ESTADÍSTICA INFERENCIAL

INTRODUCCIÓN

Durante años la estadística se ha dedicado fundamentalmente al desarrollo de la Estadística Descriptiva, cuya principal labor como hemos visto es recopilar datos, ordenarlos, representarlos gráficamente y calcular algunos parámetros que puedan darnos una representación de la totalidad de ellos. Esta concepción de la Estadística cambió durante el siglo XX con el nacimiento de la Estadística Inferencial cuyo objetivo es obtener conclusiones válidas para toda la población a partir del estudio de una muestra. Esta rama necesita apoyarse fuertemente en la teoría de las probabilidades para los cálculos y conclusiones.

La primera pregunta a la que debe contestar un investigador es la de cómo elegir una muestra y el tamaño de la misma. Hemos visto ya, en el tema de estadística descriptiva, las diferentes técnicas de muestreo y de forma muy sucinta en qué casos deberían utilizarse para que la muestra resulte lo más representativa posible de la población.

El siguiente paso será realizar inferencias sobre los parámetros poblacionales a partir de los datos y parámetros muestrales y conocer el margen de error que podemos cometer. De forma muy elemental esto es lo que vamos a estudiar en este tema, abordando únicamente los resultados que se obtienen en muestreos con reemplazamiento, en los que todos los individuos tienen la misma probabilidad de ser elegidos

A lo largo del tema representaremos por N al tamaño de la población y por n al de la muestra

DISTRIBUCIONES MUESTRALES

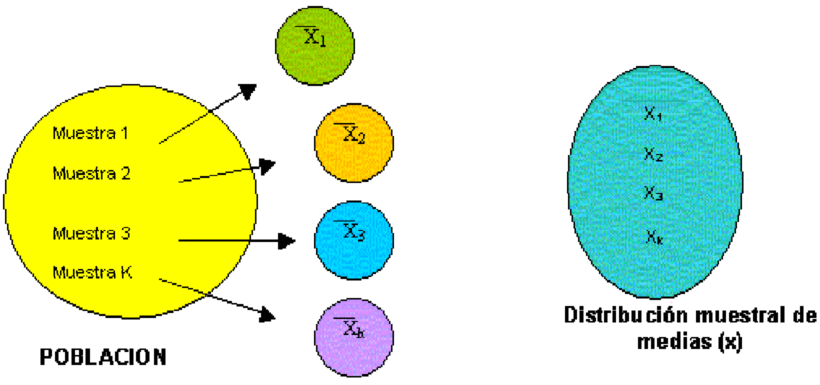
Las muestras aleatorias obtenidas de una población son, por naturaleza propia, impredecibles. No se esperaría que dos muestras aleatorias del mismo tamaño y tomadas de la misma población tenga la misma media muestral o que sean completamente parecidas. Sin embargo, pretendemos que los parámetros de las muestras nos permitan decidir sobre la aproximación más conveniente de los parámetros de la población. Por ello, es necesario comenzar estudiando el tipo de distribución que siguen los parámetros muestrales.

1 Distribución de la media muestral

Supongamos que elegimos una primera muestra de tamaño n de una población, esta dará una media $\overline{x_1}$ de la variable que se esté estudiando (pesos, alturas, etc), otra muestra diferente elegida dará otro valor medio $\overline{x_2}$ y así sucesivamente.

Si consideramos todas las muestras posibles de tamaño n que se pueden extraer de esa población, la variable aleatoria que a cada una de esas muestras le hace corresponder su media se llama media muestral y se representa por $\overline{X}$. *Dicha variable tomará los valores :*

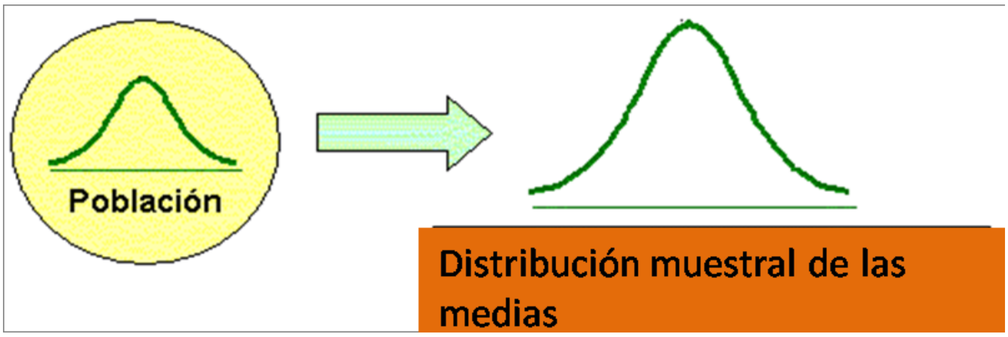
$\overline{X}.=\{\overline{x_1},\overline{x_2},......,\overline{x_k}\}$



Al ser $\overline{X}$ una variable aleatoria podemos estudiar su distribución, a la que llamaremos *distribución de la media muestral*.

Sin demostrarlas vamos a dar algunas de las propiedades de esa distribución

1. La media muestral coincide con la media de la población. Es decir $\mu_{\overline{X}} = \mu$
2. La desviación típica muestral es igual a la desviación típica de la población dividida por la raíz cuadrada de n (tamaño de la muestra).$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ (siempre que la muestra sea con reemplazamiento)
3. Si la distribución de partida se distribuye según una N(μ, σ) , la distribución de las muestras de tamaño n se distribuye según una normal N(μ, $\frac{\sigma}{\sqrt{n}}$)



4. En el caso de que la población no se distribuya según una normal, la distribución de las muestras si lo hace cuando el tamaño de estas sea por lo menos de 30 individuos. Esto queda establecido por el **teorema central del límite** que dice: Si una muestra aleatoria de tamaño n procede de una población con media μ y desviación típica σ y el tamaño de la muestra es n$\geq$ 30, las medias muestrales se distribuyen según una normal de media μy desviación típica$\frac{\sigma}{\sqrt{n}}$; N(μ, $\frac{\sigma}{\sqrt{n}}$)

Ejemplo: El gasto total semanal de los jóvenes de una ciudad tiene una media de 25€ y una desviación típica de 3€ ¿Cuál es la probabilidad de que el gasto medio de 49 jóvenes, elegidos al azar, esté comprendido entre 24 y 26€?

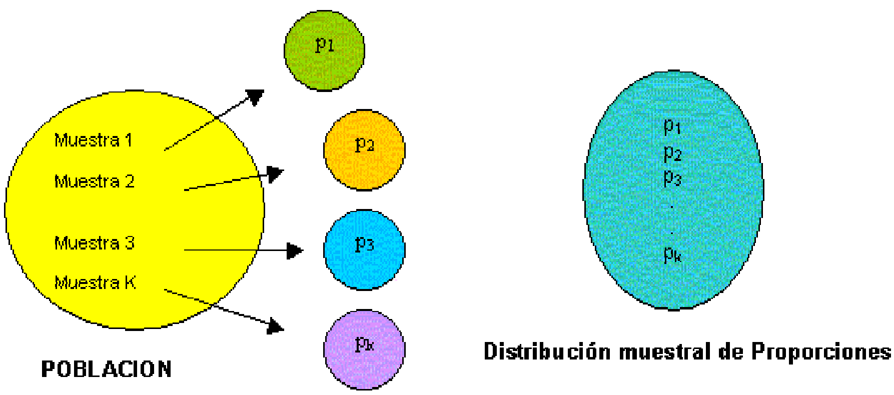
Dado que el tamaño de la muestra es > 30 la distribución de las medias es N(25,$\frac{3}{\sqrt{49}}$)= =N(25,3/7); P(24<$\overline{X}$<26)=P($\frac{24-25}{3/7} < Z < \frac{26-25}{3/7}$)=P(-2’33<Z<2’33)=P(Z<2’33)-(1-P(Z<2’33))= 2.P(Z<2’33)-1=0’9802

2. Distribución de la proporción muestral

Supongamos ahora que estudiamos si los individuos de una población tienen o no una determinada característica. Habrá una proporción p de individuos que si la tienen y el resto de individuos, una proporción q=1-p, que no la tienen. Recordemos que su estudio se puede realizar mediante distribuciones binomiales B(n,p), siendo n=número de ensayos y p=probabilidad o proporción de individuos que tienen la característica estudiada.

Si tomamos distintas muestras de tamaño n, la primera nos dará una proporción $\widehat{p_1}$, la segunda una proporción $\widehat{p_2}$ y así sucesivamente.

De manera análoga a lo hecho con las medias, podemos considerar todas las posibles muestras de tamaño n y considerar las proporciones muestrales $\hat{P}$ como una variable aleatoria. A la distribución de la variable $\hat{P}$ descrita le llamaremos *distribución muestral de las proporciones*



Propiedades

1. La media o esperanza de $\hat{P}$ coincide con la proporción de la población.
2. La desviación típica de $\hat{P}$ coincide con $\sqrt{\frac{p.q}{n}}$ (siempre que la muestra sea con reemplazamiento)
3. Si n≥ 30, teniendo en cuenta la aproximación de la binomial por la normal, la distribución muestral de las proporciones se aproxima a una distribución N(p, $\sqrt{\frac{p.q}{n}}$)

Ejemplo:

El 10% de las bolsas de pipas de una marca contiene menos peso del que anuncia. Se han seleccionado al azar 400 bolsas a) ¿Cuál es la distribución que sigue la proporción de envases no completos de la muestra?; b) Halla la probabilidad de que en la muestra haya más de 50 bolsas de pipas con menos peso del anunciado.

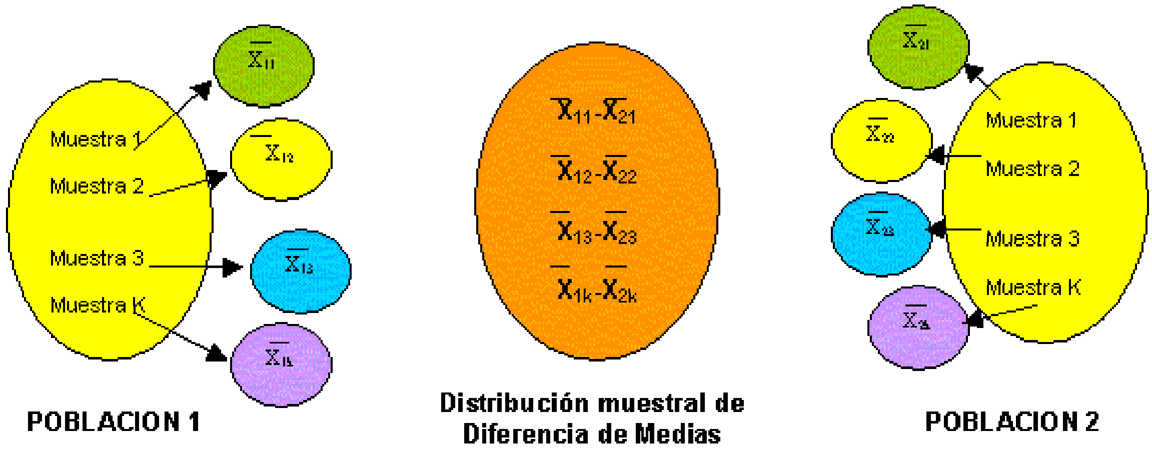
A) n=400>30, p=10%=0’1 $\hat{P}$ sigue una distribución N(0’1, $\sqrt{\frac{0'1.0'9}{400}}$)=N(0’1, 0’015)

B) P(más de 50 bolsas)=P($\hat{P}$>50/400)=P($\hat{P}$>0’125)=P(Z>$\frac{0'125-0'1}{0'015}$)=P(Z>1’67)= =1-P(Z<1’67)=1-0’9525=0’0475. Es decir, hay una probabilidad de 4’75% de que en la muestra haya más de 50 bolsas con peso inadecuado.

3. Distribución de la diferencia de las medias muestrales

Supongamos ahora que tenemos dos poblaciones de tamaños N_1 y N_2 respectivamente y que extraemos de cada una de ellas todas las muestras aleatorias de tamaños n_1 y n_2. Para la primera población tendríamos una media muestral $\overline{X_1}$ y para la segunda una media muestral $\overline{X_2}$

Consideremos la variable aleatoria que se forma restando ambas variables,se llamará diferencia de las medias muestrales y se denota por: $\overline{X}_1$- $\overline{X}_2$



PROPIEDADES

Siempre que las poblaciones de las que se extraigan las muestras sigan distribuciones normales de medias μ_1 y μ_2 y desviaciones típicas σ_1 y σ_2, respectivamente**,** o también en caso de que las muestras se extraigan de poblaciones que no siguen una distribución normal pero el tamaño de ambas muestras: n_1 y n_2, es mayor que 30 se verifica que:

La distribución de las medias muestrales sigue una distribución normal de media μ_1 - μ_2 y desviación típica $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$: $\overline{X}_1$ - $\overline{X}_2$ =N(μ_1 - μ_2, $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$)

Ejemplo:

Si dos poblaciones se distribuyen con medias 40 y 30 y desviaciones típicas 4 y 3, respectivamente, ¿cuál es la distribución de las diferencias de las medias muestrales para muestras de tamaño 36 extraidas de esas poblaciones? ¿Cuál es la probabilidad de que la diferencia sea mayor que 8?

Los tamaños de las muestras son n_1=n_2=36>30. Las poblaciones tienen unas medias y desviaciones típicas de: μ_1=40, μ_2=30; σ_1=4, σ_2=3. Por tanto:

$$\overline{X}_2 - \overline{X}_1 = N(40-30, \sqrt{\tfrac{16}{36}+\tfrac{9}{36}}) = N(10, 5/6)$$

P($\overline{X}_2$ - $\overline{X}_1$ >8)=P(Z>$\frac{8-10}{5/6)}$)=P(Z>-2’4)=1-P(Z<-2’4)=1-(1-P(Z<2’4))=P(Z<2’4)=0’9918

EJERCICIOS RESUELTOS : I DISTRIBUCIONES MUESTRALES

(solución al final del tema)

1. Las estaturas de 1200 estudiantes de un centro de enseñanza se distribuyen normalmente con una media de 1’72m y desviación típica 0’09. Si se toma al azar una muestra de 36 estudiantes calcular: a) la probabilidad de que la media sea inferior a 1’75m. b) la probabilidad de que la media esté entre 1’68m y 1’73 m
2. Una máquina fabrica piezas de precisión. En su producción habitual, fabrica un 3% de piezas defectuosas. Un cliente recibe una caja de 500 piezas procedentes de esa fábrica. Calcular la probabilidad de que: a)haya más de un 5% de piezas defectuosas en la caja, b)haya menos de 10 piezas defectuosas en la caja
3.La duración de las llamadas de teléfono en dos departamentos de atención al cliente sigue una distribución normal con desviación típica de 6 mn el el primero y 7 en el segundo. Con el fin de estimar la diferencia de medias, se elige una muestra compuesta por 42 llamadas del primer departamento y 38 del segundo.

Si se anotan los tiempos medios de conversación y se obtienen 16 y 14 mn, respectivamente, ¿cuál es la distribución para la diferencia de medias muestrales?
4. El gasto bimensual en electricidad por familia en España se distribuye según una ley normal de media 142’32€ y desviación típica 8’5€. A) Halla la probabilidad de que una muestra de 25 familias, escogidas al azar, tenga un gasto medio de electricidad superior a 144’6€, b)Halla la probabilidad de que una muestra de 100 familias, elegidas al azar, tenga un gasto medio de electricidad superior a 144’6€
5. Una fábrica de chocolate ha fabricado 800 chocolatinas con un peso medio de 150gr y una desviación típica de 20 gr. Calcula la probabilidad de que una muestra de 80 chocolatinas, elegidas al azar, tenga un peso total mayor de 12 kg y 400 gr.
6. Los pesos, em gramos, de los tornillos que fabrica una máquina se distribuyen según una variable N(100, 8). Se toman muestras de 25 tornillos. Calcular la probabilidad de que una muestra elegida al azar tenga un peso medio mayor de 101’5 gr.
7. De una población de 120 alumnos hay 48 que tienen 2 o más hermanos. Si se toma una muestra al azar de 40 alumnos ¿Cuál es la probabilidad de que haya más de un 55% de alumnos con 2 o más hermanos?
8. En un saco mezclamos judías blancas y pintas en la relación de 14 blancas por cada pinta. Extraemos al azar un puñado de 100 judías, ¿cuál es la probabilidad de que haya entre 5 y 10 judías pintas en el puñado?
9. Las pruebas de control de calidad para un modelo A de bombillas ha determinado que su duración, en horas, sigue una distribución N(4000, 270), mientras que para otro modelo B la duración sigue una N(3900,280). Si se toman muestras, al azar, de 50 bombillas de cada uno de los modelos, halla la probabilidad de que la diferencia de medias de las duraciones de las bombillas de cada modelo sea inferior a 50 horas
10. Uno de los principales fabricantes de televisores compra los tubos de rayos catódicos a dos compañías. Los tubos de la compañía A tienen una vida media de 7.2 años con una desviación estándar de 0.8 años, mientras que los de la B tienen una vida media de 6.7 años con una desviación estándar de 0.7. Determine la probabilidad de que una muestra aleatoria de 34 tubos de la compañía A tenga una vida promedio de al menos un año más que la de una muestra aleatoria de 40 tubos de la compañía B.

ESTIMACIÓN DE PARÁMETROS

El objetivo principal de la estadística inferencial es la estimación, esto es que mediante el estudio de una muestra se quiere generalizar las conclusiones al total de la población. Normalmente los parámetros poblacionales no son conocidos y hemos de estimarlos a partir de los parámetros muestrales.
Se denomina estimador a un valor que puede calcularse a partir de los datos muestrales y que proporciona información sobre un parámetro de la población. 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, etc.
Existen dos formas distintas de realizar estas estimaciones: puntuales o por intervalos.

1. Estimación puntual

Consiste en estimar para la población el mismo resultado obtenido para la muestra
Ejemplo 1: Queremos estimar la duración media de las baterías fabricadas por una determinada empresa, para ello tomamos una muestra de 150 baterías observamos su duración y calculamos la media muestral que resulta ser de 8277 horas de uso. Concluimos que 8277 es la duración media de todas las baterías fabricadas.

Ejemplo 2: Deseamos conocer la proporción de conductores que superan la velocidad máxima permitida en un punto concreto de una autopista. Tomamos una muestra de 300 conductores y observamos que el 22% de ellos supera esa velocidad, estimamos entonces que el 22% de todos los conductores superan el límite de velocidad en dicha autopista o bien 0'22 si lo escribimos en términos de proporción.

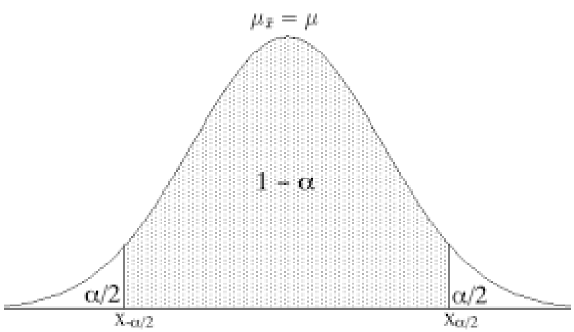
***2. Estimación por intervalos de confianza*.**

Los estimadores puntuales tienen un valor diferente según la muestra elegida, no podemos conocer el grado de precisión de la estimación que estamos realizando y, por tanto, la estimación realizada puede resultar poco fiable.
Para evitar este problema podemos estimar no el valor concreto de dicho parámetro sino un posible rango de valores o intervalo, en el que se encontrará el parámetro con una probabilidad, o nivel de confianza, prefijado.

Por ejemplo, imaginemos que queremos conocer la vida media de los focos fabricados por una empresa. Para ello, elegimos una muestra de 30 focos y calculamos la vida media de esa muestra, obteniendo una duración media de 780 horas.
Si realizamos una estimación puntual: asignar esas 780 horas como media de duración para todos los focos fabricados, sabemos que ese parámetro resultaría diferente si hubiésemos elegido otra muestra distinta. Resulta más útil calcular dos valores entre los que se encontrará dicha duración media con un nivel de confianza prefijado de antemano. En nuestro ejemplo, podríamos decidir buscar un intervalo en el que se encontrase la vida media de los focos con un nivel de confianza del 95%. Una vez hechos los cálculos, encontraríamos que ese intervalo es (765,795). Es decir, podemos esperar que la vida media de todos los focos fabricados esté comprendida entre 765 horas y 795 horas con un nivel de confianza del 95%. O lo que es lo mismo, si elegimos 100 muestras de tamaño 30 y para cada una de ellas calculamos el intervalo de confianza resultante, acertaremos en nuestro pronóstico 95 de las 100 veces.
A lo que acabamos de hacer le llamamos estimación por intervalos de confianza. Vamos a explicar el método que seguimos para llegar a las conclusiones de nuestro ejemplo.

Definiciones

Llamamos ***Intervalo de confianza*** *al intervalo que contiene al parámetro que se está estimando con una probabilidad o nivel de confianza* prefijado.



Nivel de confianza *es la "probabilidad" de que el intervalo calculado contenga al verdadero valor del parámetro.*
Se indica por 1-α y normalmente se da en términos de porcentaje (1-α)% (en el ejemplo de los focos sería 95%, por lo que α=0'05). Le llamamos nivel de confianza y no probabilidad porque una vez extraida una muestra, el intervalo de confianza puede contener o no al verdadero valor del parámetro, lo que sabemos es que si repitiésemos el proceso con muchas muestras el (1-α)% de los intervalos así construidos contendrían al verdadero valor del parámetro.
Llamamos ***nivel de significación*** *al valor de α. Representa la probabilidad de equivocarnos al estimar que el parámetro se encuentra en ese intervalo de confianza.*

Como es lógico no tiene sentido hallar intervalos de confianza con "poca confianza", los valores más comunes son: 90%, 95% y 99%.

Intervalos de confianza para la media de una distribución normal con desviación típica conocida.

En una población desconocemos la media, μ, y queremos estimarla, a partir de la media $\bar{x}$de una muestra, con un nivel de confianza 1-α.
Recordemos que la distribución de medias muestrales $\overline{X}$ sigue una distribución normal cuya media coincide con la de la población y cuya desviación típica es $\frac{\sigma}{\sqrt{n}}$; . Por lo tanto la variable normalizada Z=$\frac{\overline{\mathrm{X}}-\mu}{\sigma/\sqrt{n}}$ seguirá una distribución N(0,1).
Recordemos también que hemos visto en el tema de probabilidad como calcular los valores críticos $z_{\alpha/2}$, para un determinado nivel de confianza, verificándose que
$\mathrm{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$.
Aplicando ambos resultados obtenemos:

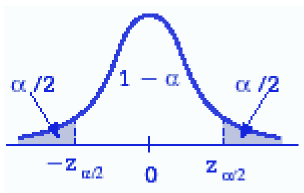
$$\mathrm{P}\left(-z_{\alpha/2} < \frac{\overline{\mathrm{X}}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha \Rightarrow P(-z_{\alpha/2}.\sigma/\sqrt{n} < \overline{\mathrm{X}}-\mu < z_{\alpha/2}.\sigma/\sqrt{n}) = 1-\alpha \Rightarrow$$

$$P(\overline{\mathrm{X}}-z_{\alpha/2}.\sigma/\sqrt{n} < \mu < \overline{\mathrm{X}}+z_{\alpha/2}.\sigma/\sqrt{n}) = 1-\alpha$$

Hemos obtenido el intervalo que contiene a la media poblacional con un nivel de confianza
N_c = 1-α : $I_c = (\overline{\mathrm{X}}-z_{\alpha/2}.\sigma/\sqrt{n}\ ,\ \overline{\mathrm{X}}+z_{\alpha/2}.\sigma/\sqrt{n})$

A la expresión E= $z_{\alpha/2}.\sigma/\sqrt{n}$ se le denomina ***margen de error*** y en ocasiones se expresa en tanto por ciento. Obsérvese que se trata del radio del intervalo.
Tamaño de la muestra Para determinar el tamaño de la muestra, se puede fijar el error máximo admisible y el nivel de confianza $1-\alpha$, a partir de esos datos podemos despejar n en la fórmula anterior para calcular el tamaño mínimo que debe tener la muestra para que se cumplan esas condiciones: E= $z_{\alpha/2}.\sigma/\sqrt{n} \Rightarrow \sqrt{n} = \frac{z_{\alpha/2}.\sigma}{E} \Rightarrow n = \left(\frac{z_{\alpha/2}.\sigma}{E}\right)^2$

NOTA: el valor crítico $z_{\alpha/2}$ asociado al nivel de confianza 1-α verifica que
$\mathrm{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$. Fijándonos en el significado gráfico vemos que entonces $\mathrm{P}(Z<z_{\alpha/2}) = 1-\frac{\alpha}{2}$ pues
$\mathrm{P}(Z<z_{\alpha/2}) = 1-\alpha+\frac{\alpha}{2} = 1-\frac{\alpha}{2}$
No tenemos más que usar las tablas para determinar dicho valor crítico.



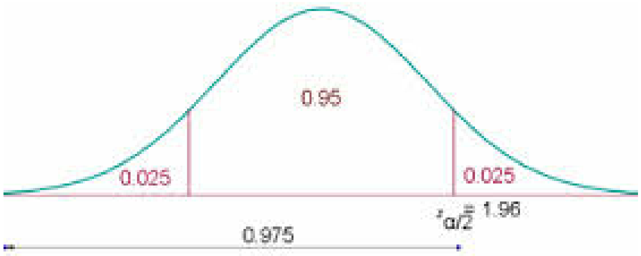
Ejemplo: Se desea estimar la media del tiempo empleado por un nadador en una prueba olímpica. Para ello, se cronometran 10 pruebas, obteniéndose una media de 41,5 minutos. Sabiendo por otras pruebas que la desviación típica de esta variable para este nadador es de 0,3 minutos

a) Obtener un intervalo en el que se encuentre la media del tiempo empleado por el nadador con un nivel de confianza del 95%.
b) ¿Cuantas pruebas habría que cronometrar para que el margen de error en la estimación de la media fuese inferior a tres segundos. (Suponemos que la variable que mide el tiempo del nadador sigue una distribución normal.)

a) Del enunciado del problema se desprenden directamente los siguientes datos:

$\overline{\mathrm{X}}$ = 41 5. σ = 0'3. n = 10 ; $1-\alpha$ = 95%=0'95

En primer lugar buscamos $z_{\alpha/2}$ para el nivel de confianza del 95%. Dado que dentro del intervalo $(-z_{\alpha/2}, z_{\alpha/2})$ ha de quedar el 95% de la población, fuera de él estará el 0'05% y como el porcentaje que queda a la derecha y a la izquierda del intervalo ha de ser el mismo, P(Z>$z_{\alpha/2}$)=0'025. Por tanto, P(Z<$z_{\alpha/2}$)=1-0'025=0'975.



(también podría utilizarse la fórmula de la nota anterior para llegar a este resultado)

Buscando ahora en las tablas de la N(0,1) encontramos que el valor crítico es $z_{\alpha/2}$=1'96
En consecuencia, el intervalo de confianza para la media de tiempo del nadador será:
$(41'15-\frac{0'3}{\sqrt{10}}.1'96, 41'15+\frac{0'3}{\sqrt{10}}.1'96) = (41'314, 41'686)$
También se puede expresar así: Se estima que la media es 41,5 más menos un margen de error del 18,59%. (Recordemos que el margen de error cometido en la estima es el radio del intervalo, es decir $\frac{0'3}{\sqrt{10}}.1'96$ =0,1859)
b) Nos piden el tamaño de la muestra para que, en las mismas condiciones, el margen de error sea inferior a 3 seg, es decir 0,05 minutos. El márgen de error venía dado por la expresión: $z_{\alpha/2}.\sigma/\sqrt{n}$, en nuestro caso $\frac{0'3}{\sqrt{n}}.1'96$, por tanto hemos de calcular n para que $\frac{0'3}{\sqrt{n}}.1'96$<0'05 . Despejando n obtenemos: $\frac{0'3}{\sqrt{n}}.1'96 < 0'05 \Rightarrow 0'588 < 0'05\sqrt{n} \Rightarrow$
$\Rightarrow \frac{0'588}{0'05} < \sqrt{n} \Rightarrow 11'76 < \sqrt{n}$ y elevando al cuadrado se obtiene: n>138,29.
Es decir, para obtener un error inferior a 0,05 minutos, deberemos tomar una muestra de al menos 139 pruebas.

NOTA:
Como es evidente, en la estimación por intervalos de confianza un dato importante es el tamaño de la muestra. Parece claro que, a igual nivel de confianza, cuanto mayor sea el tamaño de la muestra menor sera el radio del intervalo de confianza, puesto que el valor obtenido en la muestra se acercará más al valor real de la población y por tanto el margen de error cometido (radio del intervalo) se hará más pequeño.
Si el tamaño de la muestra permanece constante y variamos el nivel de confianza, el radio del intervalo será más grande cuanto mayor sea dicho nivel. Es decir, el margen de error será más grande cuanto mayor sea la precisión exigida.

Ejercicios resueltos

1. En una encuesta se pregunta a 10 000 estudiantes de Bachillerato sobre su consumo de refrescos semanal, encontrándose una media de 5 botes, con una desviación típica de 2 botes.
a) Halla los intervalos de confianza para la media 80% y al 95% de probabilidad.
b) Si aceptamos un error de 0.25 botes para la media de la población, con una fiabilidad de 0.8, ¿a cuántos estudiantes es necesario entrevistar? ¿Y si queremos un nivel de confianza del 90%?

Solución:
A) n=10 000; $\bar{x}$=5; σ=2
Para el 80% : comenzamos hallando el valor crítico $z_{\alpha/2}$; 1-α=0'8⇒ $\alpha = 0'2$ y α/2=0'1
P(Z<$z_{\alpha/2}$)=$1-\frac{\alpha}{2} = 1-0'1$ =0'9⇒$z_{\alpha/2}$=1'285.

Intervalo de confianza (5-1’285.$\frac{2}{\sqrt{10000}}$, 5+1’285.$\frac{2}{\sqrt{10000}}$)=(4’9743,5’0257)

Para el 90%: comenzamos hallando el valor crítico $z_{\alpha/2}$; 1-α=0’9⇒ $\alpha = 0'1$ y $\alpha/2$=0’05

P(Z<$z_{\alpha/2}$)=$1-\frac{\alpha}{2} = 1-0'05$ =0’95⇒$z_{\alpha/2}$=1’645

Intervalo de confianza (5-1’645.$\frac{2}{\sqrt{10000}}$, 5+1’645.$\frac{2}{\sqrt{10000}}$)=(4’9896, 5’0104)

B) Para el 80% : 1’285.$\frac{2}{\sqrt{n}} < 0'25 \Rightarrow \left(\frac{1'285.2}{0'25}\right)^2 < n \Rightarrow 105'6784 < n$ Luego hemos de tomar una muestra mínima de 106 alumnos

Para el 90%: 1’96.$\frac{2}{\sqrt{n}} < 0'25 \Rightarrow \left(\frac{1'96.2}{0'25}\right)^2 < n \Rightarrow 245'86 < n$ Luego hemos de tomar una muestra mínima de 246 alumnos

2. En una de las pruebas de acceso a la Universidad, la variable "puntuación obtenida en la materia de Matemáticas aplicadas a las Ciencias Sociales II" se distribuye normalmente con una desviación típica de 1,38. En una muestra de 50 alumnos se ha medido la misma variable y el valor obtenido para la media es de 4,93 puntos. Halla un intervalo de confianza para la media poblacional con una confianza del 92 % y explica el significado de este intervalo.

Solución:

Sea X = puntuación obtenida en MCS II. Se tiene que X es $N(\mu, 1.38)$). Además, tenemos los siguientes datos:n=50; $\bar{x}$ =4.93; 1-α=0’92⇒ $\alpha = 0'08 \Rightarrow \alpha/2 = 0'04$

P(Z<$z_{\alpha/2}$)=1-0’04=0’96. Buscando en las tablas $z_{\alpha/2}$=1’75 Por tanto el intervalo es:

(4’93-1’75.$\frac{1'38}{\sqrt{50}}$, 4’93+1’75.$\frac{1'38}{\sqrt{50}}$)=((4.5885 , 5.2715)

El intervalo obtenido significa que en el 92% de las posibles muestras de tamaño 50, la media de la nota en MCS II está en el intervalo (4.5885 , 5.2715)

3. El gasto mensual (en euros) de una familia en electricidad, para las familias de una cierta ciudad, sigue una distribución normal de media desconocida y desviación típica 25 euros.

a) A partir de una muestra de 100 familias de esa ciudad, se estableció el intervalo (45,55) como intervalo de confianza para el gasto medio mensual, por familia en electricidad. ¿cuál es el valor de $\bar{x}$? ¿con qué nivel de confianza se construyó el intervalo?

b) ¿Qué número de familias tendrías que seleccionar para que el nivel de confianza del intervalo anterior sea del 97%?

c) ¿Qué número de familias tendrías que seleccionar, como mínimo, para garantizar, con un nivel de confianza del 99 %, una estimación de ese gasto medio con un error no superior a 3 euros?

Solución

a) El intervalo que nos dan (45,55) está centrado en la media $\bar{x}$ y por tanto esta será (45+55)/2=50. Veamos ahora con qué nivel de confianza 1-α se construyó el intervalo.

Sabemos que 45=$\bar{x} - z_{\alpha/2}.\sigma/\sqrt{n} \Rightarrow$ 45=50-$z_{\alpha/2}\frac{25}{\sqrt{100}}$; Vamos a despejar $z_{\alpha/2}$

45=50-$z_{\alpha/2}\frac{25}{\sqrt{100}} \Rightarrow -5 = -z_{\alpha/2}.\frac{25}{\sqrt{100}} \Rightarrow z_{\alpha/2} = \frac{5.10}{25} = 2.$

Sabemos que P(Z<$z_{\alpha/2}$)=$1-\frac{\alpha}{2}$, buscando en las tablas :

P(Z<$z_{\alpha/2}$)= P(Z<2)=0’9773=$1-\frac{\alpha}{2}$, despejando α=0’0454 y el nivel de confianza será 1-α=1-0’054=0’9546. Es decir, 95’46%

b) Nos piden que calculemos n para que el intervalo (45,55) tenga un nivel de confianza del 97%. Entonces:

1-α=0’97⇒ $\alpha=0'03 \Rightarrow \alpha/2=0'015$. P(Z<$z_{\alpha/2}$)=$1-\frac{\alpha}{2}$=1-0’015=0’985, buscando en las tablas $z_{\alpha/2}$ =2’17

Sabemos que el radio del intervalo, en nuestro caso 5, es igual a $z_{\alpha/2}.\sigma/\sqrt{n}$. Es decir:
5=2’17.$\frac{25}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{2'17.25}{5} = 10'85 \Rightarrow n = 117'25$, como el número de familias ha de ser entero hemos de tomar 118 familias.

c) En primer lugar hemos de calcular $z_{\alpha/2}$ para 1-α=0’99

1-α=0’99⇒ $\alpha=0'01 \Rightarrow \alpha/2=0'005$. P(Z<$z_{\alpha/2}$)=$1-\frac{\alpha}{2}$=1-0’005=0’995, buscando en las tablas $z_{\alpha/2}$ =2’575

El error será E=2’575.$\frac{25}{\sqrt{n}}$<3⇒ $\frac{2'575.\ 25}{3} < \sqrt{n} \Rightarrow \left(\frac{2'575.\ 25}{3}\right)^2 < n \Rightarrow 460'46 < n$ Hemos de coger por lo tanto 461 familias.

Intervalos de confianza para la proporción

Supongamos ahora que queremos estimar la proporción de individuos de una población que cumplen una determinada característica a partir de la proporción de los que la cumplen en una muestra de tamaño n y con un nivel de confianza 1-α

Sabemos que las proporciones con tamaños muestrales n ≥30 siguen distribuciones normales con media p y desviación típica $\sqrt{\frac{p.q}{n}}$ siendo q=1-p. $\hat{P} = N(p, \sqrt{\frac{p.q}{n}})$

Entonces la varible tipificada Z=$\frac{\hat{P}-p}{\sqrt{\frac{p.q}{n}}}$ seguirá una distribución N(0,1)

Como no disponemos del valor de p tenemos que estimarla de forma puntual mediante la proporción muestral obtenida:$\hat{p}$

De este modo, por la definición de intervalo de confianza la P($-z_{\alpha/2} < \frac{\hat{P}-p}{\sqrt{\frac{\hat{p}.\hat{q}}{n}}} < z_{\alpha/2}$)$=1-\alpha \Rightarrow$

$$P(-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}} < \hat{P}-p < z_{\alpha/2})=1-\alpha \Rightarrow P(\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}} < p < \hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}})=1-\alpha$$

Hemos obtenido el intervalo que contiene a proporción poblacional con un nivel de confianza

N_c = 1-α : $I_c = (\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}\ , \hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}})$

a expresión E= $z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$ se le denomina ***margen de error*** y en ocasiones se expresa en tanto por ciento. Obsérvese que se trata del radio del intervalo.

Tamaño de la muestra Para determinar el tamaño de la muestra, se puede fijar el error máximo admisible y el nivel de confianza $1-\alpha$, a partir de esos datos podemos despejar n en la fórmula anterior para calcular el tamaño mínimo que debe tener la muestra para que se cumplan esas condiciones: E=$z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}} \Rightarrow \sqrt{n} = z_{\alpha/2}\frac{\sqrt{\hat{p}.\hat{q}}}{E} \Rightarrow n = \left(z_{\alpha/2}\frac{\sqrt{\hat{p}.\hat{q}}}{E}\right)^2$

Ejemplos

1.Para saber qué proporción de alumnos de la ESO tienen teléfono móvil con conexión de datos se selecciona una muestra de 500 alumnos, de ellos contestan afirmativamente 225. ¿Cuál es el intervalo de confianza para la proporción de los alumnos que tienen móvil con conexión de datos, con un nivel de confianza del 95%?

Solución

En primer lugar hemos de hallar la proporción en la muestra $\hat{p}$= 225/500=0’45; por tanto $\hat{q}$=1-0’45=0’55; n=500

Ahora tenemos que calcular el valor crítico $z_{\alpha/2}$ para un nivel de confianza 1-α=0’95;

$P(Z<z_{\alpha/2}) = 1-\frac{\alpha}{2} = 1-\frac{0'05}{2} = 1-0'025 = 0'975$ Buscando en las tablas hallamos $z_{\alpha/2}$=1’96

Por tanto el intervalo de confianza para la proporción que nos piden será:

$I_c =(\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}, \hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}})=(0'45-1'96.\sqrt{\frac{0'45.0'55}{500}}, 0'45+1'96.\sqrt{\frac{0'45.0'55}{500}})=$

=(0’4064, 0’4936)

2. En una encuesta hecha por los alumnos y alumnas de un Instituto a un total de 100 votantes elegidos al azar en su Ayuntamiento, se indica que el 55% volvería a votar por el alcalde actual.

a) Calcular un intervalo de confianza al 99% y otro al 99,73% para la proporción de votantes favorables al alcalde actual.

b)¿Cuáles deben ser los tamaños muestrales en el sondeo para tener, con los mismos niveles de confianza, la certeza de que el alcalde actual salga reelegido por mayoría absoluta, en el caso de arrojar la encuesta los mismos resultados?

Solución

N=100; $\hat{p}=0'55$; $\hat{q}=0'45$

a) Para calcular el intervalo de confianza para el 99% hallamos primero $z_{\alpha/2}$ para el nivel de confianza 1-α=0’99;

$P(Z<z_{\alpha/2}) = 1-\frac{\alpha}{2} = 1-\frac{0'01}{2} = 1-0'005 = 0'995$ Buscando en las tablas hallamos $z_{\alpha/2}$=2’57

Por tanto el intervalo de confianza para la proporción que nos piden será:

$I_c =(\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}, \hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}})=(0'55-2'57.\sqrt{\frac{0'55.0'45}{100}}, 0'55+2'57.\sqrt{\frac{0'55.0'45}{100}})=$

=(0’422, 0’677)

Para un nivel de confianza del 99’73%: $P(Z<z_{\alpha/2}) = 1-\frac{\alpha}{2} = 1-\frac{0'0027}{2} = 1-0'00135 = 0'99865$ Buscando en las tablas hallamos $z_{\alpha/2}$=2’98

Por tanto el intervalo de confianza para la proporción que nos piden será:

$I_c =(\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}, \hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}})=(0'55-2'98.\sqrt{\frac{0'55.0'45}{100}}, 0'55+2'98.\sqrt{\frac{0'55.0'45}{100}})=$

=(0’401, 0’698)

B) Nos piden el valor de n condicionado a que todos los valores del intervalo de confianza sean superiores a 0,5 (mayoría absoluta), es decir que, dado que la media muestral es 0,55, el radio del intervalo ha de ser necesariamente menor que 0,55-0,5 = 0,05.

Por tanto, en el caso del nivel de confianza del 99% $E=z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}} = 2'57.\sqrt{\frac{0'55.0'45}{n}}$ <0’05, despejando n resulta n>653’8, es decir hemos de tomar una muestra de al menos 654 personas.

En el caso del 99’73% se hace igual resultando n>891 personas.

Intervalos de confianza para la diferencia de medias

Supongamos que tenemos 2 poblaciones que se distribuyen normalmente con desviaciones típicas σ_1 y σ_2 conocidas y cuyas medias respectivas μ_1 y μ_2 son desconocidas. Supongamos que tomamos una muestra de tamaño n_1 de la primera población y obtenemos para dicha muestra una media $\overline{x_1}$. Asímismo, tomamos una muestra de la segunda población de tamaño n_2, obteniendo para ella una media $\overline{x_2}$. ¿cómo podemos estimar la diferencia de las medias poblacionales a partir de estas muestras, con un nivel de confianza prefijado 1-α?

Recordemos que la diferencia de las medias muestrales sigue una distribución normal : N($\mu_1 - \mu_2$, $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$) por similitud a lo explicado en los intervalos de confianza para la media y para la proporción, la variable tipificada Z=$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}}$ sigue una distribución N(0,1) y, por la definición de intervalo de confianza, P($-z_{\alpha/2}$<$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}}$<$z_{\alpha/2}$)=1-α; en consecuencia, despejando como lo hemos hecho en las dos situaciones anteriores,
P($(\overline{X}_1 - \overline{X}_2)$-$z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}$ <$(\mu_1 - \mu_2)$<$(\overline{X}_1 - \overline{X}_2)$+$z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}$)=1-α. Es decir, el intervalo de confianza para la diferencia de las medias muestrales, con un nivel de confianza 1-α será:

($(\overline{X}_1 - \overline{X}_2)$-$z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}$, $(\overline{X}_1 - \overline{X}_2)$+$z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}$)

Ejemplo: En un hospital se realiza un estudio sobre la influencia del tabaco en el peso de los recién nacidos. Se consideran 2 grupos de mujeres embarazadas: fumadoras y no fumadoras y se obtienen los siguientes datos sobre el peso de sus hijos:

	Muestra	Peso medio recién nacidos	Desviación típica
Mujeres no fumadoras	n_1=300	$\overline{x_1}$=3’6 Kg	$\sigma_1 = 0'5$
Mujeres fumadoras	n_2=220	$\overline{x_2}$=3’2 Kg	$\sigma_2 = 0'8$

Decidir como influye que la madre sea fumadora en el peso de su hijo al nacer, utilizando un nivel de confianza para la diferencia de las medias del 95%

Solución:

En primer lugar calculamos el valor crítico $z_{\alpha/2}$ para dicho nivel de confianza.

(Z<$z_{\alpha/2}$) = 1-$\frac{\alpha}{2} = 1 - \frac{0'05}{2} = 1 - 0'025 = 0'975$ Buscando en las tablas hallamos $z_{\alpha/2}$=1’96

Luego el intervalo de confianza para la diferencia de las medias será:

((3’6–3’2)-1’96$\sqrt{\frac{0'5^2}{300} + \frac{0'8'^2}{220}}$,(3’6–3’2)+1’96$\sqrt{\frac{0'5^2}{300} + \frac{0'8'^2}{220}}$)=(0’28,0’52)

Interpretación del resultado: La diferencia de medias está en el intervalo (0’28, 0’52); es decir, hay una diferencia positiva entre los pesos de los bebés 0’28<$\mu_1 - \mu_2 < 0'52$)
Así, el peso medio de un bebé de una madre no fumadora supera como mínimo en 0’28 Kg y como máximo en 0’52 Kg al de un bebé de madre fumadora, con un nivel de confianza del 95%

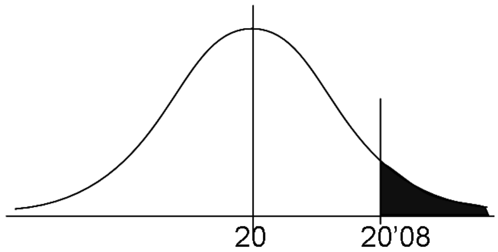
EJERCICIOS RESUELTOS: II INTERVALOS DE CONFIANZA
(solución al final del tema)

11. El número de días de permanencia de los enfermos en un hospital sigue una ley normal de media μ días y desviación típica 3 días.
a)Determinar un intervalo de confianza para estimar μ a un nivel del 97%, con una muestra aleatoria de 100 enfermos cuya media es 8'1 días.
b)¿Qué tamaño mínimo debe tener una muestra aleatoria para poder estimar μ con un error máximo de 1 día y un nivel de confianza del 92%?
12. El tiempo diario que los jóvenes pasan ante el televisor sigue una distribución normal con desviación típica 20 minutos. Una muestra aleatoria de 100 chicos ha dado un tiempo medio de 170 minutos.
a)obtener el intervalo de confianza del 90% para el tiempo medio que los jóvenes pasan ante el televisor. b)¿qué tamaño mínimo debe tener la muestra si deseamos que el error cometido al estimar la media con un nivel de confianza del 99% no exceda los 0'5 minutos?
13. Para efectuar un control de calidad sobre la duración en horas de un modelo de juguetes electrónicos se elige una muestra aleatoria de 36 juguetes de ese modelo obteniéndose una duración media de 97 horas. Sabiendo que la duración de los juguetes electrónicos de ese modelo se distribuye normalmente con una desviación típica de 10 horas,
a) encontrar el intervalo de confianza al 99,2% para la duración media de los juguetes electrónicos de ese modelo. b) Interpretar el significado del intervalo obtenido
14. El coeficiente intelectual de los individuos presentes en una sala puede suponerse que sigue una distribución normal de media μ y varianza igual a 81
a)¿cuánto vale μ si sabemos que solo un 10% de las personas en la sala sobrepasa un coeficiente intelectual de 105?
En los dos siguientes apartados supondremos que $\mu = 95$:
b)elegida una persona al azar de la sala, ¿cuál es la probabilidad de que su coeficiente intelectual esté entre 86 y 107? c)elegimos 9 personas al azar de la sala y calculamos la media de sus coeficientes intelectuales, ¿cuál es la probabilidad de que esa media esté entre 86 y 107
15. El peso de los usuarios de un gimnasio tiene una media desconocida y una desviación típica de 5.4 kg. Tomamos una muestra aleatoria de tamaño 100, obteniendo una media de 60 kg. a) Calcula con un nivel de confianza del 95 % el intervalo de confianza para el peso medio de todos los usuarios. b) Interpreta el significado del intervalo obtenido.
c) Se realiza la siguiente afirmación: "peso medio de un usuario de ese gimnasio está comprendido entre 58,5 y 61.5 kg. ¿Con qué probabilidad esta afirmación es correcta?
16. Se sabe que el nivel medio de protombina en una población normal es de 20 mg/100 ml de plasma, con una desviación típica de 4 mg/100 ml. Se toma una muestra de 40 individuos en los que la media es 18.5 mg/100 ml. ¿Es la muestra comparable con la población, con un nivel de significación de 0.05? NOTA: la muestra es comparable con la población con un cierto nivel de confianza si la media pertenece al intervalo de confianza obtenido para ese nivel.
17. En un país se selecciona aleatoriamente una muestra de 900 personas. A la salida de los colegios electorales se les preguntó si habían votado al partido X y 289 contestaron que sí y el resto que no. Determinar un intervalo que nos de el porcentaje de votos del partido X con un nivel de confianza del 95%, explicando los pasos realizados para su obtención.
18. En una población, por cada persona que fuma 4 no lo hacen. Calcular el tamaño mínimo que debe tener una muestra de dicha población para que, con un nivel de confianza del 95%, la proporción muestral y la poblacional no difieran en más de 0´04. Explicar los pasos seguidos para obtener la respuesta.

CONTRASTE DE HIPÓTESIS

Si sospechamos que una moneda ha sido trucada para que se produzcan más caras que cruces al lanzarla al aire, podríamos realizar 30 lanzamientos, tomando nota del número de caras obtenidas. Si obtenemos un valor demasiado alto, por ejemplo 25, consideraríamos que el resultado es poco compatible con la hipótesis de que la moneda no está trucada, y concluiríamos que las observaciones contradicen dicha hipótesis.
En esta sección vamos a ver como, a partir de los resultados obtenidos en las muestras, podemos aceptar o rechazar determinadas hipótesis que suponemos cumple un parámetro de la población.
Ejemplo: hace 10 años se realizó una muestra sobre la altura de los varones españoles obteniéndose que esta se distribuía normalmente con media 170cm y desviación típica 10 cm. En la actualidad se ha tomado una muestra a 100 varones adultos resultando que la media es de 173cm. ¿Podemos afirmar, con un nivel de confianza del 99% que la media ha aumentado o, por el contrario, debemos rechazar esa hipótesis?
Las hipótesis son afirmaciones que involucran al total de la población, por lo que su verdad o falsedad solo puede garantizarse evaluando a todos los individuos que la componen. Para aceptarla o rechazarla, a partir de los resultados obtenidos en las muestras, se utilizan procedimientos estadísticos que determinan la probabilidad de que esos resultados sean compatibles con la hipótesis establecida. Si es altamente improbable que, de ser cierta la hipótesis, se hubiesen producido los resultados obtenidos la rechazaremos. Si no es así, aceptamos la hipótesis al no existir razones para pensar que no es cierta.

Un **contraste o test de hipótesis** es el procedimiento estadístico mediante el cual decidimos la aceptación o rechazo de la hipótesis.
Ejemplo: Producimos un componente electrónico cuya resistencia X sigue una distribución con media $\mu = 20\Omega$(ohmios) y desviación típica $\sigma = 0'5\Omega$ Por razones imprevisibles el proceso se desajusta a veces produciendo un aumento o disminución de la resistencia media de los componentes sin variar la desviación típica. Para contrastar si en cierto momento el proceso se ha desajustado, se toma una muestra de tamaño n=5 componentes. Se mide la resistencia de los componentes obteniendo los siguientes datos: x_1=22’2; x_2=21; x_3=18’8; x_4=21’5; x_5=20’5.
¿Podemos decir que el proceso se ha desajustado?
Si el proceso no se ha desajustado, al ser X(20,0’5) la distribución de las medias muestrales sería $\overline{\mathrm{X}}(20, \frac{0'5}{\sqrt{5}})$ ya que estamos tomando muestras de tamaño 5. Calculamos la media de la muestra tomada: $\bar{x} = \frac{22'2+21+18'8+21'5+20'5}{5} = 20'8$



Como podemos ver en la gráfica, 20’08 está muy alejada de la media. Para poder verlo formalmente tipificamos la variable: $\frac{20'8-20}{0'5} = 3'58$. Es decir, se encuentra a una distancia de más de 3 veces y media la desviación típica. Por ello, podemos pensar que la afirmación de que el proceso se ha desajustado es correcta.

Nosotros vamos a estudiar únicamente contrastes de hipótesis para aceptar o rechazar una hipótesis previamente emitida sobre el valor de un parámetro desconocido de la población. Concretamente realizaremos contrastes de medias, proporciones y diferencias de las medias poblacionales.

Veamos a continuación los principales conceptos relacionados con el contraste de hipótesis y su relación con los intervalos de confiana.

CONCEPTOS:

- Llamaremos **hipótesis nula,** y la representaremos por H_0, a la hipótesis que se formula y que queremos contrastar para ser aceptarda o rechazarda. Se llama hipótesis alternativa, y se denomina H_1, a la hipótesis contraria a H_0, de tal forma que la aceptación de la hipótesis nula implica el rechazo de la hipótesis alternativa y viceversa.
  Al formular la hipótesis nula debemos tener en cuenta que esta siempre debe contener una igualdad. Es decir ha de ser de uno de los siguientes tipos: H_0=k; $H_0 \leq$k: $H_0 \geq$k

  Ejemplo 1. Un médico afirma que la proporción de recién nacidos varones en un determinado hospital es del 53%. La hipótesis nula sería H_o=" proporción p =0'53" mientras que la hipótesis alternativa sería H_1= " proporción p$\neq$0,53". En este caso, la hipótesis alternativa se llama **hipótesis alternativa bilateral** porque considera los valores menores y mayores que 0,53. (es decir a ambos lados del eje OX)

  Ejemplo 2: Una fábrica de conservas indica en su información al consumidor que el peso medio de sus latas es de 1000 gr. Una asociación de consumidores quiere aceptar o rechazar esa proposición. En este caso, como hipótesis alternativa nos interesaría tomar H_1= " peso medio menor de 1000 gr." ya que si fuese mayor no haría más que reforzar lo que interesa al consumidor, por ello tomaremos como hipótesis nula H_0=" pesos iguales o mayores a 1000gr.". H_1 es una **hipótesis alternativa unilateral** (solo contendría los valores que están por debajo de 1000 gr., es decir a la izd del eje OX).
  En resumen:

H_0=k	$H_1 \neq$k	Bilateral
$H_0 \leq$k	H_1>k	Unilateral derecho
$H_0 \geq$k	H_1<k	Unilateral izquierdo

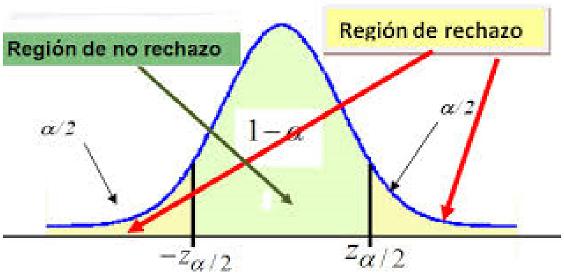
- El test o **contraste de hipótesis** es el procedimiento que me permite decidir si debo aceptar o rechazar la hipótesis nula.. Debemos tener en cuenta lo siguiente:
  - ✓ El test no sirve para demostrar que H_0 es cierta
  - ✓ El text sirve para decidir que, a partir de los datos de la muestra, no puede rechazarse H_0 , es decir que es aceptable suponer que H_0 es cierta.
  - ✓ Sirve también para rechazar H_0 y aceptar H_1 cuando los valores muestrales difieren mucho de los teóricos que se obtendrían de ser cierta H_0.
- Como en cualquier inferencia estadística, podemos cometer **errores** al inferir los resultados de la muestra a la población. En los contrastes de hipótesis estos errores pueden ser de dos tipos: **Error de tipo I**: es el que se produce al rechazar H_0 cuando en

	H_0 es verdadera	**H_0 es falsa**
Se acepta H_0	Decisión correcta	**Error tipo II**
Se rechaza H_0	**Error tipo I**	Decisión correcta

realidad era verdadera**. Error de tipo II**: es el que se produce al aceptar H_0 cuando en realidad era falsa

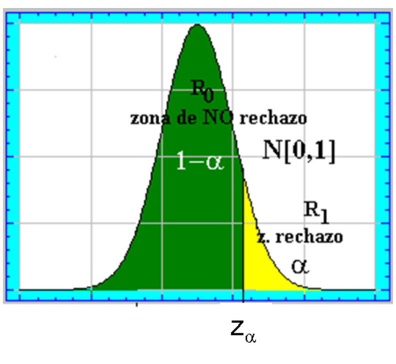
- Llamaremos **nivel de significación** α a la probabilidad de cometer un error de tipo I. Es decir α= P(RechazarH_0 /H_0 es verdadera). Llamamos **nivel de confianza** a la P(Aceptar H_0 /H_0 sea verdadera**)** que, evidentemente es igual a 1-α
  A la probabilidad de cometer un error de tipo II, P(aceptar H_0/H_0 es falsa), se le denota por Llamamos **potencia del contraste** al valor 1-β= P(rechazar H_0/H_0 es falsa).
  Cuando es necesario diseñar un contraste de hipótesis sería conveniente hacerlo de tal manera que la probabilidad de ambos tipos de error fuesen lo más pequeños posibles. Sin embargo, con una muestra de tamaño prefijado disminuir la probabilidad del error de tipo I, α, conduce al aumento de la probabilidad del error de tipo II, β. La única manera de disminuir los dos tipos de errores a la vez es aumentar el tamaño de la muestra, con el consiguiente aumento de costes del proceso.
  Usualmente, se diseñan contrastes con valores de α muy peqeños: 1%, 5% o, como mucho10%. Esto equivale a que los niveles de confianza, (1-α), sean muy altos: 99% , 95% o 90% respectivamente.
- Una vez que hemos formulado la hipótesis nula y decidido el nivel de confianza con el que queremos trabajar, es necesario encontrar un intervalo de valores dentro del cual es lógico que se pueda encontrar el parámetro muestral si es cierta la hipótesis nula, a este intervalo se le llama **zona de aceptación**.
  La zona crítica o de rechazo está formada por los valores que no serían aceptables que tomase el parámetro muestral si la hipótesis nula fuese cierta. Por tanto, si el valor del parámetro muestral perteneciese a esa región tendríamos que tomar la decisión de rechazar la hipótesis nula y aceptar, por el contrario, la alternativa.
  Estas regiones se establecerán, como veremos más adelante, mediante intervalos de confianza según un determinado nivel de significación, prefijado de antemano, que delimita las zonas de aceptación y rechazo
  Veamos estas zonas de forma gráfica para los dos tipos de contraste:

Contraste bilateral: Tendremos que encontrar dos valores,-z_α y z_α, que verifiquen P(-$\mathbf{z}_{\alpha/2}$<Z<$\mathbf{z}_{\alpha/2}$)=1-α



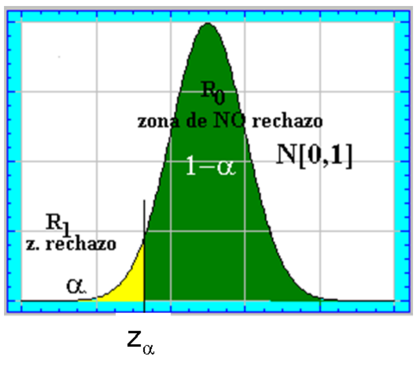
Unilateral derecho

Tendremos que encontrar un valor que verifique P(Z<$\mathbf{z}_\alpha$)=1-α



Unilateral Izquierdo

Tendremos que encontrar un valor que verifique P(Z>$\mathbf{z}_\alpha$)=1-α



Los anteriores ejemplos gráficos se refieren a distribuciones N(0,1) porque en el contraste de hipótesis, como hicimos anteriormente en la estimación de parámetros por intervalos de confianza, lo primero que haremos será tipificar la variable de forma que los valores críticos $\mathbf{z}_{\alpha/2}$ o $\mathbf{z}_{\alpha}$, según se trate de contrastes bilaterales o unilaterales, se buscarán en las tablas de la N(0,1)

Ejemplos: Se trabaja con la hipótesis de que uno de cada 10 varones manifiesta algún tipo de daltonismo. Elegidos 400 varones, 50 de ellos resultan ser daltónicos,

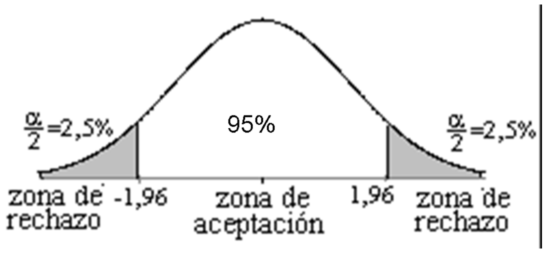
1.Se trabaja con la hipótesis de que uno de cada 10 varones manifiesta algún tipo de daltonismo. Elegidos 400 varones, 50 de ellos resultan ser daltónicos. a)Plantea el contraste de hipótesis correspondiente para decidir si la hipótesis de partida es cierta. b)Halla la zona de aceptación y rechazo de la variable tipificada, con un nivel de significación del 5%, para decidir si la hipótesis de partida es cierta?

Sol:a) Estamos haciendo una hipótesis sobre la proporción, que p=1/10=0’1. Por lo tanto el contraste que plantearemos será: H_0: p=0’1;

H_1: $p\neq$0’1. Contraste bilateral.

b) α=0’05 $\Rightarrow \alpha/2 = 0'025$ y 1- α=0’95; como vimos con anterioridad eso significa que $P(Z<z_{\alpha/2})$ = 0’95+0’025=0’975.

Buscando en las tablas hallamos $z_{\alpha/2}$=1’96



2. Hace 10 años, el consumo medio de electricidad al mes por vivienda en una ciudad era de 320KW. El año pasado se tomó una muestra de 250 viviendas y se obtuvo un consumo medio de 370 KW con una desviación típica de 80. a) Plantea el contraste de hipótesis correspondiente para decidir si el consumo medio ha aumentado. B) Halla la zona de aceptación y rechazo de la variable tipificada, con un nivel de significación del 1%, para decidir si el consumo medio ha aumentado.

Sol: a) En este caso lo que nos interesa es saber si el consumo ha aumentado o no. Si el consumo medio es de 320 o menor podemos concluir que no ha aumentado por tanto nuestro contraste será: H_0: $\mu \leq$320;

H_1: μ>320 Contraste unilateral

b)α=0’01; 1-α=0’99

$P(Z<\mathbf{z}_{\alpha})$=1-α=0’99; buscando en las tablas $\mathbf{z}_{\alpha}$=2’33

PASOS A DAR PARA REALIZAR UN CONTRASTE DE HIPÓTESIS

1. En primer lugar debemos formular la hipótesis nula y la alternativa que serán objeto de contraste. Ambas hipótesis pueden ser enunciadas para un contraste unilateral o bilateral
2. Determinar las regiones de aceptación o rechazo, con un nivel de significación α, para la variable Z normal estandar. Ello conlleva calcular $-z_{\alpha/2}$ y $z_{\alpha/2}$ si se trata de un contraste bilateral o calcular z_{α} si es un contraste unilateral y determinar después la zona en la que sería aceptada la hipótesis nula y la zona en la que sería rechazada.
3.Determinar el estadístico apropiado para el contraste que dependerá del parámetro que estemos contrastando. Teniendo en cuenta lo estudiado anteriormente utilizaremos los siguientes estadísticos:

Estadístico para la media $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$

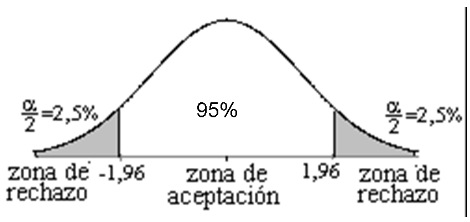
Estadístico para la proporción $Z=\frac{\hat{P}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$

Estadístico para la diferencia de medias $Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}}=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}}$ (En la diferencia de medias la diferencia de medias poblacionales se considerará cero $(\mu_1-\mu_2)=0$)

NOTA: Sabemos que las tres variables son N(0,1) por lo que podemos utilizarlas sin problemas para el contraste.
4. Determinar el valor z_0 correspondiente a sustituir $\overline{X}$, $\hat{P}$, o $(\overline{X_1}-\overline{X}_2)$ por la correspondiente media, proporción o diferencia de medias obtenidas de la muestra
5. Decidir si se acepta o se rechaza la hipótesis en función de que el valor observado z_0 pertenezca a la región de aceptación o a la de rechazo.

Ejemplo 1:
Las estaturas de 16 alumnos/as de 2º de Bachillerato son: 156, 185, 193, 164, 186, 170, 168, 174, 163, 157, 178, 168, 169, 172, 174, 180. Sabemos, además, que las estaturas del alumnado de Bachillerato sigue una ley normal de media desconocida y desviación típica 8,44. Sostenemos que la media poblacional es 171,51. Realizar el contraste de hipótesis para un nivel de significación del 5%. e interpretar el resultado.
Como no nos la dan debemos calcular previamente la media de la muestra
$\overline{x}=\frac{156+185+....+180}{16}=172'31$

Vayamos al contraste de hipótesis:
1. La hipótesis nula vendra dada por H_0: μ=171’51; luego H_1: $\mu \neq 171'51$. Es un contraste bilateral.
2. Calculamos los valores $-z_{\alpha/2}$ y $z_{\alpha/2}$ para un nivel de significación del 5%:
La región de aceptación será $P(-\mathbf{z}_{\alpha/2}<Z<\mathbf{z}_{\alpha/2})=1-\alpha$
$\alpha=0'05 \Rightarrow \alpha/2=0'025$ y $1-\alpha=0'95$; como vimos con anterioridad eso significa que $P(Z<z_{\alpha/2})$ $=1-\frac{\alpha}{2}=1-\frac{0'05}{2}=1-0'025=0'975$ Buscando en las tablas hallamos $z_{\alpha/2}=1'96$. En consecuencia la región de aceptación será (-1’96, 1’96) la de rechazo será $(-\infty,-1'96)\cup(1'96,\infty)$



3. El parámetro sobre el que hacemos la hipótesis es la media. La distribución de medias muestrales sigue una normal $N(\mu,\frac{\sigma}{\sqrt{n}})$, por tanto el estadístico que utilizaremos será

$Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$, en nuestro caso $Z=\frac{\overline{X}-171'51}{8'44/\sqrt{16}}$
4. Calculamos el valor de ese estadístico sustituyendo la media muestral por la obtenida en la prueba: $Z_0=\frac{172'31-171'51}{8'44/\sqrt{16}}=0'3791$
5. Comprobamos si el valor de z_0 pertenece a la región de aceptación o de rechazo: $0'3791\in(-1'96, 1'96)$ que es la zona de aceptación por lo que aceptamos la hipótesis nula.

Ejemplo 2:

Una encuesta a 64 profesionales de una institución reveló que el tiempo medio de empleo en dicho campo era de 5 años, con una desviación típica de 4. Considerando un nivel de significación del 0.05, Suponiendo que el tiempo de empleo se distribuye normalmente, ¿sirven estos datos para contrastar si el tiempo medio de empleo de los profesionales de esta institución está por debajo de los 6 años?

1. En este caso la hipótesis nula será H_0: $\mu \leq 6$ y la hipótesis alternativa μ>6 Se trata por tanto de un contraste unilateral
2. Calculamos la región de aceptación para un nivel de significación α=0'05
$P(Z<Z_\alpha)=1-\alpha=1-0'05=0'95$, buscando en las tablas Z_α= 1'645, por tanto la región de aceptación será $(-\infty, 1'645)$
3. El parámetro sobre el que hacemos la hipótesis es la media. La distribución de medias muestrales sigue una normal $N(\mu, \frac{\sigma}{\sqrt{n}})$, por tanto el estadístico que utilizaremos será
$Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$, en nuestro caso $Z=\frac{\overline{X}-6}{4/\sqrt{64}}$
4. Calculamos el valor de ese estadístico sustituyendo la media por la obtenida en la prueba:

$Z_0= \frac{5-6}{4/\sqrt{64}} = -2$

5. Comprobamos si el valor anterior pertenecece a la zona de aceptación. Como $-2 \in (-\infty, 1'645)$ aceptams la hipótesis nulaes decir que la media sea igual o inferior 6 años.

Ejemplo 3.

Al lanzar 5000 veces una moneda al aire salieron 3000 caras. ¿Se puede aceptar, con un nivel de significación del 0.04, que la moneda no está trucada?

1. En este caso se trata de hacer un test de contraste de hipótesis sobre la proporción. Si queremos aceptar que la moneda no está trucada la hipótesis nula será H_0: p=0'5 y la alternativa H_1: p≠0'5, es por lo tanto un contraste bilateral.
2. Calculamos los valores $-z_{\alpha/2}$ y $z_{\alpha/2}$ para un nivel de significación de 0'04:
La región de aceptación será $P(-\mathbf{z}_{\alpha/2}<Z<\mathbf{z}_{\alpha/2})=1-\alpha$; $\alpha=0'04 \Rightarrow \alpha/2 = 0'02$ y $1-\alpha=0'96$; eso significa que $P(Z<z_{\alpha/2}) =1-\frac{\alpha}{2} = 1-\frac{0'04}{2} = 1-0'02 = 0'98$
Buscando en las tablas hallamos $z_{\alpha/2}$=2'05 . En consecuencia la región de aceptación será (-2'05, 2'05) la de rechazo $(-\infty, -2'05) \cup (2'05, \infty)$

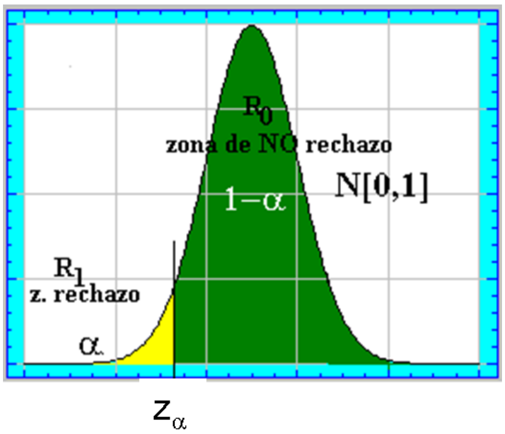
3. La distribución de proporciones sigue una variable normal $N(p, \sqrt{\frac{p.q}{n}})$ por tanto el estadístico del contraste será $Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}=\frac{\hat{p}-0'5}{\sqrt{\frac{0'5.0'5.}{5000}}}=\frac{\hat{p}-0'5}{0'0071}$
4. Calculamos el valor de ese estadístico sustituyendo la proporción por la obtenida en la prueba $z_0 =\frac{\frac{3000}{5000}-0'5}{0'0071}=\frac{0'6-0'5}{0'0071} = 14'08$

5. Como 14’08$\notin$(-2’05, 2’05) sino que pertenece a la región de rechazo, se rechaza la hipótesis nula. Es decir tenemos que aceptar, con un nivel de confianza del 96%, que la moneda está trucada.

Ejemplo 4

Un experto sostiene que si se celebran elecciones generales en este momento,
tan solo acudiría a votar el 48% de la población. No obstante, en un sondeo electoral realizado recientemente, entre 1500 personas, 800 de ellas tienen intención de votar.
¿Supone esto, con un nivel de confianza del 99%, que el experto se equivoca y la intención de voto es mayor?

1.Se trata de nuevo de un contraste aceptar o rechazar un valor de la proporción poblacional, como queremos contrastar si el experto se equivoca, la hipótesis nula será $p \geq 0'48$.
Por tanto: H_o: $p\geq 0'48$; H_1: $p<0'48$ contraste unilateral.
2. Calculamos el punto crítico $P(Z>z_\alpha)=1-\alpha=0'99 \Rightarrow P(Z<z_\alpha)=0'1$.



Dado que esa probabilidad es menor que 0’5 el valor de z_α es negativo y no viene en las tablas.
Por la simetría de la distribución normal $P(Z<z_\alpha)$ $=P(Z>-z_\alpha)=0'1$
Eso es equivalente a escribir
$P(Z<-z_\alpha)=0'9$. Buscando en las tablas encontramos que $-z_\alpha=2'33$, es decir $z_\alpha=-2'33$
En consecuencia, el intervalo de aceptación es:
$(-2'33, \infty)$

3. Elegimos el estadístico de contraste que en nuestro caso sería: por tanto el estadístico del contraste será $Z=\dfrac{\hat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}=\dfrac{\hat{p}-0'48}{\sqrt{\frac{0'48.0'52.}{1500}}}=\dfrac{\hat{p}-0'48}{0'0129}$ por seguir la distribución de proporciones una variable normal $N(p, \sqrt{\frac{p.q}{n}}$)
4. Calculamos el valor de ese estadístico sustituyendo la proporción por la obtenida en la prueba $z_0=\dfrac{\frac{800}{1500}-0'48}{0'0129}=4'1318$
5. Como $4'1318 \in (-2'33, \infty)$ aceptamos la hipótesis nula. Es decir tenemos que aceptar, con un nivel de confianza del 99%, que el experto se equivoca.

Ejemplo 5:

En una ciudad se ha observado que la venta de discos se ha reducido en el último año. El año pasado analizaron 25 tiendas y se registró una media de ventas de 1200 discos al mes, con una desviación típica de 75, mientras que este año se han estudiado 20 tiendas y la media ha sido de 1150 discos al mes, con una desviación típica de 60. Plantea un contraste de hipótesis con un nivel de significación del 10%, para decidir si la venta de discos se mntiene o, por el contrario, ha disminuido.

1. Se trata ahora de plantear un contraste de hipótesis para la diferencia de las medias. La hipótesis nula será que la venta de discos se ha mantenido. Es decir, H_0: $\mu_1 - \mu_2 = 0$; H_1: $\mu_1 - \mu_2 \neq 0$. Contraste bilateral
2. Calculamos los valores $-z_{\alpha/2}$ y $z_{\alpha/2}$ para un nivel de significación de 0'1:
La región de aceptación será $P(-z_{\alpha/2}<Z<z_{\alpha/2})=1-\alpha$; $\alpha=0'1 \Rightarrow \alpha/2 = 0'05$ y $1-\alpha=0'9$; eso significa que $P(Z<z_{\alpha/2}) = 1-\frac{\alpha}{2} = 1-\frac{0'1}{2} = 1-0'05 = 0'95$
Buscando en las tablas hallamos $z_{\alpha/2}=1'645$. En consecuencia la región de aceptación será (-1'645, 1'645)
3. El estadístico del contraste será: $Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2))}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} = \frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} = \frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{75^2}{25}+\frac{60^2}{20}}} =$
$=\frac{(\overline{X}_1-\overline{X}_2)}{20'1246}$
4. Calculamos el valor del estadístico sustituyendo los valores de la prueba:
$Z_0=\frac{1200-1150}{20'1246} = 2'48$
5. Dado que $2'48 \notin (-1'645, 1'645)$ se rechaza la hipótesis nula y hemos de aceptar la alternativa. Es decir, con un nivel de confianza del 90% decidimos que la diferencia de medias es significativa por lo que podemos deducir que la venta de discos ha disminuido en el último año.

Ejemplo 6:
Se quieren probar dos tipos de alimentos para los 50 pingüinos de un zoológico cuyo peso se distribuye normalmente. Se separan en dos grupos, uno formado por 30 pingüinos y otro por 20. Son pesados después de un mes y se obtiene, para el primer grupo un peso medio de 12 kg y desviación típica de 0'5, y, para el segundo, un peso medio de 10 kg y desviación típica de 0'8. ¿Se puede afirmar, con el nivel de confianza del 96%, que están mejor alimentados los del primer grupo que los del segundo?

1. Se trata de un contraste de hipótesis para la diferencia de medias. La hipótesis nula será que $\mu_1 \geq \mu_2$ o lo que es equivalente $H_0 : \mu_1 - \mu_2 \geq 0$; H_1: $\mu_1 - \mu_2 < 0$ Se trata de un contraste unilateral
2. Calculamos el valor z_α para el valor de confianza $1-\alpha=0'96$: $P(z>z_\alpha)=0'96=P(z<-z_\alpha)$ buscando en las tablas $-z_\alpha= 1'755 \Rightarrow z_\alpha=-1'755$. Región de aceptación $(-1'755,\infty)$
3. El estadístico del contraste será: $Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2))}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} = \frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} =$
$=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{0'5^2}{30}+\frac{0'8^2}{20}}}=\frac{(\overline{X}_1-\overline{X}_2)}{0'2008}$ (Los datos son: $n_1 = 30$; $\overline{x_1} = 12$ kg; $\sigma_1 = 0{,}5$; $n_2 = 20$; $\overline{x_1} = 10$ kg; $\sigma_1 = 0{,}8$)
4. Calculamos el valor del estadístico sustituyendo los valores de la prueba: $z_0=\frac{(12-10)}{0'2008}=9'96$
5. Como $9'96 \in (-1'755,\infty)$, se acepta la hipótesis nula y podemos concluir, con un nivel de confianza del 96%, que los del primer grupo están mejor alimentados.

(soluciones al final del tema)

19. La empresa de transportes urgentes El Rápido afirma en su publicidad que al menos el 70% de sus envíos llega al día siguiente a su destino. Para contrastar la calidad de este servicio, la asociación de consumidores selecciona aleatoriamente 100 envíos y observa que 39 no llegaron al día siguiente a su destino.
a) Con una significación del 1%, ¿se puede aceptar la afirmación de la empresa?
b) ¿Se concluiría lo mismo con un nivel de significación del 8%?
20. Un informe de la Asociación de Compañías Aéreas indica que el precio medio del billete de avión entre Canarias y la península Ibérica es, como máximo, de 120 €€ con una desviación típica de 40. €
Se toma una muestra de 100 viajeros Canarias-península Ibérica y se obtiene que la media de los precios de sus billetes es de 128 €€
a) ¿Se puede aceptar, con un nivel de significación igual a 0,1, la afirmación de partida?
b) ¿Se concluiría lo mismo si el nivel de significación fuera del 1%?
21. Se sabe que la renta anual de los individuos de una ciudad sigue una distribución normal de media desconocida y desviación típica 2000 € Se ha observado la renta anual de 35 individuos de esa localidad escogidos al azar y se ha obtenido un valor medio de 20 000 € Contrasta, a un nivel de significación del 5%, si la media de la distribución es de 18 000 €
22. Se sabe que la duración de una determinada enfermedad sigue la ley normal. Para la curación de dicha enfermedad se aplica un determinado antibiótico. Se desea comparar la duración de la enfermedad según que al enfermo se le haya aplicado o no en otra ocasión dicho antibiótico.
Observamos a 36 enfermos a los que no se había aplicado anteriormente el antibiótico y la duración media de la enfermedad ha sido de 12 días, y a 35 enfermos a los que sí se había aplicado y que han permanecido enfermos 15 días. La estimación común de la varianza es 16. ¿Qué podemos afirmar acerca de la duración de la enfermedad para un nivel de significación $a = 0{,}01$?
23. Un 43% de la población adulta de cierta ciudad sabía realizar el cambio entre euros y pesetas correctamente. Mediante una campaña informativa se ha pretendido elevar ese porcentaje y parece que se han cumplido sus objetivos a la vista del resultado de una encuesta a 110 personas: de ellas, 55 sabían realizar bien tales operaciones. Sin embargo, hay quien duda de la efectividad de la campaña.
a) Plantea un test para contrastar que la campaña no ha surtido efecto frente a que sí lo ha hecho. Si se concluye que el porcentaje se mantuvo y realmente subió, ¿cómo se llama el error cometido?
b) ¿A qué conclusión se llega en el test planteado en el apartado anterior a un nivel de significación del 1%?
24. Un hospital está probando dos tipos de medicamentos, A y B. Se toman dos grupos de pacientes de 40 y 30 individuos para probar los tipos A y B, respectivamente. El número medio de efectos secundarios en el primer grupo fue de 3 con una desviación típica de 1,5, y para los del segundo fue de 2 con desviación de 2.
¿Se puede afirmar con el nivel de confianza del 90% que el primer medicamento provoca menos efectos que el segundo?

EJERCICICIOS PAAU (solución en el propio ejercicio, la mayoría no resueltos)

2010

1. El peso en gr. de los pollos que llegan a un matadero sigue una distribución normal con una desviación típica de 320 gr.
A) Se estableció un intervalo de (2990, 3130) como intervalo de confianza para la media, a partir de una muestra de 64 pollos ¿cual es el valor de la media muestral $\bar{x}$? ¿con qué nivel de confianza se construyó el intervalo?
B) ¿cuántos pollos deberíamos pesar para que el nivel de confianza del intervalo sea del 97%?
Sol: $\bar{x}$=3060gr. Nivel de confianza 0’9198; deberíamos pesar 99 pollos

2. Las puntuaciones de un text de aptitud hecho a los alumnos de un centro de secundaria siguen una distribución N(1000, 600).
a) Calcula la probabilidad de que la puntuación media, para una muestra de 64 alumnos, esté comprendida entre 964 y 1036 puntos.
B) ¿Cuántos alumnos deberíamos seleccionar como mínimo para garantizar con un 99’5% de confianza una estimación de la puntuación media de todos los alumnos de centro con un error no superior a 150 puntos?
Sol (a P=0’3688; b 127 alumnos)

3. Un equipo de la guardia civil de tráfico hace controles de velocidad en una travesía de una determinada población. Se sabe que la variable “velocidad en travesía” en km/h sigue una distribución normal N(μ, σ)
a) Tras controlar el paso por la travesía de 100 vehículos nos dicen que “la velocidad media en travesía toma valores entre 56’08 km/h y 63’92 km/h con el 95% de confianza”. Con esta información calcula σ y el valor medio de la muestra.
B) Si tomamos como μ= 60 km/h y como σ= 20 km/h calcula el porcentaje de muestras de 64 vehículos cuya velocidad media supera los 65 km/h.
(sol: a) $\bar{x} = 60$; σ=20; b) P=0’0228 es decir 2’28%)
4.a)Si los salarios anuales de los trabajadores de una cierta empresa se distribuyen según una N(μ, σ=1200) Calcula el intervalo del 95% de confianza para el salario medio anual de los trabajadores de la empresa, si para eso se seleccionan al azar 64 trabajadores y se obtiene que su salario medio anual es de 26000 euros.
B) ¿qué tamaño de muestra habría que tomar para garantizar con un 97% de confianza, una estimación media del salario anual de los trabajadores de esa empresa con un error no superior a 200 euros?
(sol: a) Se estima que el salario anual de los trabajadores está entre 25706 euros y 26294 euros con un 95% de confianza; b)170 trabajadores o más)

2011

5. Debido a la futura fusión de dos entidades de ahorro, un estudio preliminar estima que como máximo un 5% de los clientes causará baja en la nueva entidad resultante. Un analista de mercados sospecha que la proporción de bajas será mayor y para contrastarlo realiza una encuesta a 400 clientes, elegidos al azar, sobre su intención de seguir operando con la nueva entidad después de la fusión. De ellos 370 contestan que seguir´n operando con la nueva entidad.
A) Formula un test para contrastar la hipótesis de que la proporción es la que se formula en el estudio preliminar frente a la del analista. ¿A qué conclusión se llega con un nivel de significación del 5%?
B) Explica, en el contesto del problema, lo que significan los errores de tipo I y de tipo II.
(sol: H_0: $p_0 \leq 0’05$; H_1: $p_0 > 0’05$; región de aceptación $(-\infty, 1’645)$; Estadístico de la prueba

$Z=\frac{\widehat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$; $Z_0=\frac{0'075-0'05}{\sqrt{\frac{0'05.0'95}{400}}}=2'29$. Como no pertenece a la región de aceptación se rechaza la hipótesis nula. Es decir, con los datos de la prueba y con un riesgo del 5% de equivocarnos se concluye que es cierta la hipótesis del analista.
B) Error de tipo I: error que se comete al rechazar H_0 siendo cierta. En nuestro ejemplo afirmaríamos que la proporción de clientes que causará baja es la que cree el analista, cuando realmente no es cierto.
Error de tipo II: error que se comete al aceptar H_0 siendo falsa. En nuestro ejemplo afirmaríamos qe el analista no tiene razón cuando realente si la tiene.)

6. Se conoce que la renta declarada por persona para todos los ciuddanos de un país sigue una distribución normal con media 10840 euros y desviación típica 2700 euros. Con el objeto de analizar la renta de los contribuyentes domiciliados en una cierta administración de hacienda, se tomó una muestra de 400 declaraciones obteniéndose una renta media de 10500 euros por persona. Si se supone que se mantiene la desviación típica.
A) formula un text para contrastar la hipótesis de que la renta media de las declaraciones presentadas en la administración es la misma que la global para todo el país, frente a que es menor tal y como parece indicar la muestra y explica a qué conclusión se llega con un nivel de significación del 1%
B) Calcula un intervalo del 98% de confianza para la renta media de los contribuyentes de la citada administración (Sol: a) $H_0 : \mu_0 \geq 10840$; H_1: $\mu_0 < 10840$; estadístico de la prueba Z= $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; $z_0=\frac{10500-10840}{2700/\sqrt{400}}$=-2'52. Región de aceptación (-2'33, ∞) como -2'52 no pertenece a la región de aceptción se rechaza la hipótesis nula. Es decir, con los datos de la prueba y con un riesgo del 1% de equivocarnos, concluimos que la renta media de las declaraciones presentadas en esta administración es inferior que la global para todo el país.
B) $P(\overline{X}-z_{\alpha/2}.\sigma/\sqrt{n} < \mu < \overline{X}+z_{\alpha/2}.\sigma/\sqrt{n}) = 1-\alpha$; $z_{\alpha/2}$ =2'33; I_c(10185'45,10814'55). Es decir, en base a la muestra dada, se estima con un 98% de confianza que la renta media por persona en la citada administración está entre 10185'45 y 10814'55 euros.)

7. La información que ofrece el editor de una escala de madurez para estudiantes de secundaria señala que las puntuaciones en la escala siguen una distribución normal con media 5 y desviación típica 2. La escala tiene ya 10 años lo que hace sospechar a un educador de enseñanza secundaria que el promedio de la escala podría haber aumentado en el momento actual. Para comprobarlo selecciona una muestra aleatoria de 49 estudiantes y tras pasarles la prueba obtiene una media de 5'6. Suponiendo que se mantiene la desviación típica.
A) Formula un test para contrastar que la puntuación media no aumentó frente a que si lo hizo tal y como sospecha el educador y explica la conclusión a la que se llega con una significación del 5%
B) Utilizando la muestra dada, calcula el intervalo de la puntuación media de los estudiantes de secundaria en el momento actual con una confianza del 95%
(sol: a) $H_0 \mu_0 \leq 5$; H_1: $\mu_0 > 5$; estadístico de la prueba Z=$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; $z_0=\frac{5'6-5}{2/\sqrt{49}}$=2'1. Región de aceptación (-∞, $1'645$) como 2'1 no pertenece a la región de aceptción se rechaza la hipótesis nula. Es decir, con los datos de esta muestra y con un riesgo de un 5% de equivocarnos concluimos que la puntución media actual aumentó como sospechaba el educador.

B)$P(\overline{X} - z_{\alpha/2}.\sigma/\sqrt{n} < \mu < \overline{X} + z_{\alpha/2}.\sigma/\sqrt{n}) = 1-\alpha$; $z_{\alpha/2}$=1'96; I_c(5'04,6'16). Es decir, en base a la muestra dada, se estima con un 95% de confianza que la puntuación media en el momento actual está entre 5'04 y 6'16)

8. Una compañía telefónica A afirma que la proporción de hogares que contratan su servicio de ADSL es por lo menos del 26%. Sin embargo, otra compañía de la competencia B sostiene que actualmemte la proporción de usuarios de la compañía A es menor del 26%. Para comprobarlo hace una encuesta a 400 clientes que tienen en sus hogares servicio de ADSL y 85 manifiestan que lo tienen contratado con la compañía A
A) formula un test para contrastar que la proporción es la que afirma la compañía A frente a la alternativa sostenida por la compañía B ¿A qué conclusión se llega con un nivel de significación del 5%?
B) Utilizando la información obtenida en la encuesta calcula un intervalo de confianza del 95% para la proporción de hogares que contratan anualmente los servicios de ADSL de la compañía A.
(Sol: H_0: $p_0 \geq 0'26$; $H_1 : p_0 < 0'26$; región de aceptación (-1'645,∞)); Estadístico de la prueba

Z=$\dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$; Z_0=$\dfrac{0'2125 - 0'26}{\sqrt{\frac{0'26.0'74}{400}}} = -2'16$. Como -2'16 no pertenece al intervalo de aceptación rechazamos la hipótesis nula. Es decir, con un riesgo de equivocarnos del 5% concluimos que la proporción de usuarios que contrata ADSL con la compañía A es menor del 26% como afirma la compañía B.

B) P($\hat{P} - z_{\alpha/2}\sqrt{\dfrac{\hat{p}.\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}.\sqrt{\dfrac{\hat{p}.\hat{q}}{n}}) = 1-\alpha$; $z_{\alpha/2}$=1'96; I_c=(0'1725, 0'2525). Es decir, en base a la muestra dada se estima con un 95% de confianza que la proporción de hogares que contrata ADSL con la compañía A está comprendido entre un 17'25% y un 25'25%.

2012

Se supone que el número de telespectadores (en millones) de un programa semanal de televisión se aproxima a una distribución normal con desviación típica 0'5 (millones). La dirección del programa afirma que la media semanal de telespectadores que ven el programa es de 7 millones. Para contrastar tal afirmación se observa una muestra de 10 semanas obteniéndose una media semanal de 6'54 millones de telespectadores.
A) Utilizando la muestra calcula un intervalo de confianza al 95% para la media semanal de telespectadores de ese programa.
B) Formula un test para contrastar que la media semanal de telespectadores es la que dice la dirección del programa frente a la alternativa de que es menor. ¿Cuál es la conclusión a la que se llega con un 5% de significación?
(sol: Sea X=número de telespectadores semanal del programa" X sigue una distribución N(μ, 0'5). Expresión del intervalo de confianza $P(\overline{X} - z_{\alpha/2}.\sigma/\sqrt{n} < \mu < \overline{X} + z_{\alpha/2}.\sigma/\sqrt{n}) = 1-\alpha$; $z_{\alpha/2}$=1'96; I_c(6'23,6'85). Es decir, En base a la muestra dada, se estima con un 95% de confianza, que la media semanal de telespectadores de ese programa está entre 6.230.000 y 6.850.000 espectadores.

B) H_0: $\mu_0 \geq 7$; $H_1 : \mu$<7; estadístico de la prueba Z=$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$; z_0=$\dfrac{6'45 - 7}{0'5/\sqrt{10}}$=-2'9. Región de aceptación ($-$$-1'645, \infty$) como -2'9 no pertenece a la región de aceptción se rechaza la hipótesis nula. Es decir, con los datos de esta muestra y con un riesgo de un 5% de equivocarnos concluimos que la media de telespectadores semanales de este programa es inferior a los 7 millones que dice la dirección del programa)

10. Se realiza una encuesta para conocer la intención de voto al partido político MLM. De los 2000 entrevistados 600 dicen que lo votarán.
A) Calcula un intervalo al 95% de confianza para la proporción de futuros votantes de ese partido.
B) Una afirmación publicada en la prensa dice que la proporción de votantes para ese partido es de al menos el 33%. Formula un test para contrastar dicha información frente a que la proporción de futuros votantes es inferior como parece pronosticar la encuesta ¿ A qué conclusión se llega con un nivel de significación del 1%?
So: Sea p=”proporción de futuros votantes al partido MLM (parámetro a estimar)” y $\widehat{P}$ la proporción muestral de votantes en muestras de tamaño 2000; $\widehat{p} = 600/2000 = 0'3$(estimación puntual de p. Expresión del intervalo de confianza:
P($\widehat{P} - z_{\alpha/2}\sqrt{\dfrac{\widehat{p}.\widehat{q}}{n}} < p < \widehat{P} + z_{\alpha/2}.\sqrt{\dfrac{\widehat{p}.\widehat{q}}{n}}$) $= 1-\alpha$; $z_{\alpha/2}$=1’96; I_c=(0’28, 0’32). En base a la muestra dada se estima que la proporción de futuros votantes al partido MLM está comprendida entre el 28% y el 32%, con un nivel de confianza del 95%
b)H_0: p$\geq 0'33$; $H_1 : p < 0'33$; región de aceptación (-2’33,∞); Estadístico de la prueba Z=$\dfrac{\widehat{\text{p}}-\text{p}}{\sqrt{\frac{\text{p.q}}{n}}}$; Z_0=$\dfrac{0'3-0'33}{\sqrt{\frac{0'33.0'67}{2000}}} = -2'85$. Como -2’85 no pertenece al intervalo de aceptación rechazamos la hipótesis nula. Con un riesgo del 1% de equivocarnos concluimos que la proporción de futuros votantes al partido es menor del 33% publicado en la prensa.)

11. El tiempo de espera para la realización de cierta prueba en un hospital sigue una distribución normal con desviación típica de 5 días. La gerencia afirma que “el tiempo de espera para la realización de dicha prueba es como máximo de 20 días. Para contrastar esta afirmación se tomó una muestra aleatoria de 100 pacientes que necesitaban realizar la prueba, resultando que el tiempo medio de espera fue de 21 días.
A)Formula un test para contrastar la hipótesis de la gerencia frente a que el tiempo medio es superior ¿A qué conclusión se llega con un nivel de significación del 5%? ¿Se llegaría a la misma conclusión con un nivel de significación del 1%?
B) Explica en el contesto del problema en qué consisten los errores de tipo I y de tipo II
Sol. Sea X=”tiempo de espera para la realización de la prueba en el hospital” enmuestras de 100 pacientes. H_0: $\mu_0 \leq 20$; $H_1 : \mu$>20; estadístico de la prueba Z=$\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; z_0=$\dfrac{21-20}{5/\sqrt{100}}$=2.
Región de aceptación (-$\infty, 1'645$) como 2 no pertenece a ese intervalo se rechaza la hipótesis nula. Con los datos de la muestra y con un riesgo de un 5% de equivocarnos concluimos que el tiempo medio de espera para la realización de la prueba es superior a los 20 días que afirma la gerencia.
Si el nivel de significación fuese del 1% La región de aceptación sería: (-$\infty, 2'33$) por lo que no se llegaría a la misma conclusión ya que en este caso se acepta la hipótesis de la gerencia al pertenecer 2 al intervalo de aceptación. Es decir, con un riesgo del 1% de equivocarnos no podríamos rechazar la hipótesis de que el tiempo medio de espera es superior a 20 días.
Error de tipo I: error que se comete al rechazar la hipótesis nula siendo esta verdadera. En nuestro problema consistiría en decidir que el tiempo máximo de espera es superior a los 20 días cuando no es cierto.

Error de tipo II: Consiste en sceptar la hipótesis nul cuando es falsa. En nuestro caso el que cometeríamos al aceptar que el tiempo máximo de espera es de 20 días cuando en realidad es mayor)

**12.Se quiere estimar el porcentaje de españoles que, teniendo derecho al voto, no votarán en las próximas elecciones al Parlamento Europeo.¿Cuál debe ser el tamaño de la muestra para garantizar un margen de error no superior al 2’5% con un nivel de confianza del 95%?
B) Se selecciona una muestra de 1540 españoles con derecho a voto y, de ellos, 693 aseguran que no votarán. Calcula un intervalo al 95% de confianza para el porcentaje de españoles con derecho a voto que no votarán en las citadas elecciones. ¿Qué error máximo se está cometiendo en esta estimación?
Sol: Sea p=”proporción de españoles con derecho a voto que no votarán en las elecciones (parámetro a estimar)”; $\widehat{P}$= proporción muestral. El estadístico a utilizar es Z=$\frac{\widehat{\mathrm{p}}-\mathrm{p}}{\sqrt{\frac{\mathrm{p.q}}{n}}}$ que se aproxima a una distribución N(0,1). **Como no conocemos una estimación puntual de p tomaremos el caso más desfavorable para “p”: p=1/2 (ya que la función f(p)=p(1-p) tiene un máximo cuando p=1/2. Formulamos el margen de error no superior al 2’5%:
E=$z_{\alpha/2}.\sqrt{\frac{p(1-p)}{n}}$ <0’025. Calculamos el valor $z_{\alpha/2}$ para un nivel de confianza del 95% resultando ser $z_{\alpha/2}$=1’96 y sustituimos en la fórmula anterior 1’96.$\sqrt{\frac{0'5.0'5}{n}} < 0'25$ despejando n>1536’64 por lo que debemos tomar una muestra de 1537 españoles o más.
b)Expresión del intervalo de confianza: $P(\widehat{P}-z_{\alpha/2}\sqrt{\frac{\widehat{p}.\widehat{q}}{n}} < p < \widehat{P}+z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}) = 1-\alpha$
; $z_{\alpha/2}$=1’96; $\widehat{p}$.=693/1540=0’45, sustituyendo obtenemos I_c=(0’4252, 0’4748). Se estima con un nivel de confianza del 95% que el porcentaje de españoles con derecho a voto que no votarán en esas elecciones está comprendido entre el 42’52% y el 47’48%. El error máximo cometido en esta estimación es del 2’48% (se halla el radio del intervalo: (0’4748-0’4252)/2= =0’0248 y se escribe en términos de porcentaje)

13. En el proceso industrial de envasado de un producto el peso de los envases se aproxima a una normal de media 500 gr. y desviación típica 4 gr.,los direcivos de la empresa sospechan que la máquina de envasado está estropeada y deciden cambiarla si el peso medio de los envases es superior a 500 gr. Para eso examinan una muestra aleatoria de 30 envases y obtienen una media de 501’5 gr.
A) Formula un test para contrastar la hipótesis de que no es necesario cambiar la máquina frente a que sí lo es, tal como sospechan los directivos. ¿que conclusión se extrae con un 5% de significación?
B)Explica el tipo de error que cometerían si deciden, erróneamente, no cambiar la máquina.
Sol: Sea X=”peso, en gr., de un envase” X es N(500,4) ; H_0: $\mu_0 \leq 500$; H_1 : μ>500; estadístico de la prueba Z=$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; z_0=$\frac{501'5-500}{4/\sqrt{30}}$=2’05. Región de aceptación (-∞, $1'645$) como 2’05 no pertenece a ese intervalo se rechaza la hipótesis nula. Con los datos de la muestra y con un riesgo de equivocarnos del 5%, concluimos que el peso medio de los envases supera los 500 gr., con lo que los directivos decidirían cambiar la máquina.
B)Error de tipo II: aceptar H_0 siendo falsa. Los directivos decidirían no cambiar la máquina siendo el peso medio de los envases superior a los 500 gr.

**14.Un estudio sobre el hábito de fumar entre los adultos de una ciudad informa que el intervalo de proporción de fumadores se encuentra entre el 30% y el 40%
A) Determina la proporción muestral de fumadores observados según ese estudio
B) El estudio añade que los datos se obtuvieron de una encuesta aleatoria realizada a 364 habitantes adultos de esa ciudad ¿Cuál es entonces el nivel de confianza de dicho intervalo de estimación de la proporción de fumadores?
Sol: a)Sea p=" proporción de fumadores entre los habitantes adultos de la ciudad (parámetro poblacional)" $\widehat{P}$=" proporción muestral de fumadores de entre los habitantes adultos" (estimación puntual de p). El intervalo que nos dan: (0'3,0'4) tendrá la siguiente expresión:

$(\widehat{P}-z_{\alpha/2}\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}\ ,\ \widehat{P}+z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}})$ y por lo tanto $\begin{cases}\widehat{P}-z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}=0'3\\ \widehat{P}+z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}=0'4\end{cases}$ sumando ambas expresiones obtenemos que 2$\widehat{P}$=0'7; por tanto $\widehat{P}$=0'7/2=0'35. Es decir, la proporción muestral en el estudio es de 0'35 (observar que corresonde al punto intermedio del intervalo)
B) El radio del intervalo es (0'4-0'3)/2=0'1/2=0'05. Dicho radio es igual a .$z_{\alpha/2}\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}$ =
$=z_{\alpha/2}\sqrt{\frac{.0'35.0'65}{364}}=0'05\Rightarrow z_{\alpha/2}=\frac{0'05.\sqrt{364}}{\sqrt{0'35.0'65}}=2;$ P(-2<Z<2)=1-α; Buscando en las tablas P(-2<Z<2)=2P(Z<2)-1=2.0'9772-1=0'9544 por tanto el nivel de confianza es del 95'44%.
Con un 95'44% estimamos que la proporción de fumadores entre los adultos de esa ciudad se encuentra entre el 30% y el 40% en base a los datos obtenidos por una encuesta hecha a 364 adultos de esa ciudad.

15.En cierto país la renta media familiar sigue una distribución normal de media16260 euros y desviación típica 6320 euros. Un estudio realizado con 200 familias elegidas al azar en una comarca proporcionó una renta media de 15308 euros. Suponiendo que se mantiene la desviación típica
A)Calcula un intervalo de confianza al 95% para la renta media de las familias de esa comarca.
B) Formula un test para contrastar la hipótesis de que la renta media de las familias en esa comarca es la misma que la de toda la población, frente a la hipótesis de que es menor que la del país ¿cuál es la conclusión a la que se llega con un nivel de significación del 5%? ¿se llegaría a la misma conclusión para una significación del 1%?
Sol: a) X=" renta anual de las familias en una comarca" N(μ, 6320)
$\overline{X}$ estadístico de la media muestral. $\overline{x}$=15308 (valor particular para la muestra dada)
Z=$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ estadístico de la prueba, N(0,1)
Expresión del intervalo de confianza: $P(\overline{\mathrm{X}}-z_{\alpha/2}.\sigma/\sqrt{n}<\mu<\overline{\mathrm{X}}+z_{\alpha/2}.\sigma/\sqrt{n})=1-\alpha$=0'95
$z_{\alpha/2}$=1'96; I_c=(14432'1, 16183'9). Con un 95% de confianza se estima que la renta media en esa comarca está comprendida entre 14432'1€ y 16183'9€
b)H_0: $\mu_0\geq 16260$; H_1 : μ_0<16260; estadístico de la prueba Z=$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; z_0=$\frac{15308-16260}{6320/\sqrt{200}}$=
=-2'13. Región de aceptación (-1'645,∞) como -2'13 no pertenece a ese intervalo se rechaza la hipótesis nula. Con los datos de la muestra y con un riesgo de un 5% de equivocarnos concluimos que la renta media en esa comarca es inferior a la renta media de la población total del país..

Para un nivel de significación del 1% la región de aceptación será (-2’33,∞) En este caso -2’13 si pertenecería a ese intervalo y por tanto deberíamos de aceptar la hipótesis nula; esto es, aceptar que la renta media de esa comarca es similar a la de la población total del país.

16. En un estudio sobre higiene dental, el porcentaje de niños que presentaron indicios de carie utilizando un dentrífrico tradicional fue de por lo menos un 10%. En un grupo de 500 niños elegidos aleatoriamente que utilizarn un nuevo dentífrico sufrieron carie 35
A) Formula un test para contrastar la hipótesis de que la proporción de niños con indicios de carie usando el nuevo dentífrico es el mismo que usando el tradicional, frente a la hipótesis de que se reduce. ¿A qué conclusión se llega con un nivel de significación del 1%?
B)Calcula un intervalo de confianza del 95% para la proporción de niños con indicio de carie que utilizan el mismo dentrífico.
Sol: Sea p=”proporción de niños con carie usando el nuevo dentrífico (parámetro a estimar)” y $\widehat{P}$ la proporción muestral en muestras de tamaño 500; $\widehat{p} = 35/500 = 0'07$(estimación puntual de p. a)H_0: p≥ $0'10$; $H_1 : p < 0'10$; región de aceptación (-1’645,∞); Estadístico de la prueba

$Z=\frac{\widehat{p}-p}{\sqrt{\frac{p.q}{n}}}$; $Z_0=\frac{0'07-0'1}{\sqrt{\frac{0'1.0'9}{500}}} = -2'236$. Como -2’236 no pertenece al intervalo de aceptación rechazamos la hipótesis nula. Con un riesgo del 5% de equivocarnos concluimos que en función de los resultados de la muestra, el nuevo dentrífico disminuye el riesgo de carie
B) Expresión del intervalo de confianza:
$P(\widehat{P}-z_{\alpha/2}\sqrt{\frac{\widehat{p}.\widehat{q}}{n}} < p < \widehat{P}+z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}) = 1-\alpha$; $z_{\alpha/2}$ =1’96; I_c=(0’048, 0’092). En base a la muestra dada se estima que la proporción deniños con carie usando el nuevo dentrífico está comprendida entre el 4’8% y el 9’2%, con un nivel de confianza del 95%

2014
17. Supongamos que el IMC (índice de masa corporal) en niñas de 13 años de una población sigue una distribución $N(\mu, 4)$
A) si el 6’68% de esas niñas tiene riesgo de sobrepeso, es decir su IMC es superior a 22’5, calcula el valor delIMC medio μ para las niñas de 13 años de esa población.
B) Si el IMC para niñas de 13 años de ea población sigue una distribución N(16’5,4) y se extrae una muestra aleatoria de 64 niñas calcula la probabilidad de que el IMC medio de la muestra esté por debajo de 15’3 (por debajo del peso adecuado)
Sol: a) P(X>22’5)=P(Z>$\frac{22'5-\mu}{4}$) = $0'668$.Buscando en las tablas se obtiene que $\frac{22'5-\mu}{4}$ =1’5 y despejando μ=16’5
B) La distribución de las medias muestrales sigue una N(16’5, $\frac{4}{\sqrt{64}}$) =N(16’5,0’5)
$P(\overline{X} < 15'3) = P(Z < \frac{15'3-16'5}{0'5}) = P(Z < -2'40) = 1 - P(Z < 2'40) = 1 - 0'9918 = 0'0082$

18.En un estudio reciente se afirma que hay un 5% de lesiones de rodilla entre futbolistas que juegan sobre cesped y calzan un nuevo modelo de botas de fútbol. De 250 futbolistad que juegan sobre el cesped y calzan botas de fútbol convencionales se dieron 20 de esas lesiones.
A) Formula un test para contrastar la hipótesis de que las lesiones de rodilla entre futbolistad que juegan con botas tradicionales no superan a las de tales lesiones jugando con el nuevo modelo frente a la de que si las supera.
B) ¿A qué conclusión se llega con un 5% de significación? ¿se llega a la misma conclusión con un nivel de significación del 1%?

Sol: **p:** “ proporción de lesiones de rodilla entre jugadores de fútbol que calzan botas convencionales”, **parámetro poblacional desconocido (y el que nos mandan contrastar)** $\widehat{P}$: proporción de lesiones de rodilla entre futbolistas que calzan botas convencionales, en muestras de 250 futbolistas (estimador puntual de " p "). $\widehat{p}$=20/250=0’08 estimación puntual de p.
a)H_0: $p \leq 0'05$; $H_1 : p > 0'05$
B) región de aceptación $(-\infty, 1'645)$ Estadístico de la prueba
$Z=\dfrac{\hat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$; $Z_0=\dfrac{0'08-0'05}{\sqrt{\frac{0'05..0'95}{250}}} = 2'176$. Como 2’176 no pertenece al intervalo de aceptación rechazamos la hipótesis nula. Con un riesgo del 5% concluimos que, en base a la muestra, la lesión de rodillas cn botas convencionales es mayor que con el nuevo calzado.
Para un nivel de significación del 1% la nueva región de aceptación es $(-\infty, 2'33)$. Como 2’176 pertenece a ese intervalo aceptamos la hipótesis nula. Es decir, con un riesgo del 1% no podemos concluir que el nuevo calzado mejore la proporción de lesiones de rodilla sobre los que calzaban botas convencionales.

**19. La proporción de mujeres de una población portadoras de hemofilia es desconocida.Para estimarla se elige una muestra de 500 mujeres de las que 80 resultan ser portadoras
A) Calcula un intervalo del 95% de confianza para la proporción de mujeres portadoras de hemofilia de esa población.
B)Suponiendo que aún no se tomó la muestra y queremos hacer la estimación cometiendo un error no superior a un 2% con un 95% de confianza ¿De qué tamño debería ser la muestra?
Sol: Sea p=”proporción de mujeres portadoras de hemofilia de esa población (parámetro a estimar)” y $\widehat{P}$ la proporción muestral en muestras de tamaño 500; $\widehat{p} = 80/500 = 0'16$ (estimación puntual de p. a) Expresión del intervalo de confianza:
$P(\widehat{P} - z_{\alpha/2}\sqrt{\frac{\widehat{p}.\widehat{q}}{n}} < p < \widehat{P} + z_{\alpha/2}.\sqrt{\frac{\widehat{p}.\widehat{q}}{n}}) = 1-\alpha$; $z_{\alpha/2}$=1’96; I_c=(0’128, 0’192). En base a la muestra dada se estima que la proporción mujeres portadoras de hemofilia está comprendida entre el 12’8% y el 19’2%, con un nivel de confianza del 95%
B) margen de error no superior al 2% $z_{\alpha/2}\sqrt{\frac{p.q}{n}}$ <0’02 Como aún no se tomó ninguna muestra no tenemos ningún valor puntual para estimar p, entonces tomamos el caso más desfavorable que es p=1/2 (ya que la función f(p)=p.(1-p) posee un máximo en p=1/2)
Entonces $z_{\alpha/2}\sqrt{\frac{p.q}{n}} < 0'02 \Rightarrow z_{\alpha/2}\sqrt{\frac{0'5.0'5}{n}} < 0'02$ estimamos $z_{\alpha/2}$ para un 95% de confianza y obtenemos : $z_{\alpha/2}$=1’96. Sustituimos en la fórmula y despejamos n, n>2401. Concluimos: para garantizar un error no superior al 2% con un nivel de confianza del 95% necesitamos tomar una muestra de 2401 mujeres de esa población.

20. En un estudio sociológico se afirma que el tiempo medio que los jóvenes están conectados a las redes sociales no supera las 60 horas mensuales. Se desea contrastar si actualmente sigue en vigor ese estudio y para ello se entrevistan a 400 jóvenes seleccionados al azar, obteniéndose que el tiempo medio es de 62 horas. Suponemos que el tiempo dedicado por los jóvenes a las redes sociales sigue una distribución normal de desviación típica 15 horas.
A)Formula un test para contrastar la hipótesis de que el tiempo medio mensual dedicado por los jóvenes a las redes sociales es el que afirma el estudio, frente a la alternativa de que aumentó. ¿A qué conclusión se llega con un intervalo de significación del 1%?

B)Usando la información recogida en la muestra, calcula el intervalo del 95% de confianza para el tiempo medio mensual dedicado actualmente por los jóvenes a conectarse a las redes sociales.
Sol:a)X=" tiempo dedicado por los jóvenes a las redes sociales" N(μ, 15)
$\overline{X}$ estadístico de la media muestral. $\bar{x}$=62 (valor particular para la muestra dada)
Z=$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ estadístico de la prueba, N(0,1)
H_0: $\mu_0 \leq 60$ H_1 : $\mu_0>60$; estadístico de la prueba Z=$\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$; z_0=$\dfrac{62-60}{15/\sqrt{400}}$= 2'67. Región de aceptación (-∞, 2'33) como 2'67 no pertenece a ese intervalo se rechaza la hipótesis nula. Con los datos de la muestra y con un riesgo de un 1% de equivocarnos concluimos que el tiempo dedicado actualmente a las redes sociales aumentó respecto al dado por el estudio
Para un nivel de significación del 1% la región de aceptación será (-2'33,∞) En este caso -2'13 si pertenecería a ese intervalo y por tanto deberíamos de aceptar la hipótesis nula; esto es, aceptar que la renta media de esa comarca es similar a la de la población total del país.

b) Expresión del intervalo de confianza: $P(\overline{\mathrm{X}}-z_{\alpha/2}.\sigma/\sqrt{n}\ <\mu<\overline{\mathrm{X}}+z_{\alpha/2}.\sigma/\sqrt{n}\ )=1-\alpha$=0'95
$z_{\alpha/2}$=1'96; I_c=(60'53, 63'47). Con un 95% de confianza se estima que el tiempo medio dedicado a las redes sociales por los jóvenes está comprendida entre 60'53 horas al mes y 63'47 horas al mes.

SOLUCIONES EJERCICIOS DEL TEMA

I Distribuciones muestrales:

1. Como hemos visto, la distribución de medias muestrales es N$\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$; en nuestro caso N(1'72,$\dfrac{0'09}{\sqrt{36}}$)= N(1'72, 0'015)
a) P($\bar{x}$<1'75)=P(Z<$\dfrac{1'75-1'72}{0'015}$)=P(Z<2)=0'9773
b) P(1'68<$\bar{x}$<1'73)=P($\dfrac{1'68-1'72}{0'015}<Z<\dfrac{1'73-1'72}{0'015}$)=P(-2'67<Z<0'67)=P(Z<0'67)-[1-P(2'67)=
=0'7486-[1-0'9962]=0'7448

2.La distribución muestral de proporciones sigue una N(p, $\sqrt{\dfrac{p.q}{n}}$); en nuestro caso N(0'03,$\sqrt{\dfrac{0'03.0'97}{500}}$)=N(0'03,0'0076)
A) P($\hat{P}>0'05)=P(Z>\dfrac{0'05-0'03}{0'076})=P(Z>2'63)=1-P(Z<2'63)=1-0'9957$=0'0043
B) P(haya menos de 10 piezas defectuosas)=P($\hat{P}<\frac{10}{500})=P(\hat{P}$<0'02)=P(Z<$\dfrac{0'02-0'03}{0'0076})=$
P(Z<-1'32)=1-P(Z<1'32)=1-0'9066=0'0934

3. La distribución de diferencia de medias muestrales es N($\mu_1-\mu_2.\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$); en nuestro caso:
N($16-14$, $\sqrt{\frac{36}{42}+\frac{49}{38}}$) = N(2, 1'46).

4. a)la distribución de medias muestrales es N$\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$; en nuestro caso N(142’32,$\frac{8'5}{\sqrt{25}}$)= N(142’32, 1’7). P($\overline{X}$>144’6)=P(Z>$\frac{144'6-142'32}{1'7}$)=P(Z>1’34)=1-P(Z<1’34)=1-0’9099= =0’0901; b)Ahora la distribución de medias muestrales es N(142’32, $\frac{8'5}{\sqrt{100}}$)= =N(142’32,0’85); P($\overline{X}$>144’6)=P(Z>$\frac{144'6-142'32}{0'85}$)=P(Z>2’68)=1-P(Z<2’68)=1-0’9963= =0’0037

5. la distribución de medias muestrales es N$\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$; en nuestro caso N(150,$\frac{20}{\sqrt{80}}$)= =N(150, 2’24); Si las 80 chocolatinas han de pesar más de 12400 gr, su media ha de ser superior a 12400/80=155 gr, P($\overline{X}$>155)=P(Z>$\frac{155-150}{2'24}$)= P(Z>2’23)=1-P(Z<2’23)=1-0’9871= =0’0129

6. $\overline{X}$ es N(100,$\frac{8}{\sqrt{25}}$)=N(100,1’6); P($\overline{X}$>101’5)=P(Z>$\frac{101'5-100}{1'6}$)=P(Z>0’94)=1-P(Z<0’94)= 1-0’8264=0’1736

7. Nos están diciendo que para la población total p= 48/120=0’4. Sabemos que la distribución muestral de proporciones sigue una N(p, $\sqrt{\frac{p.q}{n}}$); en nuestro caso N(0’4,$\sqrt{\frac{0'4.0'6}{40}}$) = =N(0’4,0’0775); P($\hat{P}>0'55)=P(Z>\frac{0'55-0'4}{0'0775})=P(Z>1'94)=1-P(Z<1'94=1-0'9738=0'0262$

8. P=1/15=0’066; la distribución muestral de proporciones sigue una N(p, $\sqrt{\frac{p.q}{n}}$); en nuestro caso N(0’066,$\sqrt{\frac{0'066.0'934}{100}}$) =N(0’071,0’025). Nos piden P(0’05<$\hat{P}$<0’1)= =P($\frac{0'05-0'066}{0'025}$<Z<$\frac{0'1-0'066}{0'025}$$)=P(-0'64<Z<1'36)=P(Z<1'36)-P(Z<-0'64)=$ =$P(Z<1'36)-[1-P(Z<0'64)]$ =0’9131-[1-0’7389]=0’652.

9. La distribución de diferencia de medias$\overline{X_1}-\overline{X_2}$ muestrales es N($\mu_1-\mu_2.\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$); en nuestro caso:
N($4000-3900,\ \sqrt{\frac{270^2}{50}+\frac{280^2}{50}}$) = N(100, 383’89). P($\overline{X_1}-\overline{X_2}$<50)=P(Z<$\frac{50-100}{383'89}$)=P(Z<-0’13)= =1-P(Z<0’13)=1-0’5557=0’4443

10. La distribución de diferencia de medias$\overline{X_1}-\overline{X_2}$ muestrales es N($\mu_1-\mu_2.\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$); en nuestro caso:
N($7'2-6'7,\ \sqrt{\frac{0'8^2}{34}+\frac{0'7^2}{40}}$) = N(0’5, 0’1762). P($\overline{X_1}-\overline{X_2}$>1)=P(Z>$\frac{1-0'5}{0'1762}$)=P(Z>2’84)= =1-P(Z<2’84)=1-0’9978=0’0022

II INTERVALOS DE CONFIANZA

11. a) .$\overline{x}$=8’1, σ=3; n=100; 1-α=0’97. Calculamos el valor crítico P(Z<$z_{\alpha/2}$)=1-$\frac{0'03}{2}$ = =1-0’015=0’985. Buscando en las tablas $z_{\alpha/2}$=2’17 El intervalo pedido es:
I_c= (8’1-2’17$\frac{3}{10}$, 8’1+2’17$\frac{3}{10}$)=(7’449,8’751)
b) En primer lugar hallamos el valor crítico para el nuevo nivel de confianza del 92%
P(Z<$z_{\alpha/2}$)=1-$\frac{0'08}{2}$ ==1-0’04=0’96. Buscando en las tablas $z_{\alpha/2}$=1’75

El error máximo es de un día, es decir E=$z_{\alpha/2}.\sigma/\sqrt{n}$ =1⇒ $1'75.\frac{3}{\sqrt{n}}=1\Rightarrow n=27'56$ Como ha de ser un número entero n=28 pacientes

12. a).$\bar{x}$=170, σ=20; n=100; 1-α=0'9. Calculamos el valor crítico P(Z<$z_{\alpha/2}$)=1-$\frac{0'1}{2}$ =
=1-0'05=0'95. Buscando en las tablas $z_{\alpha/2}$=1'645 El intervalo pedido es:
I_c= (170-1'645$\frac{20}{10}$, 170+1'645$\frac{20}{10}$)=(166'71,173'29)
b)En primer lugar hallamos el valor crítico para el nuevo nivel de confianza del 99%
P(Z<$z_{\alpha/2}$)=1-$\frac{0'01}{2}$ ==1-0'005=0'995. Buscando en las tablas $z_{\alpha/2}$=2'575
El error máximo es de 0'5mn, es decir E=$z_{\alpha/2}.\sigma/\sqrt{n}$ =0'5⇒ $2'575.\frac{20}{\sqrt{n}}=0'5\Rightarrow n=10609$

13. a) .$\bar{x}$=97, σ=10; n=36; 1-α=0'992. Calculamos el valor crítico P(Z<$z_{\alpha/2}$)=1-$\frac{0'008}{2}$ =
=1-0'004=0'996. Buscando en las tablas $z_{\alpha/2}$=2'65 El intervalo pedido es:
I_c= (97-2'65$\frac{10}{6}$, 97+2'65$\frac{10}{6}$)=(92'58,101'42)
B)El intervalo obtenido significa que en el 99'2% de las posibles muestras de este tamaño, la media de la duración de los juguetes estará en ese intervalo.

14. a)Tenemos una distribución normal de media μ y desviación típica σ=9, ya que nos dicen que la varianza es 81, sabemos que P(X>105)=0'1 ; tipificando la variable Z= $\frac{X-\mu}{9}$;
P(Z> $\frac{105-\mu}{9}$) $=0'1$ ⇒P(Z<$\frac{105-\mu}{9}$) $=0'9$. Buscando en las tablas $\frac{105-\mu}{9}=1'28$ y despejando μ=93'48
b) μ=95 Nos piden P(86<X<107)=P($\frac{86-95}{9}<Z<\frac{107-95}{9}$) ⇒P(-1<Z<1'33)=
=P(Z<1'33)-(1-P(Z<1))=0'9082-1+0'8413=0'7495
c)Tenemos una variable X que sigue una distribución N(95,9), sabemos que la distribución de las medias muestrales $\overline{X}$ sigue entonces una distribución normal N(95, $\frac{9}{\sqrt{n}}$) y como nos dicen que n=9 nuestra variable seguirá una N(95,3), nos piden
P(86<$\overline{X}$<107)=P($\frac{86-95}{3}<Z<\frac{107-95}{3}$) =P(-3<Z<4)=P(Z<4)-(1-P(Z<3))=1-1+0'9987=0'9987

15. .$\bar{x}$=60, σ=5'4; n=100; a)1-α=0'95. Calculamos el valor crítico P(Z<$z_{\alpha/2}$)=1-$\frac{0'05}{2}$ =
=1-0'025=0'975. Buscando en las tablas $z_{\alpha/2}$=1'96 El intervalo pedido es:
I_c= (60-1'96$\frac{5'4}{10}$, 60+1'96$\frac{5'4}{10}$)=(58'94,61'06)
b)El intervalo obtenido significa que en el 95% de las posibles muestras de este tamaño, el peso medio de los usuarios del gimnasio estará en ese intervalo.
c)El intervalo de confianza es ahora (58'5,61'5), para poder conocer la probabilidad debemos despejar $z_{\alpha/2}$; El radio del intervalo es 60-58'5=1'5 y sabemos que el radio es igual a $z_{\alpha/2}.\sigma/\sqrt{n}$
Entonces 1'5=$z_{\alpha/2}.\frac{5'4}{10}\Rightarrow z_{\alpha/2}=\frac{1'5.10}{5'4}$=2'78 Mirando en las tablas P(Z<$z_{\alpha/2}$)=0'9973, y sabemos que P(Z<$z_{\alpha/2}$)=1-$\frac{\alpha}{2}=0'9973\Rightarrow\frac{\alpha}{2}=0'0027\Rightarrow\alpha=0'0054$ Por tanto, el nivel de confianza será 1-α=1-0'0054=0'9946. Es decir la afirmación es correcta con una probabilidad del 99'46%

16. X(20,40); $\bar{x}$=18'5, σ=4; n=40; 1-α=1-0'05=0'95. Calculamos el intervalo de confianza: valor crítico P(Z<$z_{\alpha/2}$)=1-$\frac{0'05}{2}$ ==1-0'025=0'975. Buscando en las tablas $z_{\alpha/2}$=1'96 El intervalo pedido es: I_c= (18'5-1'96$\frac{4}{\sqrt{40}}$, 18'5+1'96$\frac{4}{\sqrt{40}}$)=(17'26,19'73). Lo que nos piden es ver si la media pertenece a dicho intervalo. Como no pertenece, la muestra no es comparable con la población para ese nivel de significación.

17. p=P(votar X)=289/900=0'321 q=1-p=0'678; n=900
Calculamos el valor crítico $z_{\alpha/2}$ para un nivel de confianza 1-α=0'95;
P(Z<$z_{\alpha/2}$) = 1-$\frac{\alpha}{2}=1-\frac{0'05}{2}=1-0'025=0'975$ Buscando en las tablas hallamos $z_{\alpha/2}$=1'96
Por tanto el intervalo de confianza para la proporción que nos piden será:

I_c =($\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}.\hat{q}}{n}}$, $\hat{P}+z_{\alpha/2}.\sqrt{\frac{\hat{p}.\hat{q}}{n}}$)=

=(0’321-1’96.$\sqrt{\frac{0'321.0'678}{900}}$, 0’321+1’96.$\sqrt{\frac{0'321.0'678}{900}}$)=(0’29, 0’352)
El porcentaje de votos estará entre 29% y el 35’2% con un nivel de confianza del 95%

18. P=P(fumar)=1/5=0’2; q=0’8; E<0’04; 1-α=0’95; En primer lugar debemos hallar el valor crítico. Dado que es el mismo que en el ejercicio anterior podemos ya decir que $z_{\alpha/2}$=1’96
E=$z_{\alpha/2}$.$\sqrt{\frac{p.q}{n}}$ $= 1'96.\sqrt{\frac{0'2.0'8}{n}} < 0'04 \Rightarrow 1'96.\frac{\sqrt{0'2.0'8}}{0'04} < \sqrt{n} \Rightarrow 19'60 < \sqrt{n} \Rightarrow 384'16 < n$ Dado que el tamaño de la muestra debe ser un número entero hemos de tomar como mínimo 385 personas

III: TEXT DE HIPÓTESIS
19. A) H_0: $p_0 \geq 0,7$; H_1: $p_0 < 0,7$ Þ Contraste unilateral para la proporción
Estadístico del contraste: Z=$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}$
Región de aceptación: $1-\alpha = 0,99$; P(Z>z_α)=0’99$\Rightarrow P(Z < -z_\alpha)$=0’99$\Rightarrow$ $-z_\alpha = 2,3 \Rightarrow$
$\Rightarrow z_\alpha = -2'3$; Región (–2’3,∞)
Al sustituir los datos en el estadístico: $z_0 = \frac{\frac{61}{100}-0'7}{\sqrt{\frac{0'7.0'3}{100}}}$ =-1,96
Como pertenece al intervalo se acepta la hipótesis nula. Por tanto, se acepta con un nivel de confianza del 99% que al menos el 70% de los envíos de esa compañía llegan a su destino al día siguiente.
b) H_0: $p_0 \geq$=0,7; Ha: $p_0 < 0,7$ Þ Contraste unilateral de la proporción
El estadístico del contraste es el mismo del apartado anterior
Región de aceptación: 1 –aα = 0,92 ; - z_α = 1,41 ; región: $(-1'41, \infty)$
Como -1’96$\notin (-1'41, \infty)$se rechaza la hipótesis nula. Luego se concluiría que menos del 70% de los envíos de esa compañía llegan a su destino al día siguiente, con un nivel de significación del 98%.

20. A) H_0: $\mu_0 \leq 120$; $H_1 : \mu_0 > 120$ Contraste unilateral para la media
Estadístico del contraste: Z=$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{\overline{X}-120}{40/\sqrt{100}}$
Región de aceptación: $1-\alpha = 0,9$; P(Z<z_α) =0’9; z_α = 1’285 ; Región$(-\infty, 1'285)$
Al sustituir la media de la muestra en el estadístico obtenemos: z_0=$\frac{128-120}{40/\sqrt{100}}$=2
Como 2$\notin (-\infty, 1'285)$ se rechaza la hipótesis nula. Por tanto, no se puede aceptar, con un nivel de significación del 10%, la afirmación de partida.
B) las hipótesis nula y alternativa son las mismas, también lo son el estadístico del contraste y su valor para esta muestra. Lo que varía es la región de aceptación.
Nueva región: $1-\alpha = 0,99$; P(Z<z_α) =0’99; z_α = 2’33 ; Región$(-\infty, 2'33)$
Como 2$\in (-\infty, 2$¡33) se acepta la hipótesis nula. Por tanto, se puede aceptar, con un nivel de significación del 1%, la afirmación de partida.
(Obsérvese que cuanto más alto es el nivel de significación es más fácil aceptar la hipótesis alternativa cuando en realidad es falsa, como ya dijimos antes en la teoría)
21 H_0: $\mu_0 = 18000$; $H_1 : \mu_0 \neq 18000$ Contraste bilateral para la media
Estadístico del contraste: Z=$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}=\frac{\overline{X}-18000}{2000/\sqrt{35}}$
Región de aceptación: $1-\alpha = 0,95$; P(-$z_{\alpha/z}$<Z<$z_{\alpha/z}$)=0’95; P(Z<-$z_{\alpha/z}$)=1-α/2=1-0’025=0’975

$Z_{\alpha/z}$= 1'96; Región de aceptación (-1'96,1'96)
Al sustituir la media de la muestra en el estadístico obtenemos: $z_0=\frac{20000-18000}{2000/\sqrt{35}}$=5'92
Como 5'92$\notin(-1'96,1'96)$ se rechaza la hipótesis nula. Por tanto, no se puede aceptar, con un nivel de significación del 5%, la afirmación de partida.

22. Se trata de un contraste de hipótesis para las defierencias de medias
H_0: $\mu_1-\mu_2=0$; H_1 :$\mu_1-\mu_2\neq 0$ Contraste bilateral.
Estadístico de contraste $Z=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{4^2}{36}+\frac{4^2}{35}}}=\frac{(\overline{X}_1-\overline{X}_2)}{0'9495}$
Región de aceptación: $1-\alpha$ = 0,99 ; $P(-z_{\alpha/z}<Z<z_{\alpha/z})$=0'99; $P(Z<-z_{\alpha/z})=1-\alpha/2$=1-0'005=0'995
$Z_{\alpha/z}$= 2'575; Región de aceptación (-2'575, 2'575)

Al sustituir la media de la muestra en el estadístico obtenemos: $z_0=\frac{12-15}{0'9495}=-3'1595$
Como -3'1595$\notin(-2'575,2'575)$ se rechaza la hipótesis nula. Por tanto, la duración media de la enfermedad es distinta para los enfermos a los que se les ha aplicado anteriormente el antibiótico que para los que no, concluyendo a partir de la muestra que es mayor la duración de la enfermedad si previamente ya se les había aplicado el antibiótico.

23. A) Si la campaña no ha surtido efecto la proporción de personas que conocen el cambio debe continuar siendo igual o menor. Por tanto H_0: $P_0\leq 0'43$; H_1: $P_0>0'43$. Contraste unilateral para la proporción. Si se concluye que el porcentaje se mantuvo y realmente subió estaríamos aceptando la hipótesis nula cuando es falsa, el error cometido se llama de tipo II.
El estadístico del contraste será: $Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}}=Z=\frac{\hat{p}-0'43}{\sqrt{\frac{0'43.0'57}{110}}}=Z=\frac{\hat{p}-0'43}{\sqrt{0'0472}}$
Región de aceptación: $1-\alpha$ = 0,99 ; $P(Z<z_\alpha)$ =0'99; z_α = 2'33 ; Región$(-\infty,2'33)$
Al sustituir la media de la muestra en el estadístico obtenemos: $z_0=\frac{\frac{55}{110}-0'43}{\sqrt{0'0472}}$=1'54
Como 1'54$\in(-\infty,2'33)$ se acepta la hipótesis nula, es decir la campaña no ha mejorado el conocimiento de la población.
24. H_0:$\mu_1\leq\mu_2\Rightarrow\mu_1-\mu_2\leq 0$; Por tanto H_1 : $\mu_1-\mu_2>0$
Estadístico de contraste $Z=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{1'5^2}{40}+\frac{2^2}{30}}}=\frac{(\overline{X}_1-\overline{X}_2)}{0'4354}$
Región de aceptación: $1-\alpha$ = 0,9 ; $P(Z<z_\alpha)$ =0'9; z_α = 1'285 ; Región$(-\infty,1'285)$
Al sustituir la media de la muestra en el estadístico obtenemos: $z_0=\frac{3-2}{0'4354}$ =2'3
Como 2'3$\notin(-\infty,1'285)$ se rechaza la hipótesis nula. Es decr, se rechaza que el primer medicamento tenga menos efectos secundarios que el segundo con una nivel de confianza del 90%