadísticas significativas que indiquen que la afirmación del ayuntamiento no sea correcta.

Ejemplo 2:

Un investigador, utilizando información de anteriores comicios, sostiene que, en una determinada zona, el nivel de abstención en las próximas elecciones es del 40% como mínimo. Se elige una muestra aleatoria de 200 individuos para los que se concluye que 75 estarían dispuestos a votar.
Determinar, con un nivel de significación del 1%, si se puede admitir como cierta la afirmación del investigador.

Solución:

Paso 1: Formular las hipótesis

Si p es la proporción de abstención:

$$\begin{cases} H_0 : p \geq 0'4 \\ H_1 : p < 0'4 \end{cases}$$

En este caso es un test unilateral hacia la izquierda.

Paso 2: Elegir el estadístico de contraste

Como se trata de un contraste sobre la proporción muestral, el estadístico de contraste será:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$ que sigue una N(0,1)

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Como se trata de un test unilateral hacia la izquierda, la región de aceptación será:

$$(-z_\alpha, +\infty)$$

Calculamos z_α

Como $\alpha = 0'01 \quad \Rightarrow \quad 1 - \alpha = 0'99$

Y buscamos ahora en la tabla el valor de z que deja a la izquierda una probabilidad de 0’99, obteniendo:

$z_\alpha = 2'33$

La región de aceptación será, en este caso: $\left(-2'33, +\infty\right)$

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

Como $\hat{p} = \dfrac{125}{200} = 0'625$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0'625 - 0'4}{\sqrt{\dfrac{0'4 \cdot 0'6}{200}}} = \frac{0'225}{0'035} = 6'43$$

Paso 5: Toma de decisión

Como Z = 6’43 pertenece al intervalo $\left(-2'33, +\infty\right)$, **no podemos rechazar** la hipótesis nula, es decir, no hay evidencias estadísticas significativas que indiquen que el investigador no tiene razón.

Ejercicios:

1.- Un entrenador asegura que sus jugadores en los entrenamientos encestan más del 92 % de los tiros libres. Con el fin de contrastar esta afirmación se ha elegido aleatoriamente una muestra de 60 lanzamientos de los que 42 han entrado en la canasta. Estos datos, ¿ponen en duda al entrenador o no?

Sol: con $\alpha = 0'1$ se rechaza la afirmación del entrenador

2.- Se lanza una moneda 100 veces y se obtienen 60 caras. ¿Podemos asegurar que la moneda está trucada? ¿Y si en 1000 lanzamientos se obtienen 600 caras? ($\alpha = 0'01$)

Sol: en el segundo caso la moneda está trucada.

3.- Se cree que uno de cada 10 varones manifiesta algún tipo de daltonismo. Se eligen 400 varones, de los que 50 son daltónicos.

a) Con un nivel de significación del 10 %, ¿se puede aceptar la hipótesis de partida?

b) ¿Se obtendría la misma conclusión con $\alpha = 0'02$?

Sol: *a) se rechaza que uno de cada 10 varones sea daltónico*
b) se acepta que uno de cada 10 varones sea daltónico

6.- TEST DE HIPÓTESIS PARA LA DIFERENCIA DE MEDIAS EN DISTRIBUCIONES NORMALES

Consideramos ahora dos distribuciones normales, $N(\mu_1,\sigma_1), N(\mu_2,\sigma_2)$, y queremos contrastar la hipótesis de que sus medias son iguales, es decir, que $\mu_1 = \mu_2$, lo que significa que el test de hipótesis que plantearemos en este caso será:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

Se trata por tanto de un test bilateral, en el que la región de aceptación será, como siempre, el intervalo $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$

Tomamos para el test una muestra de tamaño n_1 para la primera población y n_2 para la segunda población

Se puede demostrar que, si las poblaciones de partida son normales, la variable "diferencia de medias", $\overline{X}_1 - \overline{X}_2$, sigue una distribución normal del tipo:

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

Si la hipótesis nula fuese cierta, se tendría que:

$$\overline{X}_1 - \overline{X}_2 \rightarrow N\left(0, \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}\right)$$

Y tipificando obtendremos el estadístico de contraste:

$$\boxed{Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}}}$$ que sigue una N(0,1)

Al igual que pasa con el test de una media, en caso de desconocer las desviaciones típicas de las poblaciones, y si las muestras son suficientemente grandes, se pueden sustituir por las desviaciones típicas muestrales.

Ejemplo 1:

A los 100 alumnos de una clase se les separa en dos grupos: aquellos que practican habitualmente un deporte y los que no practican ninguno, formando cada grupo 60 y 40 alumnos, respectivamente. Les medimos la altura, obteniendo para el primer grupo una media de 1’80 m. y una desviación típica 0’08 m., y para el segundo grupo una media de 1’76 m. y una desviación típica de 0’1 m. Suponiendo que la variable aleatoria altura sigue una distribución normal en los dos grupos, ¿es posible afirmar, con un nivel de confianza del 95%, que hay diferencia de altura entre los alumnos que practican algún deporte y los que no?

Solución:

Paso 1: Formular las hipótesis

Se trata obviamente de un test de comparación de medias, y por tanto su planteamiento será:

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases} \quad \text{(bilateral)}$$

Paso 2: Elegir el estadístico de contraste

Como hemos visto anteriormente, al desconocer las desviaciones típicas poblacionales las sustituimos por las muestrales y el estadístico de contraste será:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{que sigue una N(0,1)}$$

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Al ser bilateral, la región de aceptación será: $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$

Calculamos $z_{\alpha/2}$

Como $\alpha = 0'05 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'025 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 0'975$

Y buscamos ahora en la tabla el valor de *z* que deja a la izquierda una probabilidad de 0’975, obteniendo:

$$z_{\alpha/2} = 1'96$$

La región de aceptación será, en este caso: (-1’96 , 1’96)

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

$$Z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^{\,2}}{n_1}+\frac{s_2^{\,2}}{n_2}}}=\frac{1'8-1'76}{\sqrt{\frac{0'08^2}{60}+\frac{0'1^2}{40}}}=\frac{0'04}{0'019}=2'105$$

Paso 5: Toma de decisión

Como $Z=2'105\notin(-1'96,1'96)$, **rechazamos** la hipótesis nula, es decir, existen diferencias estadísticamente significativas entre la altura media de los chicos que practican deporte y la de los que no.

Ejemplo 2:

A fin de determinan si existen diferencias significativas entre dos grupos de estudiantes, realizamos el mismo examen a 30 estudiantes del primer grupo y a 35 del segundo, obteniendo la información contenida en la tabla:

	Nota Media	**Desv. Típica**
Primer Grupo	5’5	0’5
Segundo Grupo	5’2	1

¿Qué conclusión se obtiene con un nivel de significación del 1 %?

Paso 1: Formular las hipótesis

Se trata obviamente de un test de comparación de medias, y por tanto su planteamiento será:

$$\begin{cases} H_0: \mu_1-\mu_2=0 \\ H_1: \mu_1-\mu_2\neq 0 \end{cases} \quad \text{(bilateral)}$$

Paso 2: Elegir el estadístico de contraste

Como hemos visto anteriormente, al desconocer las desviaciones típicas poblacionales las sustituimos por las muestrales y el estadístico de contraste será:

$$Z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^{\,2}}{n_1}+\frac{s_2^{\,2}}{n_2}}} \qquad \text{que sigue una N(0,1)}$$

Paso 3: Calcular, a partir del nivel de significación, la región de aceptación

Al ser bilateral, la región de aceptación será: $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$

Calculamos $z_{\alpha/2}$

Como $\alpha = 0'01 \quad \Rightarrow \quad \frac{\alpha}{2} = 0'005 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = 0'995$

Y buscamos ahora en la tabla el valor de *z* que deja a la izquierda una probabilidad de 0'995, obteniendo:

$$z_{\alpha/2} = 2'58$$

La región de aceptación será, en este caso: (-2'58, 2'58)

Paso 4: Calcular el valor concreto del estadístico de contraste a partir de la muestra

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5'5 - 5'2}{\sqrt{\frac{0'5^2}{30} + \frac{1^2}{35}}} = 1'56$$

Paso 5: Toma de decisión

Como $Z = 1'56 \in (-2'58, 2'58)$, **no podemos rechazar** la hipótesis nula, es decir, no existen diferencias estadísticamente significativas entre ambos cursos.

Ejercicios:

1.- El estudio de un test de satisfacción de usuario que rellenan los demandantes de servicios de una gran empresa revela que la nota media que otorgan es de 5'7 puntos y una desviación típica de 0'5. Posteriormente se ha realizado un muestreo a 100 usuarios de una determinada zona A y a 49 usuarios de otra zona B, obteniendo una media de 5'6 y 5'85 puntos, respectivamente. Con una confianza del 95 %, ¿se puede afirmar que existen diferencias entre las puntuaciones medias de cada zona?

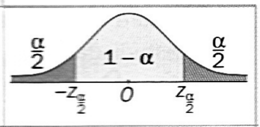
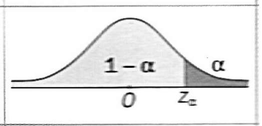
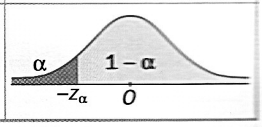
Sol: Z = -0'28, se acepta la hipótesis nula

2.- En un estudio de mercado se necesita comparar las ventas en los meses de diciembre de dos años consecutivos para decidir si existen diferencias significativas entre ambos. Se dispone de datos de ventas del primer año en 52 tiendas que proporcionan una media muestral de 942'3 euros y desviación típica de 49 euros. En el segundo año se han observado las ventas en 64 tiendas obteniendo una venta media de 981,2 euros y una desviación típica de 74 euros.
Plantea un contraste de hipótesis para decidir si existen diferencias entre las ventas de ambos años en diciembre con un nivel de confianza del 95 %.

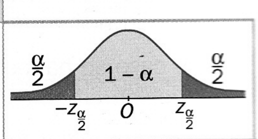
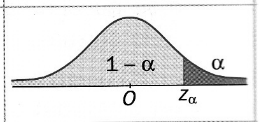
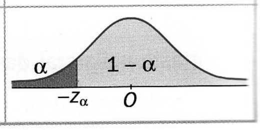
Sol: Z = -3'36, se rechaza H_0

7.- RESUMEN DE FÓRMULAS

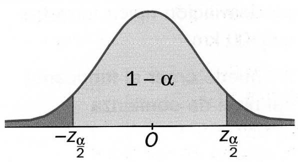
Test para la Media

Hipótesis nula H_0	Hipótesis alternativa H_a	Tipo de contraste	Estadístico del contraste	Región de aceptación	
$\mu = \mu_0$	$\mu \neq \mu_0$	bilateral	$Z = \dfrac{\bar{x} - \mu_o}{\dfrac{\sigma}{\sqrt{n}}}$ sigue una $N(0, 1)$	$(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$	
$\mu \leqslant \mu_0$	$\mu > \mu_0$	unilateral		$(-\infty, z_\alpha)$	
$\mu \geqslant \mu_0$	$\mu < \mu_0$	unilateral		$(-z_\alpha, +\infty)$	

Test para una Proporción:

Hipótesis nula H_0	Hipótesis alternativa H_a	Tipo de contraste	Estadístico del contraste	Región de aceptación	
$p = p_0$	$p \neq p_0$	bilateral	$Z = \dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}$ sigue una $N(0, 1)$	$(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$	
$p \leqslant p_0$	$p > p_0$	unilateral		$(-\infty, z_\alpha)$	
$p \geqslant p_0$	$p < p_0$	unilateral		$(-z_\alpha, +\infty)$	

Test de Comparación de Medias

Hipótesis nula H_0	Hipótesis alternativa H_a	Tipo de contraste	Estadístico del contraste	Región de aceptación	
$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	bilateral	$Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ sigue una $N(0, 1)$	$(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$	

EJERCICIOS

1.- Se sabe que la renta anual de los individuos de una localidad sigue una distribución normal de media desconocida y desviación típica 2400 euros. Se ha observado la renta anual de 16 individuos de esa localidad escogidos al azar, y se ha obtenido un valor medio de 16000 euros. Contrasta, a un nivel de significación del 5%, si la media de la distribución es de 14500 euros.

2.- El Ayuntamiento maneja el dato de que la edad a la que los hijos se independizan de sus padres es una variable normal con media 29 a˜nos y desviación típica 3 años. Aunque la desviación típica no plantea dudas, si se sospecha que la media ha descendido, sobre todo por la política de ayuda al empleo que ha llevado a cabo el Ayuntamiento. Así, de un estudio reciente sobre 100 jóvenes que se acaban de independizar, se ha obtenido una media de 28'1 años de edad.

a) Con un nivel de significación del 1 %, ¿puede defenderse que la edad media no ha disminuido frente a que sí lo ha hecho como parecen indicar los datos?
b) Explica, en el contexto del problema, en qué consisten cada uno de los errores de tipo I y de tipo II

3.- Hace 10 años, el 52% de los ciudadanos estaban en contra de una ley. Recientemente, se ha elaborado una encuesta a 400 personas y 184 se mostraron contrarios a la ley. Con estos datos y con un nivel de significación de 0'01, ¿podemos afirmar que la proporción de contrarios a la ley ha disminuido?

4.- El propietario de una pizzería sospecha que su repartidor está utilizando la moto del reparto para uso propio. Sabe, por experiencias anteriores, que el recorrido diario de la moto sigue una distribución normal de media 14 km. y desviación típica 2 km. Decide comprobar sus sospechas, y para ello contabiliza los kilómetros recorridos por la moto durante 10 días y obtiene:

14'5, 17, 16, 15, 12'5, 19, 14, 16'5, 15'5, 17 (km.)

Con una confianza del 95%, ¿debe despedir al repartidor?.

5.- Se quiere probar la eficacia de un nuevo antibiótico para ver si influye en la duración de una determinada enfermedad que sigue una ley normal de varianza 16. Observamos a 36 enfermos a los que no se les ha aplicado el antibiótico, y la duración media de la enfermedad ha sido de 15 días, y a 35 enfermos a los que sí se les ha aplicado y que han permanecido enfermos 12 días. Con un nivel de significación de 0'01, ¿se puede afirmar que el antibiótico influye en la duración de dicha enfermedad?

6.- El ayuntamiento de una ciudad hace un estudio y afirma que en fin de semana el 60 % de los jóvenes asiste a discotecas. Una plataforma juvenil afirma que dicho porcentaje es menor. Para verlo se hace una encuesta a 600 jóvenes y se obtiene que 330 van a discotecas. ¿Se puede aceptar la afirmación del ayuntamiento si se toma un nivel de confianza del 95 %?

7.- El número de reclamaciones presentadas durante la campaña de Navidad en 9 tiendas de una empresa han sido:

25, 31, 28, 30, 32, 20, 22, 34, 30

Suponiendo que el número de reclamaciones se distribuye según una ley normal, ¿es posible aceptar que el número medio de reclamaciones para las tiendas de esa empresa durante la campaña de Navidad es de 26? $(\alpha = 0'05)$

8.- Un laboratorio farmacéutico fabrica dos tipos de somníferos, A y B. Se toman dos grupos análogos de enfermos de insomnio formados por 80 y 100 individuos, respectivamente, y se suministra a los del primer grupo el somnífero A y a los del segundo grupo el B.
El número medio de horas de sueño para los enfermos del primer grupo es de 7’84 con una desviación típica de 0’9, y para los del segundo grupo es de 6’9 y 1’3, respectivamente.
¿Se puede decir que la diferencia entre los números medios de horas de sueño es significativa?

9.- Cierta marca de bebidas refrescantes pone en sus envases una etiqueta que dice: “Contenido: 33 cl.” En un departamento de defensa del consumidor han tomado una muestra aleatoria formada por 40 envases y han obtenido un contenido medio muestral de 31 cl. Con una desviación típica de 2 cl. ¿Se puede afirmar que se está engañando al público al nivel de significación del 10 %?

10.- Se sabe que la edad (en años) de los aspirantes a un puesto de trabajo en un determinado organismo oficial es una variable normal con desviación típica 5.
Se toma una muestra de 125 aspirantes y se obtiene una edad media de 22’3 años.

a) Obtener un intervalo de confianza al 95 % para la edad media de los aspirantes a dicho puesto de trabajo.
b) Plantear un test de hipótesis adecuado e indicar si se puede afirmar, con un nivel de significación del 5 %, que la edad media de los aspirantes a dicho puesto es de 21 años.
c) Explica si existe alguna relación entre el intervalo de confianza obtenido y la decisión que se toma en el test.

11.- En un informe presentado por un reportero a una revista feminista se afirma que el número medio de horas semanales de conexión a Internet es el mismo para hombres que para mujeres. Sin embargo no parece prudente publicar estos datos sin contrastarlos estadísticamente. Se selecciona para ello una muestra de 75 hombres y 50 mujeres. Los resultados muestrales se recogen en la siguiente tabla:

	Hombres	**Mujeres**
Tamaño muestral	75	50
Nº medio horas/semana	7’42	5’34
Varianza	9’08	7’24

a) Formular el contraste a realizar y calcular la región crítica para un nivel de significación del 5 %
b) Calcular el estadístico de contraste
c) ¿Existen evidencias para rechazar la hipótesis nula?
d) Explica, para este ejercicio, el significado de los distintos tipos de errores

12.- El número de averías de un determinado tipo de avión se considera una variable aleatoria con distribución Normal de media 2 averías al mes y desviación típica 1. El equipo de mantenimiento intenta reducir esta media incorporando algunas mejoras. Para comprobar si con estas medidas se reduce el número medio de averías se decide observar el número de averías en los 25 meses siguientes a la introducción de las mejoras.
Si el número medio de averías es esos 25 meses fue de 1,9. ¿Qué decisión debe adoptar el servicio técnico a un nivel de significación del 5%.

13.- El consumo de carne de pollo parece haberse visto afectada por la alarma creada con la gripe aviar. En cierta carnicería, las ventas diarias de carne de pollo seguían hasta entonces una normal de media 21 kilos y desviación típica 3 kilos. En una muestra tomada durante 35 días después de la alarma de gripe aviar, se obtuvo una media de 19 kilos de carne de pollo vendidos al día.
¿Podemos afirmar, con un nivel de confianza del 95 %, que la gripe aviar ha influido en la venta de carne de pollo? ¿Y al 99 % de confianza?

14.- En el año 2005, un estudio indicaba que un 15% de los conductores utilizaban el móvil con el vehículo en marcha.
Con el fin de investigar la efectividad de las campañas que se han realizado desde entonces para reducir esos hábitos, se ha hecho una encuesta a 120 conductores de los cuales 12 hacían un uso indebido del móvil.
¿A qué conclusión se llega con un nivel de significación del 4%?

15.- Se ha comprobado que el tiempo de espera (en minutos) hasta ser atendido, en cierto servicio de urgencias, sigue un modelo normal de probabilidad. A partir de una muestra de 100 personas que fueron atendidas en dicho servicio, se ha calculado un tiempo medio de espera de 14,25 minutos y una desviación típica de 2,5 minutos.

a) ¿Podríamos afirmar, con un nivel de significación del 5% (a = 0,05), que el tiempo medio de espera, en ese servicio de urgencias, no es de 15 minutos?

b) Calcule un intervalo de confianza con el mismo nivel de significación anterior para el tiempo medio de espera en ese servicio de urgencias y relacione el resultado con la decisión del test del apartado anterior.

16.- Se desea contrastar si los alumnos de los municipios rurales obtienen resultados semejantes a los niños urbanos en un test de percepción sensorial. Para ello se seleccionan 40 niños rurales y 100 niños urbanos del mismo curso. La media obtenida por los niños rurales es de 95 y el grupo de niños urbanos obtiene 98'5. Si sabemos que la desviación típica de los niños rurales es de 15 y de los urbanos de 20, ¿existen diferencias entre dichos grupos de niños ante el test de percepción sensorial a un nivel de confianza del 95 %? Se supone que las poblaciones se distribuyen normalmente.

17.- La duración de las bombillas de 100 w que fabrica una empresa sigue una distribución normal, con desviación típica 120 h. Su vida media está garantizada durante un mínimo de 800 h. Se escoge al azar una muestra de 50 bombillas de un lote y después de comprobarlas se obtiene una media de 750 horas. Con un nivel de significación de 0'03, ¿habría que rechazar el lote por no cumplir lo garantizado?

18.- Un dentista afirma que el 40 % de los niños de 10 años presentan indicios de caries dental. Tomada un muestra de 100 niños, se observó que 30 presentaban indicios de caries. Comprueba; al nivel de significación del 5 %, si el resultado proporciona evidencia que permita rechazar la afirmación del dentista.

19.- Una empresa asegura que unas determinadas pastillas de jabón duran más de 11 días. Para comprobarlo se realiza un encuesta en 100 casos. Éstas son las respuestas:

Duración (días)	5 a 9	10 a 14	15 a 19	20 a 24
Respuestas	24	46	19	11

¿Se puede dar como válida la afirmación de la empresa, para un nivel de significación del 5 %?

20.- Contrasta las siguientes hipótesis:

a) H_0: $\mu = 19{,}4$ ($\sigma = 2{,}6$) Nivel de significación: $\alpha = 0{,}10$
MUESTRA: $n = 114$; $\bar{x} = 18{,}6$

b) H_0: $\mu \geq 500$ ($\sigma = 31$) Nivel de significación: $\alpha = 0{,}05$
MUESTRA: $n = 300$; $\bar{x} = 495$

c) H_0: $p = 0{,}2$ Nivel de significación: $\alpha = 0{,}02$
MUESTRA: $n = 65$; $\hat{p} = 0{,}17$

d) H_0: p I 0,68 Nivel de significación: $\alpha = 0{,}02$
MUESTRA: $n = 200$; $\hat{p} = 0{,}703$

21.- Una fábrica de muebles se encargaba también del transporte y montaje de los pedidos a sus clientes. Sin embargo, recibía al menos un 16% de reclamaciones por dicho servicio.
En los últimos meses, ha contratado una empresa especializada. De 250 servicios realizados por la empresa contratada, 30 han tenido reclamación.
Plantea un test para contrastar la hipótesis de que con la empresa contratada la situación sigue igual, frente a que, como parece, ha mejorado. ¿A qué conclusión se llega para un nivel de significación del 5%?

22.- Una compañía aérea sostiene que sus precios no han variado con respecto a los del año anterior.
Para una muestra de 100 viajeros con un tipo determinado de billetes, se observó el año pasado que el precio medio era de 120 euros con una desviación típica de 40 euros.
Se realiza este año otra encuesta a otros 100 viajeros y el precio medio resulta ser de 128 euros con una desviación típica de 50 euros.
¿Se puede admitir la afirmación de la compañía aérea a un nivel de significación del 1%?