ESTADÍSTICA
GUIA No 7

INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

1. Distribuciones Muestrales

Objetivo:
El objetivo de este taller es elaborar tablas de distribuciones muestrales a partir de los datos una variable aleatoria X.

Definición:
Una muestra es la selección de un numero de observaciones de a partir de una población objeto de investigación; se denomina muestra aleatoria a una selección que sigue un método de elección impredecible. Las muestras nos permiten mediante la inferencia estadística representar los resultados de la población de donde se haya extraído, existiendo una variación en los resultados conocida como error de muestreo.

1.1 Distribución de Medias

Actividad 1:
Consideremos la variable X correspondiente a las notas de siete estudiantes en una prueba de informática. La tabla muestra los códigos de los estudiantes y las notas obtenidas por ellos:

Código	X=Notas
01	25
02	20
03	45
04	40
05	30
06	35
07	50

Consideremos todas las posibles muestras (**sin reemplazo**) de tamaño 3:

20,25,30	20,30,45	20,45,50	25,35,50	30,40,45
20,25,35	20,30,50	25,30,35	25,40,45	30,40,50
20,25,40	20,35,40	25,30,40	25,40,50	30,45,50
20,25,45	20,35,45	25,30,45	25,45,50	35,40,45
20,25,50	20,35,50	25,30,50	30,35,40	35,40,50
20,30,35	20,40,45	25,35,40	30,35,45	35,45,50
20,30,40	20,40,50	25,35,45	30,35,50	40,45,50

Definición de los estadísticos $\overline{X}$ y S^2:

Sea A= $\{x_1,x_2,x_3,...,x_n\}$ una muestra tomada de una población S= $\{x_1,x_2,x_3,...,x_n,...,x_m\}$.

a) **Media muestral**: $\overline{X}=\frac{1}{n}\sum_{k=1}^{k=n}x_k$ b) **Varianza muestral**: $S^2=\frac{1}{n-1}\sum_{k=1}^{k=n}(x_k-\overline{x})^2$.

Ejercicio 1:

Consideremos las 35 muestras resultantes, y a cada muestra, calcúlese la media $\overline{X}$:

Muestras	Medias
{ 20 , 25 , 30 }	
{ 20 , 25 , 35 }	
{ 20 , 25 , 40 }	
{ 20 , 25 , 45 }	
{ 20 , 25 , 50 }	
{ 20 , 30 , 35 }	
{ 20 , 30 , 40 }	
{ 20 , 30 , 45 }	
{ 20 , 30 , 50 }	
{ 20 , 35 , 40 }	
{ 20 , 35 , 45 }	
{ 20 , 35 , 50 }	
{ 20 , 40 , 45 }	
{ 20 , 40 , 50 }	
{ 20 , 45 , 50 }	
{ 25 , 30 , 35 }	
{ 25 , 30 , 40 }	
{ 25 , 30 , 45 }	
{ 25 , 30 , 50 }	
{ 25 , 35 , 40 }	
{ 25 , 35 , 45 }	
{ 25 , 35 , 50 }	
{ 25 , 40 , 45 }	
{ 25 , 40 , 50 }	
{ 25 , 45 , 50 }	
{ 30 , 35 , 40 }	
{ 30 , 35 , 45 }	
{ 30 , 35 , 50 }	
{ 30 , 40 , 45 }	
{ 30 , 40 , 50 }	
{ 30 , 45 , 50 }	
{ 35 , 40 , 45 }	
{ 35 , 40 , 50 }	
{ 35 , 45 , 50 }	
{ 40 , 45 , 50 }	

Ejercicio 2:

$\overline{X}$ es una variable aleatoria. En la actividad 2 se obtuvo una población de medias muestrales.

a) ¿Cuál es el recorrido de la variable aleatoria $\overline{X}$?

b) Elabore una tabla de distribución de frecuencias de la variable aleatoria $\overline{X}$.

c) Elabore una tabla de distribución de frecuencias relativas de la variable aleatoria $\overline{X}$.

Medias	Frecuencias	Frec. Relativas
25,0		
26,7		
28,3		
30,0		
31,7		
33,3		
35,0		
36,7		
38,3		
40,0		
41,7		
43,3		
45,0		

Ejercicio 3:

a) Calcule la media μ de la variable X.

b) Con base en la tabla anterior, calcule la media μ de la variable $\overline{X}$. (Agrupada)

c) ¿Qué relación hay entre las medias μ_X y $\mu_{\overline{X}}$?

d) Calcule la varianza σ^2 de la variable X, usando la tabla de X=Notas (Actividad 1).

e) Calcule la varianza σ^2 de la variable $\overline{X}$, con base en la tabla anterior (Ejercicio 2).

f) Ahora, calcule la varianza σ^2 de la variable $\overline{X}$, utilizando la fórmula $\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}\left(\frac{N-n}{N-1}\right)$

g) Compare los resultados de (e) y (f). ¿Qué relación hay entre las varianzas $\sigma^2{}_X$ y $\sigma^2{}_{\overline{X}}$?

Conclusiones:

Dada una población X, se obtienen las muestras **sin reemplazo** de tamaño n, y la población de medias muestrales $\overline{X}$, entonces:

1)	La relación entre las medias μ_X y $\mu_{\overline{X}}$ es: $\mu_X = \mu_{\overline{X}}$
2)	La relación entre las varianzas $\sigma^2{}_X$ y $\sigma^2{}_{\overline{X}}$ es: $\sigma^2{}_{\overline{X}} = \frac{\sigma^2{}_X}{n}\left(\frac{N-n}{N-1}\right)$

Actividad 2:

Consideremos la variable X correspondiente a las notas de siete estudiantes en una prueba de informática. La tabla muestra los códigos de los estudiantes y las notas obtenidas por ellos:

Códigos	O1	O2	O3	O4	O5	O6	O7
X=Notas	25	20	45	40	30	35	50

Ejercicio 1: Determinemos todas las posibles muestras (con reemplazo) de tamaño 2.

Muestas		Segundo Dato						
		25	20	45	40	30	35	50
Primer Dato	25							
	20							
	45							
	40							
	30							
	35							
	50							

(Tabla: Población de posibles muestras)

Ejercicio 2: Calculemos las medias de todas posibles muestras (con reemplazo) de tamaño 2:

Medias		Segundo Dato						
		25	20	45	40	30	35	50
Primer Dato	25							
	20							
	45							
	40							
	30							
	35							
	50							

(Tabla: Población de Medias $\overline{X}$)

Ejercicio 3: $\overline{X}$ es una variable aleatoria. Elabore la tabla de distribución de medias de $\overline{X}$.

Medias $\overline{X}$	Cantidad de Muestras
20,0	
22,5	
25,0	
27,5	
30,0	
32,5	
35,0	
37,5	
40,0	
42,5	
45,0	
47,5	
50,0	

Ejercicio 4: Calcule la media $\mu(\overline{X})$ y la varianza $\sigma^2(\overline{X})$ de la distribución muestral de $\overline{X}$.

$\mu(\overline{X}) =$	$\sigma^2(\overline{X}) =$

Ejercicio 5: Ahora, calcule la varianza $\sigma^2(\overline{X})$ utilizando la fórmula $\sigma^2_{\overline{X}} = \dfrac{\sigma^2_X}{n}$.

Conclusiones:

Dada una población X, se obtienen las muestras **con reemplazo** de tamaño n, y la población de medias muestrales $\overline{X}$, entonces:

1)	La relación entre las medias μ_X y $\mu_{\overline{X}}$ es: $\mu_X = \mu_{\overline{X}}$
2)	La relación entre las varianzas σ^2_X y $\sigma^2_{\overline{X}}$ es: $\sigma^2_{\overline{X}} = \dfrac{\sigma^2_X}{n}$

1.2 Teorema de Límite Central

Objetivo:
El objetivo de este taller es reconocer la utilidad de la distribución normal estándar en el estudio de las distribuciones muestrales y la utilidad de TLC en el desarrollo de la estadística inferencial.

Teorema:
Si X es una variable aleatoria con media μ y varianza σ^2 finita, entonces, si se toman muestras de tamaño n (suficientemente grande ($n \geq 30$)), la distribución de la variable $\overline{X}$ tiene un comportamiento aproximadamente normal con media μ y varianza $\frac{\sigma^2}{n}$.

El teorema no exige que X sea normal, y garantiza que $\overline{X}$ es aproximadamente normal, y que:
1) $\mu(\overline{X})=\mu(X)$
2) $\sigma^2(\overline{X})=\frac{\sigma^2(X)}{n}$

La construcción y formalización de este teorema empieza en 1713 con Bernoulli, luego hacen aportes Laplace, Gauss, Tchabyscheff, Markov y Liapunov, y termina con Polya y Levi en 1920.

Situaciones para resolver:

Situación 1:
El departamento de control de calidad de una fábrica de ollas de presión, establece que la presión real antes de la ruptura es una variable que se distribuye normalmente con una media $\mu=112$ libras y una varianza $\sigma^2=100$ (libras)^2. La fábrica se somete a la siguiente norma o regla: Anualmente se toma una muestra de seis ollas y se someten a una prueba, si la media de los datos recolectados es menor de 100 libras se suspende la licencia de fabricación. ¿Cuál es la probabilidad de que la licencia de esta fábrica sea suspendida injustamente?

Situación 2:
Una empacadora de camarones exporta toda pieza que pese como mínimo 35 gramos. Para decidir a qué precio compra el camarón pescado por una cooperativa, se seleccionan 50 camarones al azar y se pesan uno a uno: si el promedio de los pesos de la muestra es mayor que 33 gramos, la empacadora compra la mercancía al precio establecido para la exportación, de lo contrario, lo paga al precio del mercado nacional. ¿Cuál es la probabilidad de que les paguen el precio del mercado nacional si realmente el promedio de los pesos de los camarones es $\mu=35$ gramos y $\sigma=6$ gramos?

Situación 3:

La dieta que se utiliza en una granja para engordar pollos produce animales que pesan en promedio 1950 gramos con una desviación estándar de 220 gramos. Una franquicia de pollos rostizados ha seleccionado al azar 30 pollos y los pesa previamente antes de negociar con la granja. Si en promedio los pollos seleccionados pesan al menos 1900 gramos se realiza la negociación. ¿Cuál es la probabilidad de que no se concrete el negocio con la granja?

Situación 4:

Una procesadora de productos lácteos envasa leche en polvo en botes cilíndricos cuya etiqueta establece un contenido neto promedio de 800 gramos, empacando los productos en cajas con capacidad para 20 botes. Los inspectores oficiales hacen una visita mensual y seleccionan una caja al azar. Si el promedio de los pesos de los botes es menor de 795 gramos proceden a multar a la procesadora. La máquina envasadora coloca en los botes un promedio de 800 gramos de leche con una desviación estándar de 10 gramos. ¿Cuál es la probabilidad de que en una visita se multe injustamente a la fábrica?

Situación 5:

En una empresa empacadora de azúcar una máquina envasadora coloca en una bolsa un promedio de $\mu=500$ gramos de azúcar con una desviación estándar $\sigma=35$ gramos. Las bolsas se empaquetan en cajas para 100 unidades. ¿Cuál es la probabilidad de que el peso promedio de las bolsas de un paquete sea menor que 495 gramos? ¿Cuál es la probabilidad de que una caja de 100 bolsas pese más de 51 kg?

1.3 Distribución de Varianzas

Objetivo:
Estudiar la variabilidad de una población a partir del análisis de las variabilidades de las muestras.

Actividad 3:

Consideremos la variable X correspondiente a las notas de siete estudiantes en una prueba de informática. La tabla muestra los códigos de los estudiantes y las notas obtenidas por ellos:

Código	X=Notas
01	25
02	20
03	45
04	40
05	30
06	35
07	50

Consideremos todas las posibles muestras (**sin reemplazo**) de tamaño 3.

Ejercicio 1:
En la tabla de derecha, aparecen todas las posibles muestras. Calcúlese la varianza S^2 de cada muestra.

Muestras	Varianzas
{ 20 , 25 , 30 }	
{ 20 , 25 , 35 }	
{ 20 , 25 , 40 }	
{ 20 , 25 , 45 }	
{ 20 , 25 , 50 }	
{ 20 , 30 , 35 }	
{ 20 , 30 , 40 }	
{ 20 , 30 , 45 }	
{ 20 , 30 , 50 }	
{ 20 , 35 , 40 }	
{ 20 , 35 , 45 }	
{ 20 , 35 , 50 }	
{ 20 , 40 , 45 }	
{ 20 , 40 , 50 }	
{ 20 , 45 , 50 }	
{ 25 , 30 , 35 }	
{ 25 , 30 , 40 }	
{ 25 , 30 , 45 }	
{ 25 , 30 , 50 }	
{ 25 , 35 , 40 }	
{ 25 , 35 , 45 }	
{ 25 , 35 , 50 }	
{ 25 , 40 , 45 }	
{ 25 , 40 , 50 }	
{ 25 , 45 , 50 }	
{ 30 , 35 , 40 }	
{ 30 , 35 , 45 }	
{ 30 , 35 , 50 }	
{ 30 , 40 , 45 }	
{ 30 , 40 , 50 }	
{ 30 , 45 , 50 }	
{ 35 , 40 , 45 }	
{ 35 , 40 , 50 }	
{ 35 , 45 , 50 }	
{ 40 , 45 , 50 }	

Ejercicio 2:

S^2 es una variable aleatoria. En el ejercicio 1 se obtuvo la población de varianzas muestrales.

a) ¿Cuál es el recorrido de la variable aleatoria S^2?

b) Elabore la tabla de frecuencias de la variable aleatoria S^2.

c) Elabore la tabla de frecuencias relativas de la variable aleatoria S^2.

Varianzas	Frecuencias	Frec. Relativas
25,0		
58,3		
100,0		
108,3		
158,3		
175,0		
225,0		
233,3		
258,3		

Ejercicio 3:

a) Calcule la media $\mu = \mu(S^2)$ de la población de varianzas S^2 obtenidas en la actividad 5.

b) Calcule la varianza σ^2 de la población original X. (Actividad 4(d))

c) ¿Qué relación hay entre la varianza $\sigma^2(X)$ y la media $\mu(S^2)$?

Conclusión:

Dada una población X y su población de varianzas muestrales S^2, se concluye que: La relación entre la varianza $\sigma^2(X)$ y la media $\mu(S^2)$ es: $\sigma^2(X) = \mu(S^2)$.

1.4 Distribución de Proporciones

Objetivos:
Comprender la distribución muestral de proporciones P y calcular su media y su varianza.
Plantear y resolver problemas que involucren probabilidades y proporciones.

Definición de Proporción:
Si el tamaño de una población X es m(X) y el tamaño de un subconjunto A de X, donde sus elementos satisfacen una condición especial, es m(A), entonces, la proporción de A con respecto a X es $\rho = \frac{m(A)}{m(X)}$.

Si se selecciona al azar una muestra M de tamaño n, ¿Qué parte de los elementos de A están es M? ¿En promedio, ¿cuántos elementos de A esperamos encontrar en una muestra M de tamaño n?

Actividad 4:
Se tiene un lote de 20 artículos, entre los cuales hay 7 artículos defectuosos y 13 artículos buenos. Se selecciona al azar una muestra de 6 artículos, **sin reemplazo**.

Distinguimos los artículos de la siguiente manera:
Artículos defectuosos: {D1, D2, … , D7}. Artículos buenos: {B1, B2, … , B13}.

Ejercicio 1:
Escriba ejemplos de cuatro posibles muestras:

Muestra 1:= $\{\qquad\qquad\}$ Muestra 2:= $\{\qquad\qquad\}$

Muestra 3:= $\{\qquad\qquad\}$ Muestra 4:= $\{\qquad\qquad\}$

Ejercicio 2:

- En total ¿Cuántas muestras de tamaño 6, hay? ________________
- En cuántas muestras hay 0 artículos defectuosos? ________________
- En cuántas muestras hay 1 artículo defectuoso? ________________
- En cuántas muestras hay 2 artículos defectuosos? ________________
- En cuántas muestras hay 3 artículos defectuosos? ________________
- En cuántas muestras hay 4 artículos defectuosos? ________________
- En cuántas muestras hay 5 artículos defectuosos? ________________
- En cuántas muestras hay 6 artículos defectuosos? ________________

Ejercicio 3:
Complete la siguiente tabla: ¿En cuántas muestras hay x artículos defectuosos, para x=0, 1, 2, 3, 4, 5, 6?

Artículos defectuosos x	Artículo buenos	Cantidad de muestras con x artículos defectuosos

Ejercicio 4:
Elabore la distribución de proporciones. Si las muestras son de tamaño 6, y la muestras hay x artículos defectuosos, ¿Cuál es proporción de artículos defectuosos en la muestra?

Artículos defectuosos x	Proporción x de artículos defectuosos en las muestras	Cantidad de muestras con x artículos defectuosos

Ejercicio 5:

Las proporciones de las muestras forman una población, y constituyen una variable aleatoria P. Calcule la media de la variable P.

La media de la distribución muestral de proporciones es: $\mu_p =$ ____________________

Preguntas:

- ¿Cuál es la proporción de artículos defectuosos en el lote?
- ¿Qué relación hay entre la proporción ρ de artículos defectuosos del lote y la media μ_p de la variable aleatoria P?

Ejercicio 6:

Calcule la varianza σ_p^2 de la distribución muestral de proporciones: (Tabla - Ejercicio 4)

$\sigma_p^2 =$

Definición:
Si la población es infinita **o** se realiza el muestreo con reemplazamiento con muestras de tamaño n, entonces $\sigma_p^2 = \dfrac{\rho(1-\rho)}{n}$.

Definición:
Si la población es finita de tamaño N **y** se realiza muestreo sin reemplazamiento con muestras de tamaño n, entonces $\sigma_p^2 = \dfrac{\rho(1-\rho)}{n}\left(\dfrac{N-n}{N-1}\right)$.

Ejercicio 7:
Calcule la varianza de los datos de la actividad 1, utilizando cada una de las fórmulas anteriores:

a) Utilizando la fórmula $\sigma_p^2 = \dfrac{\rho(1-\rho)}{n}$, la varianza es: ____________________

b) Utilizando la fórmula $\sigma_p^2 = \dfrac{\rho(1-\rho)}{n}\left(\dfrac{N-n}{N-1}\right)$, la varianza es: ____________________

c) La varianza calculada en el ejercicio 6, ¿con cuál de las dos varianzas calculadas en el ejercicio 7 coincide? ¿Por qué?

____

Actividad 5:

Se tiene un lote de 12 artículos, entre los cuales hay 4 artículos defectuosos y 8 artículos buenos. Se selecciona al azar una muestra de 5 artículos, **con reemplazo**.

Distinguimos los artículos de la siguiente manera:
Artículos defectuosos: D1, D2, D3, D4.
Artículos buenos: B1, B2, B3, B4, B5, B6, B7, B8.

Ejercicio 1:

Escriba ejemplos de cuatro posibles muestras:

Muestra 1:= $\{ \qquad \}$ Muestra 2:= $\{ \qquad \}$

Muestra 3:= $\{ \qquad \}$ Muestra 4:= $\{ \qquad \}$

Ejercicio 2:

- En total ¿Cuántas muestras hay de tamaño 5? ________________
- En cuántas muestras hay 0 artículos defectuosos? ________________
- En cuántas muestras hay 1 artículo defectuoso? ________________
- En cuántas muestras hay 2 artículos defectuosos? ________________
- En cuántas muestras hay 3 artículos defectuosos? ________________
- En cuántas muestras hay 4 artículos defectuosos? ________________
- En cuántas muestras hay 5 artículos defectuosos? ________________

Ejercicio 3:

Complete la siguiente tabla: ¿En cuántas muestras hay x artículos defectuosos, para x=0, 1, 2, 3, 4, 5?

Artículos defectuosos x	Artículo buenos	Cantidad de muestras con x artículos defectuosos

Ejercicio 4:
Elabore la distribución de proporciones. Si las muestras son de tamaño 5, y en la muestra hay x artículos defectuosos, ¿Cuál es proporción de artículos defectuosos en la muestra?

Artículos defectuosos x	Proporción x de artículos defectuosos en las muestras P=p	Cantidad de muestras con x artículos defectuosos Fr(P=p)

Ejercicio 5:

Las proporciones de las muestras forman una población, y constituyen una variable aleatoria P. Calcule la media de la variable P.

La media de la distribución muestral de proporciones es: $\mu_p =$ ____________________

Preguntas:

- ¿Cuál es la proporción de artículos defectuosos en el lote?
- ¿Qué relación hay entre la proporción ρ de artículos defectuosos del lote y la media μ_p de la variable aleatoria P?

Ejercicio 6:
Calcule la varianza σ_p^2 de la distribución muestral de proporciones: (Tabla - Ejercicio 4)

$\sigma_p^2 =$

Ejercicio 7:
Calcule la varianza para los datos de la actividad 5:

a) Utilizando la fórmula $\sigma_p^2 = \frac{\rho(1-\rho)}{n}$. La varianza es: ____________________

b) Utilizando la fórmula $\sigma_p^2 = \frac{\rho(1-\rho)}{n}\left(\frac{N-n}{N-1}\right)$. La varianza es: ____________________

c) La varianza calculada en el ejercicio 6, ¿con cuál de las dos varianzas calculadas en (a) y (b) coincide? ¿Por qué?

____

Nota:
Para calcular la probabilidad de que una muestra de una población satisfaga cierta condición relacionada con proporciones de la población y la muestra, se usa la distribución normal haciendo la **tipificación de la proporción p** previamente, usando la expresión:

a) $z = \dfrac{p-\rho}{\sqrt{\dfrac{\rho(1-\rho)}{n}}} = \dfrac{p-\rho}{\sigma_p}$, si el muestreo se hace con reemplazo o la población es infinita.

b) $z = \dfrac{p-\rho}{\sqrt{\dfrac{\rho(1-\rho)}{n}}\sqrt{\dfrac{N-n}{N-1}}} = \dfrac{p-\rho}{\sigma_p}$, si el muestreo se hace sin reemplazo o la población es N-finita.

Problemas propuestos:

Problema 1:
Se ha determinado que el 80% de los estudiantes de la facultad de ingeniería de una universidad bajan los textos guías de Internet en formato PDF, lo cual es patrocinar la piratería de libros. Se toma una muestra de 250 estudiantes de esta facultad. ¿Cuál es la probabilidad de que máximo el 75% de los estudiantes de la muestra realicen esta actividad ilícita?

Ayuda:
- La población de una universidad es muy grande, estadísticamente se considera infinita.
- Tipificar p=0.75

Problema 2:
El 40% de los aspirantes que se presentan en la carrera de medicina de una universidad estatal logran superar el primer requisito que consiste en aprobar un examen de conocimientos específicos en biología y química. Si se selecciona una muestra de 30 aspirantes, ¿Cuál sería la probabilidad de que al menos el 50% de ellos aprobaran el examen específico?

Ayuda:
- La población de aspirantes es muy grande, estadísticamente se considera infinita.
- Tipificar p=0.50

Problema 3:
En las estadísticas de una universidad se tiene que el 60% de los estudiantes que ingresan reprueban una asignatura en el primer semestre. Se selecciona una muestra aleatoria de 60 estudiantes de primer semestre.

a) ¿Cuál es la probabilidad de que mas del 70% de los estudiantes seleccionados haya reprobado una asignatura?
b) ¿Cuál es la probabilidad de que la cantidad de estudiantes reprobados esté entre el 60% y el 75%, inclusive?
c) ¿Para qué valor de x, se cumple la desigualdad $P(p \leq x) = 0.05$?
d) ¿Para qué valores de a y b simétricos con respecto a p=0.62 se cumple $P(a \leq p \leq b) = 0.95$?

Problema 4:
Los 100 aspirantes al grado de bachiller de un colegio deben presentarse al ejército para definir su situación militar. Los jóvenes son sometidos a un sorteo que consiste en seleccionar una pelota de una urna donde el 40% de las pelotas son blancas y el 60% son negras. Si el aspirante saca una pelota blanca debe volver a presentarse en una próxima fecha, de lo contrario queda exento.
a) ¿Cuál es la probabilidad de que más del 50% de los aspirantes deba volver a presentarse?
b) ¿Cuál es la probabilidad de más del 62.5% de los aspirantes queden exentos?

Ayuda:
- La población de bachilleres que se presenta es muy grande, estadísticamente se considera infinita.
- Tipificar p=0.5 y p=0.625

Problema 5:
Dos compañías consultoras A y B han sido contratadas para evaluar la afirmación hecha por un partido político en el sentido de que “El 40% de los electores votará por ese partido político en las próximas elecciones”. De un muestreo aleatorio realizado por la compañía A resultó que 190 de 500 personas encuestadas manifestó la intención de votar a favor de dicho partido y el resto de votar en contra. Por su parte, la compañía B encuestó a 800 personas, en las cuales 304 tienen intención de votar a favor del partido y el resto manifestó votar en contra. Analice los resultados de los dos sondeos y determine cuál de los dos es menos probable y por lo tanto arroja más dudas con respecto a la afirmación hecha por el partido político.

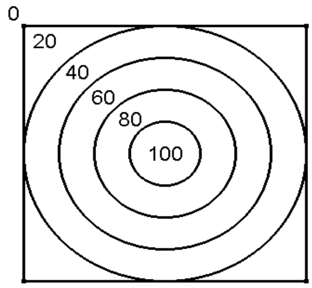
2. Estimación Estadística

2.1 Estimación Puntual

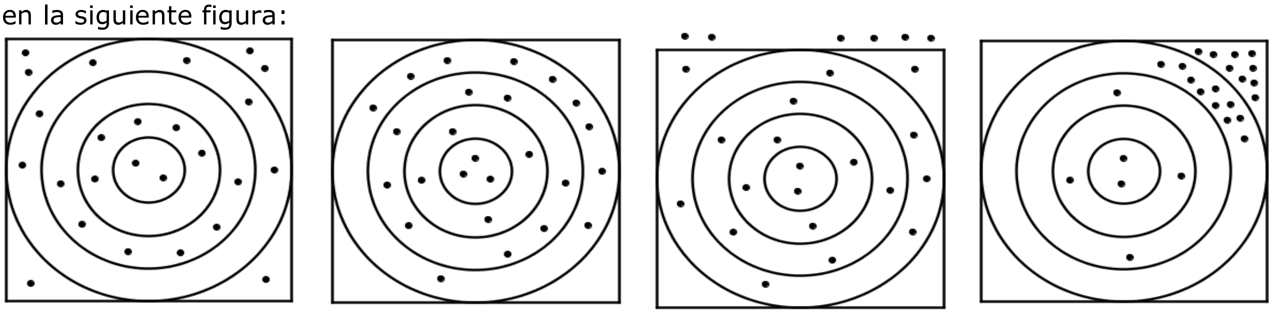
Objetivo:
El objetivo de la estimación puntual es usar una muestra para obtener números (estimaciones puntuales) que sean la mejor representación de los verdaderos parámetros de la población.

1. Sea θ un parámetro (Medida descriptiva en una población).
2. Un estimador $\hat{\theta}$ es un estadístico (Medida descriptiva en una muestra) que se usa para estimar un parámetro, a través de una variable aleatoria.
3. Los valores que toma $\hat{\theta}$ en cada muestra se llaman estimaciones puntuales.

Actividad 1:
Cuatro tiradores desconocidos llegan a una competencia en donde cada uno debe lanzar el dardo 25 veces. Dependiendo de la zona en donde quede el dardo recibirá una puntuación según la siguiente figura:



Al terminar de hacer sus lanzamientos cada tirador, se tomó una foto del objeto de tiro al blanco para ser analizada por los jueces y decidir cuál de los cuatro es el mejor. Las fotos se muestran en la siguiente figura:



Tirador 1 Tirador 2 Tirador 3 Tirador 4

Ejercicio 1:
Llene la siguiente tabla, en la primera se colocan los puntajes, y en las demás columnas se colocan la cantidad de dardos que cayó en cada zona para cada tirador.

Puntajes: X_k	**Tirador 1: F_k**	**Tirador 2: F_k**	**Tirador 3: F_k**	**Tirador 4: F_k**
0				
20				
40				
60				
80				
100				
Total				

Ejercicio 2:
Los jueces, para determinar cuál es el mejor de los cuatro competidores, calcula la media de los puntajes obtenidos por cada tirador. Calcule la media $\overline{X}$ de los puntajes X_k obtenidos por cada tirador, y escríbalos en la siguiente tabla:

Competidor	Media de los puntajes $\overline{X}$
Tirador 1	$\overline{x_1}$ =
Tirador 2	$\overline{x_2}$ =
Tirador 3	$\overline{x_3}$ =
Tirador 4	$\overline{x_4}$ =

Preguntas:

1. Desde el inicio de la competencia se podía predecir la media que obtendría cada competidor?

2. ¿Se puede decir que la media $\overline{X}$ es una variable aleatoria?

3. ¿Es $\overline{X}$ una variable aleatoria muestral?

De acuerdo con la última tabla:

4. ¿Cuál es el mejor de los cuatro competidores? (mayor exactitud)

5. ¿Cuál es el peor de los cuatro competidores?

6. ¿Cuál es el más preciso de los cuatro competidores?

Nota:

- La variable $\overline{X}$ es un estimador. $\hat{\theta} = \overline{X}$.
- Cada valor $\overline{x_k}$, es decir, la media de cada jugador es una Estimación puntual.

Actividad 2:
Con motivo de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea conocer el número de hijos que tienen los 200 obreros que ahí laboran. Para esto, se entrevistan a todos los obreros en orden alfabético, como aparecen en la nómina, obteniéndose los resultados que se muestran en la tabla que se muestra en la siguiente página.

Ejercicio 1:
Seleccione una muestra **con reemplazo** de tamaño n=30. Considere la variable X = "Número de hijos por obrero".

Ejercicio 2:
Elabore una tabla de frecuencias absolutas y relativas para su muestra, clasificando los obreros de acuerdo a la variable X.

Ejercicio 3:
Calcule $\overline{x}$ de su muestra seleccionada

Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos	Número del Obrero	Número de Hijos
1	8	51	6	101	8	151	4
2	0	52	4	102	0	152	3
3	5	53	7	103	5	153	4
4	3	54	8	104	3	154	5
5	6	55	5	105	4	155	0
6	5	56	7	106	8	156	8
7	0	57	4	107	2	157	4
8	4	58	8	108	0	158	5
9	0	59	8	109	6	159	5
10	8	60	7	110	3	160	8
11	7	61	0	111	0	161	4
12	6	62	2	112	5	162	3
13	1	63	0	113	4	163	8
14	4	64	0	114	2	164	0
15	0	65	2	115	4	165	8
16	7	66	3	116	4	166	0
17	4	67	4	117	2	167	2
18	5	68	8	118	5	168	4
19	8	69	7	119	6	169	8
20	0	70	3	120	8	170	1
21	2	71	5	121	6	171	8
22	8	72	1	122	0	172	2
23	4	73	5	123	7	173	5
24	5	74	3	124	3	174	8
25	6	75	4	125	7	175	2
26	5	76	7	126	0	176	4
27	3	77	6	127	3	177	6
28	8	78	2	128	0	178	3
29	7	79	3	129	5	179	6
30	1	80	0	130	4	180	5
31	4	81	0	131	7	181	3
32	6	82	6	132	3	182	2
33	0	83	2	133	5	183	7
34	5	84	6	134	5	184	2
35	4	85	4	135	6	185	7
36	1	86	4	136	0	186	3
37	1	87	0	137	3	187	2
38	4	88	8	138	5	188	6
39	4	89	8	139	2	189	4
40	3	90	5	140	3	190	6
41	5	91	4	141	3	191	8
42	5	92	4	142	0	192	5
43	7	93	8	143	4	193	3
44	8	94	3	144	8	194	5
45	7	95	3	145	3	195	5
46	8	96	1	146	3	196	3
47	6	97	7	147	1	197	6
48	4	98	8	148	7	198	6
49	3	99	8	149	5	199	4
50	3	100	6	150	4	200	4

Tabla: Obreros y Cantidad de hijos por obrero**.**

Preguntas finales:

Pregunta 1: ¿Cuál es el estimador?

Pregunta 2: ¿Cuál es la estimación puntual?

Pregunta 3: ¿Puede elaborarse la distribución muestral para $\bar{x}$?

2.2 Estimación po Intervalos

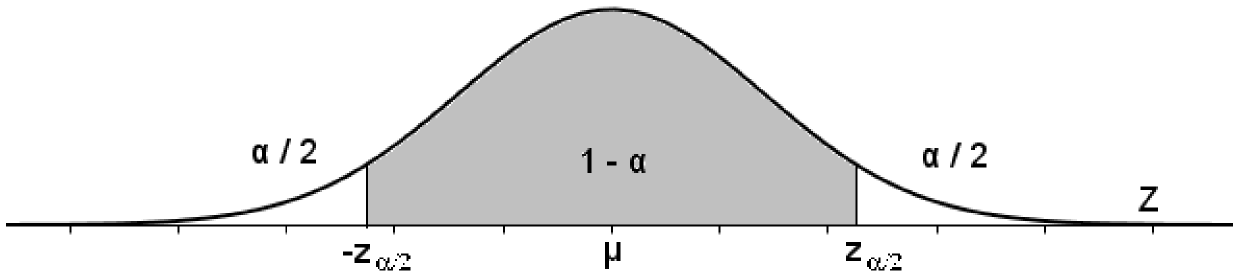
2.2.1 Estimación de la media por intervalos - Varianza conocida

Objetivo:
El objetivo de la estimación por intervalos es obtener los límites entre los cuales se encuentra el verdadero valor del parámetro de una población con un cierto nivel de confianza.

Teoría:
Un intervalo probabilístico (a,b) para el cual la probabilidad de que el intervalo contenga al parámetro θ (Medida descriptiva para la población) sea igual a 1-α se llama Intervalo de confianza al (1-α)x100% para estimar al parámetro θ. En tal caso se escribe $P(a< \theta <b) = 1- \alpha$.

El valor α se denomina Nivel de significación, y el valor (1-α) se denomina Nivel de confianza.

Si la media μ de la población es desconocida y la varianza σ^2 es conocida, se puede determinar un intervalo (a,b) talque $P(a< \mu <b) = 1- \alpha$. El valor de α se reparte simétricamente entre las dos colas de la curva Normal estándar, quedando (1-α) en la región central y $\alpha/2$ en cada cola de la curva. El problema se reduce a resolver la desigualdad $P\left(\left|\mu-\overline{x_0}\right|<Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$, donde $\overline{x_0}$ es una estimación puntual de μ y $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ es el **margen de error**. Los valores $-Z_{\alpha/2}$ y $Z_{\alpha/2}$ se llaman **coeficientes de confianza**.



Actividad 1:

Las calificaciones finales de 250 participantes en unas Olimpiadas Internacionales de Matemáticas y Estadística se presentan en la tabla 1 con sus respectivos códigos. Las pruebas se calificaron de 0 a 10 puntos. Se sabe que la varianza fue de 10.50 puntos cuadrados.

Ejercicio 1:
Seleccione una muestra con reemplazo de tamaño n=40. Considere la variable X = "Nota final del participante".

Ejercicio 2:
Calcule una estimación puntual $\overline{x_0}$ a partir de su muestra seleccionada.

Ejercicio 3:
Calcule un intervalo de confianza para la media poblacional μ con una probabilidad del 95%, tomando la estimación puntual $\overline{x_0}$ calculada por usted en el punto anterior.

Ejercicio 4:

Use la Tabla de la Normal Estándar Z, para encontrar los extremos del intervalo ($-Z_{\alpha/2}, Z_{\alpha/2}$) talque P{ ($-Z_{\alpha/2}, Z_{\alpha/2}$) } = 0.95

Explicación:

La desigualdad $\left| \mu - \overline{x_0} \right| < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ es equivalente a $\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. Por lo tanto,

$P\left(\left| \mu - \overline{x_0} \right| < Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$ = 0.95 es equivalente a $P\left(\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 0.95$

Ahora, comparando P(a < μ < b) = 0.95 y $P\left(\overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \overline{x_0} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 0.95$ se tiene

$a = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ y $b = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

Ejercicio 5:

¿Cuáles son los valores de $a = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ y $b = \overline{x_0} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$?

Ejercicio 6:
Finalmente, ¿Cuál es el intervalo de confianza (a,b) ?

Ejercicio 7:
Con base en el desarrollo anterior, complete la siguiente frase:

La probabilidad de que el verdadero valor del parámetro ______ se encuentre en el intervalo ______________ es del __________.

Ejercicio 8:
¿Cuál será el intervalo de confianza, si se toma un nivel de significación del 10% ?

Ejercicio 9:
¿Cuál será el intervalo de confianza si toma un nivel de confianza del 96.6%

Ejercicio 10:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de confianza del 95%, con un margen de error de 1 punto?

Ejercicio 11:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de confianza del 95%, con un margen de error de 2 puntos?

Ejercicio 12:
¿Cuál debe ser el tamaño de la muestra para obtener un intervalo con un nivel de significación del 10%, con un margen de error de 2 puntos?

Ejercicio 13:
Si se toma una muestra de tamaño n=20 y se desea obtener un intervalo de confianza con un nivel de confianza del 90%, entonces:
a) ¿Cuáles son los coeficientes de confianza?
b) ¿Cuál es la longitud del intervalo de confianza?

Ejercicio 14:
Ahora, seleccione una muestra sin reemplazo de tamaño n=30. Considere la variable X = "Nota final del participante".
a. Elabore la Tabla de distribución de X.
b. Calcule una estimación puntual de la media poblacional μ.
c. ¿Cuál es el margen de error para un nivel de confianza del 90%,
d. ¿Cuál es el intervalo de confianza?

Ejercicio 15:
Ahora, seleccione una muestra sin reemplazo de tamaño n=40. Considere la variable X = "Nota final del participante".
a. Elabore la Tabla de distribución de X.
b. Calcule una estimación puntual de la media poblacional μ.
c. ¿Cuál es el margen de error para un nivel de significación del 10%,
d. ¿Cuál es el intervalo de confianza?

Código	Nota	Código	Nota	Código	Nota	Código	Nota	Código	Nota
1	3	51	9	101	1	151	8	201	2
2	10	52	5	102	10	152	7	202	1
3	5	53	2	103	7	153	3	203	0
4	10	54	0	104	0	154	3	204	4
5	7	55	10	105	5	155	10	205	4
6	5	56	4	106	6	156	1	206	10
7	9	57	1	107	0	157	6	207	2
8	10	58	7	108	8	158	5	208	3
9	4	59	5	109	7	159	5	209	5
10	9	60	7	110	1	160	7	210	10
11	3	61	4	111	6	161	3	211	1
12	7	62	1	112	7	162	1	212	10
13	2	63	9	113	3	163	4	213	3
14	10	64	8	114	0	164	1	214	8
15	1	65	1	115	4	165	0	215	1
16	9	66	9	116	10	166	8	216	6
17	3	67	2	117	4	167	5	217	10
18	8	68	10	118	10	168	7	218	9
19	8	69	1	119	7	169	9	219	5
20	3	70	0	120	3	170	10	220	7
21	9	71	3	121	7	171	0	221	8
22	9	72	5	122	3	172	10	222	0
23	8	73	0	123	2	173	8	223	3
24	2	74	1	124	9	174	7	224	9
25	5	75	2	125	9	175	3	225	4
26	3	76	8	126	3	176	2	226	0
27	0	77	8	127	9	177	9	227	5
28	0	78	1	128	8	178	0	228	10
29	4	79	6	129	0	179	8	229	8
30	2	80	10	130	7	180	8	230	3
31	5	81	3	131	6	181	1	231	8
32	1	82	0	132	6	182	2	232	1
33	4	83	5	133	7	183	2	233	7
34	8	84	0	134	5	184	8	234	6
35	3	85	3	135	6	185	2	235	5
36	9	86	2	136	6	186	8	236	0
37	9	87	6	137	1	187	6	237	0
38	7	88	7	138	10	188	7	238	2
39	8	89	8	139	4	189	3	239	2
40	2	90	8	140	4	190	7	240	2
41	6	91	3	141	2	191	10	241	2
42	4	92	2	142	4	192	0	242	9
43	6	93	8	143	8	193	10	243	6
44	9	94	7	144	0	194	0	244	6
45	8	95	1	145	8	195	9	245	0
46	7	96	3	146	10	196	10	246	2
47	10	97	3	147	7	197	5	247	6
48	7	98	7	148	9	198	3	248	1
49	1	99	1	149	7	199	3	249	4
50	1	100	9	150	5	200	2	250	1

Tabla 1: Notas finales de 250 participantes.

2.2.2 Estimación de la media por intervalos - Varianza desconocida

Objetivo:

El objetivo de este taller es realizar la estimación por intervalos de la media poblacional cuando no se conoce la varianza de la población. Se usará la distribución T-Student y se trabajará con la varianza de la muestra. También veremos la utilidad del teorema central del límite.

Distribución T-Student:

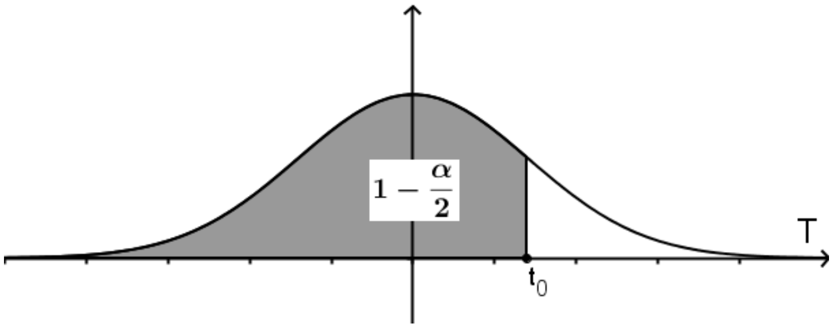
La distribución T fue creada por Goset y formalizada por Fisher. Es una familia de distribuciones semejantes a la distribución normal estándar Z. Cada miembro de la familia T está determinado por el tamaño de la muestra n. Para valores pequeños de n, se tiene que $\sigma^2(T)>\sigma^2(Z)$. A medida que aumenta el valor de n, las gráficas de f(T) se acercan a la gráfica de f(Z), esto permite calcular probabilidades de la distribución T-Student con la distribución Z-Estándar.

Supóngase que se toma una muestra de tamaño n. Entonces:

1) Cuando la varianza de X se conoce, se usa la expresión $Z=\dfrac{\overline{X}-\mu_X}{\sigma_X/\sqrt{n}}$ y la tabla de la distribución normal estándar Z.
2) Cuando la varianza de X no se conoce, se usa la expresión $T=\dfrac{\overline{X}-\mu_X}{S/\sqrt{n}}$ y la tabla de la distribución T-Student. S es la desviación estándar de la muestra $S=\sqrt{\dfrac{\sum_{k=1}^{k=n}(x_k-\overline{x})^2}{n-1}}$ con (n-1) grados de libertad.
3) La función T-Student para n grados de libertad es $f_n(x)=\dfrac{\left(\dfrac{x^2}{n}+1\right)^{-\left(\frac{n+1}{2}\right)}\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\;\Gamma\left(\dfrac{n}{2}\right)}$, donde $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$.

Estimación de μ en muestras pequeñas (n<30):

Si X es una variable distribuida normalmente con media μ y varianza desconocidas, S es la desviación estándar muestral de una muestra de tamaño n, entonces, para estimar el intervalo que contiene a la media μ de X con una seguridad del $(1-\alpha)\%$, la tabla T-Student proporciona el valor de t_0 para el cual se acumula el $\left(1-\dfrac{\alpha}{2}\right)\%$ del área bajo la curva $f_n(x)$. Es decir, $P\left(\left|\mu-\overline{x}\right|<t_0\cdot\dfrac{S}{\sqrt{n}}\right)=(1-\alpha)\%$. El intervalo de confianza es $\overline{x}-t_0\cdot\dfrac{S}{\sqrt{n}}<\mu<\overline{x}+t_0\cdot\dfrac{S}{\sqrt{n}}$.



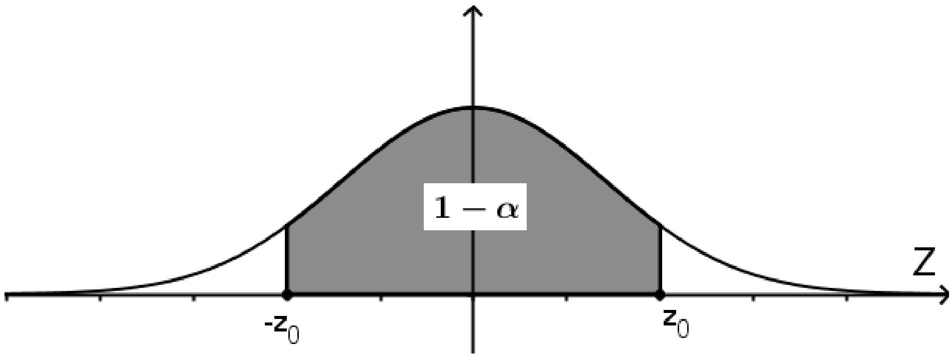
Estimación de μ en muestras grandes (n>30):

Cuando el tamaño de la muestra es n suficientemente grande, el teorema central del límite garantiza que:

1) $\overline{X}$ es aproximadamente normal sin necesidad de que la variable X sea normal.

2) $\mu(\overline{X})=\mu(X)$ y $\sigma^2(\overline{X})=\dfrac{\sigma^2(X)}{n}$.

Por lo tanto se puede utilizar la tabla de la normal estándar Z para estimar el intervalo que contiene a la media μ con una seguridad de $(1-\alpha)\%$ buscando el valor de z_0 para el cual se acumula simétricamente el $(1-\alpha)\%$ del área bajo la normal estándar. Es decir, $P\left(0<\left|\mu-\overline{x}\right|<z_0\cdot\dfrac{S}{\sqrt{n}}\right)=\left(\dfrac{1-\alpha}{2}\right)\%$. El intervalo de confianza es $\overline{x}-z_0\cdot\dfrac{S}{\sqrt{n}}<\ \mu\ <\overline{x}+z_0\cdot\dfrac{S}{\sqrt{n}}$.



Ejercicio 1:
Encuentre el valor de t_0, para el cual se cumplen las siguientes igualdades:
Nota: El índice de T corresponde al tamaño de la muestra.

a) $P(T_8<t_0)=0.75$
b) $P(T_{10}<t_0)=0.80$
c) $P(T_{12}<t_0)=0.85$
d) $P(T_{15}<t_0)=0.90$
e) $P(T_{20}<t_0)=0.95$

Ejercicio 2:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Use la tabla T-Student, para completar la siguiente tabla:

n	$\bar{x}$	$1-\alpha$	S^2	t_0
10	8.5	0.80	14.40	
15	7.5	0.90	20.25	
20	9.0	0.95	12.25	
25	10.5	0.90	72.25	
30	11.5	0.99	81.00	

Ejercicio 3:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Use la tabla T-Student, para completar la siguiente tabla:

n	$\bar{x}$	$1-\alpha$	S^2	Intervalo de confianza
10	8.5	0.80	14.40	
15	7.5	0.90	20.25	
20	9.0	0.95	12.25	
25	10.5	0.90	72.25	
30	11.5	0.99	81.00	

Ejercicio 4:
De una población X, donde no se conoce la media y la varianza, se toman muestras de tamaño n y varianza S^2. Se quiere estimar $\mu(X)$. Compare los intervalos de confianza obtenidos usando la tabla T-Student con los obtenidos usando la tabla normal estándar. Complete la siguiente tabla:

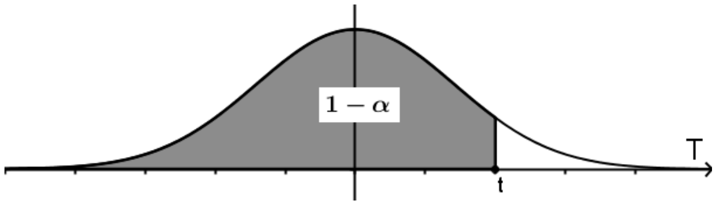
n	$\bar{x}$	$1-\alpha$	S^2	Intervalo de confianza con T	Intervalo de confianza con Z
35	8.5	0.80	14.40		
40	7.5	0.90	20.25		
45	9.0	0.95	12.25		
50	10.5	0.90	72.25		
55	11.5	0.99	81.00		

Ejercicio 5:
Un grupo de 400 estudiantes presentó una prueba de Informática calificada de 0 a 10 puntos. Se toma una muestra de 40 estudiantes. En la tabla se muestran las notas de la muestra.

4	2	10	7	2	4	3	7	1	6
2	9	10	5	9	10	10	5	5	5
1	3	6	8	5	1	4	2	5	5
4	1	0	8	9	6	8	1	4	7

Estime en que intervalo se encuentra el promedio de todo el grupo con una seguridad del 80%, 90% y 95%, utilizando las distribuciones T-Student y Z-Normal estándar.

TABLA DE LA DISTRIBUCIÓN T-STUDENT



	$1-\alpha$							
n	**0.75**	**0.80**	**0.85**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

3. Pruebas de Hipótesis

Objetivo:
El objetivo de este taller es evaluar hipótesis o proposiciones acerca de los parámetros de una población con información proporcionada por una muestra.

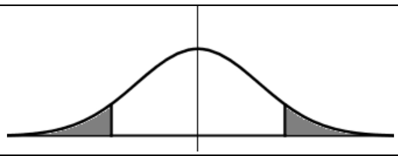
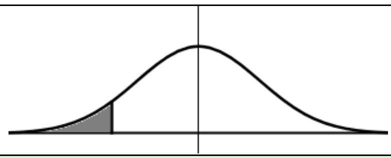
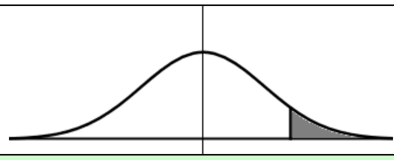
Prueba o contraste de hipótesis: Es un procedimiento o método estadístico que sirve para validar (aceptar) o invalidar (rechazar) una proposición o hipótesis acerca de un parámetro de la población. La información o evidencias requeridas para tomar la decisión es proporcionada por una muestra. Los pasos de este método o proceso son los siguientes:

Paso 1: Formular la hipótesis de investigación o **hipótesis alternativa** (Ha) y la **hipótesis nula** (Ho). El análisis conlleva a aceptar o rechazar la hipótesis nula. Las dos hipótesis se pueden formular de tres maneras:

$Ha: \theta m \neq \theta p$ $Ho: \theta m = \theta p$	$Ha: \theta m < \theta p$ $Ho: \theta m \geq \theta p$	$Ha: \theta m > \theta p$ $Ho: \theta m \leq \theta p$

Paso 2: Determinar **nivel de significación** α $(0 < \alpha < 1)$, el cual es el máximo error dispuesto a aceptar para validar la hipótesis alternativa, y se reparte entre las colas de la distribución normal estándar o se acumula en una de las colas.

Paso 3: Determinar la zona de aceptación y rechazo de la hipótesis nula Ho. Si θp es un parámetro de la población (por ej: μ) y θm es el **estadístico de prueba** o una estimación puntual calculada en una muestra (por ej: $\bar{x}$), entonces la zona de rechazo de Ho es la zona sombreada con área α (Ver figuras). La zona sombreada es determinada por Ha. La zona de aceptación de Ho es la zona no sombreada. El valor finito z_α o $z_{\alpha/2}$ donde se inician las colas se llama **valor crítico** o teórico.

		
$Ha: \bar{x} \neq \mu$ $Ho: \bar{x} = \mu$	$Ha: \bar{x} < \mu$ $Ho: \bar{x} \geq \mu$	$Ha: \bar{x} > \mu$ $Ho: \bar{x} \leq \mu$
Contraste con dos colas	Contraste con cola izquierda	Contraste con cola derecha

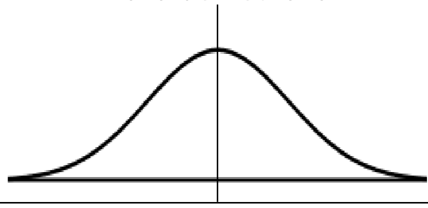
Paso 4: Elegir la **fórmula pivotal** de acuerdo al tamaño de la muestra: Si la varianza de la población es conocida y la muestra es grande (n>30), entonces $Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}}$. Si la varianza es desconocida y la muestra es pequeña $(n \leq 30)$, entonces $T = \frac{\bar{X} - \mu_X}{S / \sqrt{n}}$. El valor obtenido Z_{cal} o T_{cal} se compara con z_α o $z_{\alpha/2}$ y se observa en que región cae para tomar la decisión final.

SITUACIONES PARA RESOLVER

Situación 1:

Un examen para ingreso a la carrera de ingeniería civil ofrecida por una universidad del estado es presentado por 720 aspirantes. Se toma una muestra cuyo tamaño es equivalente al 5% del total de aspirantes. Para esta muestra el puntaje promedio fue 5.6 puntos con una varianza de 5.76 puntos^2. El decano de la carrera dice que tiene una seguridad del 95% de que el puntaje promedio en toda la prueba fue de 6.0 puntos. ¿El decano tiene la razón?

n =	$\overline{X} =$	$\sigma =$

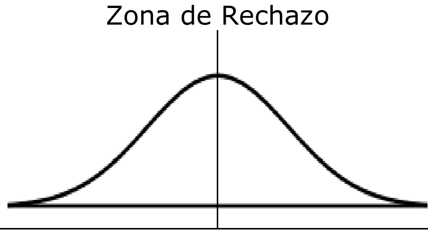
Ha:	Ho:	Zona de Rechazo 

Valor crítico Z_0:	Z calculado:	Decisión:

Situación 2:

El director financiero de la aerolínea española Iberia asegura que el costo promedio del tiquete de avión entre Madrid y Lisboa es como máximo de 120 euros con una varianza de 1600 euros^2. Una encuesta realizada a 100 viajeros frecuentes indica que el costo promedio del tiquete es 128 euros. ¿Se puede aceptar, con un nivel de significación del 1%, la afirmación del directivo?

n =	$\overline{X} =$	$\sigma =$

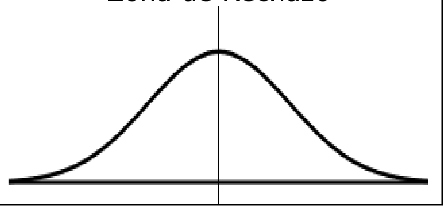
Ha:	Ho:	Zona de Rechazo 

Valor crítico Z_0:	Z calculado:	Decisión:

Situación 3:

La vida útil de las bombillas de 100 W de la fábrica Philips está distribuida normalmente con una desviación estándar de 120 horas. El fabricante garantiza que la vida media es mínimo de 800 horas. Se selecciona una muestra aleatoria de 50 bombillas de un lote particular, se instalan y se registra la duración de cada una, obteniéndose una vida media de 750 horas. Con un nivel de significación del 1%, ¿habría que rechazar el lote por no cumplir la garantía?

n =	$\overline{X}$ =	σ =

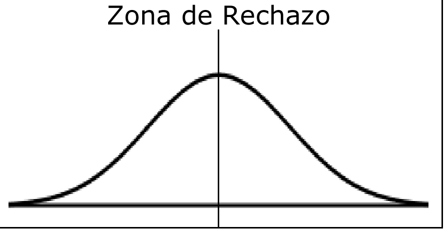
Ha:	Ho:	Zona de Rechazo
		

Valor crítico Z_0:	Z calculado:	Decisión:

Situación 4:

El departamento de control de calidad una fábrica de baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Se sabe que el tiempo de duración seguía una distribución normal con media 300 minutos y desviación típica 30 minutos. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que de una muestra de 60 baterías el tiempo medio de duración fue de 290 minutos. ¿Se puede concluir que las sospechas del departamento de control de calidad son ciertas a un nivel de significación del 2%?

n =	$\overline{X}$ =	σ =

Ha:	Ho:	Zona de Rechazo
		

Valor crítico Z_0:	Z calculado:	Decisión: