UNIDAD 7: ESTADÍSTICA INFERENCIAL

ÍNDICE DE LA UNIDAD

1.- INTRODUCCIÓN.

Durante buena parte de la historia de las Matemáticas, la Estadística se ha dedicado fundamentalmente al desarrollo de la Estadística Descriptiva, cuya principal tarea ha sido recopilar datos, clasificarlos, tabularlos y relacionarlos. Esta concepción tan limitada de la Estadística ha venido cambiando radicalmente a partir de los años 30 del siglo XX con el nacimiento de la estadística inductiva o Estadística Inferencial, con la que se buscan métodos que permitan obtener conclusiones válidas para toda la población a partir del estudio de muestras. Esta inferencia o inducción necesita de una rama tan importante y compleja como la Probabilidad para sostener los cálculos y conclusiones, lo que establece una conexión clara entre la estadística tradicional y el cálculo de probabilidades dando lugar al complejo campo de la ***Inferencia Estadística.***

Durante esta unidad, repasaremos contenidos sobre estadística descriptiva y distribuciones de probabilidad muy utilizadas como la Normal y la Binomial trabajados en cursos anteriores, para conectar después con los contenidos propios de 2º curso de Bachillerato: Técnicas de Muestreo, Inferencia Estadística y Contraste de Hipótesis.

2.- VARIABLES ESTADÍSTICAS. PARÁMETROS.

En este punto de la unidad no pretendemos, en absoluto, desarrollar todo el contenido sobre variables estadísticas y parámetros, ya que estos se han venido trabajando durante toda la ESO y el primer curso de Bachillerato. Únicamente vamos a pasar a describir de manera breve y concisa los aspectos necesarios que hay que recordar para afrontar la presente unidad. Fundamentalmente vamos a recordar las variables estadísticas y los parámetros: media, varianza y desviación típica de un conjunto de datos sin agrupar en intervalos.

Definición 1: Llamamos **variable estadística** a cualquier característica que se observa y estudia en un estudio de tipo estadístico. Al conjunto total objeto del estudio estadístico se le llama **población.** Llamaremos **muestra** de dicha población a cualquier subconjunto de la población. Atendiendo al tipo de variable, se llama:

- **Variable cualitativa** a toda variable estadística que no se puede medir con un número.
- **Cuantitativa discreta** cuando se puede medir con un número y toma valores puntuales, es decir, entre dos valores consecutivos de la variable no hay otro valor.
- **Cuantitativa continua**, cuando se puede medir con un número y puede tomar infinitos valores de un intervalo.

Ejemplo 1: Veamos algunos ejemplos de variables estadísticas:

a) El color del pelo de una persona, su sexo o el partido político al que ha votado son variables cualitativas porque no se pueden medir con un números, ya que describen una cualidad no una cantidad.
b) El número de hijos de una persona o el número materias suspensas, por ejemplo, son variables cuantitativas discretas, ya que toman "valores a saltos": 0, 1, 2, 3….
c) El tiempo que emplea un atleta en correr los 100 metros lisos o la estatura de una persona son variables cuantitativas continuas ya que entre dos valores de la variable, hay infinitos otros que puede tomar.

Definición 2: Sea X una variable estadística cuantitativa que toma los valores $X_1, X_2, ...X_n$ con frecuencias absolutas respectivas(nº de veces que se repite) $f_1, f_2, ...f_n$. Definimos:

a) La **media aritmética** como: $\boxed{\overline{X} = \dfrac{\sum_{i=1}^{n} X_i \cdot f_i}{\sum_{i=1}^{n} f_i} = \dfrac{X_1 \cdot f_1 + X_2 \cdot f_2 + ... + X_n \cdot f_n}{N}}$

b) La **varianza** como: $\boxed{S_x^{\,2} = \dfrac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \cdot f_i}{\sum_{i=1}^{n} f_i} = \dfrac{\sum_{i=1}^{n} X_i^{\,2} \cdot f_i}{N} - \overline{X}^2 = \dfrac{X_1^{\,2} \cdot f_1 + X_2^{\,2} \cdot f_2 + ... + X_n^{\,2} \cdot f_n}{N} - \overline{X}^2}$

c) La **desviación típica** como: $\boxed{S_x = +\sqrt{S_x^{\,2}}}$

c) El **coeficiente de variación** como: $\boxed{CV_x = \frac{S_x}{\overline{X}}}$. Frecuentemente se expresa en porcentajes. Para ello, basta multiplicar el resultado por 100.

Ejemplo 2: Consideremos que las calificaciones en Matemáticas de los 20 alumnos/as de un curso de 2º de Bachillerato son: 0, 0, 1, 2, 2, 2, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 8, 9, 9, 9

a) La nota media sería: $\overline{X} = \frac{0\cdot 2+1\cdot 1+2\cdot 3+3\cdot 1+4\cdot 3+5\cdot 4+6\cdot 2+8\cdot 1+9\cdot 3}{20} = \frac{89}{20} = 4,45$

b) Con una varianza de:

$$S_x{}^2 = \frac{0^2\cdot 2+1^2\cdot 1+2^2\cdot 3+3^2\cdot 1+4^2\cdot 3+5^2\cdot 4+6^2\cdot 2+8^2\cdot 1+9^2\cdot 3}{20} - 4,45^2 = \frac{549}{20} - 4,45^2 = 7,6475$$

c) La desviación típica sería: $S_x = \sqrt{7,6475} \simeq 2,77$

d) El coeficiente de variación: $CV_x = \frac{2,77}{4,45} \simeq 0,6225$. Expresado en porcentajes, sería de un 62,25%.

Se propone la **actividad 1.**

3.- DISTRIBUCIONES DE PROBABILIDAD.

Definición 3: Sea X una variable aleatoria discreta. Llamamos **función de probabilidad** a la función: $f(x_i) = P(X = x_i)$

Definición 4: Sea X una variable aleatoria continua. Llamamos **función de densidad** a la función $f(x)$ que permite hallar mediante el cálculo de áreas las probabilidades de las distribuciones continuas.

Definición 5: Sea X una variable aleatoria. Llamamos **función de distribución** a la función $F(x) = P(X \le x)$, tanto para el caso discreto como para el caso continuo.

Proposición 1: Sea X una variable aleatoria y F su función de distribución. Entonces: $P(a < X \le b) = F(b) - F(a)$

Nota 1: Los conceptos y propiedad anterior se pueden desarrollar mucho más para su completa comprensión. No obstante, como se aleja de los objetivos de la presente unidad, nos limitaremos a ver brevemente, y sin entrar en detalles, aquello que necesitamos para los objetivos de la unidad. Si conviene observar que mientras que en el caso discreto, la probabilidad de la proposición 1 corresponde a una suma de probabilidades, en el caso continuo corresponde al área encerrada por la curva de la función de densidad entre dos valores. Por ello, estas áreas las calcularemos utilizando una tabla asociada.

Definición 6: Sea X una variable aleatoria discreta. Se definen:

a) La **esperanza matemática** o **media** de la variable como:

$\boxed{\mu_x = E(X) = \sum_{i=1}^{n} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + ... + x_n \cdot p_n}$, siendo $p_i = f(x_i) = P(X = x_i)$.

b) La **varianza** de la variable como:

$$\boxed{\sigma^2{}_x = Var(X) = \sum_{i=1}^{n}(x_i - \mu_x)^2 \cdot p_i = \sum_{i=1}^{n}(x_i{}^2 \cdot p_i - \mu_x{}^2) = x_1{}^2 \cdot p_1 + x_2{}^2 \cdot p_2 + \ldots + x_n{}^2 \cdot p_n - \mu_x{}^2}$$

c) La **desviación típica** de la variable como: $\boxed{\sigma_x = +\sqrt{\sigma_x{}^2}}$

Nota 2: Cuando no haya confusión con otra variable, se puede omitir el subíndice y escribir simplemente μ o σ.

Ejemplo 3: Consideremos la variable aleatoria discreta correspondiente a la puntuación en el lanzamiento de un dado.

a) Su función de probabilidad f será: $f(1) = P(X=1) = \frac{1}{6}$, $f(2) = P(X=2) = \frac{1}{6}$,..., $f(6) = P(X=6) = \frac{1}{6}$

b) Un ejemplo de un valor de su función de distribución es: $F(3.5) = P(X \le 3.5) = \frac{3}{6} = \frac{1}{2}$

c) Su media será: $\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = \frac{7}{2} = 3{,}5$

d) Su varianza: $\sigma^2 = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \simeq 2{,}92$

e) Su desviación típica será: $\sigma = \sqrt{\frac{35}{12}} \simeq 1{,}71$

La interpretación de la media es que corresponde con el "valor esperado" al lanzar un dado. Sería el valor al que se aproximan las medias de las puntuaciones que se obtienen a medida que aumentamos el número de lanzamiento. La desviación típica sería el valor al que se acercan las desviaciones entre las puntuaciones obtenidas y el valor central de la media.

Nota 3: En el caso de las variables continuas también existen los conceptos de media, varianza y desviación típica. No vamos a ver sus expresiones porque necesitamos del cálculo integral y se escapa de los contenidos de la unidad.

3.1.- Distribución Binomial.

Definición 7: Llamamos **experimento de Bernouilli** a todo experimento en el que se pueden dar únicamente dos resultados, uno de ellos es considerado como éxito, con probabilidad p y otro como fracaso, con probabilidad q = 1-p.

Ejemplo 4: Dos experimentos de Bernouilli son los siguientes:

a) Lanzar un dado y ver si sale un 2. En este caso la probabilidad de éxito $p = 1/6$ y la de fracaso $q = 5/6$

b) En una clase de 20 alumnos/as en la que hay 12 chicas, elegir al azar a una persona y ver si es chica. En este caso la probabilidad de éxito $p = \frac{12}{20} = \frac{3}{5}$ y la de fracaso $q = \frac{8}{20} = \frac{2}{5}$

Definición 8: Se dice que una variable estadística X sigue una **distribución binomial** de parámetros n y p si X mide el número de éxitos que se obtienen al realizar n veces un experimento de Bernouilli con probabilidad de éxito p. Se expresa: $X \to Bi(n,p)$

Proposición 2: Sea una variable Binomial $X \to Bi(n,p)$. Entonces, la probabilidad de obtener k éxitos (función de probabilidad) viene dada por: $\boxed{P(X=k)=\binom{n}{k}\cdot p^k \cdot q^{n-k}}$.

Siendo $\binom{n}{k}$ el número combinatorio $\binom{n}{k}=\frac{n!}{k!\cdot(n-k)!}$

Proposición 3: Sea una variable Binomial $X \to Bi(n,p)$. Entonces:

a) $\boxed{\mu = n\cdot p}$ b) $\boxed{\sigma=\sqrt{n\cdot p\cdot q}}$

Ejemplo 5: Si retomamos el ejemplo 3, apartado b y consideramos la variable X que mide el número de chicas que se obtienen al elegir al azar 5 personas de la clase, es evidente que X sigue una distribución binomial. Concretamente: $X \to Bi\left(5,\frac{3}{5}\right)$, ya que el experimento se realiza $n=5$ veces y si consideramos como éxito elegir a una chica, hemos visto ya en el ejemplo 3b que su probabilidad es $p=\frac{3}{5}$. Calculemos algunas probabilidades asociadas a esta distribución:

a) La probabilidad de elegir 1 chica sería: $P(X=1)=\binom{5}{1}\cdot\left(\frac{3}{5}\right)^1\cdot\left(\frac{2}{5}\right)^4=\frac{5!}{1!\cdot 4!}\cdot\frac{3}{5}\cdot\frac{16}{625}=\frac{48}{625}$

b) La probabilidad de elegir 3 chicas sería: $P(X=3)=\binom{5}{3}\cdot\left(\frac{3}{5}\right)^3\cdot\left(\frac{2}{5}\right)^2=\frac{5!}{3!\cdot 2!}\cdot\frac{27}{125}\cdot\frac{4}{25}=\frac{216}{625}$

c) La media y la desviación típica serán: $\mu=5\cdot\frac{3}{5}=3$, mientras que: $\sigma=\sqrt{5\cdot\frac{3}{5}\cdot\frac{2}{5}}=\frac{\sqrt{30}}{5}\simeq 1{,}10$ que significa que el número medio de chicas que se espera escoger son 3 con una desviación de algo más de una chica.

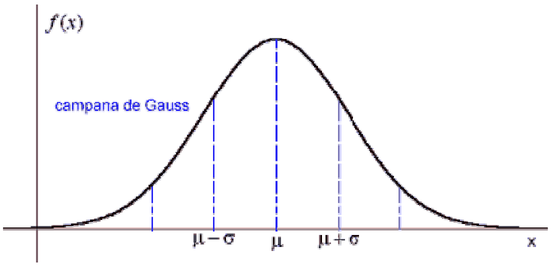
Nota 4: Como es evidente, el cálculo de probabilidades con la fórmula de la proposición 1 que acabamos de llevar a cabo, se hace bastante ardua en el caso de valores de n mayores. Por ello, existen tablas que nos permiten calcular las probabilidades anteriores con bastante precisión e incluso, muchas calculadoras actuales son capaces de hacer el cálculo muy rápidamente y con mucha precisión. De todas formas, para valores “grandes” (se considerarán grandes para $n \geq 30$), la aproximaremos por otra distribución llamada distribución normal y que veremos a continuación.

Se proponen la **actividades 2 y 3**.

3.2.- Distribución Normal.

Ejemplo 6: Se dice que una variable aleatoria continua sigue una **distribución normal** de parámetros μ y σ cuando su función de densidad viene dada por la expresión: $f(x)=\frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$.

Su gráfica se llama campana de Gauss:



Abreviadamente lo expresaremos: $X\to N(\mu,\sigma)$

Proposición 4: La función de densidad de una distribución normal cumple las siguientes propiedades:

a) La media, varianza y desviación típica de una variable normal son μ , σ^2 y σ .

b) El área encerrada bajo la campana de Gauss es 1.

c) La campana de Gauss es simétrica respecto de la recta vertical $x=\mu$, con lo que el área encerrada a izquierda y derecha de la media es igual a 0,5.

d) Al tratarse de una variable continua, la probabilidad de que la una variable normal tome un valor puntual es siempre 0.

Nota 5: Las distribuciones normales son las distribuciones más importantes, con diferencia, de toda la estadística debido a que la mayoría de los fenómenos que se estudian, son o se acercan, al modelo de distribución normal. Variables como la edad, estatura, peso,….. se comportan "normalmente" de ahí su nombre.

Nota 6: Como ya hemos visto antes, la probabilidad de que el valor de una variable normal esté entre dos corresponde con la diferencia de la función de distribución en dichos valores. Para este cálculo se suelen utilizar tablas en lugar de cálculo integral. Como es evidente, para cada valor de la media y la desviación típica, la función de densidad y, por tanto, la tabla tendría que se distinta. Como en la práctica esto es inviable, vamos a ver una propiedad que permite reducir todos los casos a uno.

Proposición 5: **(Tipificación de una distribución normal)** Sea X una variable aleatoria normal $X\to N(\mu,\sigma)$. Entonces, la variable $\boxed{Z=\frac{X-\mu}{\sigma}}$ cumple que $Z\to N(0,1)$. En resumen, el cambio de variable del recuadro, transforma cualquier variable normal de cualquier media y desviación típica en una con media 0 y desviación típica 1. Esto permite que todos los cálculos los podamos llevar a cabo con una única tabla. A esta variable se le llama **distribución normal estándar o tipificada**.

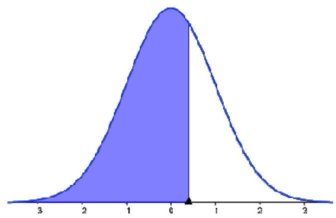
Ejemplo 7: Veamos en el siguiente ejemplo, el manejo de la tipificación y de la tabla normal. Supongamos que una variable X, que describe la nota media de 2º de Bachillerato del alumnado de un instituto sigue una distribución normal de media 6.4 años con desviación típica 2. Abreviadamente $X\to N(6.4,2)$. Elegido un alumno al azar, calculemos las probabilidades siguientes:

a) Tenga una nota media de un 7.

Evidentemente $P(X=7)=0$ ya que se trata de un valor puntual y la distribución normal es continua.

b) Tenga una nota media de menos de 7.2.

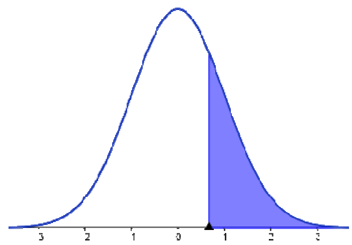
$$P(X\le 7.2)\overset{Tipificando}{=}P\left(\frac{X-6.4}{2}\le\frac{7.2-6.4}{2}\right)=P(Z\le 0.4)\overset{Tabla}{=}0.6554$$



c) Tenga más de un 7.75 de nota media.

$$P(X\ge 7.75)\overset{Tipificando}{=}P\left(\frac{X-6.4}{2}\ge\frac{7.75-6.4}{2}\right)=$$

$$P(Z\ge 0.675)=1-P(Z\le 0.675)\overset{Tabla}{=}1-0.7517=0.2483$$

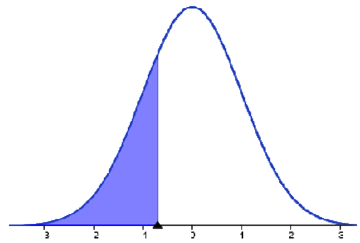


d) Sabiendo que hay 76 alumnos en total, ¿Cuantos/as alumnos/as habrá con menos de un 5 de nota media?

$$P(X\le 5)\overset{Tipificando}{=}P\left(\frac{X-6.4}{2}\le\frac{5-6.4}{2}\right)=P(Z\le -0.7)\overset{Simetría}{=}$$

$$=P(Z\ge 0{,}7)=1-P(Z\le 0.7)\overset{Tabla}{=}=1-0.7580=0.242.$$

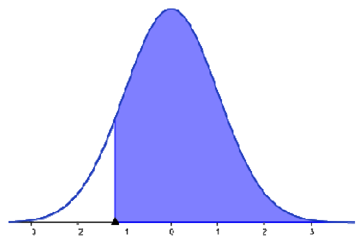
Por lo tanto, habrá $0.242\cdot 76=18.392\simeq 18$ alumnos/as



e) Tenga más de un 4 de nota media.

$$P(X\ge 4)\overset{Tipificando}{=}P\left(\frac{X-14.5}{2}\ge\frac{4-6.4}{2}\right)=P(Z\ge -1.2)\overset{Simetría}{=}$$

$$=P(Z\le 1.2)\overset{Tabla}{=}0.8849$$

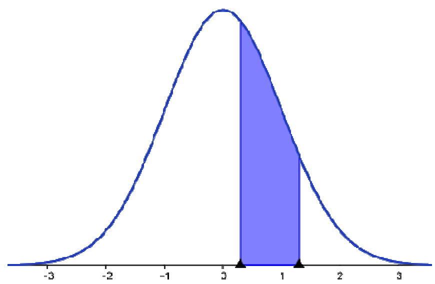


f) Tenga un notable (entre 7 y 9)

$$P(7\le X\le 9)=P\left(\frac{7-6.4}{2}\le\frac{X-6.4}{2}\le\frac{9-6.4}{2}\right)=$$

$$P(0.3\le Z\le 1.3)=P(Z\le 1.3)-P(Z\le 0.3)\overset{Tabla}{=}$$

$$=0.9032-0.6179=0.2853$$

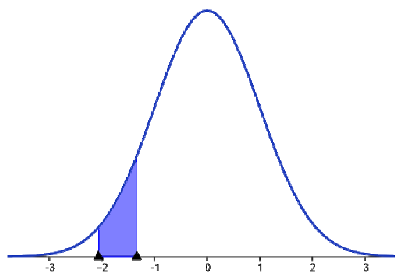


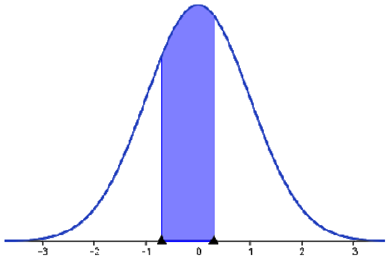
g) Tenga entre un 2.38 y un 3.72.

$$P(2.28\le X\le 3.72)=P\left(\frac{2.28-6.4}{2}\le\frac{X-6.4}{2}\le\frac{3.72-6.4}{2}\right)=$$

$$P(-2.06\le Z\le -1.34)\overset{Simetría}{=}P(2.06\le Z\le 1.34)=$$

$$P(Z\le 2.06)-P(Z\le 1.34)\overset{Tabla}{=}0.9803-0.9099=0.0704$$





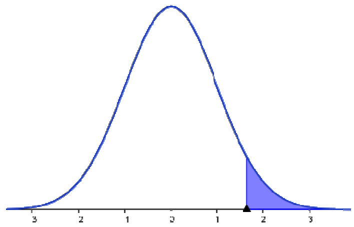
h) Tenga un suficiente o un bien (entre un 5 y un 7)

$$P(5 \le X \le 7) = P\left(\frac{5-6.4}{2} \le \frac{X-6.4}{2} \le \frac{7-6.4}{2}\right) =$$

$$P(-0.7 \le Z \le 0.3) = P(Z \le 0.3) - P(Z \le -0.7) =$$

$$\overset{Simetría}{=} P(Z \le 0.3) - P(Z \ge 0.7) = P(Z \le 0.3) - (1 - P(Z \le 0.7)) =$$

$$\overset{Tabla}{=} 0.6179 - 1 + 0.7580 = 0.3759$$



i) Se sabe que sólo un 5% del alumnado puede obtener una matrícula de honor. Halla la nota media que debe alcanzar un alumno/a para que se le conceda. Buscamos una nota k que cumpla:

$$P(X \ge k) = 0.05 \to P(X \le k) = 0.95 \overset{Tipificando}{\to} P\left(\frac{X-6.4}{2} \le \frac{k-6.4}{2}\right) = 0.95 \to$$

$$\to P\left(Z \le \frac{k-6.4}{2}\right) = 0.95 \overset{Tabla}{\to} \frac{k-6.4}{2} = 1.645 \to k = 2 \cdot 1.645 + 6.4 = 9.69$$

Los dos enlaces siguientes pueden ser de gran ayuda para entender toda la casuística que nos podemos encontrar, además de las expuestas en este ejemplo. Incluyen representaciones gráficas interactivas.

a) http://www.geogebra.org/en/upload/files/spanish/PepeTarraga/DistribucionNormal.html
b) http://www.slideshare.net/jefedo61/normalgeogebra

Se proponen las **actividades 4, 5 y 6**.

3.3.- Aproximación de la Binomial por la Normal.

*Como ya se comentó en el punto de la distribución binomial, al aumentarse el parámetro n de la binomial, los cálculos utilizando la función de probabilidad se complican sustancialmente. Para evitar este problema, vamos a proporcionar una herramienta que permite aproximar los valores de la distribución binomial por la normal en determinados casos. La siguiente proposición es consecuencia directa de un teorema más general llamado **Teorema Central del** Límite. Esto nos permitirá determinar probabilidades de la distribución binomial utilizando la distribución normal estándar.*

Proposición 6: **(Aproximación de la binomial por la normal).** Sea $X \to Bi(n,p)$ tal que $n \ge 30$; $np \ge 5$ y $nq \ge 5$. Entonces la variable X se puede aproximar por una distribución normal $X' \to N(np, \sqrt{npq})$.

Nota 7: Como es lógico, el hecho de aproximar una variable discreta como la binomial, por una continua como la normal, da lugar a un problema. El problema es que mientras sí existe la probabilidad de que una binomial tome un valor concreto, en el caso de la normal, esta probabilidad es cero. Para resolver este problema, se introducen las llamadas correcciones de Yates, que consiste en considerar los valores de la variable discreta X como marcas de clase de intervalos de la forma siguiente:

Si $X \to Bi(n,p)$, con $n \geq 30$; $np \geq 5$ y $nq \geq 5$, entonces $X' \to N\left(np, \sqrt{npq}\right)$. Las correcciones de Yates consisten en tomar: $P(X = k) \simeq P(k - 0.5 \leq X' \leq k + 0.5)$

<u>Ejemplo 8:</u> Se sabe que la probabilidad de que un piloto de Fórmula 1 sufra un reventón en un circuito es de 0.04. Si en una carrera participan 200 conductores, calcula:

a) La probabilidad de que se registren entre 12 y 18 reventones.

Si consideramos como éxito el que sufra un reventón, es evidente que la variable $X \to Bi(200, 0.04)$. Evidentemente, calcular esta probabilidad con la función de distribución o la de probabilidad de la binomial es una tarea bastante larga. Vemos si cumple las hipótesis para aproximarla por una normal: $n = 200 \geq 30$; $np = 200 \cdot 0.04 = 8 \geq 5$ y $nq = 200 \cdot 0.96 = 192 \geq 5$. Así pues, podemos aproximar la variable X por la normal $X' \to Bi\left(200 \cdot 0.04, \sqrt{200 \cdot 0.04 \cdot 0.96}\right) \simeq Bi(8, 2.77)$. Así pues, la probabilidad pedida será, aproximadamente:

$$P(12 \leq X \leq 18) \overset{Yates}{\simeq} P(11.5 \leq X' \leq 18.5) = P\left(\frac{11.5-8}{2.77} \leq \frac{X'-8}{2.77} \leq \frac{18.5-8}{2.77}\right) = P(1.26 \leq Z \leq 3.79) =$$
$$P(Z \leq 3.79) - P(Z \leq 1.26) = 0.99992 - 0.8962 = 0.10372$$

Si lo calculamos con Geogebra, el valor utilizando la binomial sale 0.107, que es bastante aproximado.

b) La probabilidad de que se registren exactamente 5 reventones.

$$P(X = 5) \overset{Yates}{\simeq} P(4.5 \leq X' \leq 5.5) \overset{Tipificando}{=} P(-1.26 \leq Z \leq -0.90) = P(1.26 \leq Z \leq 0.90) = P(Z \leq 1.26) -$$
$$-P(Z \leq 0.90) = 0.8962 - 0.8159 = 0.0803$$

Aunque este cálculo sí se podría hacer fácilmente con la calculadora:

$$P(X = 5) = \binom{200}{5} \cdot 0.04^5 \cdot 0.96^{195} \simeq 0.0906.$$

Se proponen las **actividades 7 y 8**.

<u>4.- MUESTREO.</u>

A partir de este punto, vamos a comenzar a trabajar específicamente los contenidos más relevantes de la unidad relativos a la Inferencia Estadística, comenzando por el muestreo. Ya definimos al comienzo de la unidad los conceptos de población y muestra. La primera pregunta a la que nos enfrentamos es la de cómo elegir una muestra de la población y de qué tamaño debe tomarse. A estas preguntas iremos respondiendo a lo largo de los siguientes puntos y vamos a comenzar describiendo las principales técnicas de muestreo. Hay muchas maneras de elegir una muestra de una población y lo que debemos tratar siempre es que la muestra sea lo más representativa posible de toda *la población y con un número de elementos con el que se pueda trabajar de manera cómoda y fiable. Para que una muestra sea representativa, es imprescindible que el muestreo sea aleatorio, es decir, que los individuos de la población, en la medida de lo posible, tengan las mismas posibilidades de ser elegidos. Por ello, nos centraremos en los llamados muestreos probabilísticos, en los que la persona que realiza el estudio, elige al azar a los individuos que conformarán la muestra. En lo que sigue, designaremos por N al tamaño de la población y por n al tamaño de la muestra.*

Definición 9: Un muestreo se llama:

a) **Aleatorio simple**, cuando cada individuo de la población tiene la misma probabilidad de ser elegido.

b) **Sistemático**, cuando se elige al azar a un individuo de la población y, a partir de este, se eligen los demás a intervalos constantes.

c) **Estratificado**, cuando se divide a la población en clases o estratos y se escogen aleatoriamente un número de individuos de cada estrato mediante muestreo aleatorio simple o sistemático. Puede ser:

- Con **afijación igual** si se toma el mismo número de individuos de cada estrato sin tener en cuenta su tamaño.
- Con **afijación proporcional**, si se toma el número de individuos de cada estrato proporcionalmente al tamaño del mismo.

d) **Por conglomerados**, si se divide la población en conjuntos o conglomerados, se eligen al azar algunos de estos conglomerados y se escoge una muestra únicamente de estos conglomerados por muestreo aleatorio simple.

Nota 8: No se debe confundir estrato con conglomerado. Mientras que un estrato es homogéneo (sus individuos tienen características del estudio comunes), el conglomerado es heterogéneo (sus individuos no tienen porqué tener características del estudio comunes para que representen bien a la población)

Ejemplo 9: Supongamos que de los 600 alumnos/as de ESO de un instituto, queremos escoger una muestra de 30 para hacer un estudio sobre la edad.

a) Un muestreo aleatorio simple se podría hacer utilizando un censo del alumnado (en el que aparece ordenando alfabéticamente) y escoger al azar 30 números del 1 al 600.

b) Un muestreo aleatorio sistemático se podría hacer eligiendo al azar un número del 1 al 600 y tomar a los 29 restantes separados por intervalos de $\frac{600}{30} = 20$ en 20. Si, por ejemplo, sale en el sorteo el 457, elegiríamos a los/as alumnos/ que tengan los números 477, 497, 517, 537, 557, 577, 597, 17, 37, 57, 77, 97, 117,………………, 417, 437.

c) Uno estratificado sería, por ejemplo, considerar los estratos: 1º ESO (en el que hay 192), 2º ESO (en el que hay 173), 3º ESO (en el que hay 149) y 4º ESO (en el que hay 86).

- Con afijación igual, escogeríamos a $\frac{30}{4}=7.5$ de cada nivel. Como esto es imposible, tendríamos que elegir entre 7 u 8 alumnos/as de cada nivel. Como hay más en 1º y 2º, lo lógico sería escoge 8 de 1º, 8 de 2º, 7 de 3º y 7 de 4º mediante muestreo aleatorio simple o sistemático.
- Con afijación proporcional, escogeríamos un número de alumnos/as en cada muestra proporcional al número de alumnos que tiene el estrato. Para ello, basta con hacer una regla de tres para cada estrato:

Total Muestra

$$\begin{array}{ccc} 600 & \to & 30 \\ 192 & \to & n_1 \end{array} \quad \to \quad \frac{600}{192}=\frac{30}{n_1} \to n_1=\frac{192\cdot 30}{600}=9.6 \simeq 10 \text{ alumnos/as de 1º ESO}$$

Análogamente, tomaríamos 9 de 2º ESO, 7 de 3º y 4 de 4º ESO.

d) Uno por conglomerados sería escoger, por ejemplo, los conglomerados de chicas y chicos y escoger uno de ellos al azar, por ejemplo el de las chicas. Después, se toman 30 chicas mediante muestreo aleatorio simple. Es algo similar a lo hecho en el apartado a pero sólo con las chicas.

Nota 9: Además de esto, hemos de tener en cuenta que todo muestreo se puede llevar a cabo con reemplazamiento (cuando un mismo individuo puede aparecer varias veces en la muestra. En este sentido, una población finita puede ser considerada como infinita) o sin reemplazamiento (cuando no puede aparecer más que una vez). Siguiendo las directrices de las PAU, nos centraremos siempre en muestreos con reemplazamiento y únicamente estudiaremos muestreos aleatorios simples o estratificados, en la inmensa mayoría de los casos, con afijación proporcional.

Se propone la **actividad 9**

5.- DISTRIBUCIONES MUESTRALES.

Una vez vistas las principales técnicas de muestreo, el siguiente paso en un estudio de inferencia estadística es realizar inferencias sobre ciertos parámetros poblacionales a partir de datos muestrales. Los parámetros muestrales habituales (media, desviación típica y proporción habitualmente) nos van a permitir decidir sobre la aproximación más conveniente de los correspondientes parámetros poblacionales. Por ello es necesario que estudiemos, como primer paso, las distribuciones que siguen estos parámetros. En esta unidad, nos centraremos en estudiar los parámetros media y proporción asociados a las dos distribuciones habituales: la normal y la binomial.

5.1.- Distribución muestral de medias.

Si consideramos todas las posibles muestras de tamaño n que se pueden extraer de una población, podemos considerar como una variable aleatoria a la media $\overline{X}$ cuyo valor sería la media aritmética de cada muestra.

Definición 10: Llamamos **distribución muestral de medias** a la distribución de la variable $\bar{X}$ descrita anteriormente.

Proposición 7: **(Propiedades de la distribución muestral de medias)** La distribución muestral de medias definida anteriormente verifica:

a) La media o esperanza de $\bar{X}$ coincide con la media poblacional μ, es decir, $\boxed{\mu_{\bar{X}} = \mu_X}$

b) La desviación típica de $\bar{X}$ coincide con $\frac{\sigma_X}{\sqrt{n}}$, es decir, $\boxed{\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}}$ (siempre que la muestra sea con reemplazamiento)

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

Para salvar el inconveniente de cálculos en poblaciones que no se distribuyan normalmente, es de enorme importancia el siguiente teorema, que es una consecuencia del llamado Teorema Central del Límite.

Proposición 8: Sea X una variable estadística y consideramos la distribución muestral de medias con n suficientemente grande ($n \geq 30$), entonces la distribución muestral de medias se aproxima a una distribución normal y podemos considerar en la práctica que $\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Nota 10: En las actividades de este curso consideraremos siempre distribuciones normales o no normales con tamaño muestral $n \geq 30$, por lo que siempre podremos considerar la distribución muestral de medias como una normal por el teorema anterior.

Nota 11: En algunos casos prácticos, la desviación típica poblacional σ_X es desconocida. En estos casos, si el tamaño de la muestra es suficientemente grande, normalmente $n \geq 100$, podemos tomar como aproximación la desviación típica de la muestra S_X.

Ejemplo 10: Las estaturas de 1200 estudiantes de un centro de enseñanza se distribuyen normalmente con una media de 1,72 m y desviación típica 0,09. Si se toma al azar una muestra de 36 estudiantes. Hallemos:

a) La probabilidad de que la media sea inferior a 1,75 m

Como hemos visto, la distribución muestral de medias sigue una distribución normal:

$$\bar{X} \to N\left(1.72, \frac{0.09}{\sqrt{36}}\right) = N(1.72, 0.015) \to P\left(\bar{X} \leq 1.75\right) = P\left(\frac{\bar{X} - 1.72}{0.015} \leq \frac{1.75 - 1.72}{0.015}\right) = P(Z \leq 2) = 0.9772$$

b) La probabilidad de que la media esté entre 1,68 m y 1,73 m.

$$P\left(1{,}68 \leq \bar{X} \leq 1.73\right) = P\left(\frac{1.68 - 1.72}{0.015} \leq \frac{\bar{X} - 1.72}{0.015} \leq \frac{1.73 - 1.72}{0.015}\right) = P(-2.67 \leq Z \leq 0.67) =$$

$$= P(Z \leq 0.67) - P(Z \leq -2.67) = P(Z \leq 0.67) - P(Z \geq 2.67) = P(Z \leq 0.67) - \left(1 - P(Z \leq 2.67)\right) =$$

$$= 0.7486 - 1 + 0.9962 = 0.7448$$

Se proponen las **actividades 10, 11 y 12.**

5.2.- Distribución muestral de proporciones.

De manera análoga a lo hecho con las medias, podemos considerar todas las posibles muestras de tamaño n de una población con una distribución binomial de probabilidad de éxito p y considerar las proporciones muestrales $\hat{p}$ como una variable aleatoria.

Definición 11: Llamamos **distribución muestral de proporciones** a la distribución de la variable $\hat{p}$ descrita anteriormente.

Proposición 9: **(Propiedades de la distribución muestral de proporciones)** La distribución muestral de proporciones definida anteriormente verifica:

a) La media o esperanza de $\hat{p}$ coincide con la media poblacional p, es decir, $\boxed{\hat{p} = p}$

b) La desviación típica de $\hat{p}$ coincide con $\sqrt{\frac{p \cdot q}{n}}$, es decir, $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$ (siempre que la muestra sea con reemplazamiento)

Proposición 10: Sea X una variable binomial y consideramos la distribución muestral de proporciones con n suficientemente grande ($n \geq 30$), siendo además $n \cdot p \geq 5$ y $n \cdot q \geq 5$ Entonces la distribución muestral de proporciones se aproxima a una distribución normal y podemos considerar en la práctica que $\hat{p} \to N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$.

Nota 12: Aunque es poco habitual, si en algún caso práctico, es desconocido el parámetro p, lo aproximaremos por la proporción muestral siempre que el tamaño sea suficientemente grande, normalmente $n \geq 100$.

Ejemplo 11: 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 la fábrica. Calculemos la probabilidad de que:

a) Haya más de un 5% de piezas defectuosas en la caja.
De acuerdo con lo visto anteriormente, la distribución muestral de proporciones sigue una distribución normal $\hat{p} \to N\left(0.03, \sqrt{\frac{0.03 \cdot 0.97}{500}}\right) = N(0.03, 0.0076)$. Así pues, la probabilidad pedida será:

$$P(\hat{p} > 0.05) = P\left(\frac{\hat{p} - 0.03}{0.0076} > \frac{0.05 - 0.03}{0.0076}\right) = P(Z > 2.63) = 1 - P(Z \leq 2.63) = 1 - 0.9957 = 0.0043$$

b) Haya menos de 10 piezas defectuosas en la caja.
Como 10 piezas constituyen el 2% de las piezas (a saber: $\frac{10}{500} = 0.02$), la probabilidad pedida es: $P(\hat{p} < 0.01) = P\left(\frac{\hat{p} - 0.03}{0.0076} < \frac{0.02 - 0.03}{0.0076}\right) = P(Z < -1.32) = P(Z > 1.32) =$
$= 1 - P(Z \leq 1.32) = 1 - 0.9066 = 0.0934$.

Se proponen las **actividades 13 y 14**.

6.- ESTIMACIÓN DE PARÁMETROS.

Ya hemos estudiado la base de la Estadística Inferencial, que son la teoría de muestreo y las distribuciones muestrales. El principal problema de las distribuciones muestrales es que lo normal es que no se tengan datos exactos de los parámetros poblacionales. Para ello, se utiliza la estimación de parámetros, tanto puntual como por intervalos, que vamos a abordar a continuación, y el contraste de hipótesis que veremos al final de la unidad.

6.1.- Estimación puntual.

Definición 12: Llamamos **estadístico** a toda función de los datos muestrales que asigna a cada muestra de tamaño n un valor numérico.

Definición 13: Un **estimador** para un parámetro poblacional desconocido es un estadístico que nos proporciona una aproximación de dicho parámetro poblacional.

Ejemplo 12: Un estimador para la media poblacional μ es la media muestral $\overline{X}$.

6.2.- Estimación por intervalos de confianza.

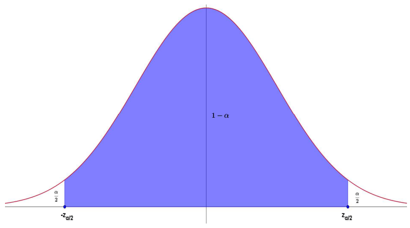
Evidentemente, en la mayoría de las ocasiones, no tiene mucho sentido llevar a cabo una estimación puntual sino más bien una estimación de entre qué dos valores se puede encontrar un cierto parámetro con una probabilidad prefijada. A esto nos dedicaremos en este punto en el que abordaremos la estimación mediante intervalos de confianza.

Definición 14: Llamamos **intervalo de confianza** para un parámetro poblacional al intervalo que contiene a dicho parámetro con una probabilidad, también llamada **nivel de confianza** $N_c = 1-\alpha$. Al valor de α se le llama **nivel de significación.**

Nota 13: En este curso únicamente abordaremos los intervalos de confianza para la media μ y la proporción p de distribuciones normales (o con tamaño suficientemente grande para aproximarlas a una normal) y binomiales (que también la aproximamos a la normal).

Nota 14: Obsérvese que el intervalo de confianza no es más que un intervalo en el que se encontraría el parámetro poblacional con una probabilidad o confianza de $N_c = 1-\alpha$. Además, como es lógico, no tiene mucho sentido hallar intervalos de confianza con "poca confianza" ya que no aportarían mucha información precisa. Por ello, los valores más comunes del nivel de confianza son 90%, 95% y 99%, que corresponden a los valores 0.90, 0.95 y 0.99 de $N_c = 1-\alpha$.

Como es lógico también, en virtud de los resultados consecuentes del Teorema Central del Límite usados para describir las distribuciones muestrales, hemos de pensar que dichos intervalos van a estar asociados a una distribución Normal. Por ello, el primer paso lógico es determinar el valor $z_{\alpha/2}$ de la tabla de la Normal asociado a cada nivel de confianza y que debe cumplir: $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$, ya que es lógico que esté centrado en la media.



Ahora bien, como pretendemos determinar un intervalo de confianza para la media poblacional a partir de la media muestral y sabemos por el punto anterior que la distribución de medias muestrales sigue una distribución normal $\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Entonces, la expresión es equivalente a:

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z_{\alpha/2}\right) = 1-\alpha \quad \Leftrightarrow \quad P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \quad \overset{Simetría}{\Leftrightarrow}$$

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

Así pues, el intervalo de confianza para la media poblacional con un nivel de confianza $N_c = 1-\alpha$, viene dado por la fórmula: $\boxed{I_c = \left(\bar{X}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)}$ que será la que usemos directamente en las actividades.

Nota 15: Como es evidente, el valor $z_{\alpha/2}$, al que llamamos **valor crítico** asociado al nivel de confianza $1-\alpha$ verifica que $P\left(Z \le z_{\alpha/2}\right) = 1-\frac{\alpha}{2} = \frac{1+N_c}{2}$. No hay más que usar la tabla de la Normal para determinar el valor crítico $z_{\alpha/2}$ que necesitamos para determinar el intervalo de confianza.

Nota 16: Según hemos visto en la construcción del intervalo de confianza, la diferencia máxima entre la media muestral y la poblacional, es $\boxed{E = z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}}$ y se le llama **error máximo admisible** y es la mitad de la **amplitud del intervalo** $\boxed{A = 2\cdot z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}}$. De la expresión del error, podemos despejar el **tamaño muestral** $\boxed{n = \left(\frac{z_{\alpha/2}\cdot\sigma}{E}\right)^2}$

Ejemplo 13: Una muestra aleatoria de 100 alumnos/as que se presenta a las PAU revela que la media de edad es de 18,1 años. Halla un intervalo de confianza del 90% para la media de edad de todos los estudiantes, sabiendo que la desviación típica de la población es de 0,4

Evidentemente, los datos son: $n = 100$; $\bar{X} = 18.1$; $N_c = 1-\alpha = 0.9$; $\sigma = 0.4$, así pues, el único dato que necesitamos para aplicar la fórmula del intervalo de confianza es el valor crítico $z_{\alpha/2}$ que cumplirá en este caso que: $P\left(Z \le z_{\alpha/2}\right) = 0.95 \xrightarrow{Tabla} z_{\alpha/2} = 1.645$. Así pues, el intervalo de confianza será $I_c = \left(18.1-1.645\cdot\frac{0.4}{\sqrt{100}}, 18.1+1.645\cdot\frac{0.4}{\sqrt{100}}\right) = (18.0342\ , 18.1658)$

cuyo significado es que la media de la edad de todo el alumnado de la población se encuentra entre 18.0342 y 18.1658 con una probabilidad de 0.9, correspondiente al 90%.

Ejemplo 14: Se sabe que la dedicación media de los jóvenes al ocio sigue una distribución normal de desviación típica 63 minutos. Hallemos la media muestral y el tamaño mínimo de la muestra de jóvenes que garantiza con una probabilidad de 0.95 que el tiempo medio de ocio está entre 382 y 418 minutos.

Los datos son $\sigma = 63$; $N_c = 0.95$; $I_c = (382\ ,\ 418)$. Como la media es el punto medio del intervalo de confianza, tenemos que $\overline{X} = \dfrac{382+418}{2} = 400$. Por otra parte, el error máximo admisible será: $E = \dfrac{418-382}{2} = 18$. Además, el valor crítico correspondiente a un 0.95 de nivel de confianza viene dado por: $P(Z \le z_{\alpha/2}) = 0.975 \xrightarrow{Tabla} z_{\alpha/2} = 1.96$. Así pues, de la expresión del error máximo admisible, obtenemos la igualdad: $18 = 1.96 \cdot \dfrac{63}{\sqrt{n}}$. Sin más que despejar n, obtenemos: $n = \left(\dfrac{1.96 \cdot 63}{18}\right)^2 = 6.86^2 = 47.0596$. Así pues, el tamaño muestral mínimo es de 48 jóvenes.

Se proponen las **actividades 15 y 16**.

*Vamos a ver ahora cómo se construye el **intervalo de confianza para la proporción** en el caso de una Binomial.*

El punto de partida, como en el caso de la Normal, es que los valores críticos han de cumplir la igualdad $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$. Ahora bien, hemos visto durante la unidad que, para poblaciones grandes $(n \ge 30)$, la distribución muestral de proporciones sigue una normal $\hat{p} \to N\left(p, \sqrt{\dfrac{p \cdot q}{n}}\right)$. Como no disponemos de p, se estima puntualmente por la proporción muestral $\hat{p}$. Entonces, la expresión es equivalente a:

$$P\left(-z_{\alpha/2} \le \dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}} \le z_{\alpha/2}\right) = 1-\alpha \quad \Leftrightarrow \quad P\left(-z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} \le \hat{p}-p \le z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}\right) = 1-\alpha \quad \overset{Simetría}{\Leftrightarrow}$$

$$\Leftrightarrow P\left(-z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} \le p-\hat{p} \le z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}\right) = 1-\alpha \Leftrightarrow P\left(\hat{p}-z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} \le p \le \hat{p}+z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}\right) = 1-\alpha$$

Así pues, el intervalo de confianza para la proporción con un nivel de confianza $N_c = 1-\alpha$, viene dado por la fórmula: $\boxed{I_c = \left(\hat{p}-z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}\ ,\ \hat{p}+z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}\right)}$ que será la que usemos directamente en las actividades.

Nota 17: Como es evidente, el **valor crítico** $z_{\alpha/2}$, asociado al nivel de confianza $1-\alpha$ verifica que $P\left(Z \le z_{\alpha/2}\right) = 1 - \frac{\alpha}{2} = \frac{1+N_c}{2}$. No hay más que usar la tabla de la Normal para determinar el valor crítico $z_{\alpha/2}$ que necesitamos para determinar el intervalo de confianza.

Nota 18: Según hemos visto en la construcción del intervalo de confianza, la diferencia máxima entre la proporción muestral y la poblacional, al que ya antes hemos llamado **error máximo admisible**, es $\boxed{E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}}$ y es la mitad de la **amplitud del intervalo** $\boxed{A = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}}$. De la expresión del error, podemos despejar el **tamaño muestral** $\boxed{n = \frac{z_{\alpha/2}{}^2 \cdot \hat{p} \cdot \hat{q}}{E^2}}$

Nota 19: Conviene tener en cuenta que, tanto en el caso de la normal, como en el de la binomial:

a) Manteniendo el mismo nivel de confianza y tamaño muestral, a medida que aumenta la desviación típica (en el caso normal) o el producto $\hat{p} \cdot \hat{q}$ (en el caso binomial), aumenta el error máximo admisible y la amplitud del intervalo de confianza; y viceversa.
b) Manteniendo la misma desviación típica (en el caso normal) o el producto $\hat{p} \cdot \hat{q}$ (en el caso binomial) y tamaño muestral, a medida que aumenta el nivel de confianza, aumenta el error máximo admisible y la amplitud del intervalo de confianza; y viceversa.
c) Manteniendo el mismo nivel de confianza y desviación típica (en el caso normal) o el producto $\hat{p} \cdot \hat{q}$ (en el caso binomial), a medida que aumenta el tamaño muestral, disminuye el error máximo admisible y la amplitud del intervalo de confianza; y viceversa.

Si lo pensamos detenidamente es lógico y natural lo que hemos expresado, ya que:

a) Si aumenta la desviación, la muestra es más dispersa y, al estar "menos concentrada" en torno a la media, es lógico que nos equivoquemos más al hacer inferencia sobre el comportamiento de la población. Lo contrario ocurre si disminuye, ya que al estar "más concentrada", se comete menos error.

b) Si aumenta el nivel de confianza lo que hacemos es aumentar la seguridad con que queremos que esté el parámetro en un intervalo. Eso hace que dicho intervalo sea menos preciso y por tanto, con mayor error de equivocarnos. Lo contrario ocurre si se disminuye "la confianza" ya que lo volvemos "más exigente" y, por tanto, con menos error.

c) Si aumenta el tamaño, la muestra "se parecerá más a la población" y, por tanto, es lógico que se cometa menor error. Lo contrario ocurre al disminuir el tamaño muestral, ya que, al ser menos representativa de la población, aumenta el error.

Ejemplo 15: Se selecciona una muestra de 500 alumnos/as de ESO y se les pregunta si tienen smartphone con conexión de datos, contestando afirmativamente 225. Veamos cuál es el intervalo de confianza para la proporción de alumnos/as que tienen smartphone con conexión de datos con un nivel de confianza del 95%

Evidentemente, los datos son: $n = 500$; $\hat{p} = \dfrac{225}{500} = 0.45$; $N_c = 1 - \alpha = 0.95$, así pues, el único dato que necesitamos para aplicar la fórmula del intervalo de confianza es el valor crítico $z_{\alpha/2}$ que cumplirá en este caso que: $P\left(Z \le z_{\alpha/2}\right) = 0.95 \xrightarrow{Tabla} z_{\alpha/2} = 1.645$. Así pues, el intervalo de confianza será $I_c = \left(0.45 - 1.96 \cdot \sqrt{\dfrac{0.45 \cdot 0.55}{500}}, 0.45 + 1.96 \cdot \sqrt{\dfrac{0.45 \cdot 0.55}{500}}\right) = \left(0.4064\ ,\ 0.4936\right)$, cuyo significado es que la proporción de todo el alumnado de la población que tiene smartphone con datos se encuentra entre 0.4064 y 0.4936 con una probabilidad de 0.95, correspondiente al 95%, o dicho de otra manera que hay entre un 40.64% y un 49.36% de alumnado que tiene smartphone con datos con una probabilidad del 95%.

Ejemplo 16: Deseamos conocer el número de niños que hay que incluir, como mínimo, en una encuesta sobre la adecuada alimentación en el desayuno, sabiendo que, con un nivel de confianza del 99.73%, ha resultado el intervalo de confianza para la proporción de niños que lleva a cabo una adecuada alimentación en el desayuno el siguiente: $\left(0.42\ ,\ 0.50\right)$

Los datos son $N_c = 0.9973$; $I_c = \left(0.42\ ,\ 0.50\right)$. Como la proporción muestral es el punto medio del intervalo de confianza, tenemos que $\overline{X} = \dfrac{0.42 + 0.50}{2} = 0.46$. Por otra parte, el error máximo admisible será: $E = \dfrac{0.50 - 0.42}{2} = 0.04$. Además, el valor crítico correspondiente a un 0.9973 de nivel de confianza viene dado por: $P\left(Z \le z_{\alpha/2}\right) = 0.9973 \xrightarrow{Tabla} z_{\alpha/2} = 3$. Así pues, de la expresión del error máximo admisible, obtenemos la igualdad: $0.04 = 3 \cdot \sqrt{\dfrac{0.46 \cdot 0.54}{n}}$. Sin más que despejar n, obtenemos: $n = \dfrac{3^2 \cdot 0.46 \cdot 0.54}{0.04^2} = 1397.25$. Así pues, el tamaño muestral mínimo es de 1398 niños.

Se proponen las **actividades 17 y 18**.

7.- CONTRASTES DE HIPÓTESIS.

El último paso en el estudio de la Inferencia Estadística que vamos a abordar es el llamado Contraste de Hipótesis, cuyo principal objetivo es tomar decisiones sobre si determinadas hipótesis o supuestos a partir de muestras, pueden extrapolarse a la población con un determinado nivel de confianza.

Definición 15: Un **contraste o test de hipótesis** es el procedimiento estadístico mediante el cual se investiga la veracidad o falsedad de una hipótesis acerca de algún parámetro poblacional. Llamaremos **hipótesis nula** H_0 a la hipótesis que se formula y que se desea contrastar, y llamaremos **hipótesis alternativa** H_1 a cualquiera otra situación que sea contraria a la hipótesis nula.

<u>Definición 16:</u> Se llaman **pruebas de hipótesis** a los procedimientos que permiten decidir si una hipótesis se acepta o rechaza o determinar si las muestras observadas difieren significativamente de los resultados esperados.

<u>Nota 20:</u> Hemos de tener en cuenta algunas consideraciones que tendremos para esta unidad.

a) En esta unidad nos centraremos en hipótesis sobre los parámetros habituales con los que hemos trabajado en la unidad: μ parar la media poblacional de variables normales o aproximables por una normal y p para la proporción poblacional en el caso binomial..

b) Abordaremos dos tipos de contrastes de hipótesis: el contraste unilateral, que implica una desigualdad en la hipótesis nula y el bilateral, que implica una igualdad. Ambos lo veremos detalladamente en lo que sigue.

c) Como en todo estudio, podemos cometer errores debidos a que la información muestral nos da a pensar en algo distinto a las hipótesis debido a que la muestra, por el motivo que sea, no es lo más representativa que debe ser de la población estudiada. Estos errores los podemos clasificar en:

	H_0 es verdadera	H_0 es falsa
Se acepta H_0	Acierto	ERROR DE TIPO II
Se rechaza H_0	ERROR DE TIPO I	Acierto

En la siguiente nota, vamos a ver la parte más importante en la práctica de este punto de la unidad, ya que vamos a describir paso a paso, las etapas en las pruebas de hipótesis.

<u>Nota 21:</u> (Etapas en las pruebas de hipótesis)

<u>ETAPA 1</u>: FORMULAR LAS HIPÓTESIS

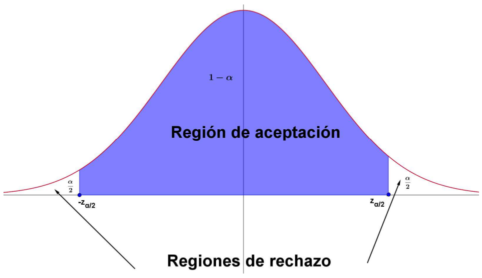
En este paso, que habitualmente nos vendrá dado en las actividades, se deben formular las hipótesis nula y alternativa que serán objeto del contraste. Ambas hipótesis, que como hemos visto deben ser excluyentes, pueden ser enunciadas para un contraste unilateral y bilateral. Los casos que abordaremos en esta unidad son:

	Contraste bilateral	Contraste unilateral	
		Derecho	Izquierdo
Media (μ)	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$
Proporción (p)	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p \leq p_0$ $H_1: p > p_0$	$H_0: p \geq p_0$ $H_1: p < p_0$

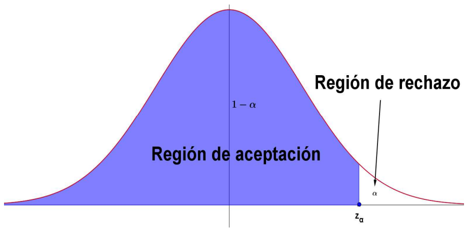
<u>ETAPA 2</u>: DETERMINAR LAS REGIONES DE ACEPTACIÓN Y RECHAZO (O CRÍTICA)

A partir de cierto **nivel de significación** $N_s = \alpha$, que es un valor muy pequeño correspondiente a la probabilidad de rechazar la hipótesis nula siendo verdadera, es decir, es la probabilidad de cometer un error de tipo I, hemos de determinar las zonas de aceptación y rechazo, que estarán limitadas por uno o dos **valores críticos** según el contraste sea unilateral o bilateral.

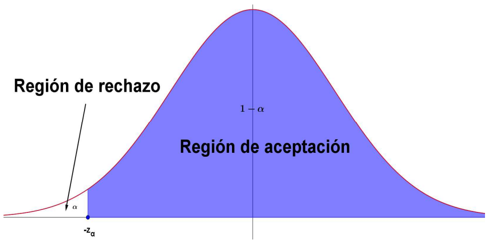
- En un **contraste bilateral**, tendremos que obtener dos valores $-z_{\alpha/2}$ y $z_{\alpha/2}$ que verifiquen $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$.



- En un **contraste unilateral derecho**, tendremos que obtener un valor z_α que verifique $P\left(Z \leq z_{\alpha/2}\right) = 1-\alpha$.



- En un **contraste unilateral izquierdo**, tendremos que obtener un valor z_α que verifique $P\left(Z \geq -z_{\alpha/2}\right) = 1-\alpha$.



<u>ETAPA 3</u>: DETERMINAR EL ESTADÍSTICO APROPIADO PARA EL CONTRASTE

Como ya hemos comentado antes, los estadísticos de contraste que vamos a utilizar son $Z = \dfrac{\bar{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ para el caso de distribuciones normales o que se aproximen a la normal y $Z = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0 \cdot q_0}{n}}}$ para el caso de distribuciones binomiales. Por lo que hemos visto ya en la unidad, sabemos que ambos siguen una distribución $N(0,1)$, podremos utilizarlo sin problemas para el contraste.

<u>ETAPA 4</u>: CALCULAR EL VALOR OBSERVADO DEL ESTADÍSTICO DE PRUEBA.

En esta etapa simplemente debemos determinar el valor z_0 correspondiente a sustituir $\bar{X}$ o $\hat{p}$ en el estadístico de prueba, según el tipo de distribución sea normal (o aproximada a la normal) o binomial.

<u>ETAPA 5</u>: TOMAR LA DECISIÓN E INTERPRETARLA.

En esta etapa, simplemente observaremos si el valor observado z_0 del estadístico de prueba pertenece a la región de aceptación (en cuyo caso se aceptará la hipótesis nula

H_0, rechazando la hipótesis alternativa H_1) o en la región de rechazo (en cuyo caso se rechazará la hipótesis nula H_0 y, por tanto, se aceptará la hipótesis alternativa H_1).

Nota 22: Una pequeña variante en el proceso del contraste de hipótesis consiste en, una vez determinada la región de aceptación para la variable Z, determinar cuál sería para la variable $\bar{X}$ *o* $\hat{p}$, según sea normal (o aproximada) o binomial. En este caso, simplemente determinaríamos la región y veríamos si el valor observado de $\bar{X}$ *o* $\hat{p}$ está dentro o fuera de dicho intervalo. De forma sencilla, utilizando los estadísticos de contraste y las regiones de aceptación y rechazo vistas en la etapa 2, obtenemos las regiones críticas ya "destipificadas" que mostramos en la tabla a continuación:

	Contraste bilateral	Contraste unilateral	
		Derecho	Izquierdo
Media (μ)	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$
REGIÓN DE ACEPTACIÓN	$\left(\mu_0 - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$	$\left(-\infty, \mu_0 + z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right)$	$\left(\mu_0 - z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}, +\infty\right)$
Proporción (p)	$H_0: p = p_0$ $H_1: p \neq p_0$	$H_0: p \leq p_0$ $H_1: p > p_0$	$H_0: p \geq p_0$ $H_1: p < p_0$
REGIÓN DE ACEPTACIÓN	$\left(p_0 - z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}, p_0 + z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right)$	$\left(-\infty, p_0 + z_{\alpha}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right)$	$\left(p_0 - z_{\alpha}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}, +\infty\right)$

Veamos, a continuación, un ejemplo de cada caso y hecho de las dos formas:

Ejemplo 17: (Contraste bilateral en distribución Normal) 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 y queremos contrastar esta hipótesis para un nivel de significación del 5%.

- 1ª FORMA:

Es evidente que el contraste de hipótesis sería: $\begin{array}{l} H_0: \mu = 171.51 \\ H_1: \mu \neq 171.51 \end{array}$

El nivel de significación es $\alpha = 0.05$. Como se trata de un contraste bilateral, hemos de hallar el valor $z_{\alpha/2}$ tal que $P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 0.95 \rightarrow P\left(Z \leq z_{\alpha/2}\right) = 0.975$.

Este valor corresponde, según la tabla de la normal, a $z_{\alpha/2} = 1.96$. Así pues, la región de aceptación correspondiente a Z sería $(-1.96, 1.96)$.

Como se trata de una distribución normal, el estadístico de contraste sería: $Z=\dfrac{\overline{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}$.

En nuestro caso: $Z=\dfrac{\overline{X}-171.51}{\dfrac{8.44}{\sqrt{16}}}=\dfrac{\overline{X}-171.51}{2.11}$

Si determinamos el valor de la media muestral: $\overline{X}=\dfrac{156+185+.......+180}{16}=172.31$.

Sustituyendo en el estadístico de contraste: $z_0=\dfrac{172.31-171.51}{2.11}\simeq 0.3791$ que es un valor dentro de la región de aceptación, ya que: $0.3791\in(-1.96\,,1.96)$, por tanto, se acepta la hipótesis nula H_0 y podemos aceptar, a un 95% que la estatura media de la población es de 171.51 m.

- 2ª FORMA:

La otra forma de enfocar el contraste es utilizando las regiones de aceptación ya "destipificadas" expresadas en la variable del problema. En nuestro caso (ver nota 21) sería: $\left(\mu_0-z_{\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}\,,\mu_0+z_{\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}\right)=\left(171.51-1.96\cdot\dfrac{8.44}{\sqrt{16}}\,,171.51+1.96\cdot\dfrac{8.44}{\sqrt{16}}\right)=(167.37\,,175.65)$

Como el valor de la media muestral calculado: $172.31\in(167.37\,,175.65)$, se acepta la hipótesis nula H_0 y podemos aceptar, a un 95% que la estatura media de la población es de 171.51 m.

Ejemplo 18: (Contraste unilateral en distribución Normal) 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?

- 1ª FORMA:

Los datos de la muestra nos hacen pensar que el tiempo de empleo puede ser menor que 6 años (ya que la media obtenida es de 5 años por ello planteamos el siguiente contraste de hipótesis: $\begin{array}{l} H_0: \mu\le 6 \\ H_1: \mu>6 \end{array}$

El nivel de significación es $\alpha=0.05$.

Como se trata de un contraste unilateral derecho, hemos de hallar el valor z_α tal que $P(Z\le z_\alpha)=0.95$. Este valor corresponde, según la tabla de la normal, a $z_{\alpha/2}=1.645$.

Así pues, la región de aceptación correspondiente a Z sería $(-\infty\,,1.645)$.

Como se trata de una distribución normal, el estadístico de contraste sería: $Z = \dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$.

En nuestro caso: $Z = \dfrac{\overline{X} - 6}{\dfrac{4}{\sqrt{64}}} = \dfrac{\overline{X} - 6}{0.5}$

Sustituyendo en el estadístico de contraste: $z_0 = \dfrac{5-6}{0.5} = -2$ que es un valor dentro de la región de aceptación, ya que: $-2 \in (-\infty\ , 1.645)$, por tanto, se acepta la hipótesis nula H_0 y podemos aceptar, a un 95% que el tiempo medio de empleo de la población está por debajo de 6 años.

- <u>2ª FORMA:</u>

La otra forma de enfocar el contraste es utilizando las regiones de aceptación ya "destipificadas" expresadas en la variable del problema. En nuestro caso (ver nota 21) sería: $\left(-\infty\ , \mu_0 + z_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}\right) = \left(-\infty\ , 6 + 1.645 \cdot \dfrac{4}{\sqrt{64}}\right) = (-\infty\ , 6.8225)$

Como el valor de la media muestral calculado: $5 \in (-\infty\ , 6.8225)$, se acepta la hipótesis nula H_0 y podemos aceptar, a un 95% que el tiempo medio de empleo de la población está por debajo de 6 años.

<u>Ejemplo 19:</u> (Contraste bilateral en una distribución Binomial) 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?

- <u>1ª FORMA:</u>

Como la moneda no trucada tiene probabilidad de salir cara 0.5, es evidente que el contraste de hipótesis sería: $\begin{matrix} H_0 : p = 0.5 \\ H_1 : p \neq 0.5 \end{matrix}$

El nivel de significación es $\alpha = 0.04$. Como se trata de un contraste bilateral, hemos de hallar el valor $z_{\alpha/2}$ tal que $P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 0.96 \rightarrow P\left(Z \le z_{\alpha/2}\right) = 0.98$. Este valor corresponde, según la tabla de la normal, a $z_{\alpha/2} = 2.05$. Así pues, la región de aceptación correspondiente a Z sería $(-2.05\ , 2.05)$.

Como se trata de una distribución binomial, el estadístico de contraste sería: $Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 \cdot q_0}{n}}}$.

En nuestro caso: $Z = \dfrac{\hat{p} - 0.5}{\sqrt{\dfrac{0.5 \cdot 0.5}{5000}}} \simeq \dfrac{\hat{p} - 0.5}{0.0071}$

Si determinamos el valor de la proporción muestral: $\hat{p} = \dfrac{3000}{5000} = 0.6$.

Sustituyendo en el estadístico de contraste: $z_0 = \dfrac{0.6-0.5}{0.0071} \simeq 14$ que es un valor fuera de la región de aceptación, ya que: $14 \notin (-2.05\ ,\ 2.05)$, por tanto, se rechaza la hipótesis nula H_0 y podemos aceptar, a un 96% que la moneda está trucada.

- 2ª FORMA:

La otra forma de enfocar el contraste es utilizando las regiones de aceptación ya "destipificadas" expresadas en la variable del problema. En nuestro caso (ver nota 21) es:

$$\left(p_0 - z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\ ,\ p_0 + z_{\alpha/2}\cdot\sqrt{\frac{p_0\cdot q_0}{n}}\right) = \left(0.5-2.05\cdot\sqrt{\frac{0.5\cdot 0.5}{5000}}\ ,\ 0.5+2.05\cdot\sqrt{\frac{0.5\cdot 0.5}{5000}}\right) = (0.4855\ ,\ 0.5145)$$

Como el valor de la proporción muestral calculado: $0.6 \notin (0.4855\ ,\ 0.5145)$, se rechaza la hipótesis nula H_0 y podemos aceptar, a un 96% que la moneda está trucada.

Ejemplo 20: (Contraste unilateral en una distribución Binomial) Un experto, basado en los anteriores comicios, 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?

- 1ª FORMA:

Como la proporción muestral es: $\hat{p} = \dfrac{800}{1500} \simeq 0.5333$, parece lógico pensar que el supuesto del experto sea cierta, por ello planteamos el contraste: $\begin{matrix} H_0 : p \geq 0.48 \\ H_1 : p < 0.48 \end{matrix}$

El nivel de significación es $\alpha = 1-0.99 = 0.01$. Como se trata de un contraste unilateral izquierdo, hemos de hallar el valor z_α tal que $P(Z \geq -z_\alpha) = 0.99 \rightarrow P(Z \leq z_\alpha) = 0.99$. Este valor corresponde, según la tabla de la normal, a $z_\alpha = 2.33$. Así pues, la región de aceptación correspondiente a Z sería $(-2.33\ ,\ +\infty)$.

Como se trata de una distribución binomial, el estadístico de contraste sería: $Z = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0\cdot q_0}{n}}}$.

En nuestro caso: $Z = \dfrac{\hat{p}-0.48}{\sqrt{\dfrac{0.48\cdot 0.52}{1500}}} \simeq \dfrac{\hat{p}-0.48}{0.0129}$

Sustituyendo la proporción muestral en el estadístico de contraste: $z_0 = \dfrac{0.5333-0.48}{0.0129} \simeq 4.1318$ que es un valor dentro de la región de aceptación, ya que: $4.1318 \in (-2.33\ ,\ +\infty)$, por tanto, se acepta la hipótesis nula H_0 y podemos aceptar, a un 99% que el experto tiene razón.

- 2ª FORMA:

La otra forma de enfocar el contraste es utilizando las regiones de aceptación ya "destipificadas" expresadas en la variable del problema. En nuestro caso (ver nota 21) es:

$$\left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\ ,\ +\infty \right) = \left(0.48 - 2.33 \cdot \sqrt{\frac{0.48 \cdot 0.52}{1500}} \right) = (0.4499\ ,\ +\infty)$$

Como el valor de la proporción muestral calculado: $0.5333 \in (0.4499\ ,\ +\infty)$, se acepta la hipótesis nula H_0 y podemos aceptar, a un 99% que el experto tiene razón.

Se proponen las **actividades de la 19 a la 21.**

8.- ACTIVIDADES.

ACTIVIDADES INTERCALADAS EN LA TEORÍA

Actividad 1: Las edades de 70 niños/as escogidos durante el recreo de infantil en un colegio son las que aparecen en la siguiente tabla:

Edad	3	4	5	6	7
Nº niños/as	15	21	28	4	2

a) Halla la edad media de los niños/as. b) Determina la desviación típica.

Actividad 2: Supongamos que estudios hechos, determinan que aprueban Matemáticas de 2º de Bachillerato un 70%. Si escogemos un grupo de 8 alumnos/as, calcula la probabilidad de que:

a) Aprueben 5 alumnos/as. b) Aprueben más de 2 alumnos/as.

Actividad 3: En una clase hay un 67% de alumnos/as que estudian inglés y el resto estudia francés. Si tomamos un grupo de 15 alumnos/as de la clase, determina la probabilidad de que :

a) Haya al menos tres alumnos de inglés.
b) Los 15 alumnos sean de inglés.
c) Haya entre 7 y 10 alumnos de inglés.
d) ¿Cuál es el número medio de alumnos de inglés?

Actividad 4: Las estaturas de 600 soldados se distribuyen de acuerdo con una normal de media 168 cm y desviación típica 8 cm. Determina redondeando a unidades:

a) Cuántos soldados miden más de 161 cm
b) Cuántos miden menos de 165 cm
c) Cuántos miden entre 166 cm y 170 cm.

Actividad 5: Se ha aplicado a 300 alumnos/as de 4º de ESO un test de agresividad y se ha observado que se distribuye normalmente con media 30 y varianza 144. Se consideran potencialmente agresivos si obtienen una puntuación superior a 42. Se pide:

a) ¿Qué porcentaje de alumnos tendrá una puntuación en dicho test de entre 20 y 35?
b) ¿Cuántos/as alumnos/as se espera que sean potencialmente agresivos?

Actividad 6: En un examen de oposiciones al que se presentan 4520 aspirantes, sólo hay plaza para 382 y los 530 restantes formarán una bolsa de trabajo para hacer las sustituciones que surjan. Se sabe que la nota de esta oposición sigue una distribución normal de media 6.58 y desviación típica 2.32.

a) ¿A partir de qué nota se aprueba la oposición obteniendo plaza?
b) ¿Entre qué notas se entra en la bolsa pero sin obtener la plaza?

Actividad 7: Se sabe, por una estadística sociológica realizada recientemente, que el nivel de aceptación de un determinado partido político es del 25% de la población. De una muestra aleatoria de 40 personas, se desea saber cuál es la probabilidad de que 15 de ellas acepten a dicho partido. Halla esta probabilidad.

Actividad 8: Después de realizar varios sondeos sobre una población con problemas socioeconómicos, se ha conseguido averiguar que únicamente un 15% de la misma es favorable a los tratamientos de psicoterapia. Elegida una muestra de 50 personas de dicha población, calcula:

a) La probabilidad de que haya más de cinco personas favorables a dichos tratamientos.
b) La probabilidad de que, a lo sumo, haya seis personas favorables a los tratamientos.

Actividad 9: En cierta población habitan 1500 niños y jóvenes, 7500 adultos y 1000 ancianos. Se desea realizar un estudio para conocer el tipo de actividades de ocio que se desean incluir en un nuevo parque en construcción. Para ello, van a ser encuestados 200 individuos elegidos al azar mediante muestreo aleatorio estratificado.

a) ¿Cuál sería el tamaño de cada muestra si se llevase a cabo con afijación igual?
b) ¿Y si fuese con afijación proporcional?

Actividad 10: El gasto medio bimensual en electricidad por familia en España se distribuye según una ley normal de media 142,32 € y desviación típica de 8,5 €.

a) Halla la probabilidad de que una muestra de 25 familias, elegidas al azar, tenga un gasto medio en electricidad superior a 144,6 €.
b) Halla la probabilidad de que una muestra de 100 familias, elegidas al azar, tenga un gasto medio en electricidad superior a 144,6 €.

Actividad 11: Una fábrica de chocolate ha fabricado 800 chocolatinas con un peso medio de 150 gramos y una desviación típica de 20 gramos. Calcula la probabilidad de que una muestra de 80 chocolatinas, elegidas al azar entre las fabricadas, tenga un peso total de 12 kilos y 400 gramos.

Actividad 12: Los pesos (en gramos) de los tornillos que fabrica una máquina se distribuyen según una variable $N(142.32, 8.5)$. Se toman muestras de 25 tornillos. Calcula la probabilidad de que una muestra elegida al azar tenga un peso medio mayor de 144.6 gramos.

Actividad 13: De una población de 120 alumnos/as, hay 48 que tienen 2 o más hermanos/as. Si de dicha población se toma una muestra al azar de tamaño 40, ¿cuál es la probabilidad de que haya más de un 55% de alumnos con 2 o más hermanos?

Actividad 14: 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?

Actividad 15: Se ha extraído una muestra de 145 alumnos/as de una escuela de artes, a los que se les ha propuesto un test de habilidad. La media y la desviación típica obtenida de la muestra son 82 y 14 respectivamente. A partir de estos datos, calcula el intervalo de confianza en el cual se hallará la media de la población al nivel de confianza del 95%.

Actividad 16: Un psicólogo quiere medir el tiempo de reacción de sus pacientes y para ello toma una muestra de 175 pacientes y realiza una estimación con un nivel de confianza del 99%. Sabiendo que la desviación típica es de 0.05 segundos, ¿qué error máximo ha cometido?

Actividad 17: Para estimar la proporción de estudiantes de una universidad a favor de la reinserción social del delincuente, se entrevistó aleatoriamente a 500 estudiantes. El 58% estaba a favor. Halla el intervalo de confianza, a un nivel del 95%, en el cual se hallará la población universitaria que se encuentra a favor.

Actividad 18: Un psicólogo quiere medir el tiempo de reacción de sus pacientes y para ello toma una muestra de 175 pacientes y realiza la estimación con un nivel de confianza del 99%, obteniendo una proporción del 72%. ¿Qué error máximo ha cometido?

Actividad 19: Una muestra representativa de tamaño 16 procede de una población normal de desviación típica 0.25, y se desea contrastar la hipótesis $H_0 : \mu = 1$ frente a $H_1 : \mu \neq 1$, sabiendo que la media muestral es 1,15:

a) A un nivel de confianza del 95%. b) A un nivel de confianza del 99%.

Actividad 20: Según la ley electoral de un país, para obtener representación parlamentaria, un partido político debe conseguir en las elecciones, al menos un 5% de votos. Próximas a celebrarse tales elecciones, una encuesta realizada sobre 1000 ciudadanos elegidos al azar revela que 36 de ellos votarán al partido A. ¿Puede estimarse, con un nivel de significación del 5%, que el partido A tendrá representación parlamentaria? ¿Y con un nivel de significación del 1%?

Actividad 21: El tiempo necesario para montar una pieza es una variable aleatoria normal de media desconocida y desviación típica 0.6 minutos. Se toma una muestra formada por 20 piezas, dando un tiempo media de 10.2 minutos. ¿Existe alguna razón para creer que el tiempo medio de montaje de la pieza es de más de 10 minutos con un nivel de confianza del 95%?

ACTIVIDADES DE DESARROLLO

Actividad 22: La estatura de 600 estudiantes sigue una ley normal con media 175 cm y desviación típica 5 cm. Halla el número de estudiantes con estatura entre 172 y 180 cm.

Actividad 23: Los 1400 alumnos/as de un centro se reparten así: 426 de 1º, 359 de 2º, 267 de 3º, 133 de 4º y 115 de 5º. ¿Cómo se elegiría una muestra de 100 alumnos mediante muestreo aleatorio estratificado con afijación proporcional?

Actividad 24: Las bolsas de azúcar envasadas por una cierta máquina tienen una media de 500 g y desviación típica 35 g. Las bolsas se empaquetan en cajas de 100 unidades. Calcula la probabilidad de que la media de los pesos de las bolsas de un paquete sea menor que 495 g.

Actividad 25: Deseamos valorar el grado de conocimientos en historia de una población de varios miles de alumnos/as. Sabemos que su varianza es de 5.29. Nos proponemos estimar μ pasando una prueba a 100 alumnos/as, de la que hemos obtenido una media de 6.32. Halla el intervalo de confianza para la media poblacional con un nivel de confianza del 95%.

Actividad 26: La desviación típica de los resultados de las distintas mediciones que se realizan para calcular la duración de un proceso es 0.5 segundos, que sigue una ley normal. ¿Cuál es el número de medidas que hay que realizar para que, con un 99% de confianza, el error de la estimación no exceda de 0.1 s?

Actividad 27: Al medir el tiempo de reacción, un psicólogo sabe que la desviación típica de las distintas mediciones del mismo es de 0.5 segundos. Se desea estimar el tiempo medio de reacción con un error máximo de 0.1 segundos, para lo cual, realiza 100 mediciones. Determina con qué nivel de confianza podrá dar el tiempo.

Actividad 28: Tomada una muestra de 300 personas mayores de 15 años en una gran ciudad, se encontró que 104 de ellas leían algún periódico regularmente. Halla, con un nivel de confianza del 90%, un intervalo para estimar la proporción de lectores de periódicos entre los mayores de 15 años.

Actividad 29: Se lanza una moneda 100 veces obteniéndose 62 caras. Pretendemos estimar la probabilidad de obtener cara con un error menor que 0.002 y un nivel de confianza del 95%. ¿Cuántas veces habremos de lanzar la moneda?

Actividad 30: Se cree que el cociente intelectual medio de los estudiantes de una universidad es 113 con una desviación típica de 7. Para contrastar la hipótesis, se extrae una muestra de 180 estudiantes y se obtiene un cociente intelectual medio de 115. ¿Podemos aceptarla hipótesis con un nivel de significación del 5%?

Actividad 31: El peso de los pollos de una granja es normal con media 2,6 kg y desviación típica 0.5. Se experimenta con un nuevo tipo de alimentación con 50 crías. Cuando se hacen adultos, se les pesa y se obtiene una media de 2,78 kg. Contrasta, con un nivel de significación del 1% la hipótesis de que el peso medio de los pollos de la población no aumenta.

Actividad 32: En las últimas elecciones, el 53% de los votantes de un pueblo estaban a favor del alcalde. Se acaba de realizar una encuesta a 360 personas elegidas al azar y 176 de ellas están a favor del alcalde. ¿Se puede afirmar, con un nivel de confianza del 90%, que el alcalde no pierde popularidad?

ACTIVIDADES DE SELECTIVIDAD

Actividad 33: (2013) Se quiere estimar la proporción de hembras entre los peces de una piscifactoría; para ello se ha tomado una muestra aleatoria de 500 peces, y en ella hay 175 hembras.

a) Calcula un intervalo de confianza para la proporción de hembras en esta población de peces, con un nivel de confianza del 94%.
b) A la vista del resultado del muestreo se quiere repetir la experiencia para conseguir un intervalo de confianza con el mismo nivel y un error máximo de 0.02, ¿cuál es el tamaño mínimo que debe tener la nueva muestra?

Actividad 34: (2013) El tiempo que los españoles dedican a ver la televisión los domingos es una variable aleatoria que sigue una distribución Normal de media desconocida y desviación típica 75 minutos. Elegida una muestra aleatoria de españoles se ha obtenido, para la media de esa distribución, el intervalo de confianza (188’18 , 208’82), con un nivel del 99%.

a) Calcula la media muestral y el tamaño de la muestra.
b) Calcula el error máximo permitido si se hubiese utilizado una muestra de 500 y un nivel de confianza del 96%.

Actividad 35: (2013) En una población próxima a un puerto deportivo se quiere estimar la proporción de habitantes que navegan al menos una vez a la semana. Se toma una muestra, al azar, de 400 habitantes de la población, de los que 160 afirman navegar al menos una vez en semana.

a) Halla el intervalo de confianza del 90% para la proporción de habitantes que navegan al menos una vez en semana.
b) A la vista del resultado, se pretende repetir la experiencia para conseguir una cota del error de 0’1 con el mismo nivel de confianza del apartado anterior. ¿Cuántos individuos debe tener al menos la muestra?

Actividad 36: (2013) Queremos estudiar la proporción de personas de una población que acceden a internet a través de teléfono móvil. Para ello hacemos una encuesta a una muestra aleatoria de 400 personas de esa población, y obtenemos que 240 de ellas acceden a internet a través del móvil.

a) Determina un intervalo de confianza, al 98.5%, para la proporción de personas de esa población que acceden a internet a través del teléfono móvil.
b) Razona el efecto que tendría sobre la amplitud del intervalo de confianza el aumento o disminución del tamaño de la muestra, suponiendo que se mantuvieran la misma proporción muestral y el mismo nivel de confianza.

Actividad 37: (2013)

a) Una población de 6000 personas se ha dividido en 3 estratos, uno con 1000 personas, otro con 3500 y otro con 1500. En esa población se ha realizado un muestreo estratificado con afijación proporcional, en el que se han elegido al azar 15 personas del tercer estrato. Determina el tamaño de la muestra total obtenida con este muestreo y su composición.

b) Dada la población {1,4,7}, construye todas las muestras posibles de tamaño 2 que puedan formarse mediante muestreo aleatorio simple, y halla la varianza de las medias muestrales de todas esas muestras.

Actividad 38: (2013) El peso de los sobres de café que fabrica una empresa sigue una ley Normal de media desconocida y desviación típica 0.3 g. Se quiere construir un intervalo de confianza para estimar dicha media, con un nivel de confianza del 98%, y para ello se toma una muestra de 9 sobres.

a) ¿Qué amplitud tendrá dicho intervalo?
b) ¿Cómo afectaría a dicha amplitud un aumento del tamaño de la muestra, manteniendo el mismo nivel de confianza?
c) Obtén el intervalo de confianza sabiendo que los pesos, en gramos, de los sobres de la muestra son: 7 , 7.1 , 7 , 6.93 ,7.02 ,7 , 7.01 , 6.5 , 7.1

Actividad 39: (2013) Se conoce que la acidez de una solución es una variable aleatoria que sigue una distribución Normal con desviación típica 0.2. Se ha tomado una muestra aleatoria de cinco soluciones y se han obtenido las siguientes medidas de la acidez: 7.92 , 7.95 , 7.91 , 7.9 , 7.94.

a) Halla el intervalo de confianza, al 99%, para la media poblacional.
b) ¿Qué error máximo se ha cometido en el intervalo anterior?
c) Para el mismo nivel de confianza, calcula el tamaño mínimo muestral que permita reducir el error anterior a la mitad.

Actividad 40: (2013)

a) Se considera la población $\{2,4,6\}$. Escribe todas las posibles muestras de tamaño dos elegidas mediante muestreo aleatorio simple y determina la desviación típica de las medias muestrales.
b) En una ciudad se seleccionó una muestra aleatoria de 500 alumnos de Bachillerato a los que se les preguntó si poseían una determinada marca de teléfono móvil, resultando que 80 de ellos contestaron afirmativamente. Obtén un intervalo de confianza, al 92%, para estimar la proporción de estudiantes de Bachillerato que poseen esa marca de teléfono móvil.

Actividad 41: (2013) El gasto mensual de las familias de un municipio se distribuye según una variable Normal con desviación típica igual a 180 euros. Seleccionadas 30 familias al azar, han tenido un gasto medio mensual de 900 euros.

a) Calcula un intervalo de confianza para el gasto medio mensual de las familias de ese municipio con un nivel de confianza del 98%.
b) Calcula el tamaño muestral mínimo necesario para estimar el medio mensual de las familias con un error no superior a 60 euros, con el mismo nivel de confianza.

Actividad 42: (2013) Un director sanitario sostiene que el índice de Masa Corporal (IMC) media de los adolescentes de su distrito no supera el nivel 25 (sobrepeso). Para contrastar su afirmación toma una muestra aleatoria de 225 adolescentes que da como resultado un IMC medio de 26. Sabiendo que el IMC sigue una distribución Normal con desviación típica 5 discuta, mediante un contraste de hipótesis con $H_0 : \mu \leq 25$, si la afirmación del director sanitario es correcta, con un nivel de significación del 5%.

Actividad 43: (2013) Los representantes de un partido político creen que la proporción de sus votantes será al menos del 35%. Para confirmarlo eligen una muestra al azar de 1200 votantes y obtienen que 336 de ellos son partidarios de votarles. Mediante un contraste de hipótesis, con $H_0 : p \geq 0.35$, y a un nivel de significación del 0.01, ¿se puede admitir como cierta la creencia de los representantes del partido político?

Actividad 44: (2013) En una bodega utilizan una maquina que debe envasar el vino en botellas con un contenido de 750 ml. Para comprobar si esa máquina funciona correctamente, se toma una muestra de 36 botellas y se observa que el contenido medio de las mismas es de 748 ml. Suponiendo que la variable "contenido" sigue una distribución Normal con varianza 25, analice mediante un contraste de hipótesis bilateral $(H_0 : \mu = 750)$ si se puede aceptar, con un nivel de significación de 0.05, que la maquina envasadora funciona correctamente.

Actividad 45: (2012) De una muestra aleatoria de 120 alumnos presentados a las Pruebas de Acceso, sólo 15 han resultado no aptos.

a) Calcula un intervalo de confianza, al 99%, para estimar la proporción de alumnos que han resultado aptos en dicha prueba.
b) Manteniendo la misma confianza, ¿cuál debe ser el tamaño mínimo de la muestra para estimar la proporción de alumnos aptos, cometiendo un error inferior al 5%?

Actividad 46: (2012) La variable “tiempo de reacción de un conductor ante un obstáculo imprevisto” sigue una distribución Normal con desviación típica 0.05 segundos. Al medir dicho tiempo en 50 conductores se ha obtenido un tiempo medio de 0.85 segundos.

a) Halla el intervalo de confianza para el tiempo medio de reacción, con un nivel de confianza del 99%.
b) ¿De qué tamaño mínimo ha de tomarse una muestra para que el error de estimación no supere 0.01 segundos, con un nivel de confianza del 95%?

Actividad 47: (2012) Una característica de una determinada población se distribuye según una variable aleatoria Normal X de media desconocida y desviación típica 0.9. Extraída al azar una muestra de tamaño 9 de esa población y observada X, dio como resultados: 10.5 , 10 , 8,5 , 10,5 , 11,5 , 13,5 , 9,5 , 13 , 12

a) Halla un intervalo de confianza, al 99%, para la media de la variable X.
b) Determina el tamaño mínimo que debe tener una muestra de esa población, para que el error máximo que se cometa en la determinación de un intervalo de confianza para la media de X sea, a lo sumo, 0.3, con un nivel de confianza del 90%.

Actividad 48: (2012) Se acepta que los rendimientos anuales, medidos en porcentajes, que producen los depósitos bancarios a plazo, se distribuyen según una ley Normal con desviación típica 1.8 y se pretende realizar una estimación del rendimiento medio de los mismos. Para ello, se tiene una muestra de 36 entidades bancarias en las que se observa que el rendimiento medio de los depósitos es del 2.5.

a) Calcula un intervalo de confianza, al 96%, para el rendimiento medio de los depósitos a plazo. ¿Cuál es el error máximo cometido en la estimación?
b) Manteniendo el mismo nivel de confianza, ¿cuál debe ser el tamaño mínimo de la muestra para estimar el rendimiento medio de los depósitos con un error máximo de 0.5?

Actividad 49: (2012)

a) En una ciudad viven 400 hombres y 320 mujeres y se quiere seleccionar una muestra de tamaño 54 utilizando muestreo estratificado por sexos, con afijación proporcional, ¿cuál sería la composición de la muestra?

b) A partir de una población de elementos 1, 2, 3, 4 se seleccionan, mediante muestreo aleatorio simple, todas las muestras de tamaño 2. Escriba dichas muestras y calcule la varianza de las medias muestrales.

Actividad 50: (2012) La velocidad a la que circulan los conductores por una autopista sigue una distribución $N(\mu,20)$. En un control efectuado a 100 conductores elegidos al azar ha resultado una velocidad media de 110 Km/h.

a) Determina el intervalo de confianza para μ , con un nivel del 99%.
b) ¿Cuál es el máximo error cometido en esta estimación?

Actividad 51: (2012) El peso de las calabazas de una determinada plantación sigue una ley Normal con desviación típica 1200 g.

a) Halla el tamaño mínimo de la muestra que se ha de elegir para, con un nivel de confianza del 95 %, estimar el peso medio con un error menor de 450 g.
b) Para el mismo nivel de confianza, indica razonando la respuesta, si el error aumenta o disminuye al aumentar el tamaño de la muestra.

Actividad 52: (2012) Se considera que, a lo sumo, el 5% de los artículos guardados en un almacén son defectuosos. Pasado un tiempo, la persona encargada del mantenimiento del almacén decide investigar si esa estimación es adecuada. Para ello, escoge aleatoriamente 300 artículos de los que 35 están defectuosos.

a) Plantea un contraste de hipótesis $(H_0 : p \leq 0{,}05)$ para determinar si ha aumentado la proporción de artículos defectuosos. Obtén la región crítica del contraste para un nivel de significación del 5%.
b) ¿Qué conclusión se obtiene con los datos muestrales observados?

Actividad 53: (2012) En una caja de ahorros se sabe que el porcentaje de los nuevos clientes que contratan un plan de pensiones no supera el 23%. El director de una de las sucursales decide hacer un regalo a cualquier nuevo cliente que contrate uno de esos planes y, tras un mes, comprueba que 110 de los 470 nuevos clientes han contratado un plan de pensiones.

a) Plantea un contraste de hipótesis, con $H_0 : p \leq 0.23$, para decidir si, con los datos dados, se puede afirmar que la medida del director ha aumentado la contratación de estos planes de pensiones. Halla la región de aceptación de este contraste de hipótesis para un nivel de significación del 5%.

b) Según el resultado del apartado anterior, ¿qué conclusión podemos obtener sobre la medida tomada por el director de esta sucursal?

Actividad 54: (2011)

a) Una población de tamaño 1000 se ha dividido en 4 estratos de tamaño 150, 400, 250 y 200. Utilizando muestreo aleatorio estratificado con afijación proporcional se han seleccionado 10 individuos del tercer estrato, ¿cuál es el tamaño de la muestra?

b) El peso de los individuos de una población se distribuye según una ley Normal de desviación típica 6 kg. Calcula el tamaño mínimo de la muestra para estimar, con un nivel de confianza del 95%, el peso medio en la población con un error no superior a 1 kg.

Actividad 55: (2011) El peso neto de las tabletas de chocolate de una determinada marca es una variable aleatoria Normal con media μ y desviación típica 7 gramos. Se sabe que 36 tabletas, elegidas al azar, han dado un peso total de 5274 gramos.

a) Calcula un intervalo con un nivel de confianza del 94% para la media μ .
b) Con el mismo nivel de confianza, ¿cuántas tabletas, como mínimo, habrá que tomar como muestra para que la amplitud del intervalo que se obtenga sea, como máximo, de 3 gramos?

Actividad 56: (2011) Se sabe que la estatura de las personas de una población es una variable aleatoria que sigue una distribución Normal cuya desviación típica es 0.04 m. Para estimar la media de esta variable se ha tomado una muestra aleatoria de 60 personas de esa población y se ha encontrado una estatura media de 1.73 m.

a) Obtén un intervalo de confianza, con un nivel del 97%, para la media de la distribución de estaturas.
b) Halla el tamaño mínimo que debe tener una muestra de esta población, para que la amplitud de un intervalo de la media con este nivel de confianza sea inferior a 0.08 m

Actividad 57: (2011) En un distrito universitario, la calificación de los alumnos sigue una distribución Normal de media 6.2 puntos y desviación típica de 1 punto. Se seleccionó, aleatoriamente, una muestra de tamaño 25.
a) Indica la distribución de la media de las muestras de tamaño 25.
b) ¿Cuál es la probabilidad de que la media de las calificaciones de los alumnos de una de esas muestras esté comprendida entre 6 y 6.6 puntos?

Actividad 58: (2011) El peso de los adultos de una determinada población sigue una distribución Normal de media 70 Kg y desviación típica 16 Kg. Si elegimos, al azar, muestras de tamaño 4,

a) ¿Cuál es la distribución de la media muestral?
b) ¿Cuál es la probabilidad de que el peso medio de una de esas muestras esté comprendido entre 65 y 72 Kg?
c) ¿Cuál es la probabilidad de que ese peso medio sea menor que 70 Kg?

Actividad 59: (2011) Con el fin de estudiar el peso medio de los perros recién nacidos de una determinada raza, se tomó una muestra en una clínica veterinaria y se obtuvieron los siguientes pesos, medidos en Kg: 1.2 0.9 1 1.2 1.1 1 0.8 1.1

Se sabe que el peso de los cachorros de esta raza se distribuye según una ley Normal con desviación típica 0.25 Kg.

a) Obtén un intervalo de confianza para estimar la media poblacional, al 95%.
b) Halla el error máximo que se cometería usando el intervalo anterior.
c) Razona cómo variaría la amplitud del intervalo de confianza si, manteniendo el mismo nivel de confianza, aumentásemos el tamaño de la muestra.

Actividad 60: (2011) Sea X una variable aleatoria Normal de media 50 y desviación típica 4. Se toman muestras de tamaño 16.

a) ¿Cuál es la distribución de la media muestral?
b) ¿Cuál es la probabilidad de que la media muestral esté comprendida entre 47.5 y 52.5?

Actividad 61: (2011) Una máquina está preparada para fabricar piezas de, a lo sumo, 10 cm de longitud. Se toma una muestra de 1000 piezas, comprobándose que la media de sus longitudes es de 10.0037 cm. La longitud de las piezas fabricadas por esa máquina sigue una ley Normal con desviación típica 0.2 cm.

a) Plantea un contraste de hipótesis unilateral para comprobar si con los datos de esa muestra es posible afirmar que la media de la longitud de las piezas fabricadas por la máquina es de más de 10 cm.
b) Determina la región de aceptación de la hipótesis nula de ese contraste para un nivel de significación $\alpha = 0.025$.
c) Con los datos de la muestra y usando el contraste de hipótesis del primer apartado, ¿qué conclusión se obtendría sobre la longitud media de las piezas fabricadas?

Actividad 62: (2011) El director de una televisión afirma que un nuevo programa que va a emitirse será visto, al menos, por un 30% de personas. Una vez emitido se realizó una encuesta a 500 personas, elegidas al azar, y ésta reveló que 130 de ellas habían visto ese programa.

a) Formula la hipótesis nula y la alternativa del contraste de hipótesis que permite determinar si los datos de la encuesta realizada son compatibles con la afirmación del director.
b) Halla la región crítica de ese contraste para un nivel de significación del 5.5%.
c) Según el dato obtenido en el apartado anterior ¿qué conclusión se obtiene sobre la afirmación realizada por el director de esa televisión?

Actividad 63: (2011) El director de un banco afirma que la cantidad media de dinero extraído, por cliente, de un cajero automático de su sucursal no supera los 120 euros. Para contrastar esta hipótesis elige al azar 100 extracciones de este cajero y obtiene una media muestral de 130 euros. Se sabe que la cantidad de dinero extraído por un cliente en un cajero automático se distribuye según una ley Normal de media desconocida y desviación típica 67 euros.

a) Plantea el contraste de hipótesis asociado al enunciado.
b) Determina la región de aceptación, para un nivel de significación $\alpha = 0.05$.
c) Con los datos muestrales tomados, ¿existe evidencia estadística para rechazar la hipótesis de este director, con el mismo nivel de significación anterior?

<u>Actividad 64:</u> (2011) Un estudio sociológico afirma que el 70% de las familias cena viendo la televisión. Se desea contrastar la veracidad de esta afirmación y, para ello, se toma una muestra de 500 familias, en las que se observa que 340 ven la televisión mientras cenan. Decide, mediante un contraste de hipótesis, si la afirmación es cierta con un nivel de significación de 0.01.

<u>Actividad 65:</u> (2011) Suponiendo que la variable "años de vida de los individuos de un país" sigue una distribución Normal con desviación típica 8.9 años, se desea contrastar la hipótesis de que la vida media de los mismos no supera los 70 años. A partir de una muestra aleatoria de 100 individuos se ha obtenido que su vida media ha sido 71.8 años.

a) Formula el contraste de hipótesis que indica el enunciado.
b) Determina la región crítica a un nivel de significación del 5%.
c) Con los datos muestrales, ¿existe evidencia estadística para rechazar la hipótesis a ese nivel de significación?

<u>Actividad 66:</u> (2010) Una empresa consultora quiere estudiar algunos aspectos de la vida laboral de los trabajadores de una ciudad. Para ello selecciona una muestra aleatoria de 500 trabajadores, de los que 118 afirman residir en otra ciudad. Con un nivel de confianza del 93%,

a) Calcula un intervalo de confianza para la proporción de trabajadores que residen fuera.
b) Calcula el error cometido en el intervalo anterior.

<u>Actividad 67:</u> (2010)

a) En una población de 2000 hombres y 2500 mujeres se quiere seleccionar una muestra de 135 personas mediante muestreo aleatorio estratificado con afijación proporcional, ¿cuál sería la composición de la muestra?

b) Dada la población $\{6, 8, 11, a\}$, ¿cuánto debe valer a sabiendo que la media de las medias muestrales de tamaño 3, obtenidas mediante muestreo aleatorio simple, es 10.3?

<u>Actividad 68:</u> (2010) De una muestra aleatoria de 350 individuos de una población, 50 son adultos.

a) Calcula un intervalo de confianza, al 98%, para la proporción de adultos de esa población.
b) ¿Puede admitirse, a ese nivel de confianza, que la proporción de adultos de esa población es $\frac{2}{15}$?

<u>Actividad 69:</u> (2010) Se desea estimar la proporción de votantes a un determinado partido político mediante una muestra aleatoria.

a) Si de una muestra de 500 personas 200 dicen que lo votan, calcula con un nivel de confianza del 97% un intervalo para la proporción de votantes a ese partido en la población.
b) Si la proporción de votantes en otra muestra ha sido 0.2 y el error cometido en la estimación ha sido inferior a 0.05, con un nivel de confianza del 99%, calcula el tamaño mínimo de dicha muestra.

Actividad 70: (2010) Se sabe que el tiempo de reacción a un determinado estímulo se distribuye según una ley Normal de media desconocida y desviación típica 0.2 segundos.

a) Observada una muestra aleatoria de tamaño 25 se ha obtenido una media muestral de 0.3 segundos. Obtén un intervalo de confianza para la media de la población con un nivel de confianza del 94%.
b) A un nivel de confianza del 90%, ¿cuál será el tamaño muestral mínimo si el error cometido es inferior a 0.05?

Actividad 71: (2010) En los individuos de una población, la concentración de una proteína en sangre se distribuye según una ley Normal de media desconocida y desviación típica 0.42 g/dl. Se toma una muestra aleatoria de 49 individuos y se obtiene una media muestral de 6.85 g/dl.

a) Obtén un intervalo de confianza, al 96%, para estimar la concentración media de la proteína en sangre de los individuos de esa población.
b) ¿Es suficiente el tamaño de esa muestra para obtener un intervalo de confianza, al 98%, con un error menor que 0.125 g/dl?

Actividad 72: (2010)

a) La altura de los alumnos de una Universidad sigue una distribución Normal de media desconocida y desviación típica 11 cm. Calcula el tamaño mínimo que ha de tener una muestra aleatoria de esos alumnos para que el error cometido al estimar la altura media sea inferior a 1 cm, con un nivel de confianza del 98%.

b) Dada la población $\{10,12,17\}$, escribe todas las muestras de tamaño 2 mediante muestreo aleatorio simple y calcula la media y la desviación típica de las medias muestrales.

Actividad 73: (2010) Un agricultor piensa que la producción media por naranjo, en su finca, es de 88 kg o más. Para confirmar su creencia selecciona, al azar, 10 de sus naranjos, pesa su producción y obtiene como resultado, en kg, para cada uno de ellos:
80 , 83 , 87 , 95 , 86 , 92 , 85 , 83 , 84 , 95. Se acepta que la producción de un naranjo sigue una distribución Normal con desviación típica 5 kg.

a) Plantea el contraste de hipótesis unilateral que responda a las condiciones del problema y determine la región crítica para un nivel de significación $\alpha = 0.05$.
b) Con los datos de esta muestra, ¿qué conclusión debe obtener el agricultor sobre la producción media por naranjo de su finca, utilizando ese mismo nivel de significación?

Actividad 74: (2010) Una máquina de envasado está diseñada para llenar bolsas con 300 g de almendras. Para comprobar si funciona correctamente, se toma una muestra de 100 bolsas y se observa que su peso medio es de 297 g. Suponiendo que la variable "peso" tiene una distribución Normal con varianza 16, y utilizando un contraste bilateral ¿es aceptable, a un nivel de significación de 0.05, que el funcionamiento de la máquina es correcto?

Actividad 75: (2010) Se sabe que los años de vida de los individuos de una población es una variable aleatoria Normal con desviación típica 8.9 años. Una muestra aleatoria de 100 individuos de esa población mostró una vida media de 71.8 años. Mediante un contraste unilateral, ¿puede afirmarse con los datos anteriores que la vida media es mayor de 70 años, a un nivel de significación $\alpha = 0.05$?

Actividad 76: (2010) El peso de los sacos de patatas de una cooperativa es una variable aleatoria Normal con desviación típica 0.25 kg. El agente de ventas de esa cooperativa afirma que el peso medio de los sacos no baja de 5 kg. Se desea contrastar estadísticamente esta hipótesis. Para ello se toma una muestra aleatoria de 20 sacos y se obtiene que su peso medio es de 4.8 kg.

a) Determina las hipótesis del contraste que se plantea en este enunciado.
b) Halla la región crítica de este contraste para $\alpha = 0.01$?
c) Con los datos de la muestra tomada, ¿puede decirse que existe evidencia estadística suficiente para rechazar la hipótesis del agente de ventas de la cooperativa, al nivel de significación $\alpha = 0.01$?

Actividad 77: (2010) En una determinada especie animal el porcentaje de mortalidad debida a una enfermedad vírica es de al menos un 40%. Se está realizando un estudio para probar la eficacia de un fármaco que permite tratar esa enfermedad y, consecuentemente, reducir el porcentaje de mortalidad en esa especie. Para ello, se suministró el fármaco a 50 sujetos enfermos, elegidos al azar, de los que murieron 14. A la vista de estos datos, y tomando como nivel de significación 0.015, ¿se puede afirmar que existe evidencia estadística suficiente para rechazar la hipótesis $H_0 : p \geq 0.4$, donde p es la proporción, y por lo tanto aceptar la eficacia del fármaco?

Actividad 78: (2009) El tiempo (en horas) que permanecen los coches en un determinado taller de reparación es una variable aleatoria con distribución Normal de desviación típica 4 horas.

a) Se eligieron, al azar, 16 coches del taller y se comprobó que, entre todos, estuvieron 136 horas en reparación. Determina un intervalo de confianza, al 98,5%, para la media del tiempo que permanecen los coches en ese taller.
b) Determina el tamaño mínimo que debe tener una muestra que permita estimar la media del tiempo que permanecen en reparación los coches en ese taller con un error no superior a una hora y media y con el mismo nivel de confianza del apartado anterior.

Actividad 79: (2009) En un estudio de mercado del automóvil en una ciudad se ha tomado una muestra aleatoria de 300 turismos, y se ha encontrado que 75 de ellos tienen motor diesel. Para un nivel de confianza del 94%.

a) Determina un intervalo de confianza de la proporción de turismos que tienen motor diesel en esa ciudad.
b) ¿Cuál es el error máximo de la estimación de la proporción?

Actividad 80: (2009) En una muestra aleatoria de 100 individuos se ha obtenido, para la edad, una media de 17.5 años. Se sabe que la edad en la población, de la que procede esa muestra, sigue una distribución Normal con una desviación típica de 0.8 años.

a) Obtén un intervalo de confianza, al 94%, para la edad media de la población.
b) ¿Qué error máximo se comete en la estimación anterior?

Actividad 81: (2009) El cociente intelectual de los alumnos de un centro educativo se distribuye según una ley Normal de media 110 y desviación típica 15. Se extrae una muestra aleatoria simple de 25 alumnos.

a) ¿Cuál es la probabilidad de que la media del cociente intelectual de los alumnos de esa muestra sea superior a 113?
b) Razona cómo se vería afectada la respuesta a la pregunta anterior si el tamaño de la muestra aumentase.

Actividad 82: (2009) Escriba todas las muestras de tamaño 2 que, mediante muestreo aleatorio simple (con reemplazamiento), se pueden extraer del conjunto $\{8,10,12\}$ y determina el valor de la varianza de las medias de esas muestras.

Actividad 83: (2009) En una empresa de gas trabajan 150 personas en mantenimiento, 450 en operaciones, 200 en servicios y 100 en cargos directivos. Con objeto de realizar una encuesta laboral, se quiere seleccionar una muestra de 180 trabajadores de esa empresa por muestreo aleatorio estratificado con afijación proporcional, ¿qué número de trabajadores se debe elegir de cada grupo?

Actividad 84: (2009) Una variable aleatoria X se distribuye de forma Normal, con media μ y desviación típica σ = 0.9.

a) Una muestra aleatoria de tamaño 9 ha proporcionado los siguientes valores de X: 7.0 , 6.4 , 8.0 , 7.1 , 7.3 , 7.4 , 5.6 , 8.8 , 7.2 . Obtén un intervalo de confianza para la media con un nivel de confianza de 97%.
b) Con otra muestra, se ha obtenido que un intervalo de confianza para μ, al 95%, es el siguiente (6.906 , 7.494). ¿Cuál es el tamaño de la muestra utilizada?

Actividad 85: (2009) Tomando, al azar, una muestra de 80 empleados de una empresa, se encontró que 20 usaban gafas. Halla, con un nivel de confianza del 90%, un intervalo de confianza para estimar la proporción de empleados de esa empresa que usan gafas.

Actividad 86: (2009) El gasto que hacen las familias españolas en regalos de Navidad sigue una ley Normal de media desconocida y desviación típica 84 euros. Para estimar esta media se seleccionó una muestra aleatoria y se obtuvo el intervalo de confianza (509.41 , 539.79), con un nivel de confianza del 97%.

a) ¿Cuál ha sido la media de la muestra escogida?
b) ¿Qué tamaño tenía la muestra?

Actividad 87: (2009) Los jóvenes andaluces duermen un número de horas diarias que se distribuye según una ley Normal de media desconocida, μ , y desviación típica 2 horas. A partir de una muestra de 64 jóvenes se ha obtenido una media de 7 horas.

a) Halla un intervalo de confianza, al 97%, para la media poblacional μ.
b) Manteniendo la misma confianza, ¿cuál debe ser el tamaño mínimo de la muestra para estimar la media de horas de sueño, cometiendo un error máximo de 0.25 horas?

Actividad 88: (2009) Se desea estimar la proporción de fumadores de una población mediante una muestra aleatoria.

a) Si la proporción de fumadores en la muestra es 0.2 y el error cometido en la estimación ha sido inferior a 0.03, con un nivel de confianza del 95%, calcula el tamaño mínimo de la muestra.
b) Si en otra muestra de tamaño 280 el porcentaje de fumadores es del 25%, determina, para un nivel de confianza del 99%, el correspondiente intervalo de confianza para la proporción de fumadores de esa población.

Actividad 89: (2009) El tiempo que se tarda en la caja de un supermercado en cobrar a los clientes sigue una ley Normal con media desconocida y desviación típica 0.5 minutos. Para una muestra aleatoria de 25 clientes se obtuvo un tiempo medio de 5.2 minutos.

a) Calcula un intervalo de confianza, al nivel del 97%, para el tiempo medio que se tarda en cobrar a los clientes.
b) Indica el tamaño muestral mínimo necesario para estimar dicho tiempo medio con un error máximo de 0.5 y un nivel de confianza del 96%.

9.- SOLUCIONES A LAS ACTIVIDADES.

Actividad 1:

a) $\overline{X} \simeq 4.39$ b) $\sigma \simeq 0.96$

Actividad 2:

a) 0.2541 b) 0.9887

Actividad 3:

a) 0.999973 b) 0.0025 c) 0.5564 d) 10.05

Actividad 4:

a) Aprox. 486 b) Aprox. 212 c) Aprox. 118

Actividad 5:

a) 45.95% b) 48 alumnos/as.

Actividad 6:

a) 9.78 b) 8.53

Actividad 7: 0.0277

Actividad 8:

a) 0.1170 b) 0.3446

Actividad 9:

a) 67 niños y jóvenes, 67 adultos y 66 ancianos

b) 30 niños y jóvenes, 150 adultos y 20 ancianos

Actividad 10:

a) 0.0901 b) 0.0037

Actividad 11: 0.0129

Actividad 12: 0.0901

Actividad 13: 0.0262

Actividad 14: 65.99%

Actividad 15: $(79.72\ ,\ 84.28)$

Actividad 16: 0.0098 segundos

Actividad 17: $(0.5408\ ,\ 0.6192)$

Actividad 18: 0.0876

Actividad 19:

a) Región de aceptación $(0.8775\ ,\ 1.1225)$. Por lo tanto, se rechaza.
b) Región de aceptación $(0.8388\ ,\ 1.1613)$. Por lo tanto, se acepta.

Actividad 20:

a) Región de aceptación $(0.0386\ ,\ +\infty)$. Por lo tanto, se rechaza.
b) Región de aceptación $(0.0339\ ,\ +\infty)$. Por lo tanto, se acepta.

Actividad 21: Región de aceptación $(10.22\ ,\ +\infty)$. Por lo tanto, se rechaza

Actividad 22: 340 estudianes.

Actividad 23: 33 de 1º, 28 de 2º, 20 de 3º, 10 de 4º y 9 de 5º.

Actividad 24: 0.0764

Actividad 25: $(5.8692\ ,\ 6.7708)$

Actividad 26: 166 medidas.

Actividad 27: Con un 39.7%

Actividad 28: $(0.302\ ,\ 0.392)$

Actividad 29: 226271 veces.

Actividad 30: La región de aceptación es $(111.98\,,114.02)$, luego se rechaza la hipótesis.

Actividad 31: La región de aceptación es $(-\infty\,,2.76)$, luego se rechaza la hipótesis y concluimos que el peso sí aumentará.

Actividad 32: La región de aceptación es $(0.496\,,+\infty)$, luego se rechaza la hipótesis y concluimos que el alcalde pierde popularidad.

Actividad 33:

a) $(0.3099,0.3901)$ b) $n = 2011$

Actividad 34:

a) $\overline{X} = 198{,}5 \;\; y \;\; n = 351$ b) $E = 6{,}89$

Actividad 35:

a) $(0.3598,0.4402)$ b) $n = 65$

Actividad 36:

a) $(0.5405,0.6595)$

b) A medida que aumenta al tamaño de la muestra, disminuye la amplitud y a medida que disminuye el tamaño, aumenta la amplitud.

Actividad 37:

a) 60 personas en total, siendo 10 del primer estrato, 35 del 2º y 15 del 3º.

b) Las muestras son: $\{1,1\},\{1,4\},\{1,7\},\{4,1\},\{4,4\},\{4,7\},\{7,1\},\{7,4\},\{7,7\} \;\; y \;\; \sigma^2_{\overline{X}} = 3$.

Actividad 38:

a) $A = 0{,}466$ b) Disminuirá. c) $(6.727,7.193)$

Actividad 39:

a) $(7.6937,8.1543)$ b) $E = 0{,}2303$ c) $n = 20$

Actividad 40:

a) $\{2,2\},\{2,4\},\{2,6\},\{4,2\},\{4,4\},\{4,6\},\{6,2\},\{6,4\},\{6,6\} \;\; y \;\; \sigma_{\overline{X}} = 1{,}15$

b) $(0.1312,0.1888)$

Actividad 41:

a) $(823.429, 976.571)$ b) $n = 49$

Actividad 42: No es correcta. Se rechaza H_0.

Actividad 43: Es falsa. Se rechaza H_0.

Actividad 44: Se rechaza H_0.

Actividad 45:

a) $(0.7973, 0.9527)$ b) $n = 291$

Actividad 46:

a) $(0.8318, 0.8682)$ b) $n = 97$

Actividad 47:

a) $(10.2275, 11.7725)$ b) $n = 25$

Actividad 48:

a) $I_c = (1.8835, 3.1165)$ y $E = 0{,}6165$ b) $n = 55$

Actividad 49:

a) 30 hombre y 24 mujeres.
b) $\{1,1\},\{1,2\},\{1,3\},\{1,4\},\{2,1\},\{2,2\},\{2,3\},\{2,4\},\{3,1\},\{3,2\},\{3,3\},\{3,4\},\{4,1\},\{4,2\},\{4,3\},\{4,4\}$ y $\sigma^2_{\overline{X}} = 0{,}625$

Actividad 50:

a) $(104.85, 115.15)$ b) $E = 5{,}15$

Actividad 51:

a) $n = 28$ b) Disminuye.

Actividad 52:

a) $(0.0706, +\infty)$ b) Se rechaza H_0 y se acepta H_1.

Actividad 53:

a) $(-\infty, 0.2619)$ b) Se acepta H_0. El director ha acertado.

Actividad 54:

a) $n = 40$ b) $n = 139$

Actividad 55:

a) $(144.30, 148.69)$ b) $n = 78$

Actividad 56:

a) $(1.7188, 1.7412)$ b) $n = 5$

Actividad 57:

a) $N(6.2, 0.2)$ b) 0,8185

Actividad 58:

a) $N(70, 8)$ b) 0,3311 c) 0,5

Actividad 59:

a) $(0.8643, 1.2107)$ b) $E = 0{,}1732$ c) Disminuye.

Actividad 60:

a) $N(50, 1)$ b) 0,9876

Actividad 61:

a) Hacer. b) $(9.98, +\infty)$ c) Se acepta H_0.

Actividad 62:

a) Hacer. b) $(0.2673, +\infty)$ c) Se rechaza H_0.

Actividad 63:

a) Hacer. b) $(-\infty, 131.02)$ c) Se acepta H_0.

Actividad 64: Es cierta.

Actividad 65:

a) Hacer. b) $(-\infty, 71.46)$ c) Se rechaza H_0.

Actividad 66:

a) $(0.202, 0.270)$ b) $E = 0{,}034$

Actividad 67:

a) 60 hombre y 75 mujeres. b) $a = 16{,}2$

Actividad 68:

a) $(0.0993, 0.1863)$ b) Sí.

Actividad 69:

a) $(0.3525, 0.4475)$ b) $n = 425$

Actividad 70:

a) $(0.2248, 0.3752)$ b) $n = 44$

Actividad 71:

a) $(6.7267, 6.9733)$ b) No es suficiente.

Actividad 72:
a) $n = 657$
b) $\{10,10\},\{10,12\},\{10,17\},\{12,10\},\{12,12\},\{12,17\},\{17,10\},\{17,12\},\{17,17\}$ *con* $\mu_{\overline{X}} = 13$ *y* $\sigma_{\overline{X}} = 2{,}08$

Actividad 73:

a) $(85.4, +\infty)$ b) Se acepta H_0.

Actividad 74: Se rechaza H_0.

Actividad 75: Se acepta H_0.

Actividad 76:

a) Hacer. b) $(4.87, +\infty)$ c) Se rechaza H_0.

Actividad 77: Se acepta H_0.

Actividad 78:

a) $(133.57, 138.43)$ b) $n = 42$

Actividad 79:

a) $(0.203, 0.297)$ b) $E = 0{,}047$

Actividad 80:

a) $(17.3492, 17.6508)$ b) $E = 0{,}1508$

Actividad 81:

a) 0,1587 b) Disminuirá.

Actividad 82: $\{8,8\},\{8,10\},\{8,12\},\{10,8\},\{10,10\},\{10,12\},\{12,8\},\{12,10\},\{12,12\}$ y $\sigma^2_{\overline{X}} = 1{,}33$

Actividad 83: 30 de mantenimiento, 90 de operaciones, 40 de servicios y 20 cargos directivos.

Actividad 84:

a) $(6.549, 7.851)$ b) $n = 36$

Actividad 85: $(0.171, 0.329)$

Actividad 86:

a) $\overline{X} = 524{,}6$ b) $n = 131$

Actividad 87:

a) $(6.4575, 7.5425)$ b) $n = 302$

Actividad 88:

a) $n = 683$ b) $(0.184, 0.316)$

Actividad 89:

a) $(4.983, 5.417)$ b) $n = 5$

NOTA IMPORTANTE: Las actividades de la 34 a la 89 son de Selectividad. En las dos páginas web siguientes se encuentran las soluciones de todos los exámenes de forma detallada:

- http://emestrada.wordpress.com/category/matematicas-aplicadas-a-las-ccss-ii/
- *http://www.iesayala.com/selectividadmatematicas/*