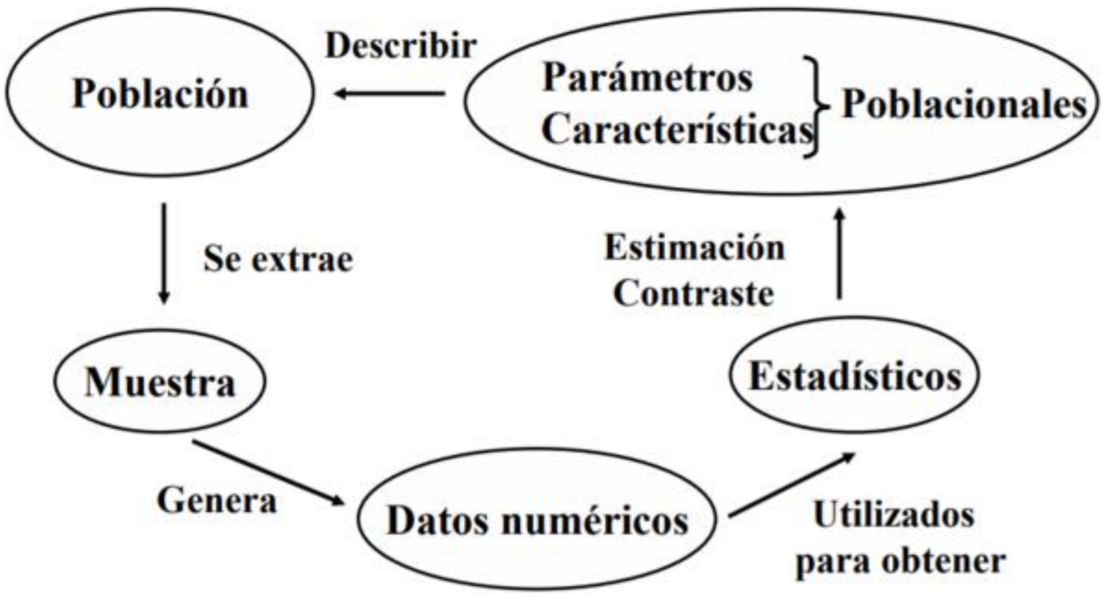
Distribuciones Muestrales



- **Población:** "Conjunto de elementos en los que se observa alguna $\underbrace{\text{característica común}}_{\text{Variables}}$"

- **Observaciones:** "Valores que toma la característica observada en cada elemento de la población"

- **Parámetro: "**Característica numérica que describe una variable observada en la población"

- **Muestra:** "Conjunto de unidades representativas de una población"

➢ **Estadístico:** "Función de los valores de la muestra"

❑ La **inferencia estadística** esta basada en el estudio de las **muestras**

❑ La muestra debe ser **representativa de la población** para extraer conclusiones validas sobre esta población

❑ La muestra debe ser **aleatoria**

Muestreo Aleatorio Simple

➢ "Cada elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra y cada muestra del mismo tamaño tiene la misma probabilidad de ser seleccionada"

Distribuciones Muestrales de la Media

muestral

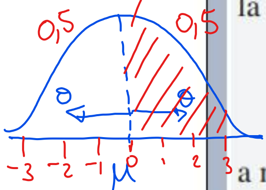
Teorema del límite central: Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n, tomada de una población con media μ y varianza finita σ^2, entonces la forma límite de la distribución de

poblacional

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

Desviación Estandar — Cantidad de Datos (muestra)

a medida que $n \to \infty$, es la distribución normal estándar $n(z; 0, 1)$.



NOTA: Los valores de Z siempre son pequeños entre 0 a 4

Ejemplos:

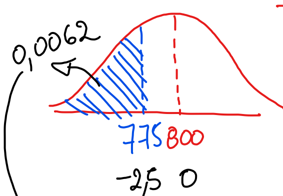
Una empresa de material eléctrico fabrica bombillas que tienen una duración que se distribuye aproximadamente en forma normal, con media de 800 horas y desviación estándar de 40 horas. Calcule la probabilidad de que una muestra aleatoria de 16 bombillas tenga una vida promedio de menos de 775 horas.

$\mu = 800$

$\bar{X} = 775$

$\theta = 40$

$n = 16$



$$Z = \frac{\bar{X} - \mu}{\frac{\theta}{\sqrt{n}}} = \frac{775 - 800}{\frac{40}{\sqrt{16}}} = -2,5$$

Busca Tabla

Area o Probabilidad $= 0,0062 \Rightarrow$ en porcentaje

$0,0062 \cdot 100 = \boxed{0,62\%}$

$P(X \leq 775) = P(Z \leq -2,5) = 0,0062$

Tipos de Muestreo

Definición Si una muestra de n elementos se selecciona de entre una población de N elementos, usando un plan muestral en el que cada una de las posibles muestras tiene la misma probabilidad de selección, entonces se dice que el muestreo es **aleatorio** y la muestra resultante es una **muestra aleatoria simple**.

Definición Un **muestreo aleatorio estratificado** comprende seleccionar una muestra aleatoria simple de cada uno de un número dado de subpoblaciones o **estratos**.

Definición Una **muestra de conglomerados** es una muestra aleatoria simple tomada de los conglomerados disponibles en la población.

Definición Una **muestra aleatoria sistemática 1 en k** involucra la selección aleatoria de uno de los primeros k elementos de una población ordenada y luego la selección sistemática de cada k-ésimo elemento de ahí en adelante.

Ejemplo

Una población está formada por $N = 5$ números: 3, 6, 9, 12, 15. Si una muestra aleatoria de tamaño $n = 3$ se selecciona sin reemplazo, encuentre las distribuciones muestrales para la media muestral $\bar{x}$ y la mediana m.

$$\bar{X} = \frac{3+6+9+12+15}{5} = 9$$

Posición

$$Me = \frac{n+1}{2} = \frac{5+1}{2} = 3$$

$$Me = 9$$

Valores de $\bar{x}$ y m para muestreo aleatorio simple cuando $n = 3$ y $N = 5$

Muestra	Valores muestrales	$\bar{x}$	m
1	3, 6, 9	~~6~~	6 ✓
2	3, 6, 12	~~7~~	6 ✓
3	3, 6, 15	~~8~~	6 ✓
4	3, 9, 12	~~8~~	9 ✓
5	3, 9, 15	~~9~~	9 ✓
6	3, 12, 15	~~10~~	12 ✓
7	6, 9, 12	~~9~~	9 ✓
8	6, 9, 15	~~10~~	9 ✓
9	6, 12, 15	11	12 ✓
10	9, 12, 15	12	12 ✓

n=10

Distribuciones muestrales para a) la media muestral y b) la mediana muestral

$\frac{f_i}{n}$

(a)

$\bar{x}$	f_i	$p(\bar{x})$
6	1	0.1
7	1	0.1
8	2	0.2
9	2	0.2
10	2	0.2
11	1	0.1
12	1	0.1
n=10		

(b)

m		$p(m)$
6	3	.3
9	4	.4
12	3	.3
	n=10	1

$$Z = \frac{\bar{X} - \mu}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{error estándar}}}$$

Definición La desviación estándar de una estadística empleada como estimador de un parámetro poblacional también se denomina **error estándar del estimador** (abreviado **SE**) porque se refiere a la precisión del estimador. Por tanto, la desviación estándar de $\bar{x}$, dada por $\sigma/\sqrt{n}$, se conoce como **error estándar de la media** (abreviada SE($\bar{x}$) o sólo SE).

$\frac{\sigma}{\sqrt{n}}$

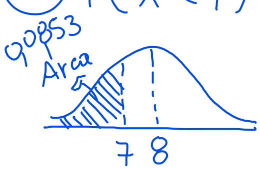
Ejemplo:

$\mu = 8$
$\bar{X} =$
$\sigma = 4$
$n = 30$

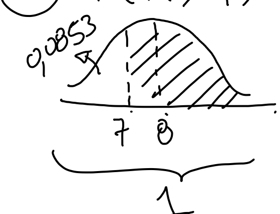
La duración de la enfermedad de Alzheimer desde el principio de los síntomas hasta el fallecimiento varía de 3 a 20 años; el promedio es 8 años con una desviación estándar de 4 años. El administrador de un gran centro médico selecciona al azar los registros médicos de 30 pacientes de Alzheimer ya fallecidos, de la base de datos del centro médico y anota la duración promedio. Encuentre las probabilidades aproximadas para estos eventos:

1. La duración promedio es menor a 7 años. ($\bar{x}$)
2. La duración promedio excede de 7 años.
3. La duración promedio está a no más de 1 año de la media poblacional $\mu = 8$.

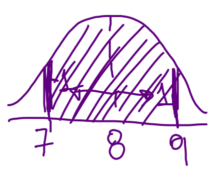
(1) $P(\bar{X} < 7) = Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7 - 8}{\frac{4}{\sqrt{30}}} = -1,37 \quad P(Z \leq -1,37) = \boxed{0,0853}$



(2) $P(\bar{X} > 7) = 1 - 0,0853 = \boxed{0,9147}$



(3)



$P(7 < \bar{X} < 9) = P(\bar{X} < 9) - P(\bar{X} < 7)$

$0,9147 - 0,0853$

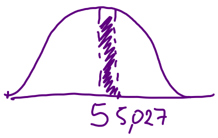
$\boxed{0,8294}$

$Z = \frac{9 - 8}{\frac{4}{\sqrt{30}}} = 1,37$

Ejemplo: $Z = \dfrac{\bar{X} - \mu}{\frac{\theta}{\sqrt{n}}}$

$\mu = 5.0$
$\bar{X} = 5{,}027$
$\theta = 0{,}1$
$n = 100$

Partes para automóviles. Un importante proceso de fabricación produce partes de componentes cilíndricos para la industria automotriz. Es importante que el proceso produzca partes que tengan un diámetro medio de 5.0 milímetros. El ingeniero implicado asume que la media de la población es de 5.0 milímetros. Se lleva a cabo un experimento donde se seleccionan al azar 100 partes elaboradas por el proceso y se mide el diámetro de cada una de ellas. Se sabe que la desviación estándar de la población es $\sigma = 0.1$ milímetros. El experimento indica un diámetro promedio muestral de $\bar{x} = 5.027$ milímetros. ¿Esta información de la muestra parece apoyar o refutar la suposición del ingeniero?



$Z = \dfrac{5{,}027 - 5}{\frac{0{,}1}{\sqrt{100}}} = 2{,}7$ Tabla $0{,}9965 - 05$

$\boxed{0{,}4965} \simeq 49{,}65\,\%$

R/ Refuta la suposición del ingeniero.

Ejemplo:

$\mu = 28$
$\bar{X} = 30$
$\theta = 5$
$n = 40$

El viaje en un autobús especial para ir de un campus de una universidad al campus de otra en una ciudad toma, en promedio, 28 minutos, con una desviación estándar de 5 minutos. En cierta semana un autobús hizo el viaje 40 veces. ¿Cuál es la probabilidad de que el tiempo promedio del viaje sea mayor a 30 minutos? Suponga que el tiempo promedio se redondea al entero más cercano.

$Z = \dfrac{\bar{X} - \mu}{\frac{\theta}{\sqrt{n}}} = \dfrac{30 - 28}{\frac{5}{\sqrt{40}}} = 2{,}53$ Tabla $= 0{,}9953$

$1 - 0{,}9943 = \boxed{0{,}0057}$

La probabilidad de que el viaje sea mayor 30 min es de 0,57 %

La distribución Muestral de la $\overbrace{\text{Proporción Muestral}}^{\hat{p}}$

$\hat{p} = \frac{x}{n}$ $\quad q = 1 - p$ $\quad Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$

$\left.\begin{array}{l} n \cdot p > 5 \\ n \cdot q > 5 \end{array}\right\}$ Garantizan Dist. Normal

Ejemplo:

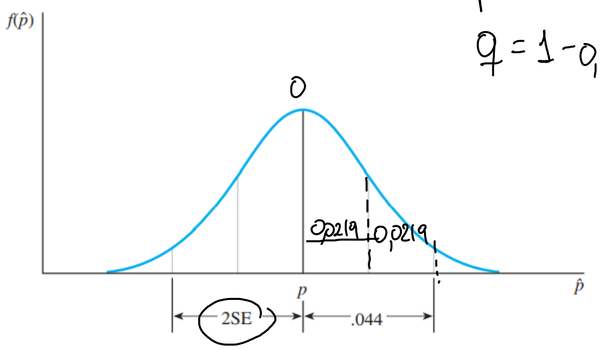
En una encuesta se preguntó a 500 madres y padres sobre la importancia del deporte para hijos e hijas. De los padres entrevistados 60% estuvo de acuerdo en que los géneros son iguales y deben tener iguales oportunidades de participar en deportes. Describa la distribución muestral de la proporción muestral $\hat{p}$ de padres que están de acuerdo en que los géneros son iguales y deben tener iguales oportunidades.

$p = 60\% = 0{,}6$ $\quad$ error $= \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0{,}6 \cdot 0{,}4}{500}}$

$q = 1 - 0{,}6 = 0{,}4$

$= 0{,}0219 \cdot 2$

$= 0{,}044$



Ejemplo:

$\hat{p}$ = muestra

p = población

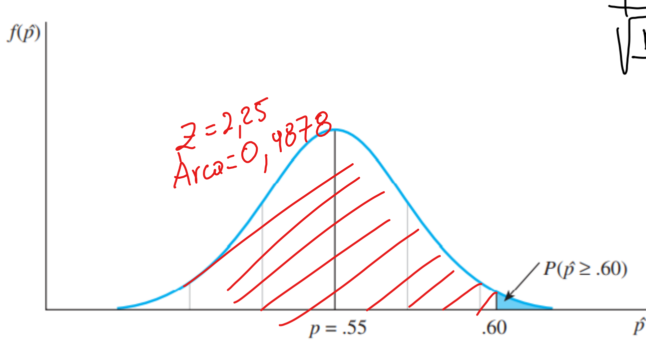
n =

Suponga que la proporción p de padres en la población es en realidad igual $p=$ a .55. ¿Cuál es la probabilidad de observar una proporción muestral igual de grande o mayor que el valor observado $\hat{p} = .60$?

$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0{,}60 - 0{,}55}{\sqrt{\frac{0{,}55 \cdot 0{,}45}{500}}} = \boxed{2{,}25}$

$= 1 - 0{,}9878$

$= \boxed{0{,}0122}$



Distribuciones Muestrales de la Varianza

Ejemplo: $\chi^2 = \frac{\sum (x_i - \bar{x})^2}{\theta^2} = \frac{(n-1)\cdot s^2}{\theta^2}$ → muestra → población   Chi al cuadrado

Un fabricante de baterías para automóvil garantiza que su producto durará, en promedio, 3 años con una desviación estándar de 1 año. Si cinco de estas baterías tienen duraciones de 1.9, 2.4, 3.0, 3.5 y 4.2 años, ¿el fabricante continuará convencido de que sus baterías tienen una desviación estándar de 1 año? Suponga que las duraciones de las baterías siguen una distribución normal.

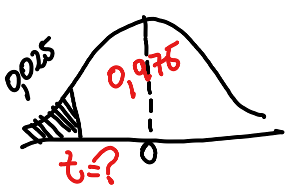
mu estra $\bar{x} = 3$

$$s^2 = \frac{n \cdot \sum (\bar{x}^2)^2 - (\sum (\bar{x}))^2}{n \cdot (n-1)} =$$

Distribución de una media con varianza poblacional no conocida.

Ejemplo: $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ } se usa cuando tengo varianza muestral "s^2" ó Desviación Estandar "S"

El valor t con $v = 14$ grados de libertad que deja una área de 0.025 a la izquierda y, por lo tanto, una área de 0.975 a la derecha, es



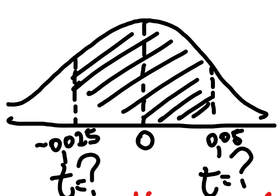
$\boxed{t = -2{,}1448}$

gL = v = grados de libertad = 14

g.L = n-1

Ejemplo:

Calcule $P(-t_{0.025} < T < t_{0.05})$.



0,5-0,025 = 0,4750   0,5-0,05 = 0,45

t = 0,4750 + 0,45 = 0,9750

Ejemplo:

Calcule k tal que $P(k < T < -1.761) = 0.045$ para una muestra aleatoria de tamaño 15 que se selecciona de una distribución normal y $\frac{\overline{X}-\mu}{s/\sqrt{n}}$.

Ejemplo:

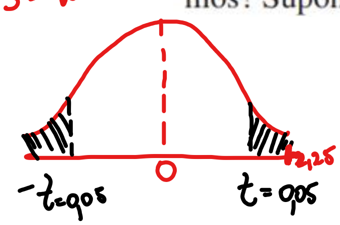
Un ingeniero químico afirma que el rendimiento medio de la población de un cierto proceso de lotes es 500 gramos por mililitro de materia prima. Para verificar dicha afirmación muestrea 25 lotes cada mes. Si el valor t calculado cae entre $-t_{0.05}$ y $t_{0.05}$, queda satisfecho con su afirmación. ¿Qué conclusión debería sacar de una muestra que tiene una media $\bar{x} = 518$ gramos por mililitro y una desviación estándar muestral $s = 40$ gramos? Suponga que la distribución de rendimientos es aproximadamente normal.

$\mu = 500$

$n = 25$

$\bar{x} = 518$

$s = 40$



$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{518-500}{\frac{40}{\sqrt{25}}} = 2{,}25$$

R/ El ingeniero puede concluir que tiene un mejor rendimiento que el promedio.

Estimación

Dado un parámetro de interés, como por ejemplo una media poblacional μ o proporción poblacional p, el objetivo de la estimación puntual es emplear una muestra para calcular un número que representa en algún sentido una buena presunción para el verdadero valor del parámetro. El número resultante se llama estimación puntual.

La inferencia estadística está casi siempre concentrada en sacar algún tipo de conclusiones acerca de uno o más parámetros (características poblacionales). Para hacerlo así se requiere que un investigador obtenga datos muestrales de cada una de las poblaciones en estudio. Por ejemplo, denotemos por μ (un parámetro) el verdadero promedio de resistencia a la ruptura de conexiones de alambres utilizados para unir obleas de semiconductores. Podrí hacerse una muestra aleatorio de n=10 conexiones y determinar la resistencia de la ruptura de cada una, por lo resultan resistentes observadas $x_1, x_2, \ldots\ldots, x_{10}$. La media muestral de la resistencia a la ruptura $\bar{x}$ se podría emplear entonces para sacar una conclusión acerca del valor de μ. Análogamente, si σ^2 es la varianza de la distribución de resistencia a la ruptura (varianza de población, otros parámetro), el valor de la varianza muestral s^2 se podía utilizar para inferir algo acerca de σ^2.

Al analizar conceptos generales y métodos de inferencia, es conveniente tener un símbolo genérico para el parámetro de interés. Utilizaremos la letra griega θ para este propósito. El objetivo de la estimación puntual es seleccionar un solo número, con base en dos muestrales, que represente el valor más razonable de θ.

Ejemplo 1:

Que el parámetro de interés es μ, el verdadero promedio de duración de baterías de cierto tipo de calculadora. Una muestra aleatoria de n=3 baterías podría dar duraciones observadas (en horas) de:

$$x_1 = 5.0$$

$$x_2 = 6.4$$

$$x_3 = 5.9$$

$$\bar{X} = \frac{5.0 + 6.4 + 5.9}{3} = 5.77$$

El valor calculado de la duración media muestral es $\bar{x} = 5.77$, y es razonable considerar 5,77 como el valor más razonable de μ.

Definición:

Una **estimación puntual** de un parámetro θ es un solo número que se puede considerar como el valor más razonable de θ. Se obtiene una estimación puntual al seleccionar una estadística apropiada y calcular su valor a partir de la información muestral dada. La estadística seleccionada se llama **estimador puntual** de θ.

Libro de Jay L. Devore.

Ejemplo 2:

Un fabricante de automóviles ha desarrollado un nuevo tipo de defensas, que se supone absorbe impactos con menos averías que las defensas anteriores. El fabricante ha utilizado esta defensa en una secuencia de 25 choques controlados contra un muro, cada uno a 10 millas por hora, usando uno de sus modelos compactos. Sea X= el número de choques que resultan sin daño visible al automóvil. El parámetro que se va estimar es p=la proporción de todos los choques que resultan sin daño. Si se observa que x es x=15, el estimador y la estimación más razonable son:

$$estimador\ \hat{p} = \frac{X}{n}, estimación = \frac{x}{n} = \frac{15}{25} = 0{,}60$$

Si por cada parámetro de interés hubiera sólo un estimador puntual razonable, no habría mucho para la estimación puntual. En muchos problemas, en cambio, habrá más de un estimador razonable.

No se espera que un estimador realice la estimación del parámetro poblacional sin error. No esperamos que $\bar{X}$ estime μ exactamente, sino que en realidad esperamos que no esté muy alejado.

Estimador Insesgado

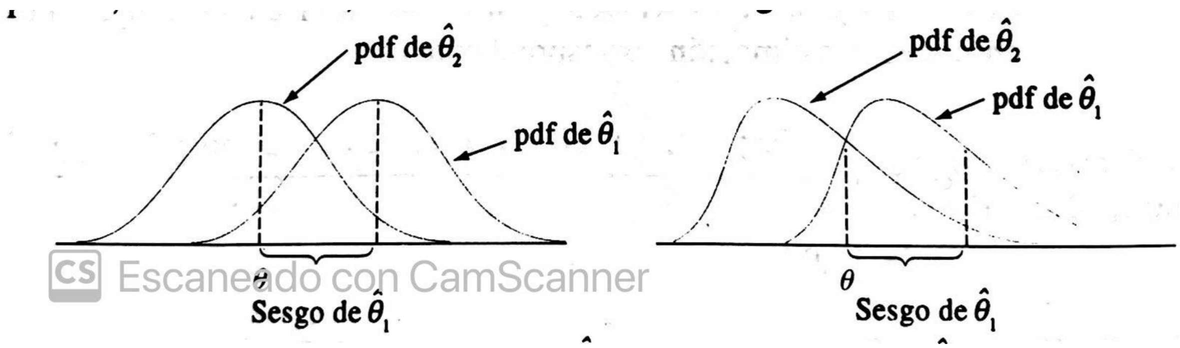
¿Cuáles son las propiedades deseables de una “buena” función de decisión que influirán sobre nosotros para elegir un estimador en lugar de otro? Sea $\widehat{\Theta}$ un estimador cuyo valor $\hat{\theta}$ es un estimador puntual de algún parámetro poblacional desconocido θ. Ciertamente desearíamos que la distribución muestral de $\widehat{\Theta}$ tuviera una media igual al parámetro estimado. Se dice que un estimador que posee esta propiedad es **Insesgado**.

Definición:

Se dice que una estimación $\widehat{\Theta}$ es un estimador **Insesgado** del parámetro θ si

$$\mu_{\widehat{\Theta}} = E(\widehat{\Theta}) = \theta$$

Por tanto, $\hat{\theta}$ es Insesgado si su distribución de probabilidad (es decir, de muestreo) está siempre centrada en el valor verdadero del parámetro. La siguiente figura ilustra las distribuciones de varios estimadores sesgados e insesgados. Observe que “centrada” en este caso quiere decir que el valor esperado, no la mediana, de la distribución de $\hat{\theta}$ es igual a θ.



Puede parecer como si, para ver si $\hat{\theta}$ es Insesgado, es necesario conocer el valor de θ (en cuyo caso la estimación no es necesaria). Éste suele no ser el caso porque muchas veces se puede utilizar un argumento general de valor esperado para verificar la condición de insesgo.

Ejemplo 3:

Suponga que X, el tiempo de reacción a cierto estímulo, tiene una distribución uniforme en el intervalo de 0 a un límite superior θ desconocido (de modo que la función de densidad de X es de forma rectangular con altura $\frac{1}{\theta}$ para $0 \leq x \leq \theta$). Se desea estimar θ con base a una muestra aleatoria $X_1, X_2, \ldots, X_n$ de tiempos de reacción. Como θes el tiempo máximo posible en toda la población de tiempos de reacción, consideremos como un primer estimador el tiempo máximo de reacción muestral: $\widehat{\theta_1} = máx(X_1, \ldots, X_n)$. Si n=5 y $x_1 = 4{,}2$, $x_2 = 1{,}7$, $x_3 = 2{,}4$, $x_4 = 3{,}9$, $x_5 = 1{,}3$ la estimación puntual de θ es $\hat{\theta}_1 = máx(4{,}2\ ,\ 1{,}7\ ,\ 2{,}4\ ,\ 3{,}9\ ,\ 1{,}3) = \boxed{4{,}2}$

$\hat{\theta}_1 = \text{media} = \frac{4{,}2+1{,}7+2{,}4+3{,}9+1{,}3}{5} =$ quien es el valor máximo

La condición de insesgo implica que algunas muestras darán estimaciones que excedan θ y otras muestran darán estimaciones menores que θ, porque de otro modo θ no podría ser posiblemente el centro de la distribución de las $\hat{\theta}_1$.

Proposición:

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con μ media y varianza σ^2.

Entonces el estimador $\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ es un estimador Insesgado de σ^2.

Ejemplo 4:

Las duraciones de ocho baterías cargadas de computadora de marca Dutec son 151, 153, 175, 134, 170, 172, 156 y 114 minutos. Entonces una estimación puntual de la duración promedio de una batería Dutec es:

$$\bar{x} = \frac{151 + 153 + 175 + 134 + 170 + 172 + 156 + 114}{8} = 153{,}125$$

Suponga que la empresa Dutec asegura que la duración promedio de sus baterías es de 200 minutos. De acuerdo con la estimación obtenida se puede sospechar que la afirmación de Dutec es poco probable.

En resumen con lo estudiado anteriormente, dada una muestra observaciones $(x_1, x_2, \dots., x_n)$ de una muestra aleatoria M se tienen la siguientes estimaciones puntuales de los respectivos parámetros:

Parámetro	Estimador	Estimación Puntual
μ	$\bar{X}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
σ^2	S^2	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$
σ	S	$s = \sqrt{s^2}$
p	$\hat{P}$	$\hat{p} = \frac{b}{n}, b\ es\ el\ número\ de\ éxitos\ en\ la\ muestra$

Suponga que se tiene una población X_1 con media población μ_1 y una población X_2 con media poblacional μ_2. Si cada una se toman muestras de tamaño $n_1 y\ n_2$ respectivamente, entonces un estimador Insesgado de la diferencia de promedios $(\mu_1 - \mu_2)$ es: $\bar{X}_1 - \bar{X}_2$. Si se toman valores de las muestras aleatorias: $(x_1, x_2, \ldots\ldots x_{n_1})\ y\ (y_1, y_2, \ldots.\, y_{n_2})$, entonces una estimación puntual de la diferencia de promedios $(\mu_1 - \mu_2)$

$$\bar{x}_1 - \bar{x}_2 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} - \frac{\sum_{i=1}^{n_2} y_i}{n_2}$$

Similarmente, considerando la misma notación, un estimar Insesgado para la diferencia de dos proporciones $(p_1 - p_2)$ es $\hat{P}_1 - \hat{P}_2$ y una estimación puntual es:

$$\hat{p}_1 - \hat{p}_2 = \frac{b_1}{n_1} - \frac{b_2}{n_2}$$

Ejercicio 1

Una bebida afirma en su publicidad por televisión que su consumo diario durante un mes produce una pérdida promedio de al menos cinco libras de peso. Para analizar esta afirmación, se toma un grupo control de ocho personas y se le suministró el producto diariamente por un mes; así, se obtiene los siguientes datos:

n=8

$\overline{x_i} = 174{,}75$

$\overline{x_F} = 171{,}25$

Peso inicial (lb)	165	195	188	170	185	163	155	177
Peso Final (lb)	164	190	187	163	185	159	148	174

Determine, a partir de las muestras, una estimación puntual de la diferencia entre el peso promedio inicial y el peso promedio final. R/3,5

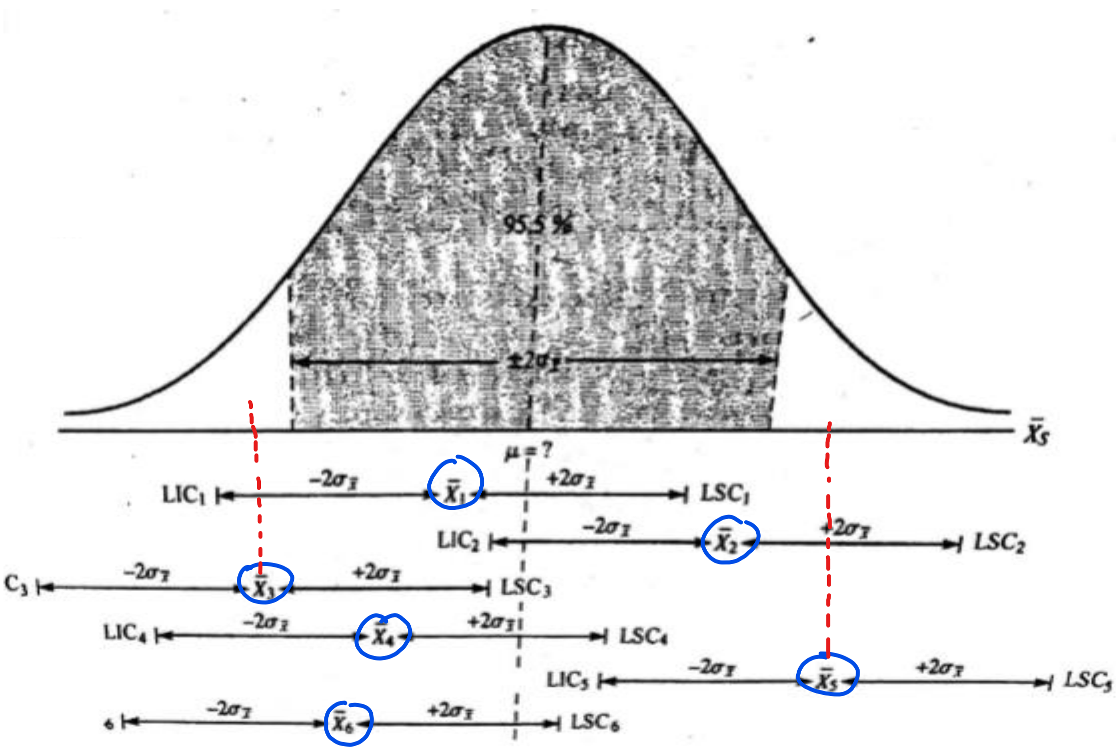
$\overline{X_i} - \overline{X_F} = 174{,}75 - 171{,}25 = 3{,}5$

¿Considera correcta la información hecha en publicidad? NO es cierta la publicidad

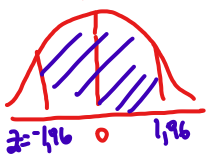
El fundamento de un intervalo de confianza

Un intervalo de confianza tiene un **límite inferior de confianza (LIC)** y un **límite superior de confianza (LSC).** Estos límites se hallan calculados primero la media muestral, $\bar{X}$. Luego se suma una cierta cantidad de $\bar{X}$ para obtener el LSC, y la misma cantidad se resta de $\bar{X}$ para obtener el LIC. La determinación de dicha cantidad es el tema que se va explicar a continuación.

Las distribuciones de muestreo mostraron que toda población se pueden obtener muchas muestras diferentes de un tamaño dado, cada una con su propia medida. En la siguiente figura se muestra seis de estas medidas muéstrales posibles. Si la muestra de $\bar{X}_1$, un intervalo que se extiende dos errores estándares por encima y dos errores estándar por debajo de $\bar{X}_1$ todavía incluye el valor desconocido de la media poblacional. De igual forma, si la muestra hubiese dado una media de $\bar{X}_2$, el intervalo resultante también incluiría la media poblacional. Vale la pena destacar que sólo $\bar{X}_3$ y $\bar{X}_5$ quedan tan lejos de la media poblacional que un intervalo de ± 2 errores estándares no incluye la media poblacional. Todas las otras muestras consideradas producirán un intervalo que contiene la media poblacional. Entonces, la clave para recordar es: como la media poblacional está a lo más a dos errores estándar para el 95,5% de todas las medias muestrales, entonces dada una media muestral cualquiera, se puede estar 95,5% seguro de que el intervalo de dos errores estándar alrededor de dicha media muestral contiene media poblacional desconocida.



confianza
95%



Si se desea construir un intervalo más convencional del 95% (en lugar del 95,5%), ¿cuántos errores estándar se debe mover por encima y por debajo de la media muestral? Como se sabe la tabla de z contiene valores sólo para el área que está por encima o por debajo de la media, se debe dividir el 95% por 2, produciendo 0,4750. Luego, se halla el valor de z, correspondiente a un área de 0,4750, el cual es z=1,96. Así, para construir un intervalo de confianza del 95%, simplemente se especifica un intervalo de 1,96 errores estándar por encima de la media y por debajo de la media muestral. Este valor del 95% es llamado **coeficiente de confianza.**

Intervalo de confianza para la media poblacional – Muestras grandes

Uno de los usos más comunes de los intervalos de confianza es timar la media poblacional.

Se debe de recordar que el intervalo se forma utilizando la media muestral como una estimación puntual para el cual se adiciona y se resta un cierto valor para obtener los límites superior e inferior del intervalo de confianza.

Intervalo de confianza para estimar μ cuando σ es desconocido

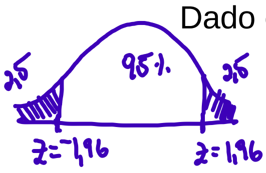
$$\mu = \bar{X} \pm Z_{\sigma_{\bar{x}}}$$

$\bar{X} - Z_{\alpha/2} \leq \mu \leq \bar{X} + Z_{\alpha/2}$

Ejemplo:

$n = 100$
$\bar{X} = 35\,500$
$\sigma = 7\,200$

Consideremos el caso de un promotor inmobiliario quien intenta construir un gran centro comercial. Puede estimar en el área el ingreso promedio por familia como indicador de las ventas esperadas. Una muestra de n=100 familias da una media de $\bar{X} = \$35\,500$. Se asume que la desviación estándar poblacional es $\sigma = \$7\,200$. Dado que $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ se estima un intervalo del 95% como:



$X - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq X + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$\mu = 35\,500 \pm (1{,}96) * \frac{7\,200}{\sqrt{100}}$$

$\frac{2{,}5}{100} = 0{,}0250$
$\downarrow$
Tabla $z = -1{,}96$

$$35\,500 - 0{,}4750 * \frac{7\,200}{\sqrt{100}} \leq \mu \leq 35\,500 + 0{,}4750 * \frac{7\,200}{\sqrt{100}}$$

$35.500 - 1{,}96 \cdot \frac{7200}{\sqrt{100}} \leq \mu \leq 35\,500 + 1{,}96 \cdot \frac{7\,200}{\sqrt{100}}$

$34\,088{,}80 \leq \mu \leq 36\,911{,}20$

$$34\ 088{,}80\ \leq \mu \leq 36\ 911{,}20$$

Interpretación: El promotor establece que con un 95% de confianza en que la media poblacional real desconocida este entre $34 088,80 y $36 911,20.

Esto por supuesto significa que el 5% de todos los intervalos estarían errados; no contendrían la media poblacional. Este 5%, hallado como (1- coeficiente de confianza), es denominado el valor alfa y representa la probabilidad de error.

Valor alfa: Es la probabilidad de error o la probabilidad de que un intervalo dado no contenga la media poblacional desconocida.

Estimación de Media de una sola muestra

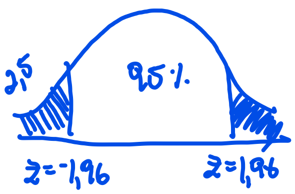
Intervalo de confianza de μ cuando se conoce σ^2	Si $\bar{x}$ es la media de una muestra aleatoria de tamaño n de una población de la que se conoce su varianza σ^2, lo que da un intervalo de confianza de $100(1-\alpha)\%$ para μ es $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

$n = 36$
$\bar{X} = 2{,}6$
$\sigma = 0{,}3$

Se encuentra que la concentración promedio de zinc que se obtiene en una muestra de mediciones en 36 sitios diferentes de un río es de 2.6 gramos por mililitro. Calcule los intervalos de confianza del 95% y 99% para la concentración media de zinc en el río. Suponga que la desviación estándar de la población es de 0.3 gramos por mililitro.

Confianza 95%



$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$2{,}6 - 1{,}96 \cdot \frac{0{,}3}{\sqrt{36}} \leq \mu \leq 2{,}6 + 1{,}96 \cdot \frac{0{,}3}{\sqrt{36}}$$

$$2{,}502 \leq \mu \leq 2{,}698$$

R/ Con un intervalo de confianza del 95% el valor promedio de zinc dista entre 2,502 a 2,698 gramos por mililitro

Confianza del 99%

$\frac{0{,}5}{100} = 0{,}005$ Tabla

$z = 2{,}58$

$$2{,}6 - 2{,}58 \cdot \frac{0{,}3}{\sqrt{36}} \leq \mu < 2{,}6 + 2{,}58 \cdot \frac{0{,}3}{\sqrt{36}}$$

$$2{,}471 \leq \mu \leq 2{,}729$$

Teorema 9.1:	Si utilizamos $\bar{x}$ como una estimación de μ, podemos tener $100(1-\alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. error

Ejemplo

¿Qué tan grande debe ser la muestra del ejemplo 9.2 si queremos tener 95% de confianza en que nuestra estimación de μ diferirá por menos de 0.05? $z = 1{,}96$

e

$$n = \left[\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{e}\right]^2 = \left[\frac{1{,}96 \cdot 0{,}3}{0{,}05}\right]^2 = 138{,}3 \approx \boxed{139}$$

Límites de Confianza unilaterales

Límites de confianza unilaterales de μ cuando se conoce el valor de σ^2	Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n a partir de una población con varianza σ^2, los límites de confianza unilaterales del $100(1-\alpha)\%$ para μ son dados por límite unilateral superior: $\bar{x} + z_\alpha \sigma/\sqrt{n}$; límite unilateral inferior: $\bar{x} - z_\alpha \sigma/\sqrt{n}$.

Ejemplo:

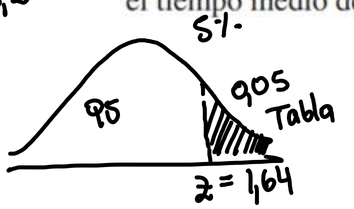
$n = 25$

$s^2 = 4$

$s = \sqrt{4} = 2$

$\bar{x} = 6,2$

En un experimento de pruebas psicológicas se seleccionan al azar 25 sujetos y se miden sus tiempos de reacción, en segundos, ante un estímulo particular. La experiencia sugiere que la varianza en los tiempos de reacción ante los diferentes tipos de estímulos es de 4 s² y que la distribución del tiempo de reacción es aproximadamente normal. El tiempo promedio para los sujetos fue de 6.2 segundos. Calcule un límite superior del 95% para el tiempo medio de reacción.



$$\bar{X} + Z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

$$6,2 + 1,64 \cdot \frac{2}{\sqrt{25}}$$

$$= 6,856 \text{ s.}$$

Cual es error máximo

$$Z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 1,64 \cdot \frac{2}{\sqrt{25}} = 0,656$$

$$= 6,56\%$$

En caso que se desconozca la desviación estándar

Intervalo de confianza para μ cuando se desconoce σ^2	Si $\bar{x}$ y s son la media y la desviación estándar de una muestra aleatoria de una población normal de la que se desconoce la varianza σ^2, un intervalo de confianza del $100(1-\alpha)\%$ para μ es $$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$ donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad que deja una área de $\alpha/2$ a la derecha.

$s =$ muestral

$n \leq 30$

$s =$ muestral

Ejemplo:

$n = 7$

$\bar{x} = 10$

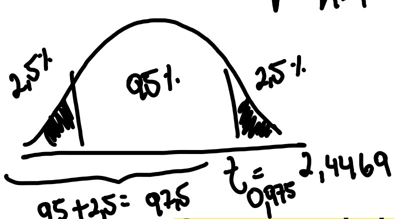
$s = 0{,}283$

$t_{95\%}$

El contenido de ácido sulfúrico de 7 contenedores similares es de 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, y 9.6 litros. Calcule un intervalo de confianza del 95% para el contenido promedio de todos los contenedores suponiendo una distribución aproximadamente normal.

$$\bar{x} = \frac{9{,}8 + 10{,}2 + 10{,}4 + 9{,}8 + 10 + 10{,}2 + 9{,}6}{7} = 10$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 0{,}283$$



$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$10 - 2{,}4469 \cdot \frac{0{,}283}{\sqrt{7}} \leq \mu \leq 10 - 2{,}4469 \cdot \frac{0{,}283}{\sqrt{7}}$$

$$\boxed{9{,}7383 \leq \mu \leq 10{,}2613}$$

Concepto de intervalo de confianza para muestras grandes

Con frecuencia los estadísticos recomiendan que incluso cuando no sea posible suponer la normalidad, se desconozca σ y $n \geq 30$, σ se puede reemplazar con s para poder utilizar el intervalo de confianza

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

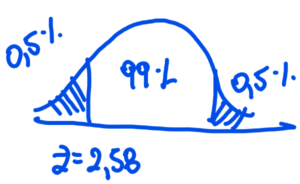
Se obtienen las calificaciones de matemáticas del Examen de Aptitudes Escolares (SAT, por sus siglas en inglés) de una muestra aleatoria de 500 estudiantes del último año de preparatoria del estado de Texas. Se calculan la media y la desviación estándar muestrales, que son 501 y 112, respectivamente. Calcule un intervalo de confianza del 99% de la calificación promedio de matemáticas en el SAT para los estudiantes del último año de preparatoria del estado de Texas.

$n = 500$

$\bar{x} = 501$

$s = 112$

$\alpha = 99\%$



$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$501 - 2{,}58 \cdot \frac{112}{\sqrt{500}} \leq \mu \leq 501 + 2{,}58 \cdot \frac{112}{\sqrt{500}}$$

$$\boxed{488{,}1 \leq \mu \leq 513{,}9}$$

Error Estándar de una Estimación Puntual

Límites de confianza para μ cuando se desconoce σ^2	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}$ e.e.$(\bar{x})$

$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ se escribe como $\bar{x} \pm z_{\alpha/2}$ e.e.$(\bar{x})$,

Intervalos de Predicción

Intervalo de predicción para una observación futura cuando se conoce σ^2	Para una distribución normal de mediciones con media μ desconocida y varianza σ^2 conocida, un **intervalo de predicción** del $100(1-\alpha)\%$ de una observación futura x_0 es $$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+1/n},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

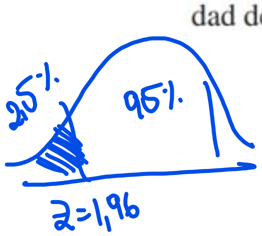
Ejemplo:

Debido a la disminución en las tasas de interés el First Citizens Bank recibió muchas solicitudes para hipoteca. Una muestra reciente de 50 créditos hipotecarios dio como resultado un promedio en la cantidad de préstamos de \$257,300. Suponga una desviación estándar de la población de \$25,000. En el caso del siguiente cliente que llena una solicitud de crédito hipotecario calcule un intervalo de predicción del 95% para la cantidad del crédito.

$n = 50$

$\bar{x} = \$257300$

$\sigma = \$25000$



$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{1+\frac{1}{n}} \leq x_0 \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{1+\frac{1}{n}}$$

$$257300 - 1{,}96 \cdot 25000\sqrt{1+\frac{1}{50}} \leq x_0 \leq 257300 + 1{,}96 \cdot 25000\sqrt{1+\frac{1}{50}}$$

$$207\,812{,}43 \leq x_0 \leq 306\,787{,}57$$

Intervalo de predicción de una observación futura cuando se desconoce σ^2	Para una distribución normal de mediciones cuando la media μ y la varianza σ^2 se desconocen, un **intervalo de predicción** del $100(1-\alpha)\%$ de una observación futura x_0 es $$\bar{x} - t_{\alpha/2}s\sqrt{1+1/n} < x_0 < \bar{x} + t_{\alpha/2}s\sqrt{1+1/n},$$ donde $t_{\alpha/2}$ es el valor t con $v = n-1$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

Un inspector de alimentos seleccionó aleatoriamente 30 paquetes de carne de res 95% magra. La muestra dio como resultado una media de 96.2% con una desviación estándar muestral de 0.8%. Calcule un intervalo de predicción del 99% para la condición baja en grasa de un paquete nuevo. Suponga normalidad.

$n = 30$

$\bar{X} = 96,2\%$

$S = 0,8\%$

$t_{0,995} = 2,7564$

$\underbrace{\qquad}_{99+0,5=99,5}$

$g.L = n-1$
$30-1$
29

$$\bar{X} - t_{\frac{\alpha}{2}} \cdot s \cdot \sqrt{1+\frac{1}{n}} < x_0 < \bar{X} + t_{\frac{\alpha}{2}} \cdot s \cdot \sqrt{1+\frac{1}{n}}$$

$$96,2 - 2,7564 \cdot 0,8 \cdot \sqrt{1+\frac{1}{30}} < x_0 < 96,2 + 2,7564 \cdot 0,8 \cdot \sqrt{1+\frac{1}{30}}$$

$$93,96\% < x_0 < 98,45\%$$

Estimación de una Proporción

En una sola muestra

Intervalos de confianza para p de una muestra grande	Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1-\hat{p}$, un intervalo de confianza aproximado del $100(1-\alpha)\%$ para el parámetro binomial p se obtiene por medio de (método 1) $$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$ o mediante (método 2) $$\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} - \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}} < p < \frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} + \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

En una muestra aleatoria de $n = 500$ familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que $x = 340$ están suscritas a HBO. Calcule un intervalo de confianza del 95% para la proporción real de familias que tienen televisores en esta ciudad y están suscritas a HBO.

Teorema 9.3: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1-\alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

Selección del tamaño de la muestra

Teorema 9.4: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1-\alpha)\%$ de confianza en que el error será menor que una cantidad específica e cuando el tamaño de la muestra sea aproximadamente

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

Ejemplo:

¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener un 95% de confianza en que la estimación de p esté dentro de 0.02 del valor verdadero?

Teorema 9.5:	Si utilizamos $\hat{p}$ como un estimado de p, podemos tener, **al menos**, un $100(1-\alpha)\%$ de confianza en que el error no excederá a una cantidad específica e cuando el tamaño de la muestra sea $$n = \frac{z_{\alpha/2}^2}{4e^2}.$$

Ejemplo:

¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener al menos un 95% de confianza en que nuestra estimación de p está dentro de 0.02 del valor verdadero?

Estimación de la diferencia entre dos proporciones

Intervalo de confianza para $p_1 - p_2$ de una muestra grande	Si $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxitos en muestras aleatorias de tamaños n_1 y n_2, respectivamente, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$, un intervalo de confianza aproximado del $100(1-\alpha)\%$ para la diferencia de dos parámetros binomiales $p_1 - p_2$ es dado por $$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

Se considera hacer un cierto cambio en el proceso de fabricación de partes componentes. Para determinar si el cambio en el proceso da como resultado una mejora, se toman muestras de partes fabricadas con el proceso nuevo y con el actual. Si se encuentra que 75 de 1500 artículos manufacturados con el proceso actual están defectuosos y 80 de 2000 manufacturados con el proceso nuevo también lo están, calcule un intervalo de confianza del 90% para la diferencia verdadera en la proporción de partes defectuosas entre el proceso actual y el nuevo.

Estimación de Media de una sola muestra

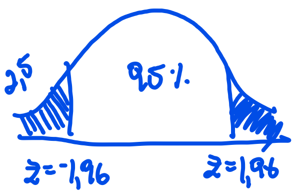
Intervalo de confianza de μ cuando se conoce σ^2	Si $\bar{x}$ es la media de una muestra aleatoria de tamaño n de una población de la que se conoce su varianza σ^2, lo que da un intervalo de confianza de $100(1-\alpha)\%$ para μ es $$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

$n = 36$
$\bar{X} = 2{,}6$
$\sigma = 0{,}3$

Se encuentra que la concentración promedio de zinc que se obtiene en una muestra de mediciones en 36 sitios diferentes de un río es de 2.6 gramos por mililitro. Calcule los intervalos de confianza del 95% y 99% para la concentración media de zinc en el río. Suponga que la desviación estándar de la población es de 0.3 gramos por mililitro.

Confianza 95%



$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$2{,}6 - 1{,}96 \cdot \frac{0{,}3}{\sqrt{36}} \leq \mu \leq 2{,}6 + 1{,}96 \cdot \frac{0{,}3}{\sqrt{36}}$$

$$2{,}502 \leq \mu \leq 2{,}698$$

R/ Con un intervalo de confianza del 95% el valor promedio de zinc dista entre 2,502 a 2,698 gramos por mililitro

Confianza del 99%

$$2{,}6 - 2{,}58 \cdot \frac{0{,}3}{\sqrt{36}} \leq \mu < 2{,}6 + 2{,}58 \cdot \frac{0{,}3}{\sqrt{36}}$$

$$2{,}471 \leq \mu \leq 2{,}729$$

$\frac{0{,}5}{100} = 0{,}005$ Tabla

Teorema 9.1:	Si utilizamos $\bar{x}$ como una estimación de μ, podemos tener $100(1-\alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. error

Ejemplo

¿Qué tan grande debe ser la muestra del ejemplo 9.2 si queremos tener 95% de confianza en que nuestra estimación de μ diferirá por menos de 0.05? $z = 1{,}96$

e

$$n = \left[\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{e}\right]^2 = \left[\frac{1{,}96 \cdot 0{,}3}{0{,}05}\right]^2 = 138{,}3 \approx \boxed{139}$$

Límites de Confianza unilaterales

Límites de confianza unilaterales de μ cuando se conoce el valor de σ^2	Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n a partir de una población con varianza σ^2, los límites de confianza unilaterales del $100(1-\alpha)\%$ para μ son dados por límite unilateral superior: $\bar{x} + z_\alpha \sigma/\sqrt{n}$; límite unilateral inferior: $\bar{x} - z_\alpha \sigma/\sqrt{n}$.

Ejemplo:

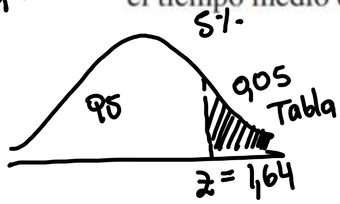
En un experimento de pruebas psicológicas se seleccionan al azar 25 sujetos y se miden sus tiempos de reacción, en segundos, ante un estímulo particular. La experiencia sugiere que la varianza en los tiempos de reacción ante los diferentes tipos de estímulos es de 4 s² y que la distribución del tiempo de reacción es aproximadamente normal. El tiempo promedio para los sujetos fue de 6.2 segundos. Calcule un límite superior del 95% para el tiempo medio de reacción.

$n = 25$

$s^2 = 4$

$s = \sqrt{4} = 2$

$\bar{x} = 6{,}2$



$$\bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

$$6{,}2 + 1{,}64 \cdot \frac{2}{\sqrt{25}}$$

$$= 6{,}856 \text{ s}$$

Cual es error máximo

$$z_\alpha \cdot \frac{\sigma}{\sqrt{n}} = 1{,}64 \cdot \frac{2}{\sqrt{25}} = 0{,}656$$

$$= 6{,}56\%$$

En caso que se desconozca la desviación estándar

Intervalo de confianza para μ cuando se desconoce σ^2	Si $\bar{x}$ y s son la media y la desviación estándar de una muestra aleatoria de una población normal de la que se desconoce la varianza σ^2, un intervalo de confianza del $100(1-\alpha)\%$ para μ es $$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$ donde $t_{\alpha/2}$ es el valor t con $v = n-1$ grados de libertad que deja una área de $\alpha/2$ a la derecha.

S = muestral

$n \leq 30$

S = muestral

Ejemplo:

$n=7$

$\bar{x}=10$

$S=0{,}283$

$t_{95\%}$

El contenido de ácido sulfúrico de 7 contenedores similares es de 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, y 9.6 litros. Calcule un intervalo de confianza del 95% para el contenido promedio de todos los contenedores suponiendo una distribución aproximadamente normal.

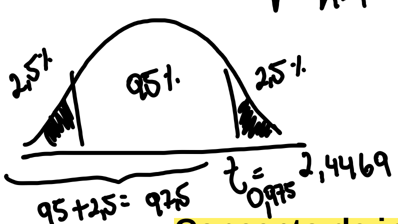
$$\bar{X} = \frac{9{,}8+10{,}2+10{,}4+9{,}8+10+10{,}2+9{,}6}{7} = 10$$

$$S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = 0{,}283$$

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$10 - 2{,}4469 \cdot \frac{0{,}283}{\sqrt{7}} \leq \mu \leq 10 - 2{,}4469 \cdot \frac{0{,}283}{\sqrt{7}}$$

$$\boxed{9{,}7383 \leq \mu \leq 10{,}2613}$$



Concepto de intervalo de confianza para muestras grandes

Con frecuencia los estadísticos recomiendan que incluso cuando no sea posible suponer la normalidad, se desconozca σ y $n \geq 30$, σ se puede reemplazar con s para poder utilizar el intervalo de confianza

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

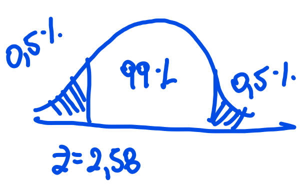
Se obtienen las calificaciones de matemáticas del Examen de Aptitudes Escolares (SAT, por sus siglas en inglés) de una muestra aleatoria de 500 estudiantes del último año de preparatoria del estado de Texas. Se calculan la media y la desviación estándar muestrales, que son 501 y 112, respectivamente. Calcule un intervalo de confianza del 99% de la calificación promedio de matemáticas en el SAT para los estudiantes del último año de preparatoria del estado de Texas.

$n=500$

$\bar{x}=501$

$S=112$

$\alpha=99\%$



$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

$$501 - 2{,}58 \cdot \frac{112}{\sqrt{500}} \leq \mu \leq 501 + 2{,}58 \cdot \frac{112}{\sqrt{500}}$$

$$\boxed{488{,}1 \leq \mu \leq 513{,}9}$$

Error Estándar de una Estimación Puntual

Límites de confianza para μ cuando se desconoce σ^2	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2}\ \text{e.e.}(\bar{x})$

$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ se escribe como $\bar{x} \pm z_{\alpha/2}\ \text{e.e.}(\bar{x})$,

Intervalos de Predicción

Intervalo de predicción para una observación futura cuando se conoce σ^2	Para una distribución normal de mediciones con media μ desconocida y varianza σ^2 conocida, un **intervalo de predicción** del $100(1-\alpha)\%$ de una observación futura x_0 es $$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+1/n},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

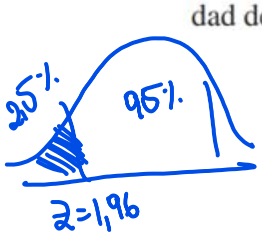
Ejemplo:

Debido a la disminución en las tasas de interés el First Citizens Bank recibió muchas solicitudes para hipoteca. Una muestra reciente de 50 créditos hipotecarios dio como resultado un promedio en la cantidad de préstamos de \$257,300. Suponga una desviación estándar de la población de \$25,000. En el caso del siguiente cliente que llena una solicitud de crédito hipotecario calcule un intervalo de predicción del 95% para la cantidad del crédito.

$n = 50$

$\bar{x} = \$257300$

$\sigma = \$25000$



$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{1+\frac{1}{n}} \leq x_0 \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \sigma \cdot \sqrt{1+\frac{1}{n}}$$

$$257300 - 1{,}96 \cdot 25000\sqrt{1+\frac{1}{50}} \leq x_0 \leq 257300 + 1{,}96 \cdot 25000\sqrt{1+\frac{1}{50}}$$

$$207\ 812{,}43 \leq x_0 \leq 306\ 787{,}57$$

Intervalo de predicción de una observación futura cuando se desconoce σ^2	Para una distribución normal de mediciones cuando la media μ y la varianza σ^2 se desconocen, un **intervalo de predicción** del $100(1-\alpha)\%$ de una observación futura x_0 es $$\bar{x} - t_{\alpha/2}s\sqrt{1+1/n} < x_0 < \bar{x} + t_{\alpha/2}s\sqrt{1+1/n},$$ donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

Un inspector de alimentos seleccionó aleatoriamente 30 paquetes de carne de res 95% magra. La muestra dio como resultado una media de 96.2% con una desviación estándar muestral de 0.8%. Calcule un intervalo de predicción del 99% para la condición baja en grasa de un paquete nuevo. Suponga normalidad.

$n = 30$

$\bar{x} = 96{,}2\%$

$S = 0{,}8\%$

$t_{0,995} = 2{,}7564$

$\underbrace{\qquad\qquad}_{99+0{,}5 = 99{,}5}$

$g.L = n-1$
$30-1$
29

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot s \cdot \sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + t_{\frac{\alpha}{2}} \cdot s \cdot \sqrt{1+\frac{1}{n}}$$

$$96{,}2 - 2{,}7564 \cdot 0{,}8 \cdot \sqrt{1+\frac{1}{30}} < x_0 < 96{,}2 + 2{,}7564 \cdot 0{,}8 \cdot \sqrt{1+\frac{1}{30}}$$

$$93{,}96\% < x_0 < 98{,}45\%$$

Estimación de una Proporción

$\% = p$

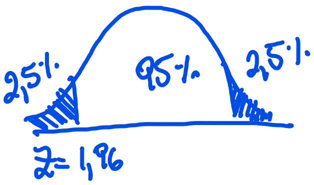
no se conoce σ (Des. Estandar)

En una sola muestra

Intervalos de confianza para p de una muestra grande	Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado del $100(1-\alpha)\%$ para el parámetro binomial p se obtiene por medio de (método 1) $$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$ o mediante (método 2) $$\frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} - \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}} < p < \frac{\hat{p}+\frac{z^2_{\alpha/2}}{2n}}{1+\frac{z^2_{\alpha/2}}{n}} + \frac{z_{\alpha/2}}{1+\frac{z^2_{\alpha/2}}{n}}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z^2_{\alpha/2}}{4n^2}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo: $\hat{p} = \frac{x}{n}$ $n = 500$ $x = 340$ $\hat{p} = \frac{340}{500} = 0{,}68$ $\hat{q} = 1 - \hat{p} = 1 - 0{,}68 = 0{,}32$

En una muestra aleatoria de $n = 500$ familias que tienen televisores en la ciudad de Hamilton, Canadá, se encuentra que $x = 340$ están suscritas a HBO. Calcule un intervalo de confianza del 95% para la proporción real de familias que tienen televisores en esta ciudad y están suscritas a HBO.



$$\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$0{,}68 - 1{,}96 \cdot \sqrt{\frac{0{,}68 \cdot 0{,}32}{500}} < p < 0{,}68 + 1{,}96 \cdot \sqrt{\frac{0{,}68 \cdot 0{,}32}{500}}$$

$$0{,}6391 < p < 0{,}7208$$

R/ La proporción real de las familias que tienen TV suscritos HBO se encuentran entre 63,91% a 72,08%

Teorema 9.3: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1 - \alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

error $= 1{,}96 \cdot \sqrt{\frac{0{,}68 \cdot 0{,}32}{500}} = 0{,}0409$

4,09%

0,04089

Selección del tamaño de la muestra $n = ?$

Teorema 9.4: Si $\hat{p}$ se utiliza como un estimado de p, podemos tener un $100(1 - \alpha)\%$ de confianza en que el error será menor que una cantidad específica e cuando el tamaño de la muestra sea aproximadamente

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

→ que tan grande quiero el error.

Ejemplo:

¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener un 95% de confianza en que la estimación de p esté dentro de 0.02 del valor verdadero?

$Z = 1{,}96$

$$n = \frac{1{,}96^2 \cdot 0{,}68 \cdot 0{,}32}{0{,}02^2} = 2089{,}83 \approx 2090$$

Teorema 9.5:	Si utilizamos $\hat{p}$ como un estimado de p, podemos tener, **al menos**, un $100(1-\alpha)\%$ de confianza en que el error no excederá a una cantidad específica e cuando el tamaño de la muestra sea $$n = \frac{z^2_{\alpha/2}}{4e^2}.$$

confianza

error

se desconoce n, $\hat{p}$, $\hat{q}$, $\bar{x}$

Ejemplo:

¿Qué tan grande debe ser una muestra en el ejemplo 9.14 si queremos tener al menos un 95% de confianza en que nuestra estimación de p está dentro de 0.02 del valor verdadero?

$$n = \frac{1{,}96^2}{4 \cdot 0{,}02^2} = 2401$$

Estimación de la diferencia entre dos proporciones

Intervalo de confianza para $p_1 - p_2$ de una muestra grande	Si $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxitos en muestras aleatorias de tamaños n_1 y n_2, respectivamente, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$, un intervalo de confianza aproximado del $100(1-\alpha)\%$ para la diferencia de dos parámetros binomiales $p_1 - p_2$ es dado por $$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}},$$ donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo:

Se considera hacer un cierto cambio en el proceso de fabricación de partes componentes. Para determinar si el cambio en el proceso da como resultado una mejora, se toman muestras de partes fabricadas con el proceso nuevo y con el actual. Si se encuentra que 75 de 1500 artículos manufacturados con el proceso actual están defectuosos y 80 de 2000 manufacturados con el proceso nuevo también lo están, calcule un intervalo de confianza del 90% para la diferencia verdadera en la proporción de partes defectuosas entre el proceso actual y el nuevo.

Nuevo

$\hat{p}_2 = \frac{80}{2000} = 0,04$

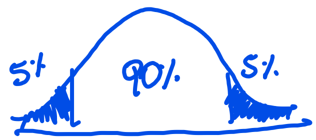
$\hat{q}_2 = 1 - 0,04 = 0,96$

Actual

$\hat{p}_1 = \frac{75}{1500} = 0,05$

$\hat{q}_1 = 1 - 0,05 = 0,95$

$\alpha = 90\%$



$\frac{5}{100} = 0,05$

$z_{\frac{\alpha}{2}} = 1,64.$

$$(\hat{p}_1 - \hat{p}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$(0,05 - 0,04) - 1,64 \cdot \sqrt{\frac{0,05 \cdot 0,95}{1500} + \frac{0,04 \cdot 0,96}{2000}} < (p_1 - p_2) < (0,05 - 0,04) + 1,64 \sqrt{\frac{0,05 \cdot 0,95}{1500} + \frac{0,04 \cdot 0,96}{2000}}$$

$$-0,0017 < (p_1 - p_2) < 0,0217$$

$$\boxed{-0,17\% < (p_1 - p_2) < 2,17\%}$$

Estimación de la diferencia de dos Muestras

Intervalo de confianza para $\mu_1 - \mu_2$ cuando se conocen σ_1^2 y σ_2^2

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaños n_1 y n_2, de poblaciones que tienen varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ es dado por

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

varianzas

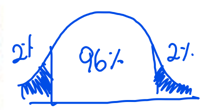
donde $z_{\alpha/2}$ es el valor z que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.10: Se llevó a cabo un experimento donde se compararon dos tipos de motores, el A y el B. Se midió el rendimiento de combustible en millas por galón. Se realizaron 50 experimentos con el motor tipo A y 75 con el motor tipo B. La gasolina utilizada y las demás condiciones se mantuvieron constantes. El rendimiento promedio de gasolina para el motor A fue de 36 millas por galón y el promedio para el motor B fue de 42 millas por galón. Calcule un intervalo de confianza del 96% sobre $\mu_B - \mu_A$, donde μ_A y μ_B corresponden a la media de la población del rendimiento de millas por galón para los motores A y B, respectivamente. Suponga que las desviaciones estándar de la población son 6 y 8 para los motores A y B, respectivamente.

Motor A
$n = 50$
$\bar{x} = 36$
$\theta_1 = 6$

Motor B
$n = 75$
$\bar{x} = 42$
$\theta_2 = 8$

$\alpha = 96\%$



$\frac{2}{100} = 0{,}02$

$z_{\frac{\alpha}{2}} = 2{,}05$

$$(\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}}$$

$$(42-36) - 2{,}05 \cdot \sqrt{\frac{8^2}{75} + \frac{6^2}{50}} < \mu_1 - \mu_2 < (42-36) + 2{,}05 \cdot \sqrt{\frac{8^2}{75} + \frac{6^2}{50}}$$

$$\boxed{3{,}4286 < \mu_B - \mu_A < 8{,}5714}$$

Se puede obtener una estimación puntual de la varianza común desconocida σ^2 agrupando las varianzas muestrales. Si representamos con S_p^2 al estimador agrupado, obtenemos lo siguiente,

Desviaciones estandares muestrales

Estimado agrupado de la varianza	$S_p^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$. $= S_p = \sqrt{\dfrac{(n_1-1)\cdot S_1^2+(n_2-1)\cdot S_2^2}{n_1+n_2-2}}$

varianzas desconocidas pero iguales

Intervalo de confianza para $\mu_1-\mu_2$, $\sigma_1^2=\sigma_2^2$ cuando se desconocen ambas varianzas	Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes con tamaños n_1 y n_2, respectivamente, tomadas de poblaciones más o menos normales con varianzas iguales pero desconocidas, un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_1-\mu_2$ es dado por $$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},$$ donde s_p es la estimación agrupada de la desviación estándar de la población y $t_{\alpha/2}$ es el valor t con $v=n_1+n_2-2$ grados de libertad, que deja una área de $\alpha/2$ a la derecha. g.l.

Ejemplo 9.11: En el artículo "Estructura comunitaria de los macroinvertebrados como un indicador de la contaminación de minas ácidas", publicado en el *Journal of Environmental Pollution*, se informa sobre una investigación realizada en Cane Creek, Alabama, para determinar la relación entre parámetros fisioquímicos seleccionados y diversas mediciones de la estructura de la comunidad de macroinvertebrados. Una faceta de la investigación consistió en evaluar la efectividad de un índice numérico de la diversidad de especies para indicar la degradación del agua debida al desagüe ácido de una mina. Conceptualmente, un índice elevado de la diversidad de especies macroinvertebradas debería indicar un sistema acuático no contaminado; mientras que un índice bajo de esta diversidad indicaría un sistema acuático contaminado.

Se eligieron 2 estaciones de muestreo independientes para este estudio: una que se localiza corriente abajo del punto de descarga ácida de la mina y la otra ubicada corriente arriba. Para 12 muestras mensuales reunidas en la estación corriente abajo el índice de diversidad de especies tuvo un valor medio de $\bar{x}_1 = 3.11$ y una desviación estándar de $s_1 = 0.771$; mientras que 10 muestras reunidas mensualmente en la estación corriente arriba tuvieron un valor medio del índice $\bar{x}_2 = 2.04$ y una desviación estándar de $s_2 = 0.448$. Calculemos un intervalo de confianza del 90% para la diferencia entre las medias de la población de los dos sitios, suponiendo que las poblaciones se distribuyen de forma aproximadamente normal y que tienen varianzas iguales.

Abajo	Arriba
$n=12$	$n=10$
$\bar{X}_1=3{,}11$	$\bar{X}=2{,}04$
$S_1=0{,}771$	$S_2=0{,}448$

$t_{0,95} = 1{,}7247$

$g.l = n_1+n_2-2$
$= 12+10-2$
$= 20$

$$S_p = \sqrt{\frac{(12-1)\cdot 0{,}771^2 + (10-1)\cdot 0{,}448^2}{12+10-2}} = \boxed{0{,}6460}$$

$$(\bar{X}_1-\bar{X}_2) - t_{\frac{\alpha}{2}}\cdot S_p\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+t_{\frac{\alpha}{2}}\cdot S_p\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(3{,}11-2{,}04) - 1{,}7247\cdot 0{,}6460\cdot\sqrt{\frac{1}{12}+\frac{1}{10}} < \mu_1-\mu_2 < (3{,}11-2{,}04)+1{,}7247\cdot 0{,}6460\cdot\sqrt{\frac{1}{12}+\frac{1}{10}}$$

$$\boxed{0{,}5929 < \mu_1-\mu_2 < 1{,}5471}$$

Interpretación del intervalo de confianza

Para el caso de un solo parámetro el intervalo de confianza simplemente produce límites de error del parámetro. Los valores contenidos en el intervalo se deberían ver como valores razonables, dados los datos experimentales. En el caso de una diferencia entre dos medias, la interpretación se puede extender a una comparación de las dos medias. Por ejemplo, si tenemos gran confianza en que una diferencia $\mu_1 - \mu_2$ es positiva, sin duda inferiremos que $\mu_1 > \mu_2$ con poco riesgo de incurrir en un error. Así, en el ejemplo 9.11 tenemos un 90% de confianza en que el intervalo de 0.593 a 1.547 contiene la diferencia de las medias de la población para valores del índice de diversidad de especies en las dos estaciones. El hecho de que ambos límites de confianza sean positivos indica que, en promedio, el índice para la estación que se localiza corriente abajo del punto de descarga es mayor que el índice para la estación que se localiza corriente arriba.

Muestras de tamaños iguales

El procedimiento para construir intervalos de confianza para $\mu_1 - \mu_2$ cuando $\sigma_1 = \sigma_2 = \sigma$ pero ésta se desconoce, requiere suponer que las poblaciones son normales. Desviaciones ligeras de la suposición de varianzas iguales o de normalidad no alteran seriamente el grado de confianza en nuestro intervalo. (En el capítulo 10 se estudia un procedimiento para probar la igualdad de dos varianzas poblacionales desconocidas con base en la información que proporcionan las varianzas muestrales). Si las varianzas de la población son considerablemente diferentes, aún obtenemos resultados razonables cuando las poblaciones son normales, siempre y cuando $n_1 = n_2$. Por lo tanto, al planear un experimento se debería hacer un esfuerzo por igualar el tamaño de las muestras.

Varianzas son desconocidas pero diferentes

Intervalo de confianza para $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ y ambas varianzas se desconocen

Si $\bar{x}_1$ y s_1^2 y $\bar{x}_2$ y s_2^2 son las medias y varianzas de muestras aleatorias independientes de tamaños n_1 y n_2, respectivamente, tomadas de poblaciones aproximadamente normales con varianzas desconocidas y diferentes, un intervalo de confianza aproximado del $100(1-\alpha)\%$ para $\mu_1 - \mu_2$ es dado por

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

donde $t_{\alpha/2}$ es el valor t con

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]} = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\left[\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}\right] + \left[\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}\right]}$$

grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.12: El Departamento de zoología de Virginia Tech llevó a cabo un estudio para estimar la diferencia en la cantidad de ortofósforo químico medido en dos estaciones diferentes del río James. El ortofósforo se mide en miligramos por litro. Se reunieron 15 muestras de la estación 1 y 12 muestras de la estación 2. Las 15 muestras de la estación 1 tuvieron un contenido promedio de ortofósforo de 3.84 miligramos por litro y una desviación estándar de 3.07 miligramos por litro; en tanto que las 12 muestras de la estación 2 tuvieron un contenido promedio de 1.49 miligramos por litro y una desviación estándar de 0.80 miligramos por litro. Calcule un intervalo de confianza de 95% para la diferencia en el contenido promedio verdadero de ortofósforo en estas dos estaciones. Suponga que las observaciones provienen de poblaciones normales con varianzas diferentes.

Estación 1

$n=15$

$\bar{x}=3{,}84$

$S_1=3{,}07$

Estación 2

$n=12$

$\bar{x}=1{,}49$

$S_2=0{,}8$

$\alpha=95\%$

2,5 — 95% — 2,5

$t_{0{,}975}=2{,}1199$

$$V=\frac{\left[\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right]^2}{\left[\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1}\right]+\left[\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}\right]}=\frac{\left[\frac{3{,}07^2}{15}+\frac{0{,}8^2}{12}\right]^2}{\left[\frac{\left(\frac{3{,}07^2}{15}\right)^2}{15-1}\right]+\left[\frac{\left(\frac{0{,}8^2}{12}\right)^2}{12-1}\right]}=\frac{0{,}4647}{0{,}0285}$$

$\boxed{V=16{,}30}$ Redondea normal

$\boxed{V=16}$

$$(\bar{x}_1-\bar{x}_2)-t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$(3{,}84-1{,}49)-2{,}1199\sqrt{\frac{3{,}07^2}{15}+\frac{0{,}8^2}{12}}<\mu_1-\mu_2<(3{,}84-1{,}49)+2{,}1199\sqrt{\frac{3{,}07^2}{15}+\frac{0{,}8^2}{12}}$$

$$\boxed{0{,}5998<\mu_1-\mu_2<4{,}1002}$$

No lo ven

Intervalo de confianza para $\mu_D = \mu_1 - \mu_2$ para observaciones pareadas	Si $\bar{d}$ y s_d son la media y la desviación estándar, respectivamente, de las diferencias distribuidas normalmente de n pares aleatorios de mediciones, un intervalo de confianza del $100(1-\alpha)\%$ para $\mu_D = \mu_1 - \mu_2$ es $$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$ donde $t_{\alpha/2}$ es el valor t con $v = n - 1$ grados de libertad, que deja una área de $\alpha/2$ a la derecha.

Ejemplo 9.13: Un estudio publicado en *Chemosphere* reporta los niveles de la dioxina TCDD en 20 veteranos de Vietnam de Massachusetts, quienes posiblemente estuvieron expuestos al agente naranja. En la tabla 9.1 se presentan los niveles de TCDD en plasma y tejido adiposo.

Calcule un intervalo de confianza del 95% para $\mu_1 - \mu_2$, donde μ_1 y μ_2 representen las medias verdaderas de los niveles de TCDD en plasma y en tejido adiposo, respectivamente. Suponga que la distribución de las diferencias es casi normal.

Tabla 9.1: Datos para el ejemplo 9.13.

Veterano	Niveles de TCDD en plasma	Niveles de TCDD en tejido adiposo	d_i	Veterano	Niveles de TCDD en plasma	Niveles de TCDD en tejido adiposo	d_i
1	2.5	4.9	−2.4	11	6.9	7.0	−0.1
2	3.1	5.9	−2.8	12	3.3	2.9	0.4
3	2.1	4.4	−2.3	13	4.6	4.6	0.0
4	3.5	6.9	−3.4	14	1.6	1.4	0.2
5	3.1	7.0	−3.9	15	7.2	7.7	−0.5
6	1.8	4.2	−2.4	16	1.8	1.1	0.7
7	6.0	10.0	−4.0	17	20.0	11.0	9.0
8	3.0	5.5	−2.5	18	2.0	2.5	−0.5
9	36.0	41.0	−5.0	19	2.5	2.3	0.2
10	4.7	4.4	0.3	20	4.1	2.5	1.6

Práctica de la semana 7

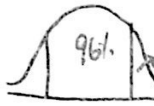
1) Una empresa eléctrica fabrica focos que tienen una duración aproximadamente distribuida de forma normal con una desviación estándar de 40 horas. Si una muestra de 30 focos tiene una duración promedio de 780 horas, encuentre un intervalo de confianza de 96% para la media poblacional de todos los focos que producen esta empresa.

θ = 40 horas
n = 30 focos.
$\bar{X} = 780$
confianza: 96%

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$780 - 2{,}05 \cdot \frac{40}{\sqrt{30}} < \mu < 780 + 2{,}05 \cdot \frac{40}{\sqrt{30}}$$

$$765{,}03 < \mu < 794{,}97$$



$\frac{2\%}{100} = 0{,}020$
$0{,}5 - 0{,}020$
$0{,}4800$
$z = 2{,}05$

R/ Con un nivel de confianza del 96% la media poblacional de la duración de los focos dista de 765,03 a 794,97 horas.

2) A muchos pacientes con problemas cardiacos se les implantó un marcapasos para controlar su ritmo cardiaco. Se monta un módulo conector de plástico sobre la parte superior del marcapasos. Suponga una desviación estándar de 0,0015 y una distribución normal, encuentre un intervalo de confianza del 95% para la media de todos los módulos conectores que fabrica cierta compañía. Una muestra aleatoria de 75 módulos tiene un promedio de 0,310 pulgadas.

θ = 0,0015
confianza = 95%
n = 75
$\bar{X} = 0{,}310$

$$0{,}310 - 1{,}96 \cdot \frac{0{,}0015}{\sqrt{75}} < \mu < 0{,}310 + 1{,}96 \cdot \frac{0{,}0015}{\sqrt{75}}$$

$$0{,}3097 < \mu < 0{,}3103$$

$\frac{2{,}5\%}{100} = 0{,}0250$
$0{,}5 - 0{,}0250$
$0{,}4750$
$z = 1{,}96$

R/ Con un nivel de confianza del 95% la media de los módulos conectores distan entre 0,3097 a 0,3103 pulgadas.

3) La estatura de una muestra aleatoria de 50 estudiantes universitarios muestra una media de 174,5 centímetros y una desviación estándar de 6,9 centímetros.

n = 50
$\bar{X} = 174{,}5$ centímetros
s = 6,9

a) Construya un intervalo de confianza del 98% para la estatura media de todos los estudiantes de la universidad.

$$\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

$$174{,}5 - 1{,}2991 \cdot \frac{6{,}9}{\sqrt{50}} < \mu < 174{,}5 + 1{,}2991 \cdot \frac{6{,}9}{\sqrt{50}}$$

$$\cancel{173{,}2323} < \mu < \cancel{175{,}7677}$$

172,15   176,846

$\frac{1\%}{100}$
$t = 0{,}01$
$-V = n-1$
$V = 49$
2,4049
$t_{\frac{\alpha}{2}} = \cancel{1{,}2991}$

b) ¿Qué podemos afirmar con 98% de confianza sobre el tamaño posible de nuestro error si estimamos que la estatura media de todos los estudiantes de la universidad es de 174,5 centímetros?

$z = 1{,}2991$ $\quad \theta_{\bar{x}}^2 = \frac{\theta^2}{n}$

$\bar{x} = 174{,}5$

$s = 6{,}9$ $\quad \theta_{\bar{x}}^2 = \frac{s}{\sqrt{n}} = \frac{6{,}9}{\sqrt{50}} = 0{,}9758$

8 4) ¿De qué tamaño se necesita una muestra en el ejercicio 1 si deseamos tener 96% de confianza que nuestra media muestral esté dentro de 10 horas de la media real?

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \theta}{e}\right)^2 = \left(\frac{2{,}05 \cdot 40}{10}\right)^2 = \boxed{67{,}43} \approx 68 \text{ muestras}$$

$z = 2{,}05$

$\theta = 40$ horas

confianza: 96% $\quad e = 10$

10 5) Un experto en eficiencia desea determinar el tiempo promedio que toma perforar tres hoyos en cierta placa metálica. ¿De qué tamaño se necesitará una muestra para tener el 95% de confianza de que esta media muestral esté dentro de los 15 segundos de la media real? Suponga que se sabe de estudios previos que la desviación estándar es de 40 segundos.

95% $\quad \frac{2{,}5\%}{100} = 0{,}025$

$\mu = 15$

$\theta = 40s$

$z = 0{,}5 - 0{,}025 = 0{,}450$

$z = 1{,}96$

$e = 15s$

$$n = \left(\frac{1{,}96 \cdot 40}{15}\right)^2 = \boxed{27{,}32\ldots} \approx 28$$

6) Una muestra aleatoria de tamaño $n_1 = 25$ que se toma de una población normal con una desviación estándar $\sigma_1 = 5$ tiene una media $\bar{x}_1 = 80$. Una segunda muestra aleatoria de tamaño $n_2 = 36$, que se toma de una población normal diferente con una desviación estándar $\sigma_2 = 3$, tiene una media $\bar{x}_2 = 75$. Encuentre un intervalo de confianza de 95% para $\mu_1 - \mu_2$.

$n_1 = 25 \quad n_2 = 36$

$\theta_1 = 5 = 25 \quad \theta_2 = 3 = 9$

$\bar{x}_1 = 80 \quad \bar{x}_2 = 75$

$$(\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}}$$

$\mu_1 - \mu_2 = 80 - 75 = 5$

$$5 - 1{,}96 \cdot \sqrt{\frac{9}{36} + \frac{25}{25}} < \mu_1 - \mu_2 < 5 + 1{,}96 \cdot \sqrt{\frac{9}{36} + \frac{25}{25}}$$

$$2{,}81 < \mu_1 - \mu_2 < 7{,}19$$

95% $\quad 2{,}5$

$z = 1{,}96$

258

IPRI

Practica Semana 6

7.15 Se seleccionaron muestras aleatorias de tamaño n de poblaciones con las medias y varianzas dadas aquí. Encuentre la media y desviación estándar de la distribución muestral de la media muestral en cada caso:

a. $n = 36$, $\mu = 10$, $\sigma^2 = 9$
b. $n = 100$, $\mu = 5$, $\sigma^2 = 4$
c. $n = 8$, $\mu = 120$, $\sigma^2 = 1$

a) $n=36$ $\mu=10$ $\theta^2=9$

$\mu=10$ $\theta = \frac{\sqrt{9}}{\sqrt{36}} = 0,5$

b) $n=100$ $\mu=5$ $\theta^2=4$

$\mu=5$ $\theta=\frac{\sqrt{4}}{\sqrt{100}} = 0,2$

c) $n=8$ $\mu=120$ $\theta^2=1$

$\mu=120$ $\theta = \frac{\sqrt{1}}{\sqrt{8}} = 0,3536$

7.19 Una muestra aleatoria de n observaciones se selecciona de una población con desviación estándar $\sigma = 1$. Calcule el error estándar de la media (SE) para estos valores de n:

a. $n = 1$ **b.** $n = 2$ **c.** $n = 4$

$\theta = \frac{\theta}{\sqrt{n}}$

a) $\theta = \frac{1}{\sqrt{1}}$

(SE) = 1

b) (SE) = $\frac{1}{\sqrt{2}}$

(SE) = 0,7071

c) (SE) = $\frac{1}{\sqrt{4}}$

(SE) = 0,5

7.21 Se selecciona una muestra aleatoria de tamaño $n = 49$ de una población con media $\mu = 53$ y desviación estándar $\sigma = 21$.

a. ¿Cuál será la forma aproximada de la distribución muestral de $\bar{x}$?

b. ¿Cuáles serán la media y la desviación estándar de la distribución muestral de $\bar{x}$?

a) apróximadamente normal

b) $n=49$ $\mu=53$ $\theta=21$

$\mu=\boxed{53}$ $\theta = \frac{21}{\sqrt{49}} = \boxed{3}$

7.26 **Salarios de profesorado** Suponga que el profesorado de una universidad, con el rango de profesor en instituciones públicas que imparten carreras de dos años, ganan un promedio de 71 802 dólares por año[7] con una desviación estándar de 4 000 dólares. En un intento por verificar este nivel de salario se seleccionó una muestra aleatoria de 60 profesores de entre una base de datos del personal para todas las instituciones de dos años en Estados Unidos.

a. Describa la distribución muestral de la media muestral $\bar{x}$.

b. ¿Dentro de qué límites se esperaría que esté el promedio muestral, con probabilidad .95?

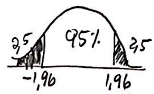
c. Calcule la probabilidad de que la media muestral $\bar{x}$ sea mayor que 73 000 dólares.

d. Si su muestra aleatoria en realidad produjo una media muestral de 73 000 dólares, ¿consideraría usted que esto es poco común? ¿Qué conclusión obtendría?

$\mu = 71\,802 \quad \theta = 4000 \quad n = 60$

a) Es opróximadamente normal

b) $P(X_1 < X < X_2) = 0{,}95$



$$-1{,}96 = \frac{X_1 - 71802}{\frac{4000}{\sqrt{60}}}$$

$$-1{,}96 \cdot \frac{4000}{\sqrt{60}} + 71802 = X_1$$

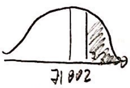
$$70.789{,}8604 = X_1$$

$$1{,}96 = \frac{X_2 - 71802}{\frac{4000}{\sqrt{60}}}$$

$$1{,}96 \cdot \frac{4000}{\sqrt{60}} + 71802 = X_2$$

$$72\,814{,}1396 = X_2$$

c) $P(X > 73000)$.



$$Z = \frac{73000 - 71802}{\frac{4000}{\sqrt{60}}}$$

$Z = 2{,}32$

$1 - P(Z \leq 2{,}32)$

$1 - 0{,}9898$

$\boxed{0{,}0102}$

d) Si sería poco común ya que se puede concluir que una gran mayoría de los profesores tienen salarios superiores al promedio que se reporto.

7.35 Se seleccionaron muestras aleatorias de tamaño n de poblaciones binomiales con parámetros poblacionales p dados aquí. Encuentre la media y la desviación estándar de la distribución muestral de la proporción muestral $\hat{p}$ en cada caso:

a. $n = 100, p = .3$

b. $n = 400, p = .1$

c. $n = 250, p = .6$

$SE = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$

a) $n = 100$, $p = 0{,}3$

$\hat{p} = 0{,}3$

$SE = \sqrt{\frac{0{,}3 \cdot 0{,}7}{100}}$

$SE = 0{,}0458$

b) $n = 400$, $p = 0{,}1$

$\hat{p} = 0{,}1$

$SE = \sqrt{\frac{0{,}1 \cdot 0{,}9}{400}}$

$SE = 0{,}0150$

c) $n = 250$, $p = 0{,}6$

$\hat{p} = 0{,}6$

$SE = \sqrt{\frac{0{,}6 \cdot 0{,}4}{250}}$

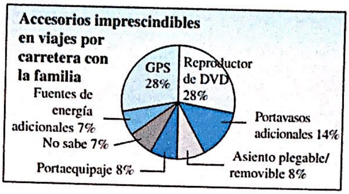
$SE = 0{,}0310$

7.44 **¡Viaje por carretera!** Los matrimonios con hijos listan un sistema GPS (28%) y un reproductor de DVD (28%) como accesorios "imprescindibles" para un viaje por carretera.[12] Suponga que se selecciona aleatoriamente una muestra de $n = 1\,000$ padres y se les pregunta cuáles dispositivos les gustaría tener para un viaje familiar por carretera. Sea $\hat{p}$ la proporción de padres en la muestra que eligen ya sea un sistema GPS o un reproductor de DVD.

$n=1000$

$z = \dfrac{p-\hat{p}}{\sqrt{\frac{p\cdot q}{n}}}$

Para sobrevivir a viajes por carretera con la familia, los padres consideran esenciales un dispositivo de navegación GPS y un reproductor de DVD



a. Si $p = .28 + .28 = .56$, ¿cuál es la distribución exacta de $\hat{p}$? ¿Cómo puede aproximar la distribución de $\hat{p}$?

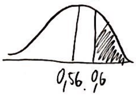
b. ¿Cuál es la probabilidad de que $\hat{p}$ exceda a .6?

c. ¿Cuál es la probabilidad de que $\hat{p}$ se encuentre entre .5 y .6?

d. ¿Un porcentaje muestral de $\hat{p} = .7$ contradice el valor reportado de .56?

a) $p = 0{,}28 + 0{,}28 = 0{,}56$
tiende a ser una distribución Normal.

b) $P(p > 0{,}6)$



$z = \dfrac{0{,}6 - 0{,}56}{\sqrt{\frac{0{,}56\cdot 0{,}44}{1000}}}$

$z = 2{,}55$

$1 - P(z > 2{,}55)$

$1 - 0{,}9946$

$0{,}0054$

c) $P(0{,}5 < p < 0{,}6)$

$z = \dfrac{0{,}5 - 0{,}56}{\sqrt{\frac{0{,}56\cdot 0{,}44}{1000}}} = -3{,}82$

$P(-3{,}82 < z < 2{,}55)$

$= P(z > 2{,}55) - P(z < -3{,}82)$

$= 0{,}9946 - 0{,}0001$

$= \boxed{0{,}9945}$

d) $\hat{p} = 0{,}7$

Si lo contradice ya que el valor de z es muy superior y se sale por completo de la tabla de distibución Normal.

$z = \dfrac{0{,}7 - 0{,}56}{\sqrt{\frac{0{,}56\cdot 0{,}44}{1000}}} = 8{,}92$