MODULO INSTRUCCIONAL PARA ESTADISTICA GENERAL

ELABORADO POR: ESP. VICTOR ANTONIO VEGAS R

ÍNDICE DE CONTENIDO

CORRELACIÓN

Como es sabido, en cualquier disciplina se encuentran variables que de una u otra forma se encuentran relacionadas, por lo cual resulta importante más importante examinar esta relación que medir el resultado de cada una de forma independiente .

En este sentido, los modelos de correlación, son medidas estadísticas de asociación entre variables, que expresan el sentido y la magnitud de la relación existente entre dos o más variables. Las medidas de correlación son de gran importancia y tienen diversas aplicaciones en el estudio y análisis de fenómenos en las diversas ciencias, por cuanto permite conocer la influencia de algunos factores respecto a otro u otros.

A través de la aplicación de los modelos de correlación, se puede tener una información bastante aceptable de los diferentes aspectos que inciden y condicionan un fenómeno.

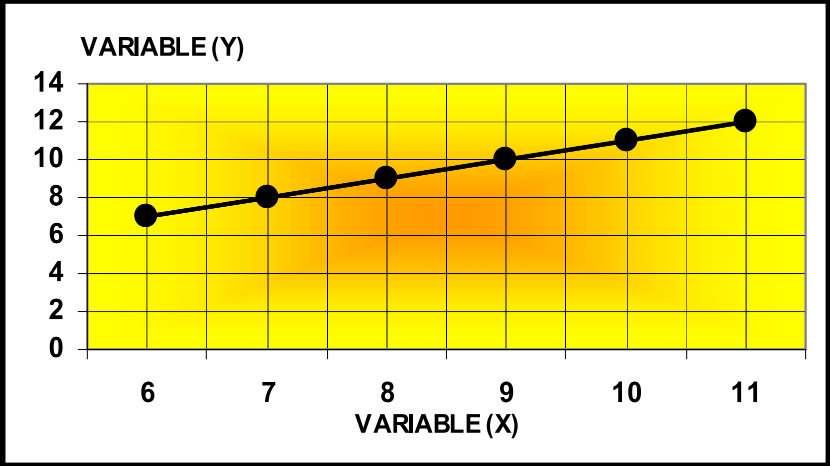
Los modelos de correlación entre variables, se miden o calculan aplicando lo que conoce como coeficiente de correlación, el cual se denota generalmente por las letras: R, r, p,∅, etc., este coeficiente está representado por un valor que está comprendido entre los valores de: -1 y 1, esto dependiendo del sentido e intensidad de la relación que se presente entre las variables en estudio.

CARACTERÍSTICAS GENERALES DE LOS MODELOS DE CORRELACIÓN

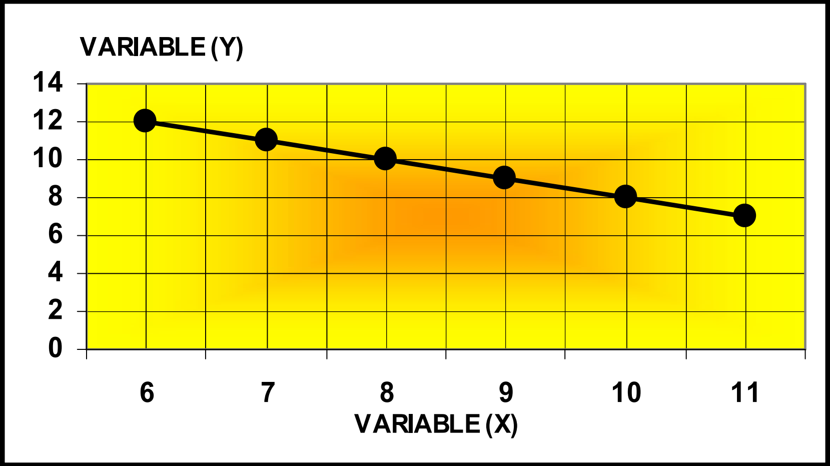
- Los modelos de correlación permiten determinar el sentido y la magnitud de la relación existente entre las variables que se estudian.
- El resultado de una correlación no debe ser mayor que 1, sea este de sentido negativo o positivo.
- Ningún modelo de correlación determina la relación de causa-efecto entre las variables consideradas.
- La magnitud de una correlación, depende en gran medida de la dispersión que presenten los valores de las variables en estudio.
- Para poder realizar una correlación en el verdadero sentido de la palabra, debe existir una relación lógica que una a las variables en estudio.

SENTIDO DE UNA CORRELACIÓN:

El sentido de una correlación, puede ser negativa o positiva, a este respecto se dice que una correlación entre variables es de sentido positivo, cuando se presenta una correspondencia directa entre las magnitudes de los valores que dichas variables toman, en otras palabras, cuando se presentan valores altos en la variable (X), se presentan valores altos en la variable (Y) y viceversa. Tal como se puede apreciar en las siguientes gráficas:



SENTIDO POSITIVO



SENTIDO NEGATIVO

COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON

Este modelo de correlación, es una de las medidas de asociación entre variables más utilizadas, y el mismo se basa en el supuesto de que la relación entre los valores de las variables consideradas se establece a través de una argumento lineal, es decir que la relación entre las variables se puede representar a través de una línea recta. Este modelo de correlación, se suele aplicar cuando las variables están medidas en una escala de intervalos.

CONTINUACIÓN COEFICIENTE DE CORRELACIÓN LINEAL

FORMULA DE CÁLCULO:

Existen varias formulas para el cálculo del modelo de correlación lineal, sin embargo en este caso explicaremos la más difundida, la cual se basa en los datos originales tanto de (X) como de (Y). Sin embargo en esta curso utilizaremos la siguiente

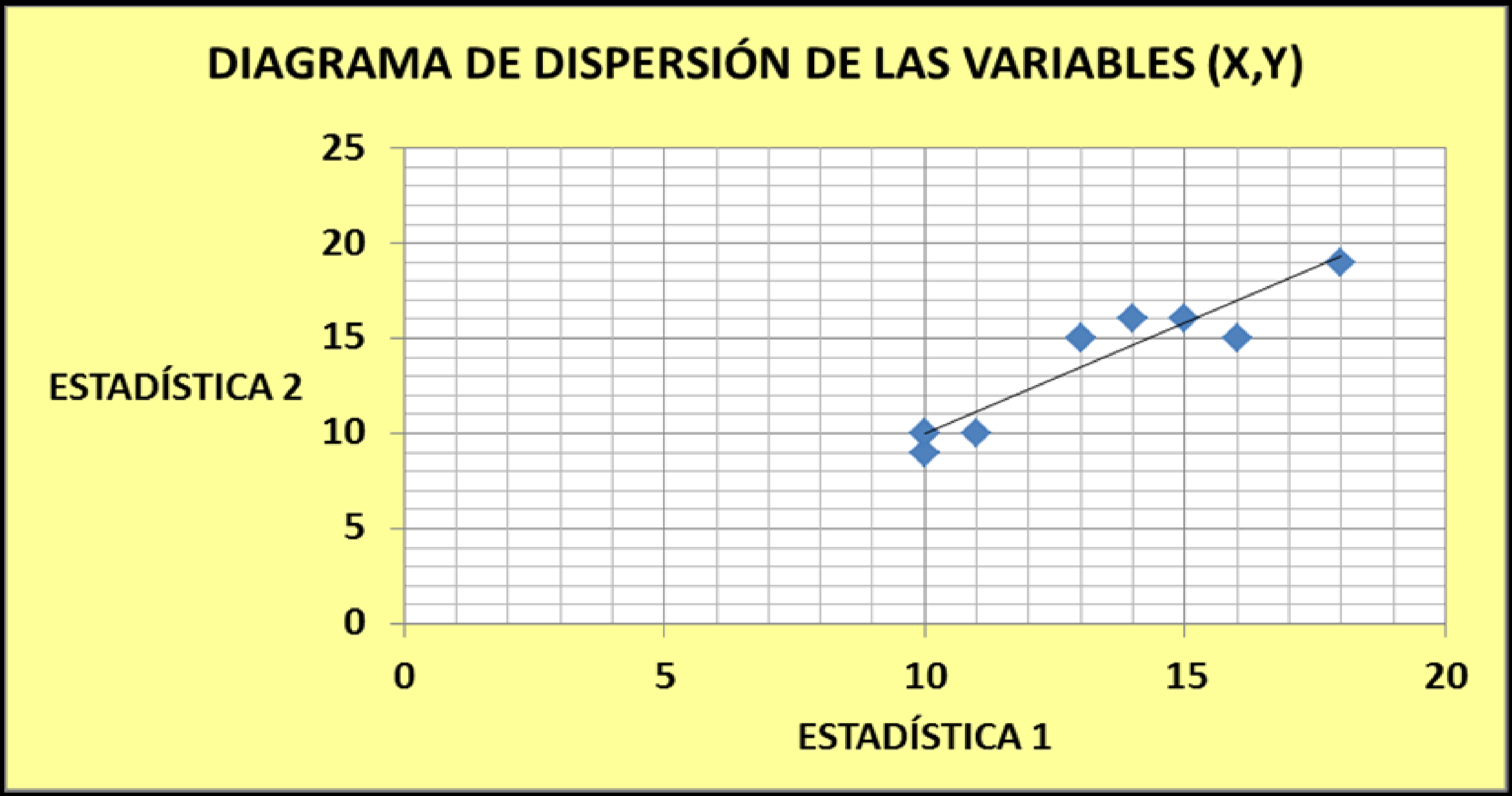
$$r = \frac{(n) * (\Sigma X * Y) - (\Sigma X) * (\Sigma Y)}{\sqrt{\left[(n) * (\Sigma X^2) - (\Sigma X)^2\right] * \left[(n) * (\Sigma Y^2) - (\Sigma Y)^2\right]}}$$

EJEMPLO DE CALCULO

Los datos que se presentan en la tabla siguiente, corresponden a las calificaciones finales obtenidas por un grupo de 10 alumnos de la U.C.V. en ESTADÍSTICA I Y ESTADÍSTICA II, en escala de 1 a 10 puntos. Antes de proceder a los cálculos hay que aclarar, que aun cuando se debería trabajar con no menos de 30 datos, para efectos de este ejercicio se tomarán 10 datos tal como aparecen en la tabla:

ESTADÍSTICA 1	10	10	14	11	13	13	18	10	15	16
ESTADÍSTICA 2	9	10	16	10	15	15	19	10	16	15

DIAGRAMA DE DISPERSIÓN (SENTIDO DE LA CORRELACIÓN)



Como se puede apreciar el diagrama de dispersión describe una tendencia o sentido positivo.

N	X EST 1	Y EST 2	X.Y	X^2	Y^2
1	10	9	90	100	81
2	10	10	100	100	100
3	14	16	224	196	256
4	11	10	110	121	100
5	13	15	195	169	225
6	13	15	195	169	225
7	18	19	342	324	361
8	10	10	100	100	100
9	15	16	240	225	256
10	16	15	240	256	225
$\sum$	130	135	1836	1760	1929

$$r = \frac{(n) * (\Sigma X * Y) - (\Sigma X) * (\Sigma Y)}{\sqrt{\left[(n) * (\Sigma X^2) - (\Sigma X)^2\right] * \left[(n) * (\Sigma Y^2) - (\Sigma Y)^2\right]}}$$

$$r = \frac{(n) * (\Sigma X * Y) - (\Sigma X) * (\Sigma Y)}{\sqrt{\left[(n) * (\Sigma X^2) - (\Sigma X)^2\right] * \left[(n) * (\Sigma Y^2) - (\Sigma Y)^2\right]}}$$

$$r = \frac{(10) * (1836) - (130) * (135)}{\sqrt{\left[(10) * (1760) - (130)^2\right] * \left[(10) * (1929) - (135)^2\right]}}$$

$$r = \frac{(18360) - (17550)}{\sqrt{\left[(17600) - (16900)\right] * \left[(19290) - (18225)\right]}}$$

$$r = \frac{(810)}{\sqrt{\left[700\right] * \left[(1065\right]}}$$

$$r = \frac{(810)}{\sqrt{745500}}$$

$$r = \frac{(810)}{863{,}42} = 0{,}93$$

En función de la magnitud resultante del coeficiente de correlación de Pearson calculado en la página anterior, el cual dio como resultado (0,93), se puede decir que existe una correlación positiva muy alta casi perfecta entre los valores de las variables consideradas (LAS CALIFICACIONES FINALES EN ESTADÍSTICA 1 Y 2). Lo cual indica que a valores en (X) se están presentando valores altos en (Y) y viceversa. (VER EL RESULTADO EN LA TABLA) (0,93)

TABLA DE VALORES PARA INTERPRETACIÓN DEL COEFICIENTE DE CORRELACIÓN	
ENTRE 0,01 ± 0,19	CORRELACIÓN MUY BAJA (INSIGNIFICANTE)
ENTRE 0,20 ± 0,39	CORRELACIÓN BAJA (DÉBIL)
ENTRE 0,40 ± 0,69	CORRELACIÓN MODERADA (SIGNIFICATIVA)
ENTRE 0,70 ± 0,89	CORRELACIÓN ALTA (FUERTE)
ENTRE 0,90 ± 0,98	CORRELACIÓN MUY ALTA (CASI PERFECTA)
ENTRE 0,99 ± 1,00	CORRELACIÓN PERFECTA

MODELO DE REGRESIÓN LINEAL CUANDO NO SON SERIES DE TIEMPO

El objetivo fundamental de muchas de las investigaciones estadísticas, consiste en establecer relaciones que hagan posible predecir una o más variables en términos de otras variables conocidas. Por ejemplo en educación se hacen estudios para predecir, la posible calificación obtenida por un estudiante en una asignatura afín con otra, tal podría ser el caso de predecir la calificación de una alumno en finales de Estadística II, tomando como referencia la calificación que este ha obtenido Estadística I.

CONTINUACIÓN DEL MODELO DE REGRESIÓN LINEAL

Sin embargo, no se puede perder de vista de que con la expresión Reresión Lineal Simple, se hace referencia al procedimiento matemático que sirve para ajustar una línea recta, a través de un conjunto de datos bivariantes, asentados en una diagrama de dispersión, y esta línea se conoce como Línea de Regresión.

Cuando Y es la variable independiente la ecuación es la siguiente:

$$Yc = a + b*(X)$$

$$aY/X = \frac{(\sum Y)*(\sum X^2) - (\sum X)*(\sum X*Y)}{(n)*(\sum X^2) - (\sum X)^2}$$

$$bY/X = \frac{(n)*(\sum X*Y) - (\sum X)*(\sum Y)}{(n)*(\sum X^2) - (\sum X)^2}$$

CONTINUACIÓN DEL MODELO DE REGRESIÓN LINEAL

A continuación, se presentan una serie de calificaciones obtenidas por un grupo de 10 estudiantes en dos evaluaciones, una de Estadística 1 y otra Estadística 2: En función de los datos, se desea calcular la ecuación de la recta de regresión de Y/X y la de X/Y, y en función de ellas estimar la calificación que pudo obtener un alumno en Estadística 2 (y), que haya obtenido 12 puntos en Estadística 1.

(X) HABILIDAD VERBAL	(Y) COMPRENSIÓN LECTORA	(X.Y)	X^2	Y^2
11	14	154	121	196
10	11	110	100	121
13	15	195	169	225
14	16	224	169	256
9	12	108	81	144
11	16	176	121	256
10	11	110	100	121
10	11	110	100	121
17	19	323	289	361
18	20	360	324	400
123	145	1870	1601	2201

ECUACIÓN DE LA RECTA CUANDO (X) ES LA VARIABLE INDEPENDIENTE

Yc = a + b (X)

$$aY/X = \frac{(\sum Y) * (\sum X^2) - (\sum X) * (\sum X * Y)}{(n) * (\sum X^2) - (\sum X)^2}$$

$$bY/X = \frac{(n) * (\sum X * Y) - (\sum X) * (\sum Y)}{(n) * (\sum X^2) - (\sum X)^2}$$

$$a\ Y/X = \frac{(\Sigma Y) * (\Sigma X^2) - (\Sigma X) * (\Sigma X*Y)}{(n) * (\Sigma X^2) - (\Sigma X)^2}$$

$$b\ Y/X = \frac{(n) * (\Sigma X \,.\, Y) - (\Sigma X) * (\Sigma Y)}{(n) * (\Sigma X^2) - (\Sigma X)^2}$$

$$a\ Y/X = \frac{(145) * (1601) - (123) * (1870)}{(10) * (1601) - (123)^2} = (2{,}42)$$

$$b\ Y/X = \frac{(10) * (1870) - (123) * (145)}{(10) * (1601) - (123)^2} = (0{,}98)$$

$$Yc = a + b\,(X) = Yc = 2{,}42 + 0{,}98\,(X)$$

Nótese que de esta forma queda elaborada la ecuación de la recta de Y/X, la cual servirá para realizar, estimaciones para la variable (Y) en función de valores de la variable (X)

Supongamos que se quiere estimar la calificación que pudo haber obtenido un estudiante en Estadística 2 si ha obtenido 12 puntos en Estadística 1, entonces lo primero que se debe hacer es identificar cual es la ecuación que permite estimar para la variable Estadística 2. en este caso la variable Estadística 2 corresponde a (y), por lo tanto tendríamos que utilizar la siguiente ecuación:

Yc = a + b (X) Y LA ESTIMACIÓN PEDIDA SERIA:

Yc = 2,42 + 0,98 (12) = 14,18 PUNTOS

Esto indica que un estudiante que haya obtenido 12 puntos en Estadística 1 tendría una calificación aproximada de 14,18 puntos en Estadística 2.

REGRESIÓN PARA SERIES CRONOLÓGICAS (SERIES DE TIEMPO)

En diversas disciplinas sociales es menester de los investigadores, realizar estudios acerca del comportamiento de una o más variables durante un período de tiempo determinado, así a través de la cuantificación de los fenómenos dentro de las ciencias sociales se pueden hacer uso de procedimientos prospectivos o proyecciones de estos fenómenos a futuro y tomar decisiones al respecto.

Una serie cronológica no es otra cosa que un conjunto de observaciones que se llevan a cabo en un período de tiempo determinado generalmente en intervalos iguales. Las series cronológicas o de tiempo representan una actividad particular que generalmente se utiliza para la organización en diversos niveles: industrial, económico, político, educacional etc.

En nuestro caso se tratarán sólo dos modelos para el análisis y proyección de series cronológicas, el modelo lineal por los mínimos cuadrados, y el modelo no lineal por parábola de segundo grado.

VEAMOS UN EJEMPLO DE APLICACIÓN

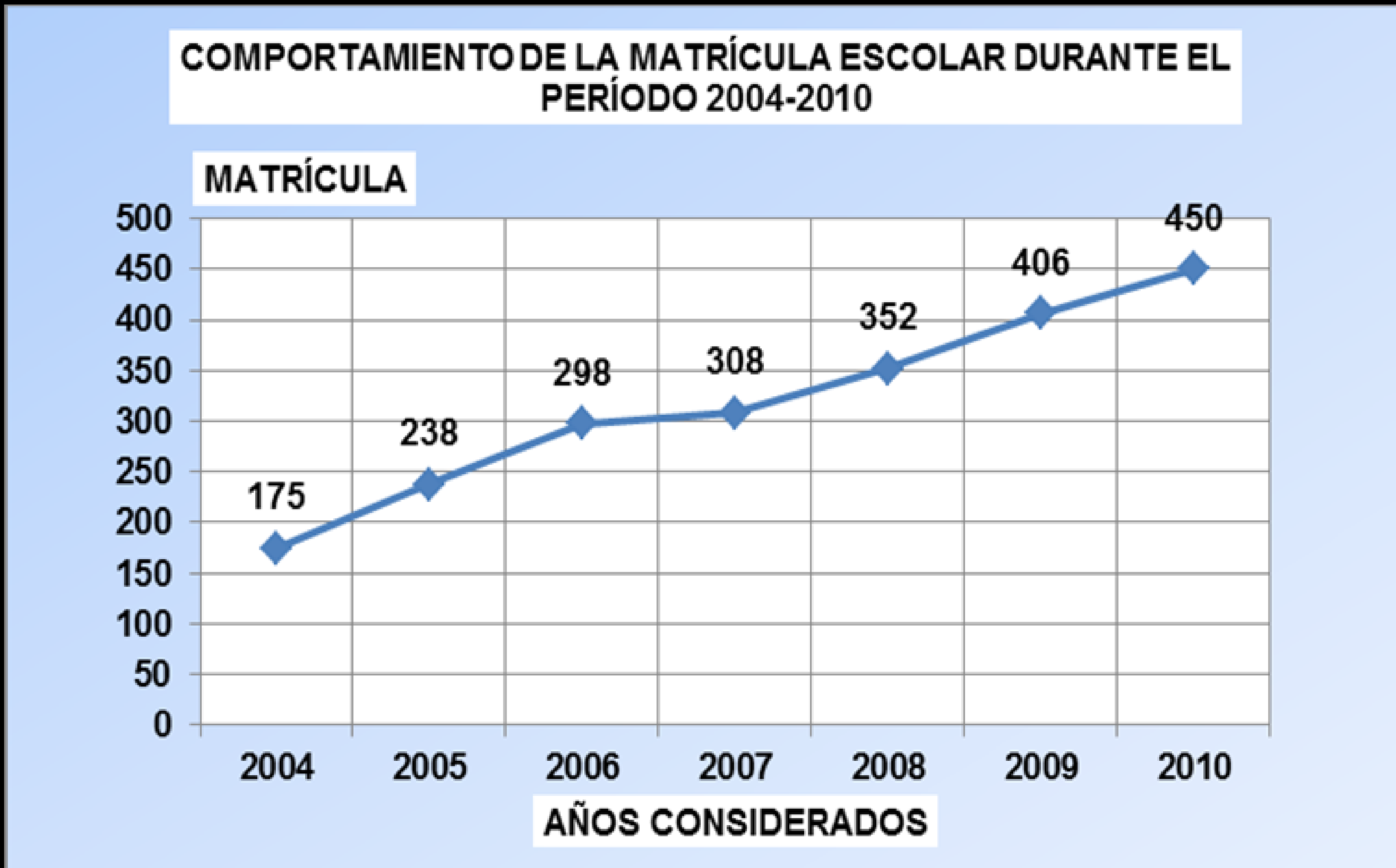
Los datos de la serie histórica corresponden al comportamiento de la matrícula escolar de una institución educativa, durante los últimos 7 años.

AÑOS	2004	2005	2006	2007	2008	2009	2010
MATRÍCULA	175	238	298	308	352	406	450

En función de esta data anterior se pide lo siguiente:

- ✓ Elabore el gráfico de tendencia correspondiente a la serie histórica
- ✓ Calcule la ecuación de la recta de regresión por mínimos cuadrados
- ✓ En función de la ecuación obtenida calcule el error de estimación
- ✓ Calcule le ecuación de la curva de regresión por parábola de 2do grado
- ✓ En función de la ecuación obtenida calcule el error de estimación
- ✓ Tomando en cuenta la tendencia que tenga menor error realice una proyección de la matrícula escolar para los próximos 3 años
- ✓ Comente los resultados de la proyección

GRÁFICO DE TENDENCIA CORRESPONDIENTE A LA SERIE



Cálculo de la ecuación de la recta de regresión por mínimos cuadrados. La ecuación de la recta viene dada por la siguiente expresión:

$$Yc = a + b\,(x)$$

$$a = \frac{\Sigma\, Y}{n} \qquad b = \frac{(\Sigma X * Y)}{(\Sigma X^2)}$$

CÁLCULO DE LA ECUACIÓN DE LA RECTA

AÑOS	MATRÍCULA (Y)	(X)	(X*Y)	X^2
2004	175	-3	-525	9
2005	238	-2	-476	4
2006	298	-1	-298	1
2007	308	0	0	0
2008	352	1	352	1
2009	406	2	812	4
2010	450	3	1350	9
Σ	2227	0	1215	28

$$a = \frac{2227}{7} = 318{,}14$$

$$b = \frac{1215}{28} = 43{,}39$$

LA ECUACIÓN DE LA RECTA SERÍA:

$$Yc = 318{,}14 + 43{,}39 * (x)$$

NOTA: La asignación de los valores de (X), tiene que ver con el tipo de serie que se está analizando, en tal sentido cuando la serie es un número impar de años como en este caso, se busca el año central y se le coloca como valor cero (0) y a partir de allí se colocan números negativos ascendentes hacia los años inferiores, y números positivos ascendentes hacia los años superiores. **(SE LLAMA VARIACIÓN INTERANUAL)**

Cuando la serie de años es par, como no existe un año central, nos colocamos entre los dos años centrales, y a partir de allí se colocan números negativos impares ascendentes hacia los años inferiores, y números positivos impares hacia los años superiores. **(SE LLAMA VARIACIÓN INTERSEMESTRAL O EN MEDIOS AÑOS)**

CÁLCULO DEL ERROR CUADRÁTICO DE REGRESIÓN A PARTIR DE LA ECUACIÓN DE LA RECTA

LA ECUACIÓN ES:

$$Yc = 318,14 + 43,39 * (x)$$

CÁLCULO EL ERROR CUADRÁTICO A PARTIR DE LA ECUACIÓN DE LA RECTA.

AÑOS	MATRÍCULA	(X)	X2	X*Y	YC	(Y-YC)2
2004	175	-3	9	-525	187,97	168,22
2005	238	-2	4	-476	231,36	44,09
2006	298	-1	1	-298	274,75	540,56
2007	308	0	0	0	318,14	102,82
2008	352	1	1	352	361,53	90,82
2009	406	2	4	812	404,92	1,17
2010	450	3	9	1350	448,31	2,86
Σ	2227	0	28	1215	2227	950,54

$$Sy/x = \sqrt{\frac{\sum(Y - Yc)^2}{n-1}} = \sqrt{\frac{950,54}{6}} = 12,58$$

Cálculo de la ecuación de la curva de regresión por parábola de segundo grado. La ecuación de la viene dada por la siguiente expresión:

$$Yc = a + b(x) + c(x2)$$

$$a = \frac{(\Sigma Y) - c * (\Sigma X^2)}{n}$$

$$c = \frac{(n) * (\Sigma X^2 * Y) - (\Sigma X^2) * (\Sigma Y)}{(n) * (\Sigma X^4) - (\Sigma X^2)^2}$$

$$b = \frac{(\Sigma X * Y)}{(\Sigma X^2)}$$

CALCULO DE LA ECUACIÓN DE LA CURVA O PARÁBOLA DE REGRESIÓN

AÑOS	MATRÍCULA	(X)	X2	X4	X*Y	X2*Y	YC	(Y-YC)2
2004	175	-3	9	81	-525	1575	184,57	91,58
2005	238	-2	4	16	-476	952	231,36	44,09
2006	298	-1	1	1	-298	298	276,79	449,86
2007	308	0	0	0	0	0	320,86	165,38
2008	352	1	1	1	352	352	363,57	133,86
2009	406	2	4	16	812	1624	404,92	1,17
2010	450	3	9	81	1350	4050	444,91	25,91
Σ	2227	0	28	196	1215	8851	2227	911,86

$$c = \frac{(n) * (\Sigma X^2 * Y) - (\Sigma X^2) * (\Sigma Y)}{(n) * (\Sigma X^4) - (\Sigma X^2)^2}$$

$$c = \frac{(7) * (8851) - (28) * (2227)}{(7) * (196) - (28)^2} = -0,68$$

$$a = \frac{(\Sigma Y) - c * (\Sigma X^2)}{n}$$

$$a = \frac{(2227) - (-0,68 * 28)}{7} = 320,86$$

$$b = \frac{(\Sigma X * Y)}{(\Sigma X^2)}$$

$$b = \frac{(1251)}{(28)} = 43,39$$

$$Yc = 320,86 + 43,39(x) + -0,68(x2)$$

CÁLCULO DEL ERROR CUADRÁTICO DE REGRESIÓN A PARTIR DE LA ECUACIÓN DE LA CURVA O PARÁBOLA DE REGRESIÓN

LA ECUACIÓN ES:

$$Yc = 320,86 + 43,39(x) + -0,68(x2)$$

AÑOS	MATRÍCULA	(X)	X2	X4	X*Y	X2*Y	YC	(Y-YC)2
2004	175	-3	9	81	-525	1575	184,57	91,58
2005	238	-2	4	16	-476	952	231,36	44,09
2006	298	-1	1	1	-298	298	276,79	449,86
2007	308	0	0	0	0	0	320,86	165,38
2008	352	1	1	1	352	352	363,57	133,86
2009	406	2	4	16	812	1624	404,92	1,17
2010	450	3	9	81	1350	4050	444,91	25,91
Σ	2227	0	28	196	1215	8851	2227	911,86

$$Sy/x = \sqrt{\frac{\Sigma(Y - YC)^2}{n-1}}$$

$$Sy/x = \sqrt{\frac{911,86}{7-1}} = 12,32$$

NOTA: EN ESTE CASO, COMO EL ERROR CUADRÁTICO POR PARÁBOLA DE SEGUNDO GRADO ES INFERIOR AL ERROR POR RECTA, SE DEBE PROYECTAR A PARTIR DE ESTA ECUACIÓN.

POR OTRA PARTE ES NECESARIO ACLARAR QUE CADA ESTIMACIÓN HECHA A PARTIR DE CUALESQUIERA QUE SEA LA ECUACIÓN ESCOGIDA ARRASTRA CONSIGO UN ERROR ESTIMADO O LO QUE ES LO MISMO EL ERROR CUADRÁTICO PROPIO DE ELLA.

PROBABILIDAD

Probabilidad es la posibilidad de ocurrencia o no de un suceso o hecho en términos de incertidumbre o azar. Generalmente encontraremos las palabras probabilístico o aleatorio asociadas a la definición de probabilidad, pero recordemos que esos términos no son otra cosa que azar, el cual es la imposibilidad de predecir el resultado de un ensayo o evento con certeza absoluta.

El cálculo de una probabilidad no podrá entonces tomar valores ni menores que cero (0) ni mayores que (1).

EVENTO O ENSAYO: Por evento se entiende a cualquier experimento aleatorio que se realice con la intensión de calcular una probabilidad, de manera que si se lanza un dado 10 veces, cada lanzamiento sería un evento, el cual puede dar lugar a un resultado diferente en cada caso.

ESPACIO MUESTRAL: El espacio muestral no es otra cosa que todos los posibles resultados que se pueden dar en un experimento probabilístico.

Por ejemplo, si se lanza un dado, el espacio muestral estaría conformado por los seis posibles resultados (1, 2, 3, 4, 5 o 6), pero si por el contrario se lanza una moneda, los posibles resultados son dos, cara o sello.

SUCESO ELEMENTAL: Este suceso no es otra cosa que lo que definimos como espacio muestral, ya que hace referencia a todos los posibles resultados de un experimento probabilístico. Por ejemplo en el lanzamiento de un Bolívar sólo se presentarán dos posibles resultados (CARA O SELLO), y este sería el suceso elemental.

SUCESO COMPUESTO: Es una subdivisión o subconjunto del suceso elemental. Por ejemplo en el lanzamiento de un dado en el cual espera que se obtenga como resultado un número impar, entonces el suceso compuesto estaría conformado por los números (1,3 y 5)

SUCESOS RELACIONADOS: Un suceso puede estar relacionado a otro: Los posibles resultados del primer suceso lo son del segundo, pero este segundo suceso tiene sus propios resultados, que son independientes del primero.

IGUALDAD DE SUCESOS: Esto se presenta cuando el resultado del (1er) evento es igual al (2do).

EJEMPLO: Si se lanza un dado y se espera (1ro) que salga un número cualquiera, y (2do) que se obtenga el número 5, veremos que ambos eventos esperados son iguales ya que en 5 es un número comprendido entre los posibles resultados de lanzar un dado.

UNIÓN DE SUCESOS: La unión de dos o más sucesos, no es otra cosa que otro suceso formado por todos los elementos de los sucesos independientes que se unen.

EJEMPLO: Lanzamos un dado al aire y esperamos dos sucesos: (1ro) que salga un número impar y (2do) que sea mayor que 1. La unión de estos dos sucesos estaría formada por los resultados: (3 y 5), que vendrían a ser los dos números impares del dado que además cumplen con la condición de ser mayores que 1.

INTERSECCIÓN DE SUCESOS: Es aquel suceso integrado por los elementos afines o comunes de dos o más sucesos simples o independientes que se intersectan.

EJEMPLO: Lanzamos un dado al aire, y analizamos dos sucesos: (1ro) que salga número par, y (2do) que sea mayor que 2. La intersección de estos dos sucesos tienen dos elementos, los números (4 y 6) es el único resultado común a ambos sucesos: ambos son mayores que 2 y a su vez son pares.

INCOMPATIBILIDAD DE SUCESOS: Se presenta en aquellos casos cuyos resultados no pueden darse al mismo tiempo ya que no tienen elementos comunes (su intersección es vacía).

EJEMPLO: Si lanzamos un dado al aire y analizamos dos sucesos: (1ro) que salga un número mayor que 6, y (2do) que sea par, evidentemente que ambos no se pueden dar al mismo tiempo, por cuanto son excluyentes, ya que en un dado no hay ningún número mayor que 6.

COMPLEMENTARIEDAD DE SUCESOS: Se da en aquellos en los cuales si no ocurre uno, obligatoriamente tiene que ocurrir el otro.

EJEMPLO: Si lanzamos una moneda al aire y esperamos que salga cara, podemos comprobar, que su complemento es que salga sello y viceversa, ya que cuando no se da el suceso esperado se da el otro.

¿CÓMO SE MIDE UNA PROBABILIDAD? Uno de los métodos más utilizados es aplicando la Regla de Laplace: la cual define la probabilidad de un suceso como el resultado entre casos favorables y casos posibles.

$$P(X) = \frac{\text{CASOS FAVORABLES}}{\text{CASOS POSIBLES}}$$

REQUISITOS PARA APLICAR LA REGLA DE LAPLACE

1. El número de resultados posibles (sucesos) tiene que ser finito. Si hubiera infinitos resultados, al aplicar la regla "casos favorables entre casos posibles" el cociente siempre sería cero.
2. Si al lanzar un dado, algunas caras tuvieran mayor probabilidad desalir que otras, no podríamos aplicar esta regla. A la regla de Laplace también se le denomina "probabilidad a priori", ya que para aplicarla hay que conocer antes de realizar el experimento cuales son los posibles resultados y saber que todos tienen las mismas probabilidades

QUÉ SE HACE SI EL EXPERIMENTO NO CUMPLE CON LOS REQUISITOS ANTES SEÑALADOS

En este caso podemos acudir a otro modelo de cálculo de probabilidades que se basa en la experiencia (modelo frecuentista). Cuando se realiza un experimento aleatorio un número muy elevado de veces, las probabilidades de los diversos posibles sucesos empiezan a converger hacia valores determinados, que son sus respectivas probabilidades.

EJEMPLO: Si se lanza una vez una moneda al aire y sale "cara", quiere decir que el suceso "cara“ ha aparecido el 100% de las veces y el suceso "cruz" el 0%.x

Si se lanza diez veces la moneda al aire, es posible que el suceso "cara" salga 7 veces y el suceso "cruz" las 3 restantes. En este caso, la probabilidad del suceso "cara" ya no sería del 100%, sino que se habría reducido al 70%.

PROBABILIDAD DE SUCESOS: Al definir los sucesos hablamos de las diferentes relaciones que pueden guardar dos sucesos entre sí, así como de las posibles relaciones que se pueden establecer entre los mismos. Vamos a ver ahora cómo se refleja esto en el cálculo de probabilidades.

a) Un suceso puede estar contenido en otro: Entonces, la probabilidad del primer suceso será menor que la del suceso que lo contiene.

Ejemplo: Lanzamos un dado y analizamos dos sucesos: a) que salga el número 6, y b) que salga un número par. Dijimos que el suceso a) está contenido en el suceso b).

$P(A) = 1 / 6 = 0{,}166$

$P(B) = 3 / 6 = 0{,}50$

Por lo tanto, podemos ver que la probabilidad del suceso contenido, suceso a), es menor que la probabilidad del suceso que lo contiene, suceso b).

b) Dos sucesos pueden ser iguales: En este caso, las probabilidades de ambos sucesos son las mismas.

Ejemplo: lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que salga múltiplo de 2. Las soluciones coinciden en ambos casos.

P(A) = 3 / 6 = 0,50

P(B) = 3 / 6 = 0,50

Ejemplo: Lanzamos un dado al aire y analizamos dos sucesos: a) que salga un número par, y b) que sea mayor que 3. La intersección de estos dos sucesos tiene dos elementos: el 4 y el 6. Su probabilidad será por tanto:

P (A Y B) = 2 / 6 = 0,33

d) Unión de dos o más sucesos: La probabilidad de la unión de dos sucesos es igual a la suma de las probabilidades individuales de los dos sucesos que se unen, menos la probabilidad del suceso intersección

Ejemplo: Lanzamos un dado al aire y analizamos dos sucesos: a) que salga número par, y b) que el resultado sea mayor que 3. El suceso unión estaría formado por los siguientes resultados: 2, el 4, el 5 y el 6.

P(A) = 3 / 6 = 0,50

P(B) = 3 / 6 = 0,50

P (A U B) = 2 / 6 = 0,33

Por lo tanto, P (A U B) = (0,50 + 0,50) - 0,33 = 0,666

f) Sucesos complementarios: la probabilidad de un suceso complementario (A) es igual a 1 – P(A).

Ejemplo: Si lanzamos un dado al aire. el suceso (A) es que salga un número par, luego su complementario, suceso (B), es que salga un número impar.

La probabilidad del suceso (A) es igual a :

$$P(A) = 3 / 6 = 0,50$$

Luego, la probabilidad del suceso (B) es igual a:

$$P(B) = 1 - P(A) = 1 - 0,50 = 0,50$$

Se puede comprobar aplicando la regla de "casos favorables / casos posibles": $P(B) = 3 / 6 = 0,50$

e) Sucesos incompatibles: la probabilidad de la unión de dos sucesos incompatibles será igual a la suma de las probabilidades de cada uno de los sucesos (ya que su intersección es el conjunto vacío y por lo tanto no hay que restarle nada).

Ejemplo: Si lanzamos un dado al aire y analizamos dos sucesos: a) que salga un número menor que 3, y b) que salga el número 6.

La probabilidad del suceso unión de estos dos sucesos será igual a:

P(A) = 2 / 6 = 0,333

P(B) = 1 / 6 = 0,166

Por lo tanto, P (A U B) = 0,33 + 0,166 = 0,50

g) Unión de sucesos complementarios: la probabilidad de la unión de dos sucesos complementarios es igual a 1.

Ejemplo: seguimos con el ejemplo anterior: a) que salga un número par, y b) que salga un número impar. La probabilidad del suceso unión de estos dos sucesos será igual a:

$$P(A) = 3 / 6 = 0,50$$

$$P(B) = 3 / 6 = 0,50$$

Por lo tanto,

$$P(AUB) = 0,50 + 0,50 = 1$$

COMBINACIONES, Y PERMUTACIONES:

Para aplicar la Regla de Laplace, el cálculo de los sucesos favorables y de los sucesos posibles a veces no plantea ningún problema, ya que son un número reducido y se pueden calcular con facilidad.

Sin embargo, a veces calcular el número de casos favorables y casos posibles es complejo y hay que aplicar reglas matemáticas:

Ejemplo: 5 grupos se sientan aleatoriamente a trabajar y se quiere calcular la probabilidad de que al menos los miembros de un grupo se sienten juntos. En este caso, determinar el número de casos favorables y de casos posibles es complejo. Las reglas matemáticas que nos pueden ayudar son el cálculo de combinaciones, variaciones y de permutaciones.

COMBINACIONES: Determinan el número de subgrupos de 1, 2, 3, etc. elementos que se pueden formar con los "n" elementos de una nuestra. Cada subgrupo se diferencia del resto en los elementos que lo componen, sin que influya el orden.

Ejemplo: calcular las posibles combinaciones de 2 elementos que se pueden formar con los números 1, 2 y 3. Se pueden establecer 3 parejas diferentes: (1,2), (1,3) y (2,3).

En el cálculo de combinaciones las parejas (1,2) y (2,1) se consideran idénticas, por lo que sólo se cuentan una vez.

PERMUTACIONES: Permiten calcular las posibles agrupaciones que se pueden establecer con todos los elementos de un grupo, por lo tanto, lo que diferencia a cada subgrupo del resto es el orden de los elementos.

Ejemplo: calcular las posibles formas en que se pueden ordenar los número 1, 2 y 3.

Hay 6 posibles agrupaciones: (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2) y (3, 2, 1)

Para calcular el número de combinaciones se aplica la siguiente fórmula:

$$Cm,n = \frac{m!}{n! * (m - n)!}$$

El termino " n ! " se denomina "factorial de n" y es la multiplicación de todos los números que van desde "n" hasta 1.

Ejemplo: 4 ! = 4 * 3 * 2 * 1 = 24. La expresión "Cm,n" representa las combinaciones de "m" elementos, formando subgrupos de "n" elementos.

Ejemplo: C8,4 son las combinaciones de 8 elementos agrupándolos en subgrupos de 4 elementos:

$$Cm, n = \frac{m!}{n! * (m - n)!}$$

$$C10{,}4 = \frac{8!}{4! * (8 - 4)!} = \frac{40320}{576} = 70$$

Es decir, podríamos formar 70 subgrupos diferentes de 4 elementos, a partir de los 8 elementos.

Para calcular el número de permutaciones se aplica la siguiente fórmula:

$$Pm = m!$$

La expresión "Pm" representa las permutaciones de "m" elementos, tomando todos los elementos. Los subgrupos se diferenciaran únicamente por el orden de los elementos.

Ejemplo: P10 son las permutaciones de 10 elementos: P10! = 3.628.800

Es decir, tendríamos 3.628.800 formas diferentes de agrupar 10 elementos.

COMBINACIONES CON REPETICIÓN

Para calcular el número de combinaciones con repetición se aplica la siguiente fórmula:

$$Cm,n = \frac{(m+n-1)}{n! * (m-1)!}$$

Ejemplo: C'10,4 son las combinaciones de 10 elementos con repetición, agrupándolos en subgrupos de 4, en los que 2, 3 o los 4 elementos podrían estar repetidos:

$$C10,4 = \frac{(10+4-1)}{4! * (10-1)!} = 715$$

Es decir, podríamos formar 715 subgrupos diferentes de 4 elementos.

Para calcular el número de permutaciones con repetición se aplica la siguiente fórmula:

$$P'm^{x1.x2,\ldots xk} = m' / x1{*}x2{*}\ldots xk$$

Son permutaciones de "m" elementos, en los que uno de ellos se repite " x1 " veces, otro " x2 " veces y así ... hasta uno que se repite " xk " veces.

Ejemplo: Calcular las permutaciones de 10 elementos, en los que uno de ellos se repite en 2 ocasiones y otro se repite en 3 ocasiones:

$$P'm^{x1.x2,\ldots xk} = m' / x1{*}x2{*}\ldots xk$$

$$P'10^{2,3} = 10! / (2! * 3!) = 302.400$$

Es decir, tendríamos 302,400 formas diferentes de agrupar estos 10 elementos

TEOREMA DE BAYES

El Teorema de Bayes viene a ser el proceso inverso al que hemos visto en el Teorema de la probabilidad total:

FORMULA DE BAYES

$$P(B1/A) = \frac{P(B1) * P(A/B1)}{P(B1) * P(A/B1) + P(B2) * P(A/B2) + \ldots P(BN) * P(A/BN)}$$

Ejemplo: Se sabe por experiencias pasadas, que las probabilidades de que un estudiante se gradúe de ingeniero en la universidad es del 25%, de que se gradúa de licenciado en educación es del 45% y de que se gradúe de licenciado en administración es del 30%. Además se conoce que las probabilidades de que sean sobresalientes durante la carrera son del 2%, 4% y 3% respectivamente. Calcular las siguientes probabilidades:

1.- Si se toma un estudiante al azar, cual será la probabilidad de que sea de ingeniería y sobresaliente.

2.- Si se toma un estudiante al azar, cual será la probabilidad de que sea sobresaliente y de la carrera de educación.

CARRERA	INGENIERIA	EDUCACION	ADMINISTRACION
GRADUARSE	25% = 0,25	45% = 0,45	30% = 0,30
SOBRESALIENTE	2% = 0,02	4% = 0,04	3% = 0,03

CASO 1: Si se toma un estudiante al azar, cual será la probabilidad de que sea de ingeniería y sobresaliente.

$$P(ING / SOB) = \frac{P(ING) * P(ING / SOB)}{P(ING) * P(ING / SOB) + P(ED) * P(ED / SOB) + P(ADM) * P(ADM / SOB)}$$

$$P(ING / SOB) = \frac{(0,25) * P(0,02)}{(0,25) * P(0,02) + (0,45) * (0,04) + (0,30) * P(0,03)} = \frac{0,005}{(0,005 + 0,018 + 0,009)} = 0,15625$$

NOTA: SI SE ANALIZA CON DETENIMIENTO ESTA ECUACIÓN SE VERÁ QUE PARA CUALQUIER OTRO DE LOS CASOS LO QUE HABRÍA QUE CAMBIAR SON LOS VALORES CORRESPONDIENTES AL DIVIDENDO DEL CASO QUE SE ESTÉ CALCULANDO

DISTRIBUCIONES DISCRETAS DE PROBABILIDADES.
(DISTRIBUCIÓN BINOMIAL)

En la mayoría de las ciencias a menudo los investigadores se encuentran en presencia de situaciones dicotómicas cuando se realizan experimentos de probabilidades, cuyos resultados se clasifican en dos categorías cualesquiera, no aceptándose categorías intermedias. Ej, si un experimento ha tenido éxito o fracaso. En todo caso el modelo binomial de probabilidades, puede aplicarse en casos en los cuales se producen resultados independientes. En todo caso es necesario aclarar que este modelo, se aplica en distribuciones de tipo discreta.

CARACTERÍSTICAS DEL MODELO BINOMIAL

1.- El resultado de cada experimento es independiente de los demás

2.- (P) representa en todo caso la proporción a favor de la ocurrencia del suceso esperado, y (Q) la proporción en contra.

3.- En cualquier caso P y Q son proporciones complementarias, da tal manera que P es igual a 1-Q y viceversa.

4.- Las proporciones o probabilidades P y Q son siempre las mismas, a menos que se cambien las condiciones del experimento.

5.- El experimento se repite un número determinado de veces pero en las mismas condiciones.

7.- La suma de todos los posibles resultados de un experimento es igual a (1) (ESPACIO MUESTRAL).

CONTINUACIÓN CARACTERÍSTICAS DEL MODELO BINOMIAL:

8.- El promedio esperado a favor es igual a N*P

9.- La varianza es igual a N*P*Q, y la Desviación Estándar es igual a la raíz cuadrada de la varianza

FORMULA DE CÁLCULO PARA BINOMIAL

$$P(x) = \frac{n!}{X!*(n-X)!} * P^{x} * Q^{(n-x)}$$

EJEMPLO DE CÁLCULO: El Departamento de Control de Estudios de una determinada universidad ha demostrado que sólo 4 de cada 10 estudiantes que ingresan logran graduarse. Si se toma una muestra al azar de 5 estudiantes, ¿cuál será la probabilidad de que se gradúen más de 3 ?

Nótese que en este caso el espacio muestral es de 5 estudiantes de manera que la probabilidad de que se gradúen más de 3, significa que se gradúen 4 o 5 estudiantes.

DATOS:

P= 4/10 = 0,4

Q= 1-P = 1-0,4 = 0,6

n = 5

$$P(X=4)=\frac{5!}{(4!)*(5-4!)}*0{,}4^{4}*0{,}6^{(5-4)}\ \frac{120}{24}*0{,}0256*0{,}6=0{,}0768$$

$$P(X=5)=\frac{5!}{(5!)*(5-5!)}*0{,}4^{5}*0{,}6^{(5-5)}\ \frac{120}{120}*0{,}01024*1=0{,}01024$$

luego la probabilidad buscada será igual a la suma de: p(x=4) + p(x=5) en otras palabras el resultado será de: 0,0768 + 0,01024 = 0,08704

DISTRIBUCIÓN DE POISSON

Cuando en una distribución binomial se realiza el experimento un número "n" muy elevado de veces y la probabilidad de éxito "p" en cada ensayo es reducida, entonces se aplica el modelo de distribución de Poisson:

Se tiene que cumplir que

" p " < 0,10

" n * p " < 10

$$P(X) = \frac{\lambda^{X}}{(X\,!) * (e)^{\lambda}}$$

DONDE:

λ = n.p (Es decir, al tamaño de la muestra multiplicada por la probabilidad “p” de éxito en cada evento)

e = 2,72 (Base de los logaritmos neperianos)

X = (Es la probabilidad que se pretende calcular).

VEAMOS UN EJEMPLO DE CÁLCULO

Una empresa ensambladora de motores ha determinado a través de un estudio , que 10 de cada 1000 motores ensamblados, no aprueban el control de calidad, si se toma una muestra al azar de 800 motores al azar, cuál será la probabilidad de que 4 de ellos no aprueben el control de calidad.

DATOS:

n = 800

p = 10/1000 = 0.01

X = 4

λ = n*p = 800*0,01 = 8

$$P(X = 4) = \frac{8^4}{(4\,!) * (2{,}72)^8} = \frac{4096}{24 * 2996} = 0{,}0569$$

De manera que 0,0569 sería la probabilidad buscada.

LA DISTRIBUCIÓN NORMAL

La distribución Normal, es una de las distribuciones de probabilidades más utilizadas en el ámbito de la estadística, la misma presenta las siguientes características:

- **La curva normal tiene forma de campana con un solo apuntamiento justo en el centro de la distribución, por lo tanto presenta una curtosis mesocurtica o normal.**
- **La distribución normal es simétrica respecto a su media, por lo tanto la moda, la mediana y el promedio aritmético son iguales.**

➢La mitad del área bajo la curva está a la derecha del apuntamiento, y la otra mitad está a la izquierda, en todo caso cada una de estas mitades presenta el mismo valor proporcional (0,5) para cada lado lo cual representa un 50% para cada lado por encima y por debajo del promedio aritmético.

➢La distribución normal es asintótica la curva se acerca cada vez más al eje x pero en realidad nunca llega a unirse.

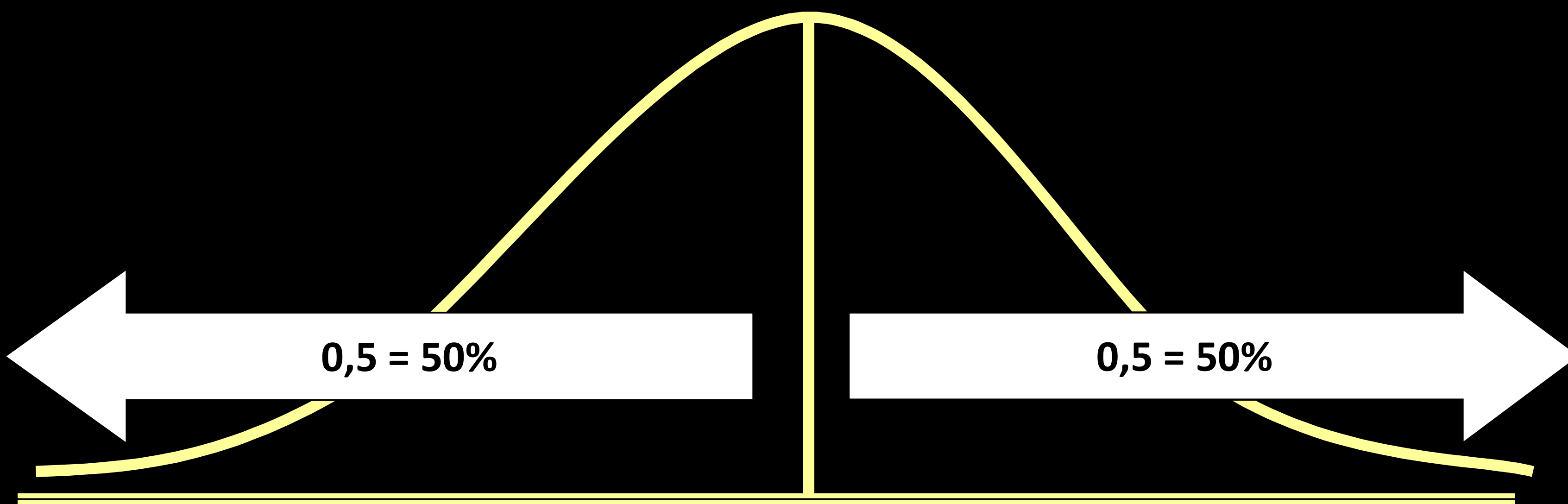
La distribución normal viene dada por la siguiente formula:

$$Z = \frac{X - \mu}{S}$$

La distribución Normal, es una de las distribuciones de probabilidades más utilizadas en el ámbito de la estadística, la misma presenta las siguientes características:

- La curva normal tiene forma de campana con un solo apuntamiento justo en el centro de la distribución, por lo tanto presenta una curtosis mesocurtica o normal.
- La distribución normal es simétrica respecto a su media, por lo tanto la moda, la mediana y el promedio aritmético son iguales.

LA DISTRIBUCION NORMAL



LA MODA, MADIANA Y MEDIA ARITMETICA COINCIDEN

TABLA DE AREAS BAJO LA CURVA NORMAL TIPIFICADAS DE 0 A Z

Z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1225	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3454	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4663	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767

TABLA DE AREAS BAJO LA CURVA NORMAL TIPIFICADAS DE 0 A Z										
Z	0	1	2	3	4	5	6	7	8	9
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4993	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

TABLA DE LA DISTRIBUCION T STUDENT MUESTRAS PEQUEÑAS ≤ 29 DATOS					
GL	VALORES CRITICOS UN COLA UNILATERAL				
	0,25	0,05	0,025	0,01	0,005
	VALORES CRITICOS DOS COLAS BILATERAL				
	0,50	0,10	0,05	0,02	0,01
1	1,000	6,31	12,70	31,8	63,7
2	0,816	2,92	4,30	6,97	9,92
3	0,765	2,35	3,18	4,54	5,84
4	0,741	2,13	2,78	3,75	4,60
5	0,727	2,02	2,57	3,37	4,03
6	0,718	1,94	2,45	3,14	3,71
7	0,711	1,89	2,36	3,00	3,50
8	0,706	1,86	2,31	2,90	3,36
9	0,703	1,83	2,26	2,82	3,25
10	0,700	1,81	2,23	2,76	3,17
11	0,697	1,80	2,20	2,72	3,11
12	0,695	1,78	2,18	2,68	3,05
13	0,694	1,77	2,16	2,65	3,01
14	0,692	1,76	2,14	2,62	2,98
15	0,691	1,75	2,13	2,60	2,95

CONTINUACION

TABLA DE LA DISTRIBUCION T STUDENT MUESTRAS PEQUEÑAS ≤ 29 DATOS					
GL	VALORES CRITICOS UN COLA UNILATERAL				
	0,25	0,05	0,025	0,01	0,005
	VALORES CRITICOS DOS COLAS BILATERAL				
	0,50	0,10	0,05	0,02	0,01
16	0,690	1,75	2,12	2,58	2,92
17	0,689	1,74	2,11	2,57	2,90
18	0,688	1,73	2,10	2,55	2,88
19	0,688	1,73	2,09	2,54	2,86
20	0,687	1,72	2,09	2,53	2,85
21	0,686	1,72	2,08	2,52	2,83
22	0,686	1,72	2,07	2,51	2,82
23	0,685	1,71	2,07	2,50	2,81
24	0,685	1,71	2,06	2,49	2,80
25	0,684	1,71	2,06	2,49	2,79
26	0,684	1,71	2,06	2,48	2,78
27	0,684	1,70	2,05	2,47	2,77
28	0,683	1,70	2,05	2,47	2,76
29	0,683	1,70	2,05	2,46	2,76
Z	0,674	1,65	1,96	2,33	2,58

VEAMOS ALGUNOS EJEMPLOS DE APLICACIÓN

El Departamento de Control de estudios de una universidad, ha determinado a través de un estudio que las calificaciones obtenidas por 150 estudiantes en una escala de 1 a 20 puntos, siguen una distribución aproximadamente normal, con Promedio de calificación de 12,2 Puntos y una Desviación Típica de 2,3 Puntos.

En función de los datos se desea que Usted calcule:

CASO Nº 1

DATOS:

μ = 12,2 Puntos

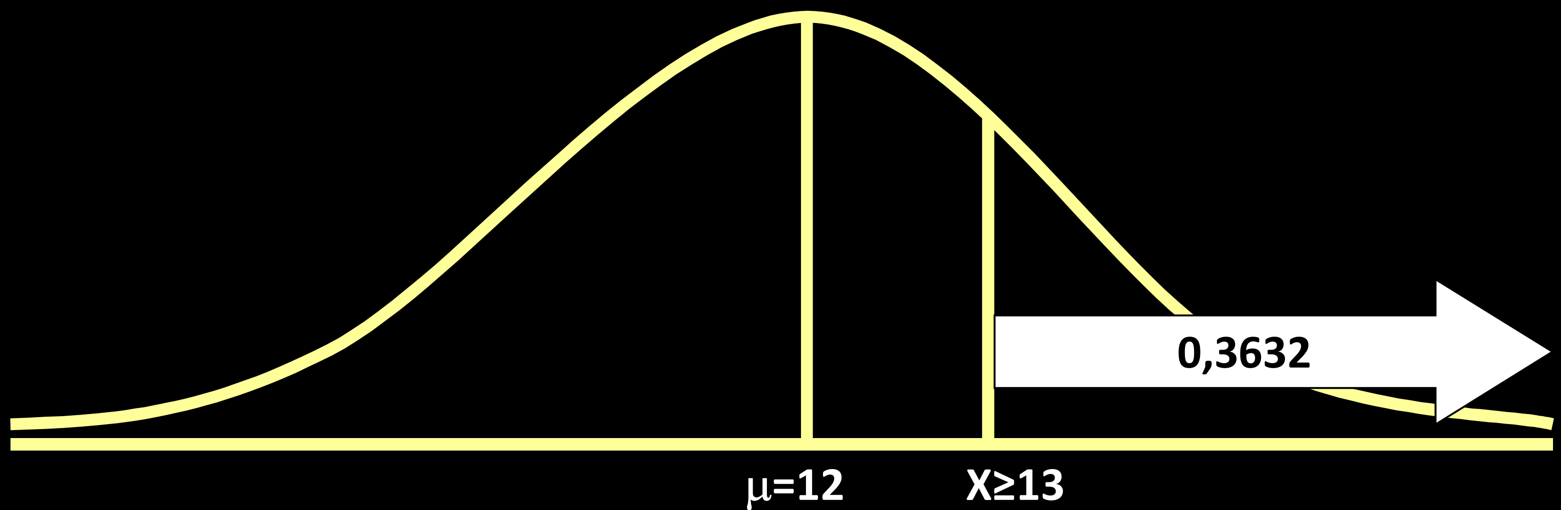
S= 2,3 Puntos

n = 150 Estudiantes.

Probabilidad que un estudiante haya obtenido una calificación igual o mayor a 13 Puntos.

$$Z = \frac{(X - \mu)}{S}$$

$$Z = \frac{(13 - 12{,}2)}{2{,}3} = 0{,}35 = 0{,}1368 = (0{,}5 - 0{,}1368) = 0{,}3632$$



CASO Nº 2

DATOS:

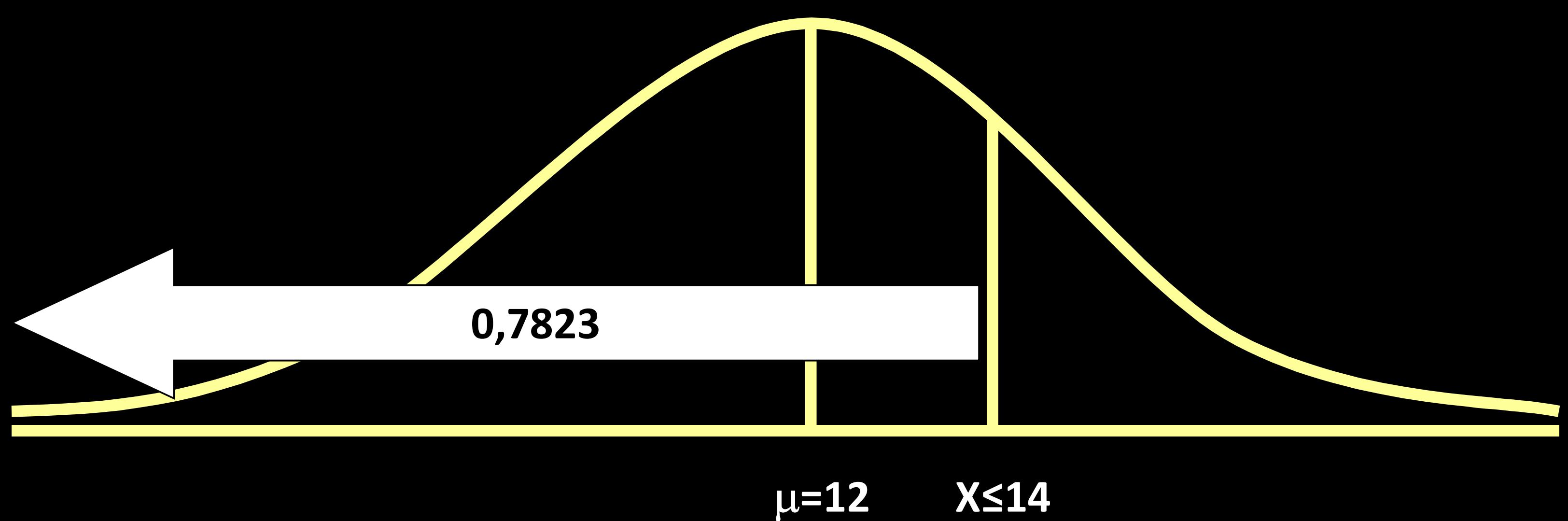
μ = 12,2 Puntos

S= 2,3 Puntos

n = 150 Estudiantes.

Probabilidad que un estudiante haya obtenido una calificación igual o menor a 14 Puntos.

$$Z = \frac{(X-\mu)}{S} \qquad Z = \frac{(14-12,2)}{2,3} = 0,78 = 0,2823 = (0,5 + 0,2823) = 0,7823$$



CASO Nº 3

DATOS:

μ = 12,2 Puntos

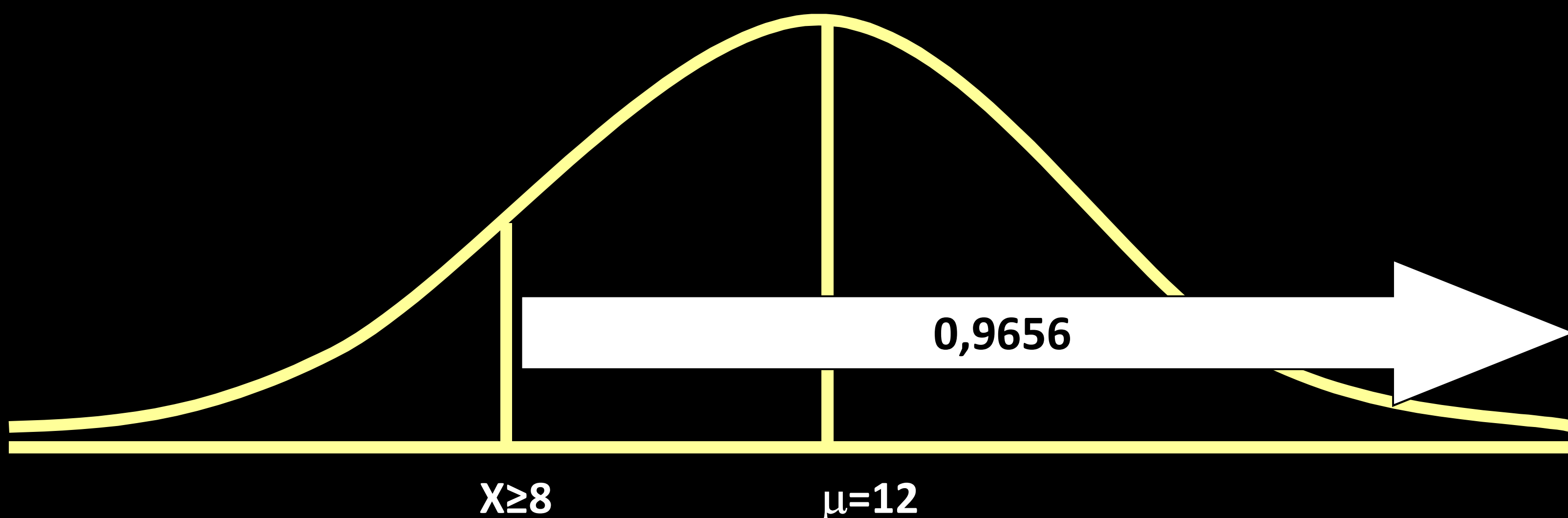
S= 2,3 Puntos

n = 150 Estudiantes.

Probabilidad que un estudiante haya obtenido una calificación igual o mayor a 8 Puntos.

$$Z = \frac{(X-\mu)}{S}$$

$$Z = \frac{(8-12,2)}{2,3} = -1,82 = 0,4656 = (0,5 + 0,4656) = 0,9656$$



CASO Nº 4

DATOS:

μ = 12,2 Puntos

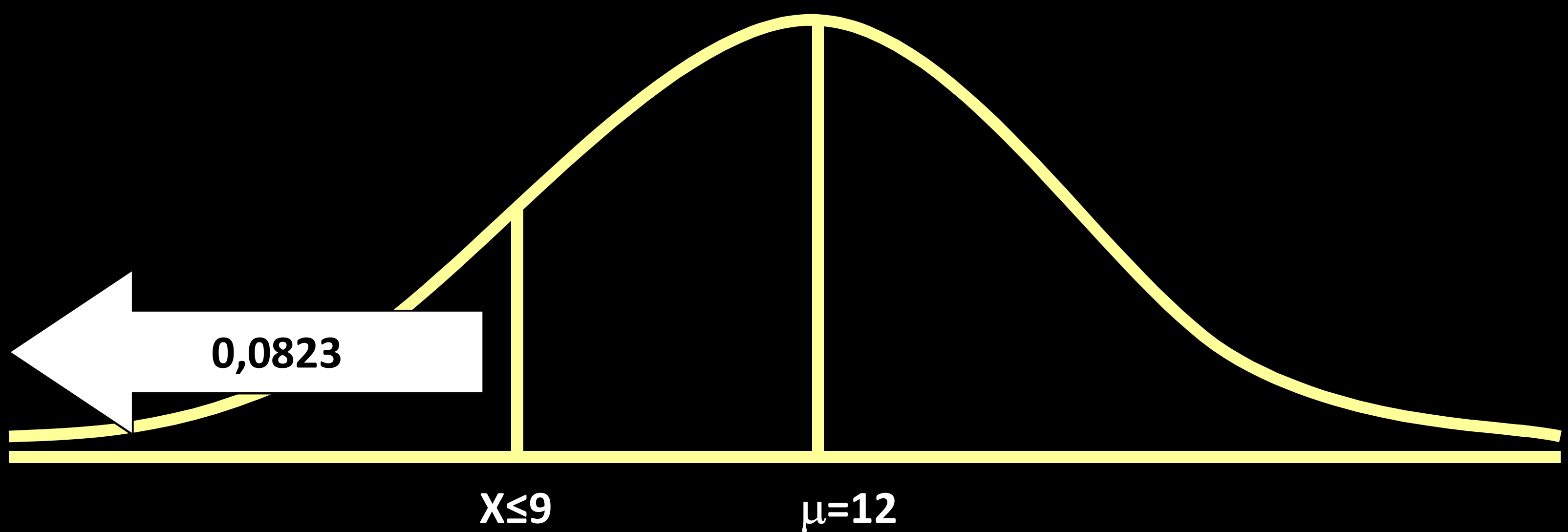
S= 2,3 Puntos

n = 150 Estudiantes.

Probabilidad que un estudiante haya obtenido una calificación igual o menor a 9 Puntos.

$$Z = \frac{(X - \mu)}{S}$$

$$Z = \frac{(9 - 12,2)}{2,3} = -1,39 = 0,4177 = (0,5 - 0,4177) = 0,0823$$



CASO Nº 5

DATOS:

μ = 12 Puntos

S= 2,5 Puntos

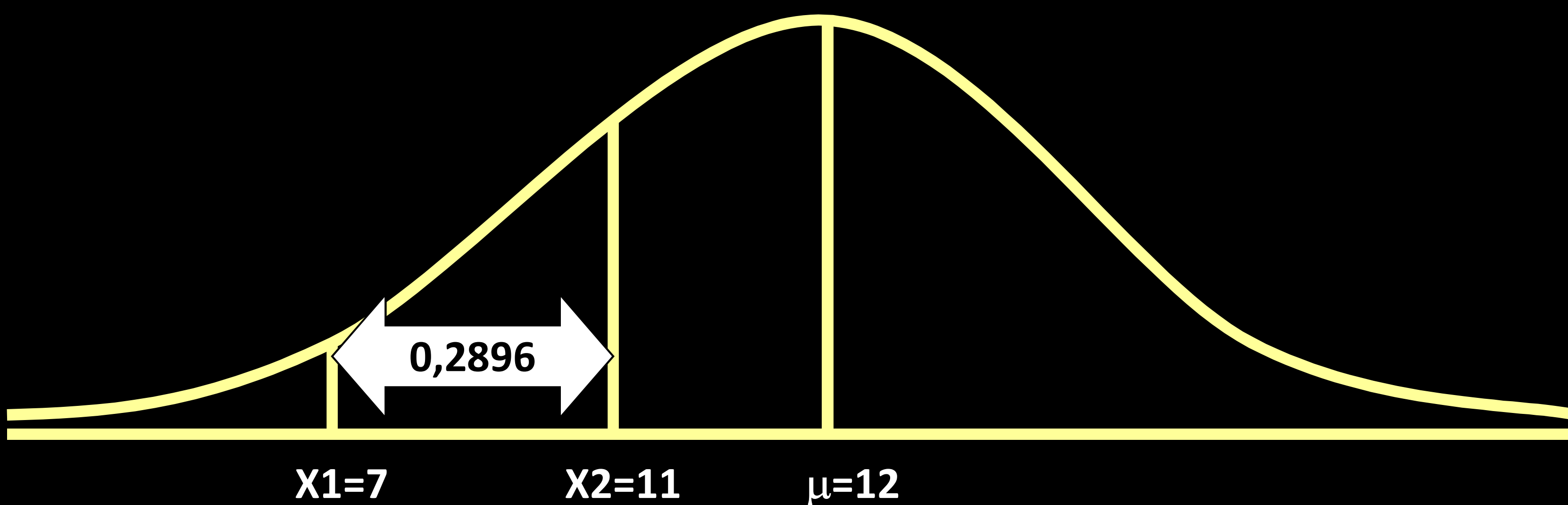
n = 150 Estudiantes.

Probabilidad de que un estudiantes haya obtenido entre 7 y 11 puntos.

$$Z = \frac{(X - \mu)}{S}$$

$$Z1 = \frac{(7 - 12,2)}{2,3} = -2,26 = 0,4881$$

$$Z2 = \frac{(11 - 12,2)}{2,3} = -0,52 = 0,1985 = (0,4881 - 0,1985) = 0,2896$$



CASO Nº 6

DATOS:

μ = 12 Puntos

S= 2,5 Puntos

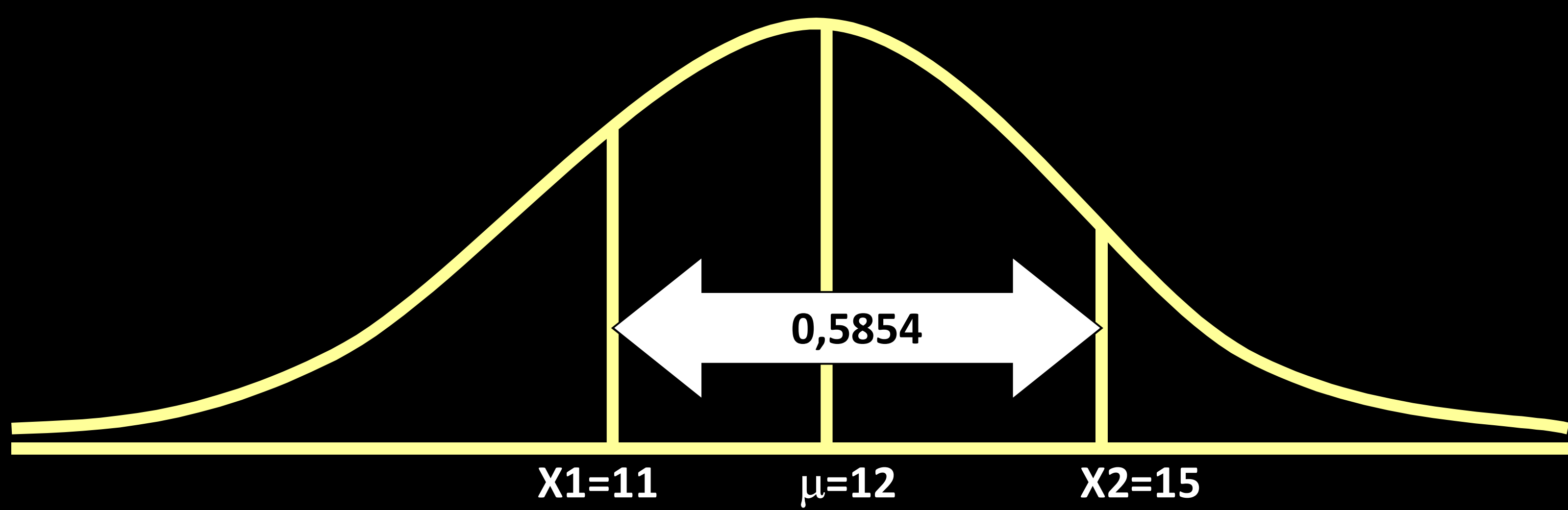
n = 150 Estudiantes.

Probabilidad de que un estudiantes haya obtenido entre 11 y 15 puntos.

$$Z = \frac{(X - \mu)}{S}$$

$$Z1 = \frac{(11 - 12,2)}{2,3} = -0,52 = 0,1985$$

$$Z2 = \frac{(15 - 12,2)}{2,3} = 1,21 = 0,3869 = (0,1985 + 0,3869) = 0,5854$$



CASO Nº 7

$$Z = \frac{X - \mu}{S}$$

DATOS:

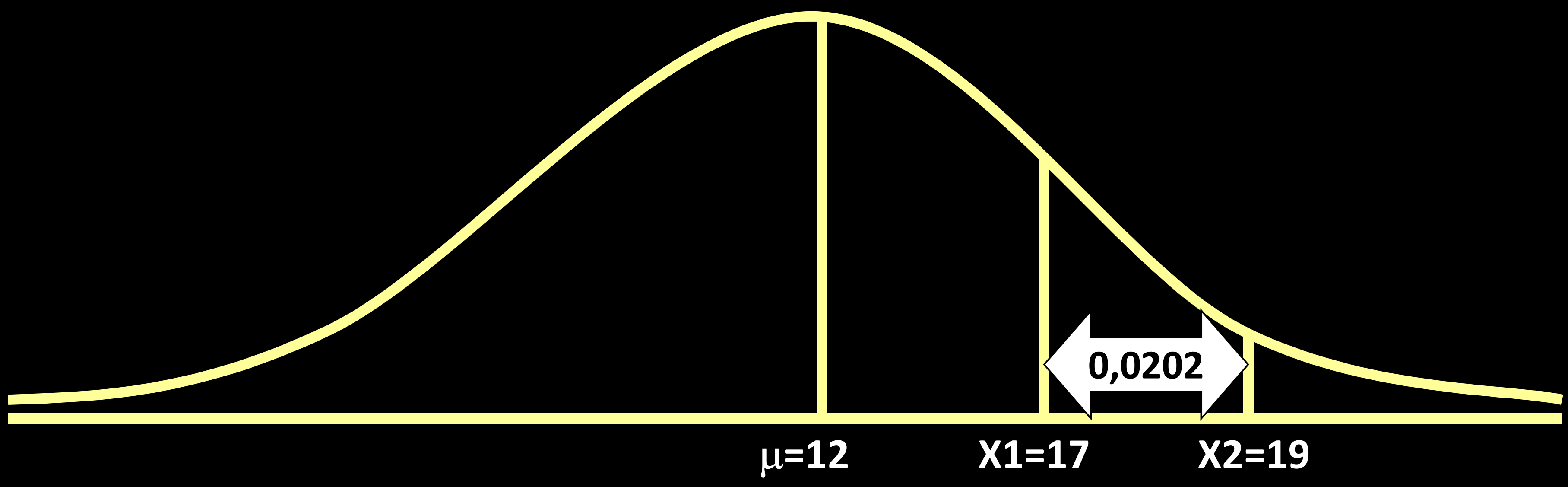
μ = 12 Puntos

S= 2,5 Puntos

n = 150 Estudiantes.

Probabilidad (X Entre 17 y 19 Puntos)

SOLUCIÓN: Z1=(17–12)/2,5=2,00. El valor por tabla sería de: 0,4772. Z2=(19–12)/2,5 = 2,80. El valor por tabla sería de: 0,4974 y la probabilidad buscada será igual a (0,4974–0,4772) = 0,0202 . En porcentaje será de 2,02%



CASO Nº 8

$$X = \mu \pm (z * s)$$

DATOS:

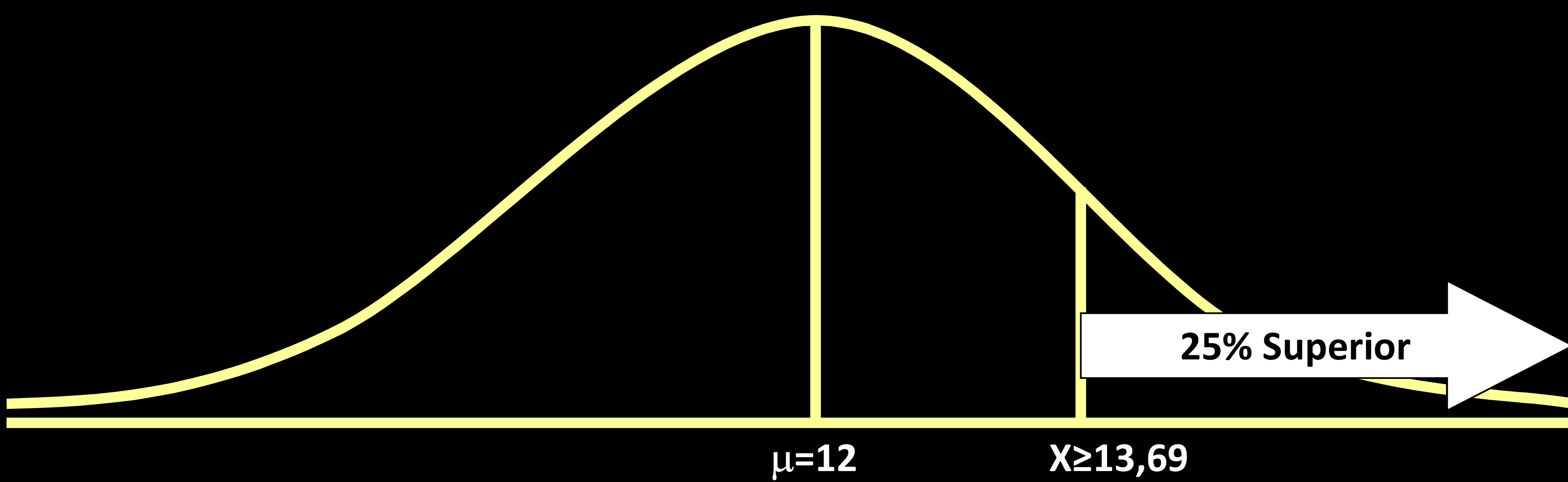
μ = 12 Puntos

S= 2,5 Puntos

n = 150 Estudiantes.

X = Calificación Mínima correspondiente para el 25% superior:

SOLUCIÓN: Primero se determina el valor de Z correspondiente al 25%, el cual sería 25%/100=0,2500, este valor se busca en la tabla de la curva normal, en caso de no aparecer exactamente se ubica el valor más próximo. Nótese que en este caso el valor de 0,2500 no aparece exactamente, sin embargo el más próximo es: 0,2486, cuyo valor Z es de (0,67), luego se procede el cálculo: X = 12 + (0,67*2,5) = 13,67 Puntos.



CASO Nº 9

DATOS:

μ = 12 Puntos

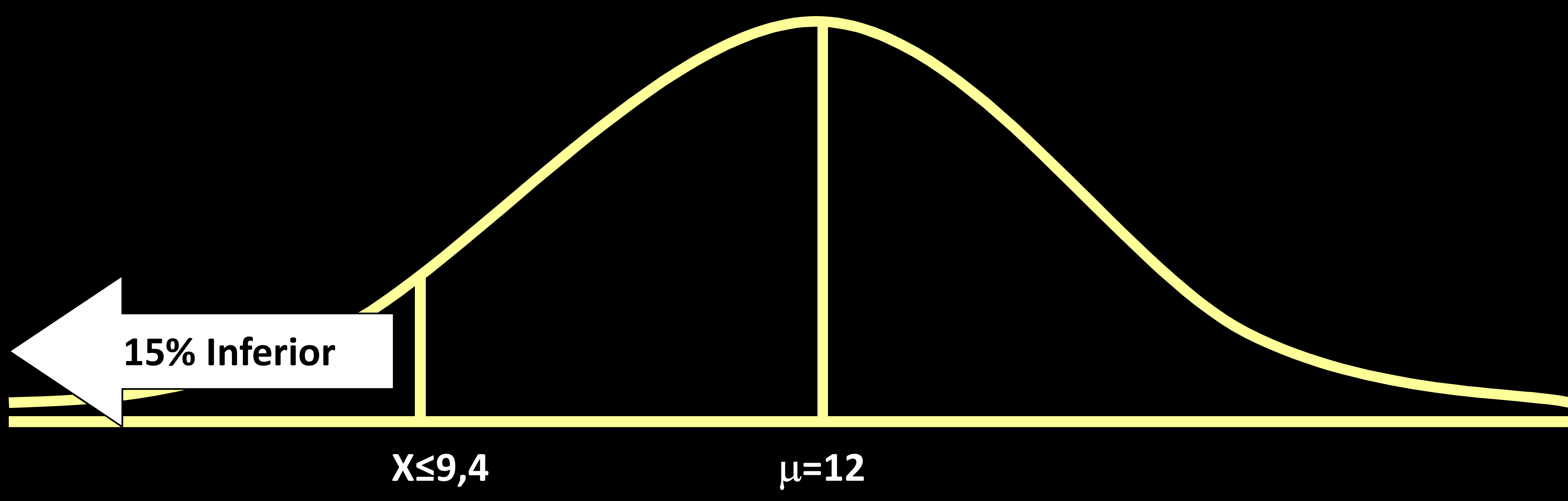
S= 2,5 Puntos

n = 150 Estudiantes.

X = Calificación Máxima correspondiente para el 15% inferior:

$$X = \mu \pm (z * s)$$

SOLUCIÓN: Primero se determina el valor de Z correspondiente al 35%, el cual sería 35%/100=0,3500, luego se busca en la tabla la curva normal, en caso de no aparecer exactamente se ubica el valor más próximo. Nótese que en este caso el valor de 0,3500 no aparece exactamente, se toma el más próximo, 0,3508, cuyo valor Z es de (1,04), una vez obtenido el valor Z se calcula: X = 12 – (1,04*2,5) = 9,4 Puntos..



CASO Nº 10

$$Z = \frac{X - \mu}{S}$$

DATOS:

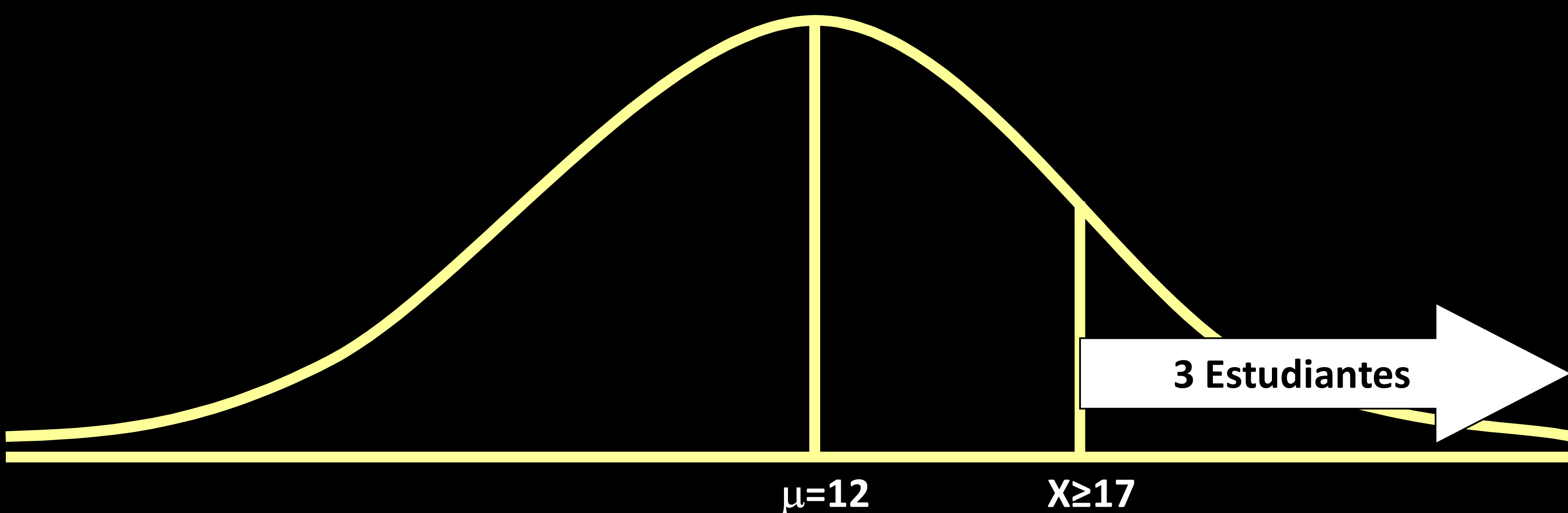
μ = 12 Puntos

S= 2,5 Puntos

n = 150 Estudiantes.

Cuantos estudiantes tendrían una calificación igual o mayor a 17 puntos

SOLUCIÓN: Z=(17–12)/2,5 = 2,00 El valor por tabla sería de: 0,4772, y la probabilidad buscada será igual a (0,5–0,4772) = 0,0228*150=3,42=3 estudiantes



CASO Nº 11

$$Z = \frac{X - \mu}{S}$$

DATOS:

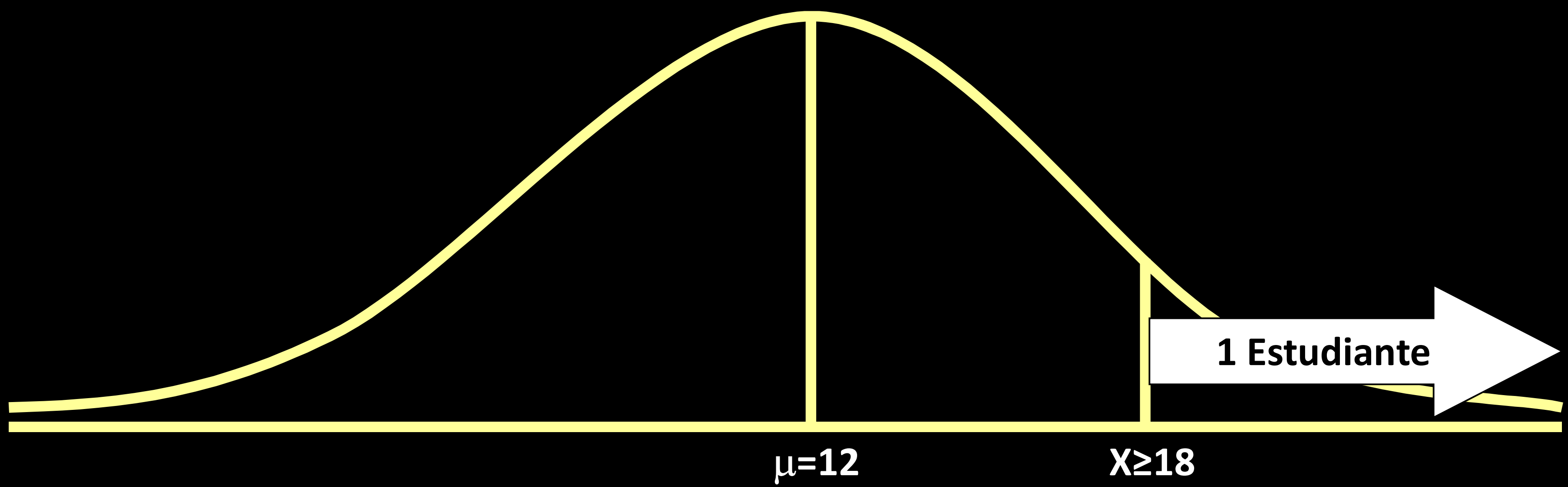
μ = 12 Puntos

S= 2,5 Puntos

n = 150 Estudiantes.

Si la universidad decide becar a los alumnos con calificaciones iguales o mayores a 18 puntos cuantos serían becados.

SOLUCIÓN: Z=(18–12)/2,5 = 2,40. El valor por tabla sería de: 0,4918, y la probabilidad buscada será igual a (0,5–0,4918) = 0,0082*150 = 1,23 = 1 estudiante



CASO Nº 12

DATOS:

μ = 12 Puntos

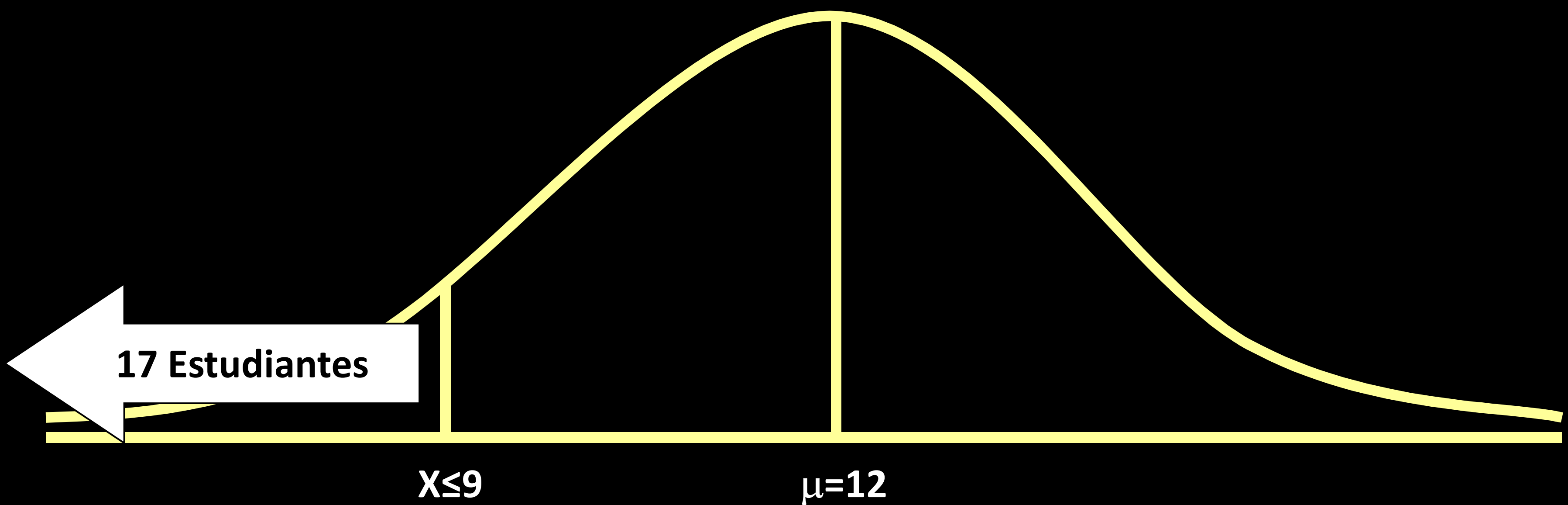
S= 2,5 Puntos

n = 150 Estudiantes.

$$Z = \frac{X - \mu}{S}$$

Si la universidad decide suspender a los alumnos con calificaciones iguales o menores a 9 puntos cuantos serían suspendidos.

SOLUCIÓN Z=(9–12)/2,5 = 1,20 El valor por tabla sería de: 0,3849, y la probabilidad buscada será igual a (0,5–0,3849) = 0,1151*150 = 17,26 =17 estudiantes



ALGUNAS DEFINICIONES PARTINENTES A LOS PROXIMOS MODELOS

UNIVERSO: Es el conjunto de seres, elementos u objetos que poseen una o más características comunes, que pueden ser observables.

POBLACIÓN: Es un subconjunto del universo de elementos o seres que poseen características comunes las cuales pueden ser sometidas a estudio y análisis.

MUESTRA: Es un subconjunto de la población de elementos o seres, que poseen características comunes que pueden ser observables y sometidas a estudio y análisis. En tal sentido existen muchos tipos de muestra, pero la clasificación más importante para nosotros, es la que distingue entre muestras probabilísticas y no probabilísticas, según se puede calcular la probabilidad o no, de que un elemento de la población sea incluido en la muestra.

MUESTRAS NO PROBABILÍSTICAS: Son todas aquellas que son escogidas aplicando procedimientos que no permiten calcular la probabilidad de que un elemento de la población pueda pertenecer a la muestra escogida. En este grupo entran las llamadas muestras opináticas a base de un experto, y las intencionales.

Sin embargo es importante señalar, que para poder hacer inferencia estadística (que tiene base probabilística), es imprescindible que se tengan muestras probabilísticas

MUESTRAS ALEATORIAS SIMPLES: Son todas aquellas en las cuales todos los elementos de observación de la población, tienen igual probabilidad de pertenecer a la muestras escogida, en otras palabras son muestras escogidas al azar.

PARÁMETROS: Son valores representativos de la distribución en toda la población, tales como la media que se designa por μ, la desviación típica que designamos por σ. Etc.

ESTADÍSTICOS: Son valores representativos de la distribución en una muestra, tales como: la media que se deseigna por X, la desviación típica que se designa por S, etc. En todo caso a los estadísticos se deseiganan con letras latinas mayúsculas, y a los parámetros con letra griegas minúsculas.

DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO

Como ya hemos visto el objeto de la Inferencia Estadística, es obtener conclusiones válidas acerca de los valores representativos de una población **(PARÁMETROS)** a partir del conocimiento de los valores representativos de una muestra **(ESTDÍSTICOS).** Para lograr esto se hace necesario que se conozca el comportamiento de esos estadísticos cuando se toman muestras de la población de la cual proceden.

Igualmente hay que destacar, que una población puede dar lugar a muchas posibles muestras de un cierto tamaño o cantidad.

DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO

Ahora bien para poder estudiar el comportamiento de los estadísticos al tomar muestras de una población, supondremos que dicha población es infinita y que las muestras que se extraen son estrictamente aleatorias, es decir, que todos y cada uno de los elementos de la población tienen la misma probabilidad de aparecer o pertenecer a una muestra cualesquiera.

Al hacer esta suposición inicial es obvio que el número de muestras de un determinado tamaño extraídas de una población dada es infinito y por lo tanto los valores de un estadístico cualquiera calculado en todas esas posibles muestras, no formarán una distribución de frecuencias, sino una función de probabilidad (o función de densidad, si se considera que la variable en estudio es continua, que es lo más común que ocurra)

En tal sentido, llamaremos **DISTRIBUCIÓN MUESTRAL DE UN ESTADÍSTICO**, a esa función de probabilidad (o densidad) de los valores de ese estadístico calculado en todas las posibles muestras de cierto tamaño que pueden extraerse de la población considerada

VEAMOS EL SIGUIENTE EJEMPLO: Supongamos que tenemos la siguiente población; (Xa=3; Xb=5; Xc=6; Xd=7 y Xe=9), y queremos construir la DISTRIBUCIÓN MUESTRAL DE LA MEDIA, con muestras al azar de n=2 de dicha población. Para esto hay que considerar todas las posibles muestras de tamaño 2 que pueden extraerse de la población a las cuales se les calculará la media aritmética entonces se tendrían las siguientes muestras:

(3-5);(3-6);(3-7);(3-9);(5-6);(5-7);(5-9);(6-7);(6-9);(7-9)

La Media de cada muestra sería: (3+5)/2=4,0 (3+6)/2=4,5 (3+7)/2=5,0 (3+9)/2=6,0 (5+6)/2=5,5 (5+7)/2=6,0 (5+9)/2=7,0 (6+7)/2=6,5 (6+9)/2=7,5 (7+9)/2=8,0

CONTINUACIÓN

A continuación calcularemos la MEDIA Y LA DSESVIACION TIPICA de ambas distribuciones, para lo cual utilizaremos datos no agrupados, y le asignaremos letras griegas por tratarse de parámetros:

_MEDIA DE LA DISTRIBUCIÓN POBLACIONAL:

μ = (3+5+6+7+9) /5= (6)

DESVIACIÓN TÍPICA DE LA DISTRIBUCIÓN POBLACIONAL:

$\sigma = \sqrt{}\ (3-6)^2 + (5-6)^2 + (6-6)^2 + (7-6)^2 + (9-6)^2 / 5 = \sqrt{}\ 20/5 = (2)$

CONTINUACIÓN

MEDIA DE LA DISTRIBUCIÓN MUESTRAL DE MEDIAS:

μ_X = (4+4,5+5+6+5,5+6+7+6,5+7,5+8) / 10 = 60 / 10 = (6)

DESVIACIÓN TÍPICA DE LA DISTRIBUCIÓN MUESTRAL DE MEDIAS:

$\sigma X = \sqrt{}$ (4-6)2 + (4,5-6)2 + (5-6)2 + (6-6)2 + (5,5-6)2 + (6-6)2 + (7-6)2 + (6,5-6)2 + (7,5-6)2 + (8-6)2 / 5 = $\sqrt{}$ 15/10 = (1,22)

Las conclusiones q1ue se obtienen al comparar los resultados nos indican , que la MEDIA DE LA DISTRIBUCIÓN MUESTRAL DE MEDIAS ES IGUAL A LA MEDIA POBLACIONAL, lo cual indica de modo general que $\mu = \mu X$, y LA DESVIACION TIPICA DE LA DISTRIBUCION MUESTRAL DE MEDIAS ES MENOR QUE LA DE LA POBLACION.

EJERCICIOS DE DISTRIBUCIÓN MUESTRAL DE LA MEDIA ARTIMÉTICA

Un estudio acerca del rendimiento académico de los estudiantes de una universidad dio como resultado una distribución aproximadamente normal, con una MEDIA de 15 puntos y una DESVIACIÓN TÍPICA de 8 puntos; si se toman muestras al azar de 50 estudiantes calcule la probabilidad de que en una de esas muestras escogidas al azar se obtenga una MEDIA:

1.- Igual o mayor a 16 puntos.

2.- Igual o menor a 12 puntos.

3.- Comprendida entre 16 y 17 puntos.

4.- Comprendida entre 13 y 14 puntos.

5.- Igual o mayor a 18 puntos.

6.- Igual o menor a 15 puntos

7.- Igual o mayor a 12 puntos.

8.- En cuantas muestras se puede esperar que la media sea igual o menor 13 puntos.

CASO Nº 1

DATOS:

μ_O= 15 Puntos.

n = Muestras de 50 Estudiantes.

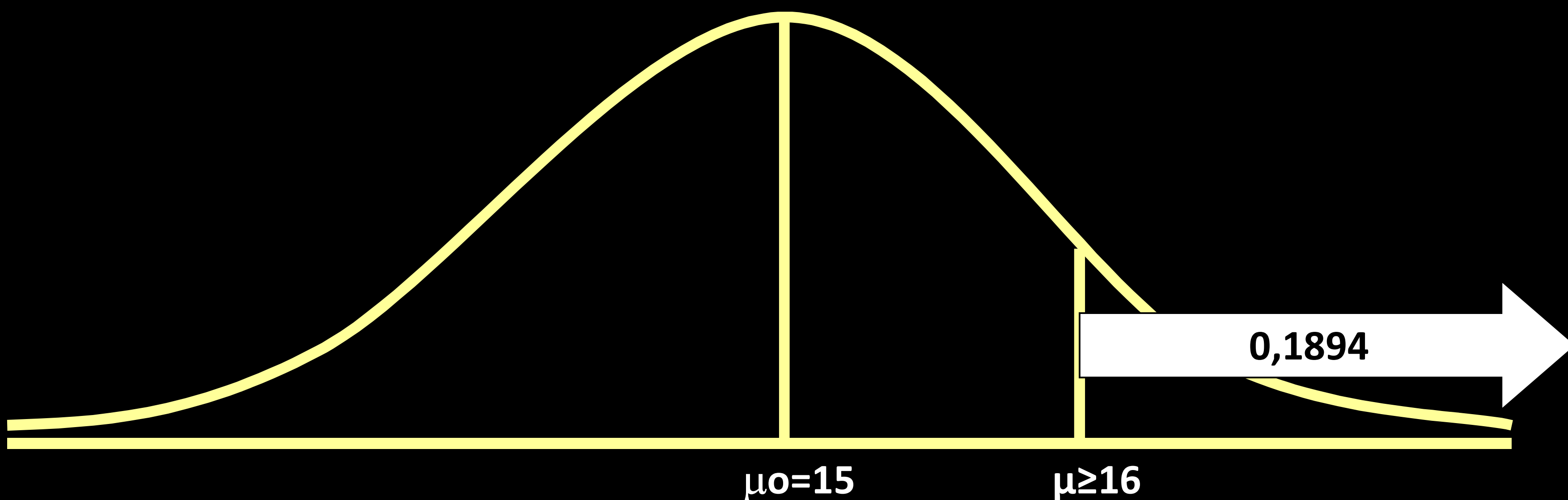
S = 8 Puntos

S/ $\sqrt{n}$ = 8/ $\sqrt{50}$ = 8/7,07 = (1,13)

Probabilidad: ($\mu \geq$ 16 Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z=(16–15)/1,13 = 0,88. El valor en tabla sería de: 0,3106, y la probabilidad buscada será de (0,5–0,3106) = 0,1894, es la probabilidad o del 18,94%.



CASO Nº 2

DATOS:

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

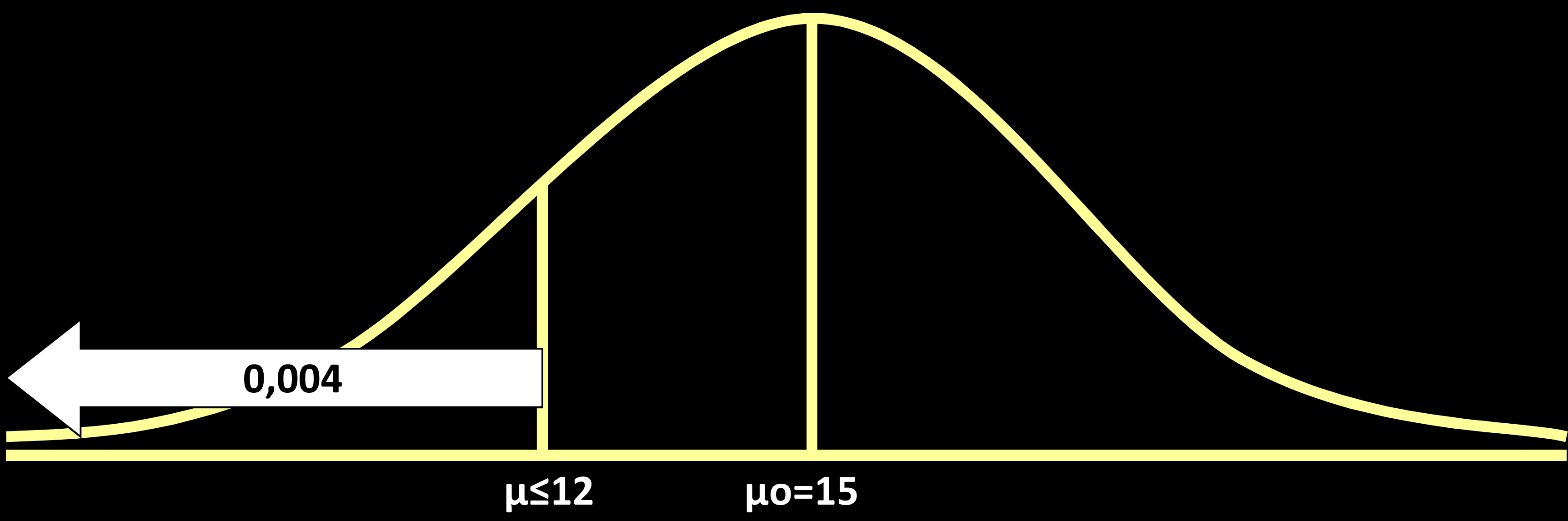
S = 8 Puntos

S/ $\sqrt{n}$ = 8/ $\sqrt{50}$ = 8/7,07 = (1,13)

Probabilidad: ($\mu \leq$ 12 Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z=(12–15)/1,13 = – 2,65. El valor en tabla sería de: 0,4960, y la probabilidad buscada será de (0,5–0,4960) = 0,004, en porcentaje 0,4%



CASO Nº 3

DATOS

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

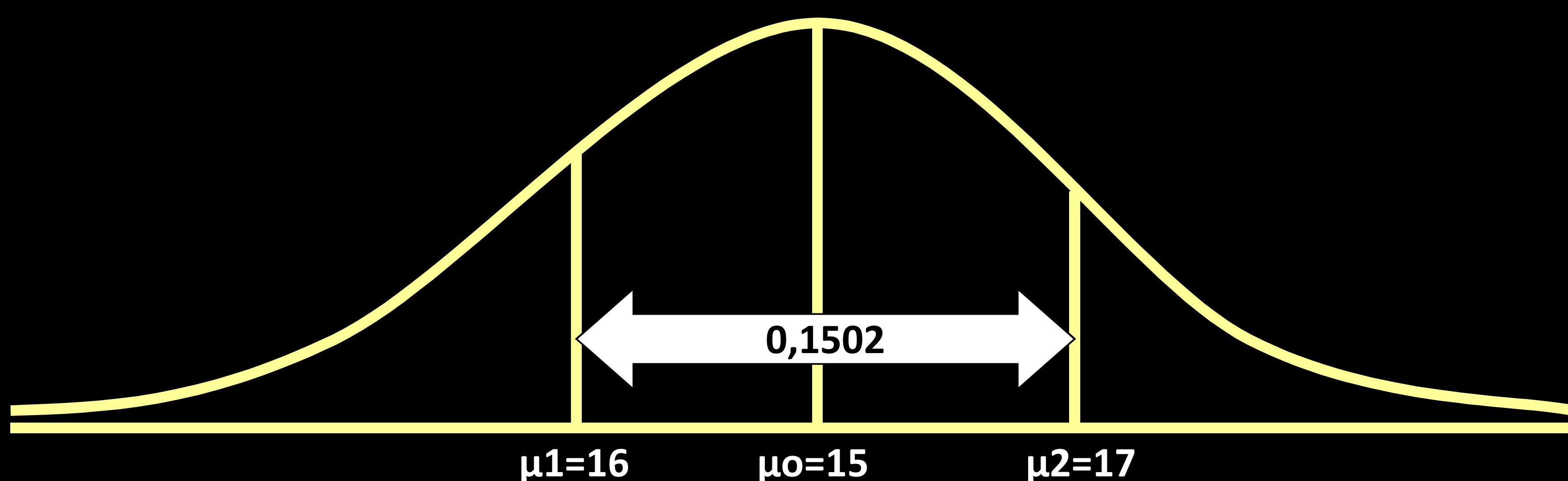
S = 8 Puntos

S/ $\sqrt{}$n = 8/ $\sqrt{}$50 = 8 / 7,07 = (1,13)

P (μ Entre 16 y 17 Puntos)

$$Z = \frac{\mu - \mu o}{s / \sqrt{n}}$$

SOLUCIÓN: Z1=(16–15)/1,13=0,88. El valor por tabla sería de: 0,3106, Z2=(17–15)/1,13=1,76. El valor por tabla sería de: 0,4608 y la probabilidad buscada será igual a (0,4608–0,3106) = 0,1502, en porcentaje 15,02%



CASO Nº 4

DATOS:

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

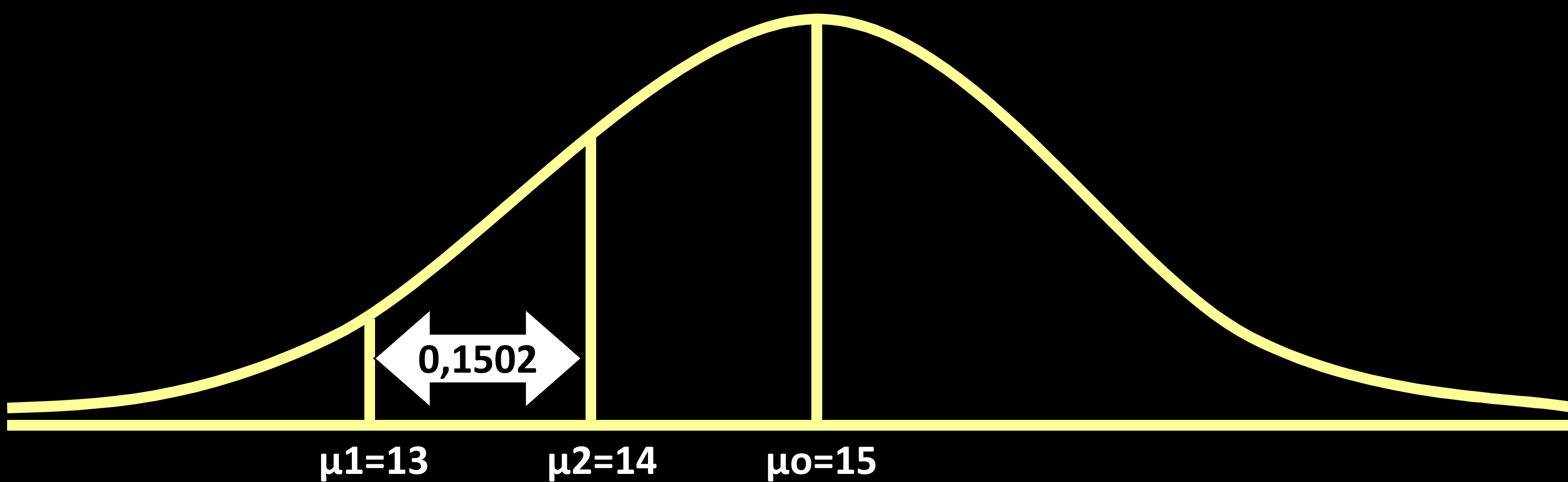
S = 8 Puntos

$S/\sqrt{n} = 8/\sqrt{50} = 8 / 7{,}07 = (1{,}13)$

Probabilidad: (μ Entre 13 y 14 Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z1=(13–15)/1,13 = - 1,76. El valor por tabla sería de: 0,4608, Z2=(14–15)/1,13 = - 0,88. El valor por tabla sería de: 0,3106 y la probabilidad buscada será igual a (0,4608 – 0,3106) = 0,1502 , en porcentaje 15,02%



CASO Nº 5

$$Z = \frac{\mu - \mu o}{s / \sqrt{n}}$$

DATOS:

μ_O= 15 Puntos.

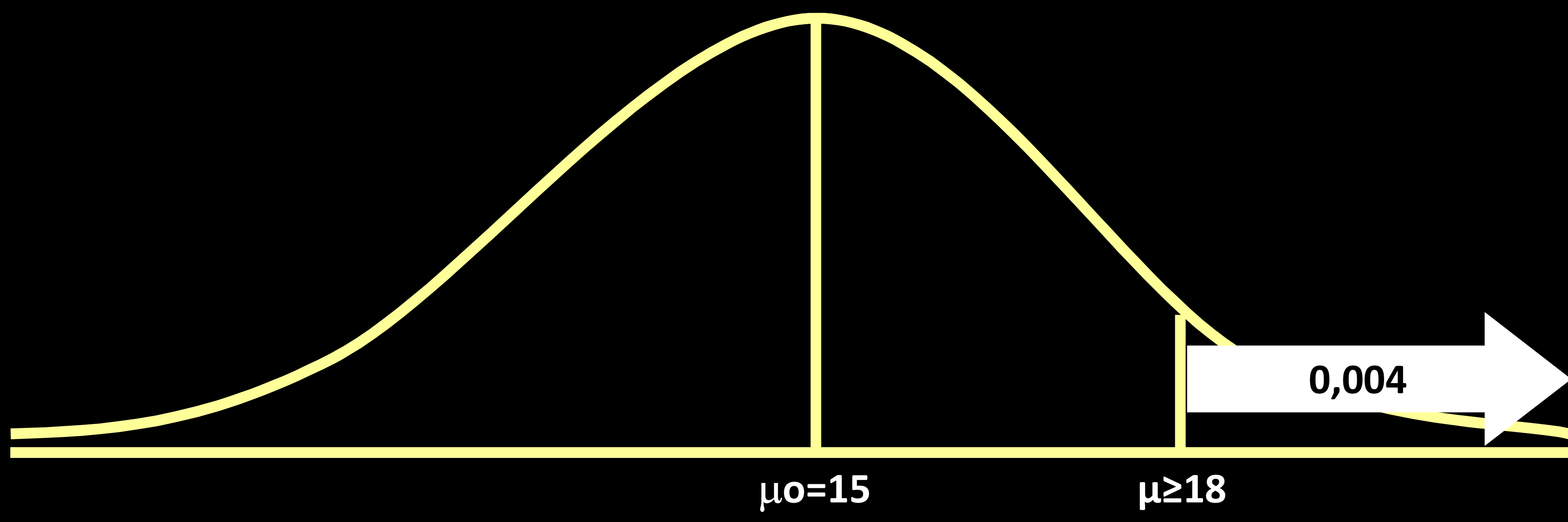
n = Muestras de 50 Estudiantes.

S = 8 Puntos

S/ $\sqrt{}$n = 8/ $\sqrt{}$50 = 8 / 7,07 = (1,13)

Probabilidad: ($\mu \geq$ 18 Puntos)

SOLUCIÓN: Z=(18–15)/1,13 = 2,65. El valor en tabla sería de: 0,4960, y la probabilidad buscada será de (0,5 – 0,4960) = 0,004, en porcentaje 0,4%



CASO Nº 6

DATOS:

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

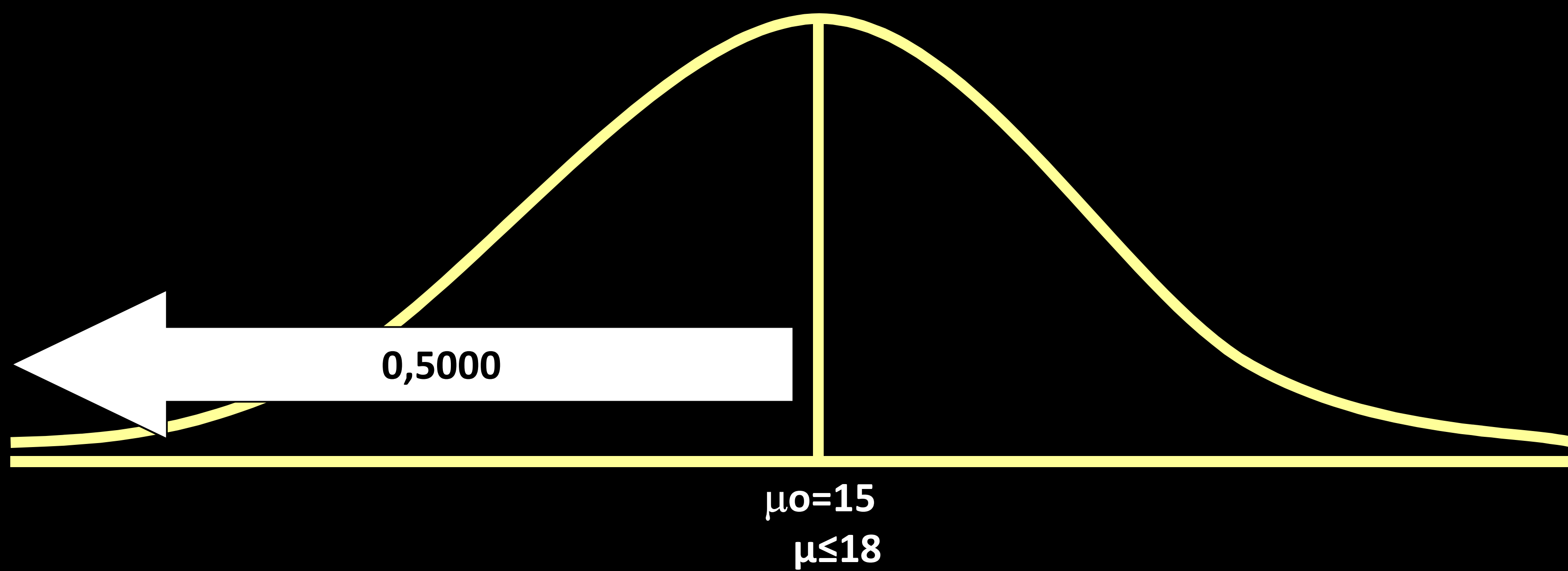
S = 8 Puntos

$S/\sqrt{n} = 8/\sqrt{50} = 8 / 7{,}07 = (1{,}13)$

Probabilidad: ($\mu \leq 15$ Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z=(15–15)/1,13 = 0,00 El valor en tabla sería de: 0,0000, y la probabilidad buscada será de 0,5 – 0,0000 = 0,5, en porcentaje 50%



CASO Nº 7

DATOS:

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

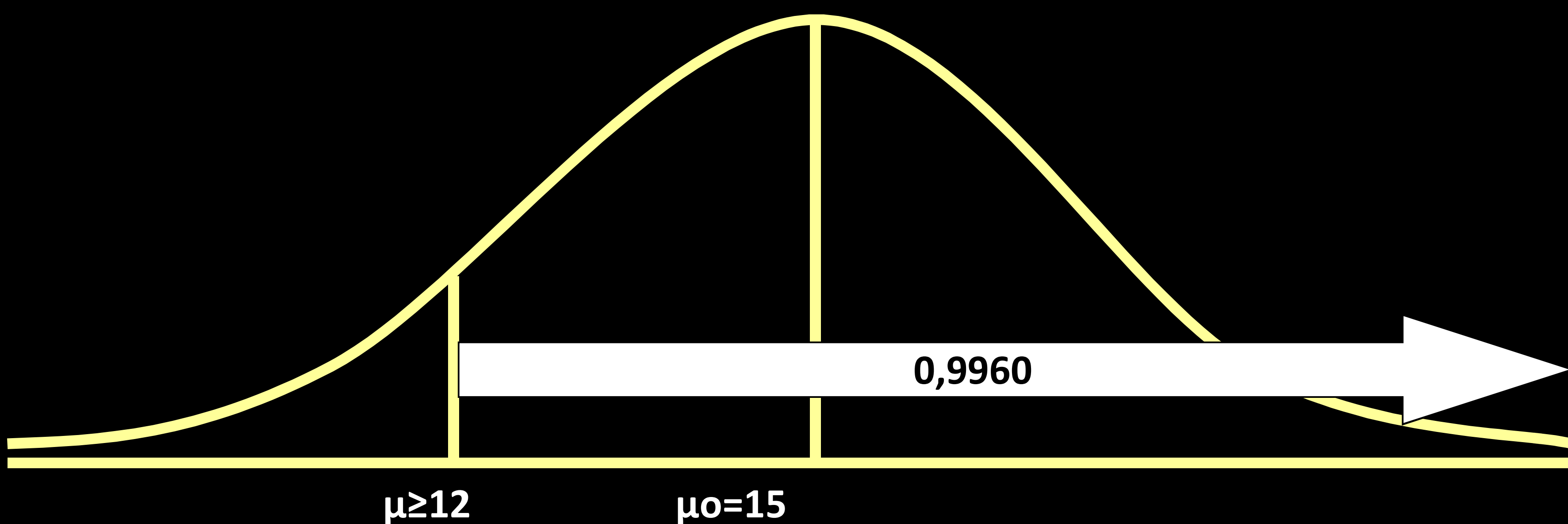
S = 8 Puntos

S/ √n = 8/ √50 = 8 / 7,07 = (1,13)

Probabilidad: (μ ≥ 12 Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z=(12–15)/1,13 = 2,65. El valor en tabla sería de: 0,4960, y la probabilidad buscada será de (0,5+0,4960) = 0,9960, en porcentaje 99,60%



CASO Nº 8

DATOS:

μ_o= 15 Puntos.

n = Muestras de 50 Estudiantes.

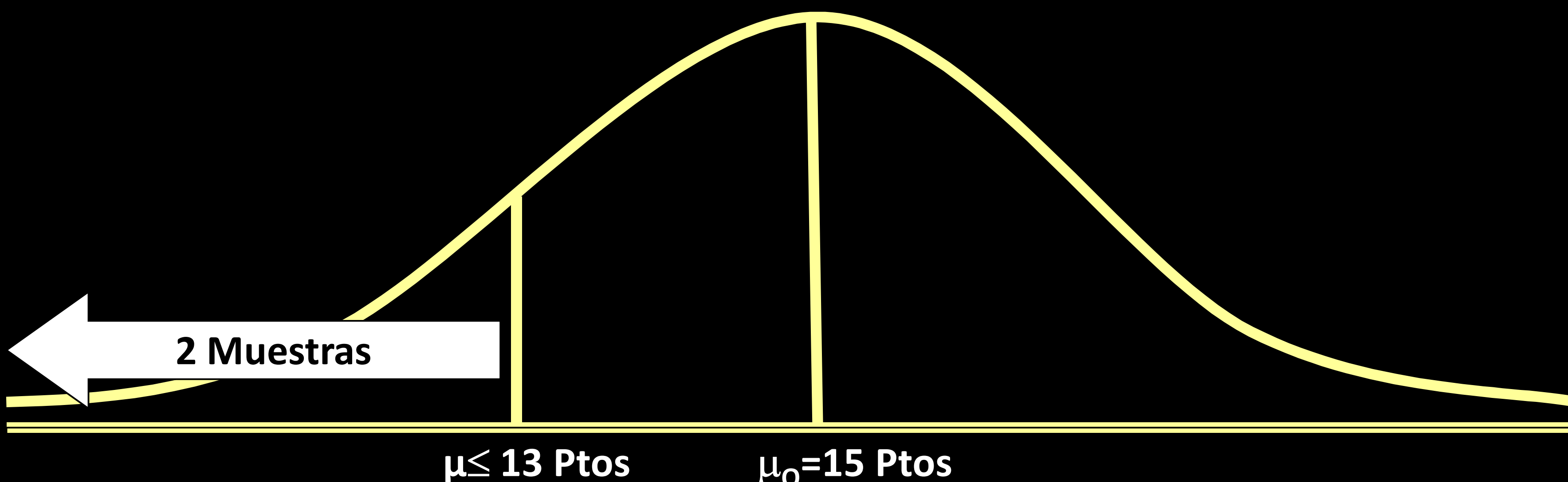
S = 8 Puntos

$S/\sqrt{n} = 8/\sqrt{50} = 8 / 7{,}07 = (1{,}13)$

Probabilidad: (En cuantas muestras se espera que la media sea $\mu \leq$ 13 Puntos)

$$Z = \frac{\mu - \mu o}{s/\sqrt{n}}$$

SOLUCIÓN: Z=(13–15)/1,13 = – 1,76. El valor en tabla sería de: 0,4608, la probabilidad buscada será de (0,5–0,4608) = 0,0392, y el número de muestras será de 0,0392*50 = 1,96 $\cong$ 2



EJERCICIOS DE DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN O PORCENTAJE:

A través de un estudio, la Universidad Gran Mariscal de Ayacucho, ha determinado que sólo el 25% de sus estudiantes, pueden clasificarse como sobresalientes. Si se toman muestras al azar de 250 estudiantes calcule la probabilidad de que en una de esas muestras escogidas al azar se obtenga un PORCENTAJE DE:

1.- 28% o más sobresalientes.

2.- 20% o menos sobresalientes.

3.- Entre el 19% y el 31% de sobresalientes.

4.- Entre el 29% y el 32% de sobresalientes.

5.- En cuantas muestras se puede esperar que porcentaje de sobresalientes sea del 30% o más

LA FORMULA A UTILIZAR ES LA SIGUIENTE:

$$Z = \frac{P - Po}{\sqrt{\frac{Po * Qo}{n}}}$$

DONDE:

- ✓ (p) representa la probabilidad a calcular
- ✓ (Po) la proporción o porcentaje a favor del suceso que se está calculando y (Q) la proporción en contra . En todo caso P y Q son probabilidades o porcentajes complementarios, de manera que en porcentaje (Q) es igual a (100 – P), y (P) es igual a (100 – Q). En caso de que se este trabajandp con proporción, entonces (Q) es igual a (1- P) y (P) es igual a (1-Q). En todo caso la suma de ambos debe ser 100% o en su defecto (1).

PRIMER CASO

DATOS:

$$Z = \frac{P - Po}{\sqrt{\frac{Po * Qo}{n}}}$$

Po = 25%

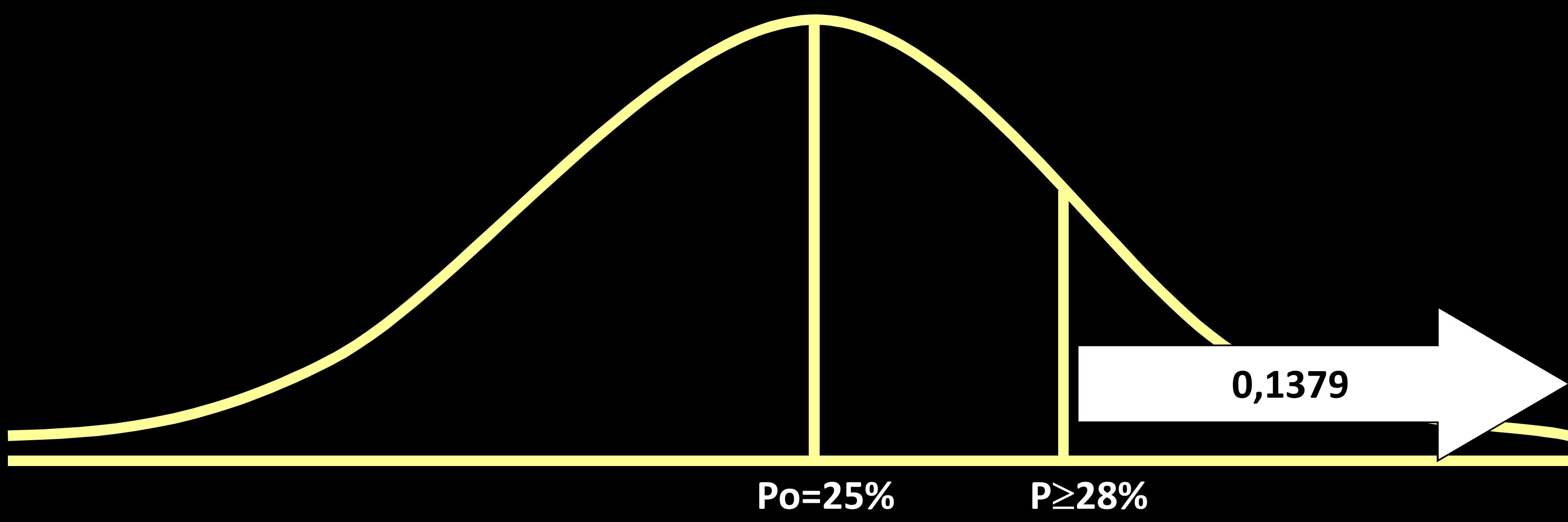
Qo = 100 – 25= 75%

n = 250 Estudiantes

√(Po*Qo / n) = √ (25%*75% /250) = 2,73

Probabilidad: (≥ 28% sobresalientes)

SOLUCIÓN: Z=(28%–25%)/2,74 = 1,09. El valor en tabla sería de: 0,3621, y la probabilidad buscada será de (0,5 – 0,3621) = 0,1379, en porcentaje 13,79%



SEGUNDO CASO.

$$Z = \frac{P - Po}{\sqrt{\frac{Po * Qo}{n}}}$$

DATOS:

Po = 25%

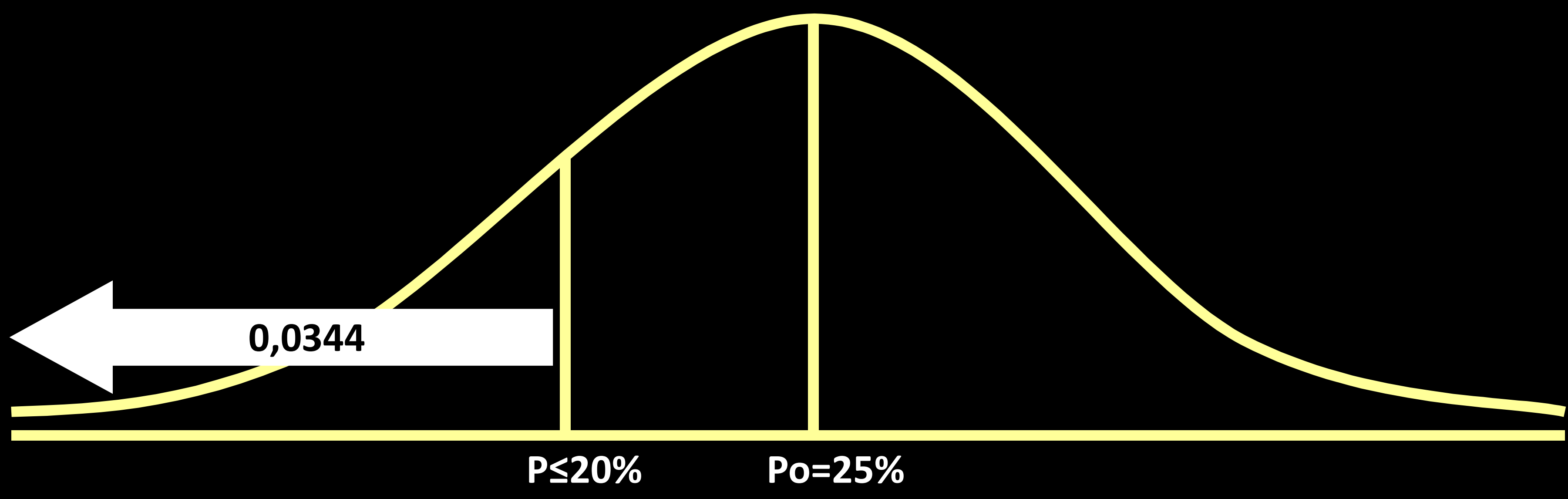
Qo = 100 – 25= 75%

n = 250 Estudiantes

√(Po*Qo / n) = √ (25%*75% /250) = 2,73

Probabilidad: (≤ 20% sobresalientes)

SOLUCIÓN: Z=(20%–25%)/2,74 = - 1,82. El valor en tabla sería de: 0,4656, y la probabilidad buscada será de (0,5 – 0,4656) = 0,0344, en porcentaje 3,44%



TERCER CASO

DATOS:

$$Z = \frac{P - Po}{\sqrt{\frac{Po \ * \ Qo}{n}}}$$

Po = 25%

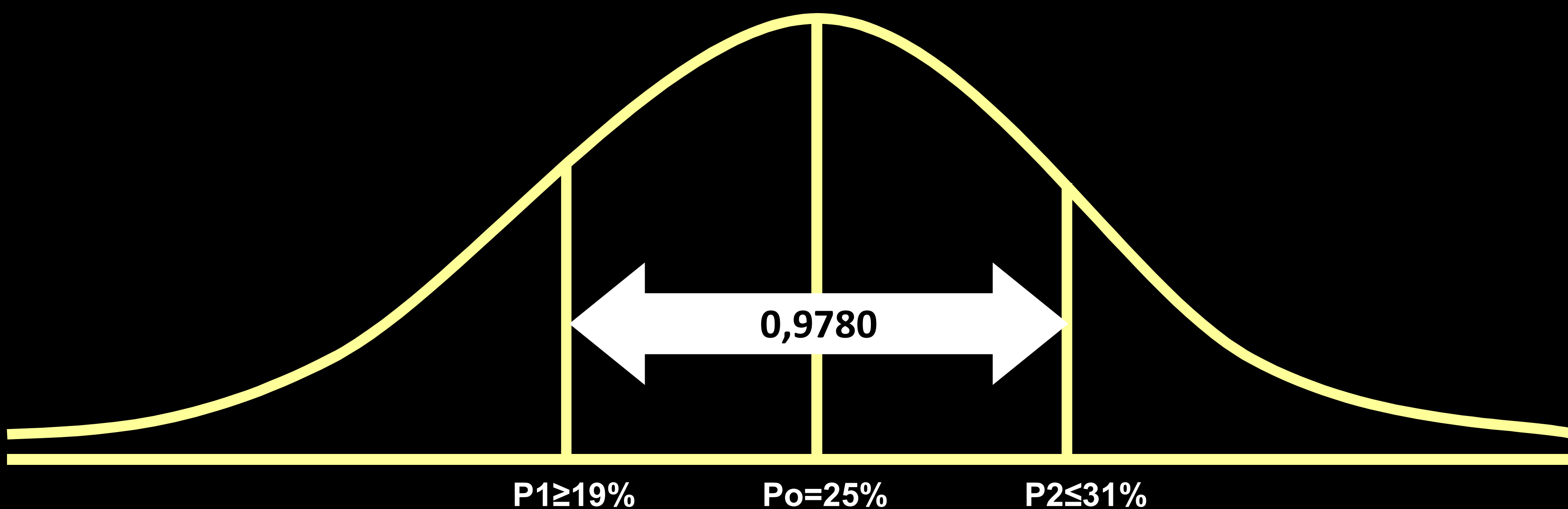
Qo = 100 – 25= 75%

n = 250 Estudiantes

√(Po*Qo / n) = √(25%*75% /250) = 2,73

Probabilidad: Entre 19% y 31% Sobresalientes

SOLUCIÓN: Z1=(19%–25%)/2,74= - 2,18. El valor en tabla sería de: 0,4854, Z2=(31%–25%)/2,74 = (2,18) El valor en tabla sería de: 0,4854, y la probabilidad buscada será de (0,4854+0,4854=0,9708, en porcentaje 97,08%



CUARTO CASO

DATOS:

Po = 25%

Qo = 100 – 25= 75%

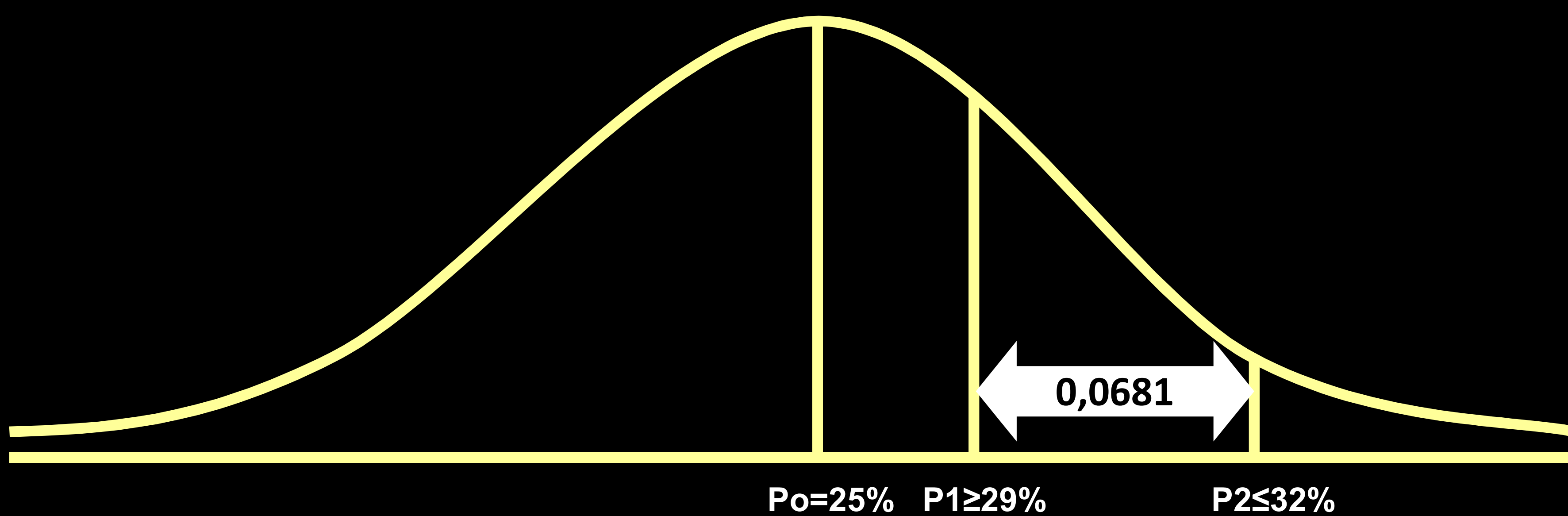
n = 250 Estudiantes

√(Po*Qo / n) = √(25%*75% /250) = 2,73

Probabilidad: (Entre 29% y 32% Sobresalientes)

$$Z = \frac{P - Po}{\sqrt{\dfrac{Po * Qo}{n}}}$$

SOLUCIÓN: Z1=(29%–25%)/2,74= 1,45. El valor en tabla sería de: 0,4265, Z2=(32%–25%)/2,74 = 2,55. El valor en tabla sería de: 0,4946, y la probabilidad buscada será de (0,4946–0,4265) = 0,0681, en porcentaje 6,81%



ESTIMACIÓN POR INTERVALOS

Consiste en la obtención de un intervalo dentro del cual estará el valor del parámetro estimado con una cierta probabilidad. En la estimación por intervalos se usan los siguientes conceptos:

INTERVALO DE CONFIANZA

El intervalo de confianza es una expresión del tipo [$\vartheta 1$, $\vartheta 2$] ó $\vartheta 1 \leq \vartheta \leq \vartheta 2$, donde θ es el parámetro a estimar. Este intervalo contiene al parámetro estimado con una determinada certeza o nivel de confianza establecido de antemano.

VARIABILIDAD DEL PARÁMETRO

Si no se conoce, puede obtenerse una aproximación en los datos aportados por la literatura científica o en un estudio piloto. También hay métodos para calcular el tamaño de la muestra que prescinden de este aspecto. Habitualmente se usa como medida de esta variabilidad la desviación típica poblacional y se denota σ.

ERROR DE LA ESTIMACIÓN

Es una medida de su precisión que se corresponde con la amplitud del intervalo de confianza. Cuanta más precisión se desee en la estimación de un parámetro, más estrecho deberá ser el intervalo de confianza y, si se quiere mantener o disminuír el error, más ocurrencias deberán incluirse en la muestra estudiada. En caso de no incluir nuevas observaciones para la muestra, más error se comete al aumentar la precisión. Se suele llamar E, según la fórmula $E = \vartheta 2 - \vartheta 1$.

LÍMITE DE CONFIANZA

Es la probabilidad de que el verdadero valor del parámetro estimado en la población se sitúe en el intervalo de confianza obtenido. El nivel de confianza se denota por (1-α), aunque habitualmente suele expresarse con un porcentaje ((1-α)·100%). Es habitual tomar como nivel de confianza un 95% o un 99%, que se corresponden con valores α de 0,05 y 0,01 respectivamente.

VALOR ALFA (α)

También llamado nivel de significación. Es la probabilidad (en tanto por uno) de fallar en nuestra estimación, esto es, la diferencia entre la certeza (1) y el nivel de confianza (1-α). Por ejemplo, en una estimación con un nivel de confianza del 95%, el valor α es (100-95)/100 = 0,05.

VALOR CRÍTICO

Se representa por Z α/2. Es el valor de la abscisa en una determinada distribución que deja a su derecha un área igual a α/2, siendo 1-α el nivel de confianza. Normalmente los valores críticos están tabulados o pueden calcularse en función de la distribución de la población. Por ejemplo, para una distribución normal, de media 0 y desviación típica 1, el valor crítico para α = 0,05 se calcularía del siguiente modo: se busca en la tabla de la distribución ese valor (o el más aproximado), bajo la columna "Área"; se observa que se corresponde con -0,64. Entonces Zα/2 = 0,64. Si la media o desviación típica de la distribución normal no coinciden con las de la tabla, se puede realizar el cambio de variable t =(X− μ)/s para su cálculo.

ESTIMACIÓN DE PARÁMETROS

ESTIMACIÓN PUNTUAL: El problema más directo que se presenta en la inferencia estadística, es el de estimar el valor de un parámetro a partir de un estadístico, para lo cual debe hacerse uso de los conocimientos acerca de la distribución muestral de ese estadístico. En general decimos que el estadístico escogido es un "ESTIMADOR PUNTUAL" del parámetro requerido, y los valores particulares de dicho estadístico son estimaciones puntuales del valor del parámetro.

ESTIMACIÓN DE PARÁMETROS

En realidad se pueden escoger muchos estadísticos, y la selección entre ellos se debe hacer en base a ciertas propiedades deseables en el estimador. La primera y más importante de ellas es que dicho estimador debe ser INSESGADO, es decir que la esperanza matemática de la distribución muestral de estadístico usado como estimador debe ser igual al parámetro, la segunda es que sea EFICIENTE en comparación con otros, es decir que su error típico sea menor a los errores típicos de los demás posibles estimadores, y la tercera es que sea CONSISTENTE, es decir que su error típico disminuye a medida que aumenta el tamaño de la muestra considerada.

ESTIMACIÓN POR INTERVALO
(NIVEL DE CONFIANZA DE UNA ESTIMACIÓN)

Para estimar el valor de un parámetro poblacional, se da un intervalo en el cual puede encontrarse el parámetro con una cierta confianza que se establece de antemano.

EJEMPLO: Para estimar la edad media de los habitantes de una población, se realizó un estudio en una muestra de 100 habitantes, obtuvo una distribución aproximadamente normal con una media aritmética de 24 años con una desviación típica de 6 años. En función de esta información realice los siguientes cálculos:

1.- Estime la edad promedio para toda la población con un 99% de confianza.

2.- Estime la edad promedio para toda la población con un 95% de confianza.

FORMULA DE CÁLCULO:

$$\mu \pm Z * \left(\frac{S}{\sqrt{n}} \right)$$

DONDE:

μ = a la Media Aritmética de la Muestra (Estadístico)

S = a Desviación Típica de la Muestra (Estadístico)

Z = Al Valor Tipificado para la Confianza Previamente Establecida (Area Bajo la Curva Normal)

n = Al tamaño de la Muestra escogida para el Estudio.

CASO Nº 1: Estimar le edad promedio para toda la población con un 99% de confianza.

DATOS:

μ = 24 Años

S = 6 Años

n = 100 Habitantes

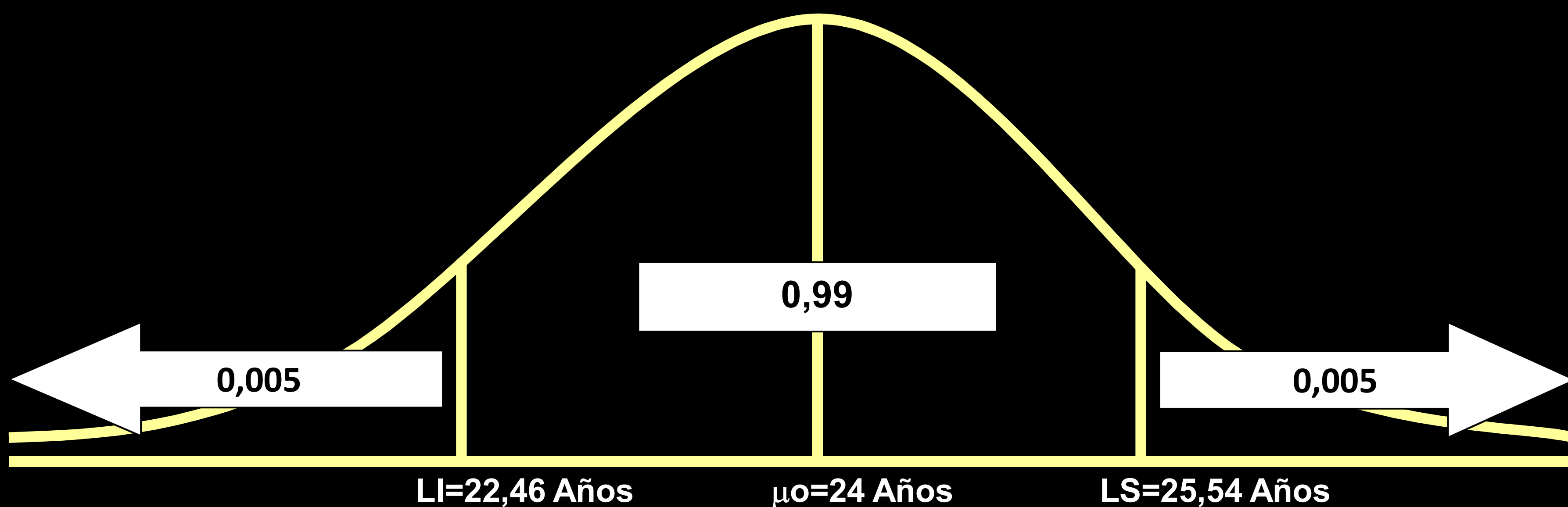
$$\mu \pm Z * \left(\frac{S}{\sqrt{n}} \right) = 24 \pm 2.58 * \left(\frac{6}{\sqrt{100}} \right)$$

$$24 + (2.58 * 0.6) = 24.54 \text{ Años } = \text{Límite Superior}$$

$$24 - (2.58 * 0.6 = 22.46 \text{ Años } = \text{Límite Inferior}$$

Z = (99%/2) = (49,5/100) = 0,4950 = (2,58) (Ver Tabla de Áreas bajo la Curva Normal)

Entonces se puede aceptar que existe un 99% de confianza de que la edad promedio se encuentre entre 22,46 y 25,54 años



CASO Nº 2: Estimar le edad promedio para toda la población con un 95% de confianza.

DATOS:

$$\mu \pm Z * \left(\frac{S}{\sqrt{n}}\right) = 24 \pm 1{,}96 * \left(\frac{6}{100}\right) =$$

$$24 + (1{,}96 * 0{,}6) = 25{,}17 \text{ Años } = \text{Límite Superior}$$

$$24 - (1{,}96 * 0{,}6) = 22{,}82 \text{ Años } = \text{Lïmite Inferior}$$

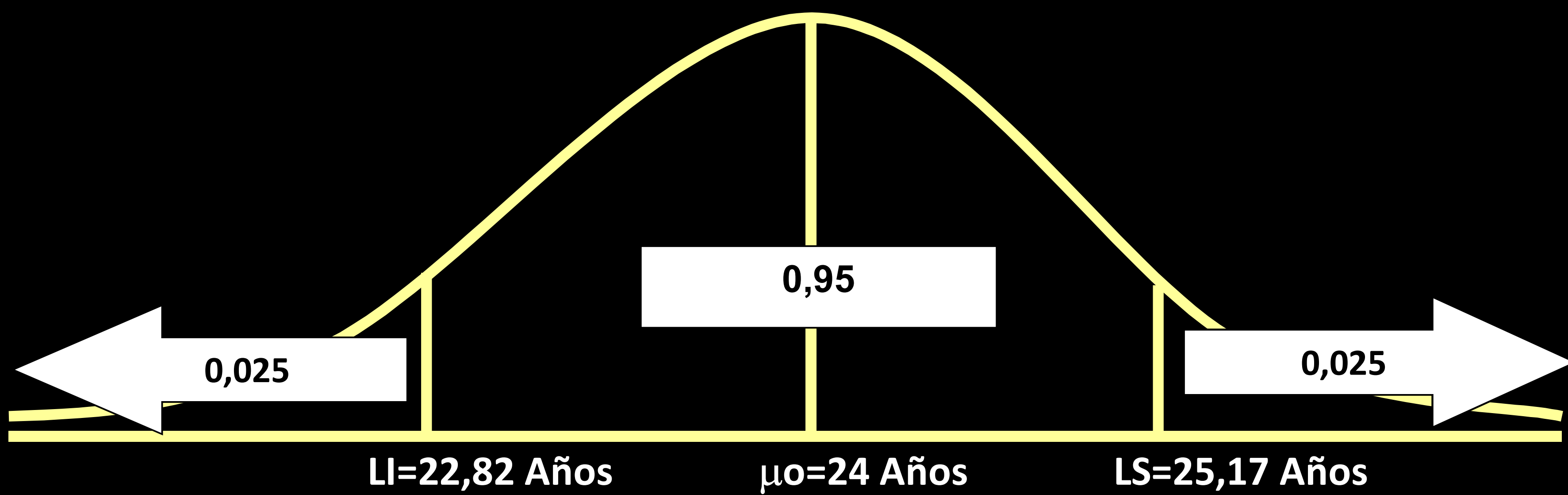
μo = 24 Años

S = 6 Años

n = 100 Habitantes

Z = (95%/2) = (47,5/100) = 0,4750 = (1,96) (Ver Tabla de Áreas bajo la Curva Normal)

Entonces se puede aceptar que existe un 95% de confianza de que la edad promedio se encuentre entre 22,82 y 25,17 años



ESTIMACIÓN POR INTERVALO ACERCA DE LA MEDIA ARITMÉTICA PARA MUESTRAS PEQUEÑAS (29 O MENOS DATOS)

En este tipo de ejercicios se conocen tanto la media como la desviación típica de la muestra, pero no conocemos la desviación típica de la población además de que la muestra es pequeña, lo cual implica que se debe recurrir a la distribución muestral denominada "T" de Student.

La forma de esta distribución es análoga a la normal: (simétrica y unimodal), además de tener esperanza matemática cero y desviación típica uno; sin embargo para cada valor de n-1 que son los grados de libertad hay una función "t" que en estos casos es la que hay que aplicar.

VEAMOS UN EJEMPLO DE APLICACIÓN PARA ESTIMACIÓN POR INTERVALOS DE LA MEDIA ARITMÉTICA PARA MUESTRAS PEQUEÑAS

En una muestra de diez estudiantes se aplicó un examen objetivo, y dio como resultado una media de 12 puntos con una desviación típica de 3 puntos. Se pide calcular la verdadera calificación promedio aplicando un intervalo de confianza del 95%

FORMULA DE CÁLCULO:

$$\mu \pm Tc * \left(\frac{S}{\sqrt{n-1}} \right)$$

DONDE:

μ = a la Media Aritmética de la Muestra (Estadístico) = 12 Puntos

S = a Desviación Típica de la Muestra (Estadístico) = 3 Puntos

Tc = Al Valor Tipificado para la Confianza Previamente Establecida para Tc

n-1 = Al tamaño de la Muestra escogida para el Estudio menos uno = 10-1 = 9

SOLUCIÓN DEL EJERCICIO ANTERIOR:

Para determinar el valor que corresponde a Tc debemos tener en cuenta los grados de libertad que resultan de restar n-1, en nuestro caso son 10-1= 9 grados de libertad con 95% de confianza. Para una estimación bilateral es de 2,26

DATOS:

μ= 12 Puntos

S= 3 Puntos

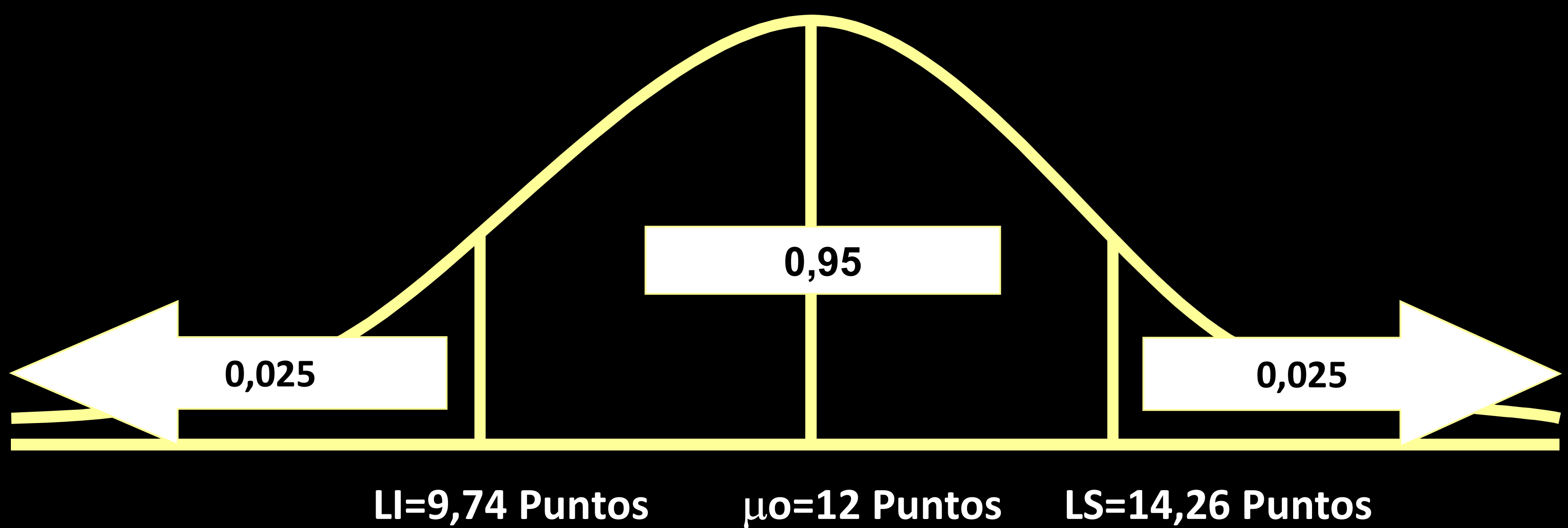
n= 10 Estudiantes

Tc= 2,26

$$\mu \pm Tc * \left(\frac{S}{\sqrt{n-1}}\right) = 12 \pm 2,26 * \left(\frac{3}{\sqrt{(10-1)}}\right) = 12 \pm 2,26$$

$$12 + 2,26 = 14,26 \text{ Puntos Límite Superior}$$

$$12 - 2,26 = 9,74 \text{ Puntos Límite Inferior}$$



ESTIMACIÓN POR INTERVALO ACERCA DE UNA PROPORCIÓN O PORCENTAJE

En este tipo de ejercicios se procede de forma similar a los anteriores de estimaciones por intervalo sólo que en estos casos se trata de proporciones o porcentajes y no de la media aritmética.

VEAMOS UN EJEMPLO DE APLICACIÓN:

Dentro de la población de adultos de una ciudad se ha elegido una muestra de 150 personas adultos de los cuales 80 resultaron ser analfabetos. Se pide que Usted determine la proporción de analfabetos adultos de esa ciudad con un 80% de confianza en la estimación. Sin embargo hay que considerar que para hallar los límites de confianza del parámetro debemos conocer si podemos aproximar la distribución muestral P (que es binomial) a una distribución normal. Para ello calculamos primero la proporción de analfabetos de la forma siguiente: 80/150= 0,533 = 53,3%, luego se multiplican por P y Q por n , y si el resultado es mayor que 5 se procede con la aproximación. Veamos: n.P = 150.0,533= (79,95) y (1-0,533) = 0,467. 150 =(70,05), como ambos son mayores que 5 procede la aproximación

FORMULA DE CÁLCULO

$$P \pm Z * \left(\sqrt{\frac{P * Q}{n}} \right)$$

DONDE:

P = Proporción a estimar (Estadístico)

Q = 1-P (Estadístico)

Zc = Al Valor Tipificado para la Confianza Previamente Establecida. (Área Bajo la Curva Normal)

n = Al tamaño de la Muestra escogida para el Estudio

DATOS:

Po = 0,533 = 53,3%

Q = 1 – 0,533 = 0,467 = 46,7%

n = 150

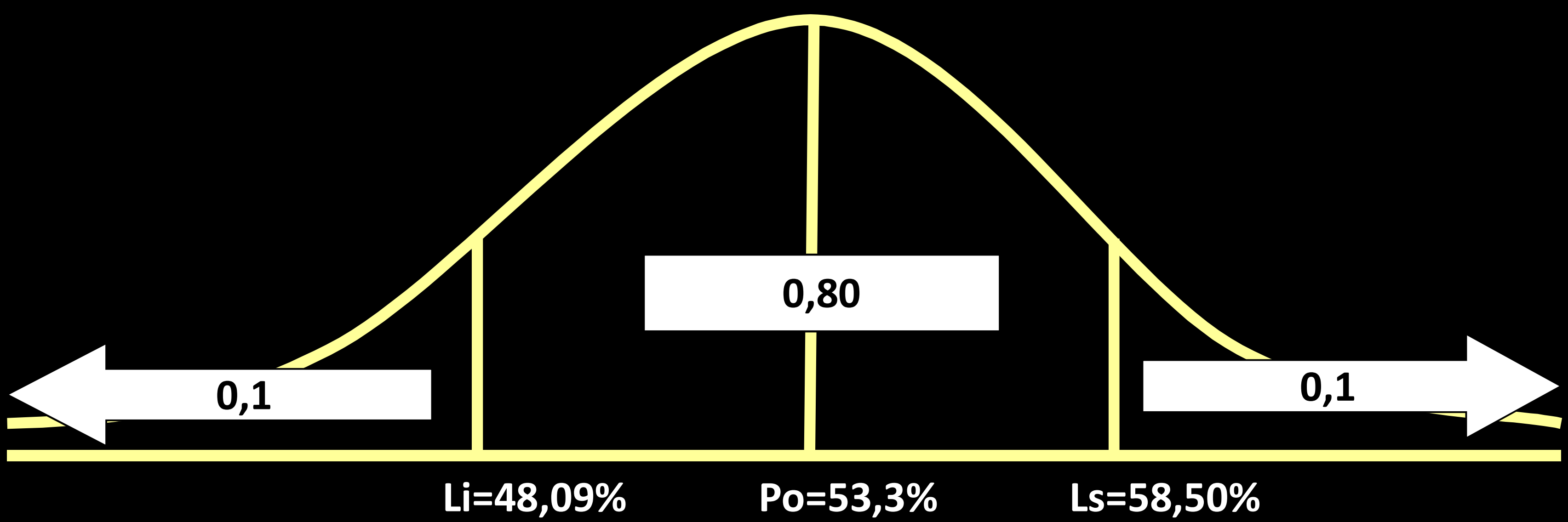
Zc = 80% / 2 /100 = 0,4000 = Zc = 1,28

$$P \pm Z * \left(\sqrt{\frac{P * Q}{n}} \right) = 53{,}3 \pm 1{,}28 * \left(\sqrt{\frac{53{,}3 * 46{,}7}{150}} \right)$$

$$53{,}3 + (1{,}28 * 4{,}07) = \text{Límite Superior } 58{,}50\%$$

$$53{,}3 - (1{,}28 * 4{,}07) = \text{Límite Inferior } 48{,}09\%$$

Luego podemos aceptar que existe un 80% de confianza de que la proporción de adultos analfabetos para toda la población está comprendida entre el 48,09% y el 58,50%.



EJERCICIOS DE ESTIMACIÓN PARA UNA DIFERENCIA DE MEDIAS MUESTRAS GRANDES:

Al igual que hemos venido procediendo para las estimaciones en los ejercicios anteriores, se puede hacer estimaciones para una diferencia de medias o de proporciones, sólo que en estos casos se trabajan con dos muestras diferentes las cuales presentan datos diferentes, sin embargo los procedimientos de cálculo son análogos.

VEAMOS UN EJEMPLO DE APLICACIÓN:

En un test de conocimientos generales aplicado a 215 estudiantes de las escuelas publicas de Caracas, se obtuvo una media de puntuaciones de 14,3 puntos con una desviación típica de 3,6 puntos; en otra muestra de 200 estudiantes de las escuelas privadas de la misma ciudad se obtuvo una media de puntuaciones de 13,2 puntos con una desviación típica de 2,4 puntos respectivamente. Se pide calcular la diferencia entre las medias de puntuaciones de ambas muestras con un 95% de confianza en la estimación. La fórmula a utilizar es la siguiente:

$$\mu 1 - \mu 2 \pm Z * \left(\sqrt{\frac{S1^2}{n1} + \frac{S2^2}{n2}} \right)$$

DATOS:

$\mu 1 = (14{,}3)$; $\mu 2 = (13{,}2)$

n1 = (215); n2 = (200)

S1 = (3,6); S2 = 2,4

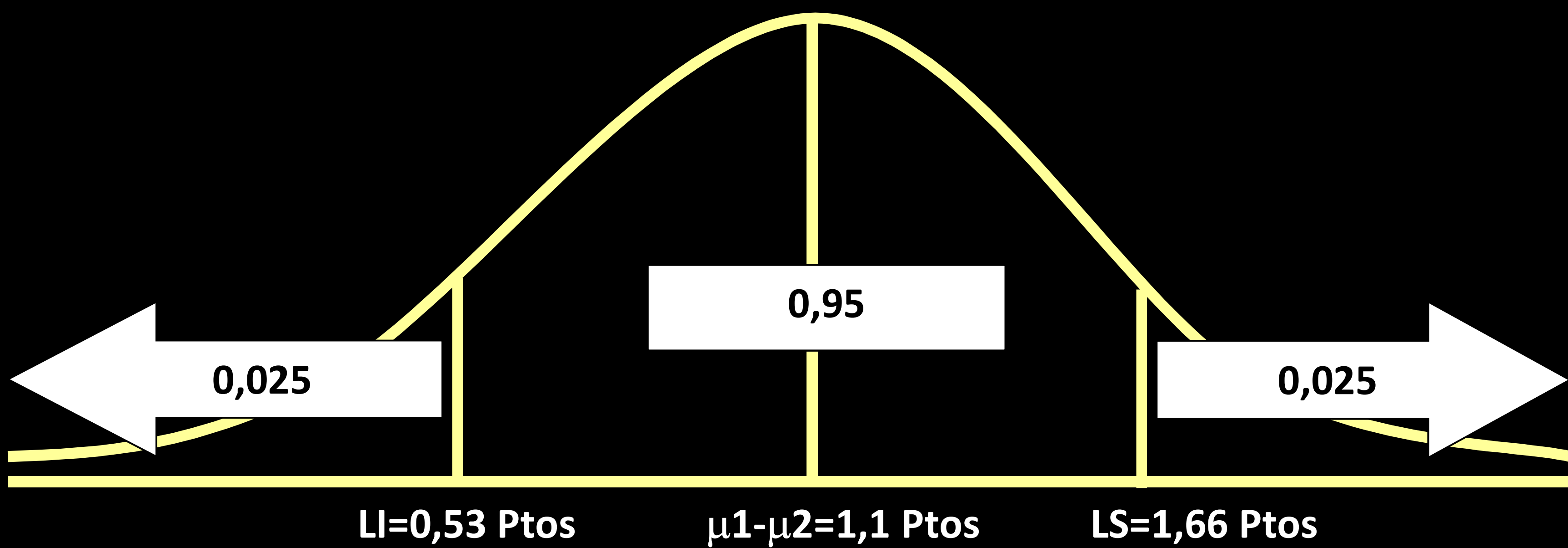
Z = 95% = 1,96

$$\mu 1 - \mu 2 \pm Z * \left(\sqrt{\frac{S1^2}{n1} + \frac{S2^2}{n2}}\right) = 14{,}3 - 13{,}2 \pm 1{,}96 * \left(\sqrt{\frac{(3{,}6)^2}{215} + \frac{(2{,}4)^2}{200}}\right)$$

$$1{,}1 + (1{,}96 * 0{,}29) = \text{Límite Superior } 1{,}66 \text{ Puntos}$$

$$1{,}1 - (1{,}96 * 0{,}29) = \text{Límite Inferior } 0{,}53 \text{ Puntos}$$

Entonces existe un 95% se confianza de que las medias de puntuaciones entre las muestras consideradas difieran entre 0,53 y 1,66 puntos respectivamente.



EJERCICIOS DE ESTIMACIÓN PARA UNA DIFERENCIA DE PROPORCIONES MUESTRAS GRANDES:

Supongamos que en el ejemplo anterior de los 215 estudiantes de las escuelas públicas, 150 eran varones y de los 200 de las escuelas privadas 84 eran varones. Se desea estimar con un 95% de confianza la diferencia entre el porcentaje de varones de las escuelas públicas y privadas. La fórmula de cálculo es la siguiente:

$$P1 - P2 \pm Z * \left(\sqrt{\frac{P1 * Q1}{n1} + \frac{P2 * Q2}{n2}} \right)$$

DATOS:

P1 = 150/215/*100 = 69,76%

P2 = 84/200*100 = 42,00%

N1 = 215; N2 = 200

Q1 = 100-69,7 6= 30,24%

Q2 = 100-42.00 = 58,00%

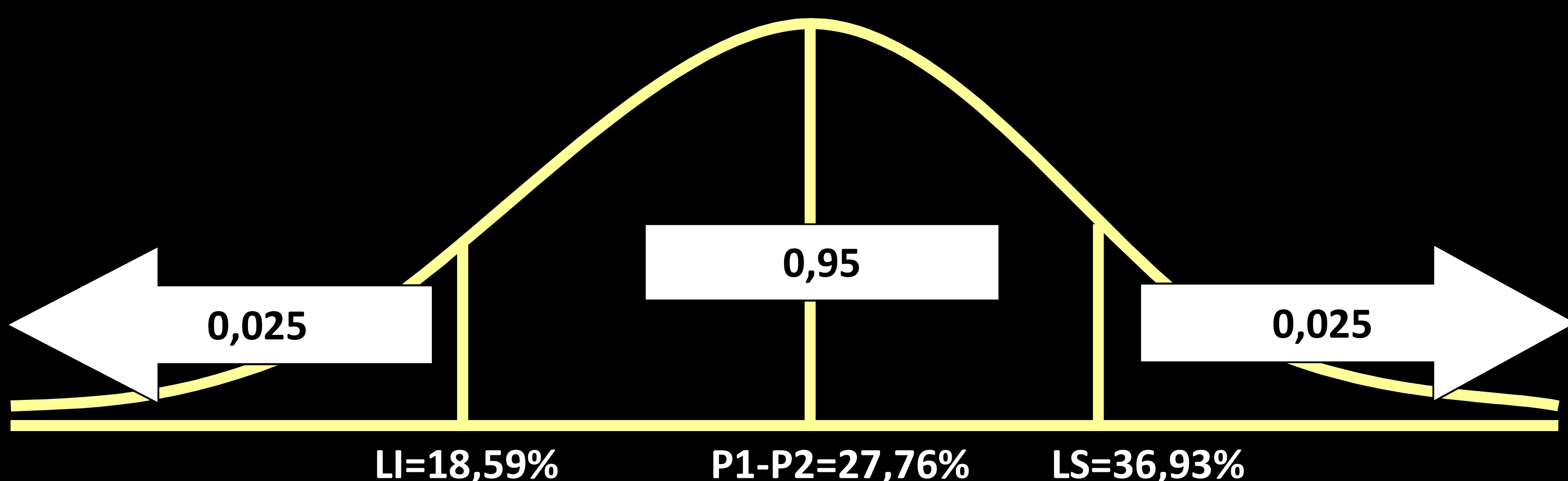
Zc = 95%/2/100 = 0,4750 = 1,96

$$P1 - P2 \pm Z * \left(\sqrt{\frac{P1 * Q1}{n1} + \frac{P2 * Q2}{n2}} \right) = 69{,}76 - 42{,}00 \pm 1{,}96 * \left(\sqrt{\frac{69{,}76 * 30{,}24}{215} + \frac{42{,}00 * 58{,}00}{200}} \right)$$

$$27{,}76 + (1{,}96 * 4{,}68) = \text{Límite Superior } 36{,}93\%$$

$$27{,}76 - (1{,}96 * 4{,}68) = \text{Límite Inferior } 18{,}59\%$$

Entonces existe un 95% se confianza de que el porcentaje de varones entre las escuelas publicas y privadas presente una diferencias entre el 18,59% y el 36,93%



ESTIMACIÓN DEL TAMAÑO DE UNA MUESTRA

En estadística el tamaño de una muestra no es otra cosa que el número de unidades de observación (sujetos u objetos) extraídos de la población objeto de estudio de forma tal que el tamaño muestral sea representativo de dicha población.

Ahora bien hay que tener presente que existen diferentes formulas para determinar o calcular el tamaño de una muestra, y esto va a depender primero de los parámetros que se conozcan, y segundo del tipo de estudio o investigación que se está realizando.

FORMULAS DE CÁLCULO

Estimación del tamaño de una muestra cuando se trata de una proporción o porcentaje y la población es infinita:

$$n = \frac{Z^2 * P * Q}{e^2}$$

Estimación del tamaño de una muestra cuando se trata de una proporción o porcentaje y la población es finita:

$$n = \frac{(Z)^2 * P * Q * NP}{(NP - 1) * (e)^2 + (Z)^2 * P * Q}$$

EJEMPLO DE CÁLCULO PARA EL ESTUDIO DE LA MEDIA ARITMÉTICA Y LA POBLACIÓN ES INFINITA

Se quiere estimar la estatura media de un grupo numeroso de estudiantes con un error no mayor de 2 centímetros y un 95% de confianza en la estimación.

Por estudios anteriores se conoce que la desviación típica de la distribución de estaturas en de aproximadamente 10 Cm. ¿Cuál será en tamaño muestral ideal para realizar el estudio respectivo?

DATOS:

$Zc^2 = (1,96)^2 = 3,84$

$\sigma^2 = (10)^2 = 100$

$e^2 = (2)^2 = 4$

SOLUCIÓN:

$$n = \frac{(1,96)^2 * (10)^2}{(2)^2} = 96,04 \approx 96 \text{ Estudiantes}$$

VEAMOS EL MISMO EJEMPLO DE CÁLCULO PARA UNA POBLACIÓN FINITA

Se quiere estimar la estatura media de un grupo numeroso de 600 estudiantes con un error no mayor de 2 centímetros y un 95% de confianza en la estimación.

Por estudios anteriores se conoce que la desviación típica de la distribución de estaturas en de aproximadamente 10 Cm. ¿Cuál será en tamaño muestras ideal para realizar el estudio en cuestión?. (NOTESE QUE EN ESTE CASO LA POBLACIÓN ES FINITA 600 ESTUDIANTES)

DATOS:

$Zc^2 = (1{,}96)^2 = 3{,}84$

$\sigma^2 = (10)^2 = 100$

$e^2 = (2)^2 = 4$

NP= 600

NP-1 = 599

SOLUCION:

$$n = \frac{(1{,}96)^2 * (10)^2 * 600}{(600-1) * (2)^2 + (1{,}96)^2 * (10)^2} = \frac{230496}{2396 + 384{,}16} = 82{,}9 \approx 83 \text{ Estudiantes}$$

EJEMPLO DE CÁLCULO PARA EL ESTUDIO DE UNA PROPORCIÓN O PORCENTAJE Y LA POBLACION ES INFINITA

Se desea estimar el porcentaje de estudiantes que estaría a favor de un determinado candidato rectoral de una universidad del país, con un error no mayor de 5% y con un 95% de confianza.

Determine el tamaño muestral que se debe escoger si puede estimarse que el porcentaje a favor del candidato estará alrededor del 40%

FORMULA A UTILIZAR

$$n = \frac{Z^2 * P * Q}{e^2}$$

DATOS:

$Zc^2 = (1{,}96)^2 = 3{,}84$

$P = (40\%)$

$Q = (60\%)$

$e = (5\%)^2 = 25$

$$n = \frac{3{,}84 * 40 * 60}{25} = 368{,}64 \cong 369$$

(369 estudiantes sería el tamaño muestral mínimo que debería escogerse para el estudio

CONTRASTE DE UNA HIPÓTESIS ESTADÍSTICA

Habiendo estudiado el primer aspecto de Inferencia Estadística como es la estimación de parámetros a partir de valores muestrales (Estadísticos), debemos considerar otro aspecto que tiene que ver con la decisión de aceptar o rechazar una hipótesis estadística de la población basándonos en los datos de una muestra de esa población.

Una hipótesis estadística es una suposición, pero acerca del valor de un parámetro de una población definida previamente. De aquí se deduce que para plantear una hipótesis estadística es necesario que se derive o deduzca de la hipótesis de investigación , el valor de un parámetro de una población bien definida que no necesariamente tiene que corresponder con una población en un sentido físico o gráfico del termino, sino que es totalmente conceptual (un conjunto de valores de una variable medidos en un universo físicamente bien determinado.

CONTINUACIÓN

Por otro lado hay que tener presente, que generalmente de hipótesis estadística que se contrasta es la de aquel valor del parámetro, que representa lo contrario de lo esperado según la hipótesis de investigación, por lo cual se llama (HIPOTESIS NULA), de manera que aceptar la hipótesis nula equivale a rechazar la hipótesis planteada originalmente en la investigación (HIPOTESIS ALTERNATIVA) y viceversa.

Cuando se contrasta una hipótesis estadística, no se demuestra o se prueba nunca la verdad o falsedad de una hipótesis planteada, sino que se da una regla para decidir la aceptación o rechazo de de una hipótesis basados en el valor de un estadístico (calculado a partir de una muestra de la población en estudio), y fijados de antemano la máxima posibilidad de error al rechazarla y la alternativa que aceptamos en ese caso.

TIPOS DE ERRORES EN UN CONTRASTE DE UNA HIPÓTESIS ESTADÍSTICA

Como cada contraste de hipótesis se refiere a la región de aceptación o rechazo de la HIPOTESISI NULA, se pueden presentar dos tipos de errores al tomar la decisión:

a.- Si se rechaza la HIPOTESIS NULA, habiendo razones para aceptarla (CIERTA) (ERROR TIPO 1)

b.- Si se acepta la HIPOTESIS NULA, habiendo razones para rechazarla (FALSA) (ERROR TIPO 2)

La probabilidad de cometer ERROR TIPO 1 se denomina con la letra (α) y a la probabilidad de cometer ERROR TIPO 2, se denomina (β) en tal sentido tendríamos:

CONTINUACIÓN

NIVEL DE CONFIANZA 1 – α	Probabilidad de no rechazar una hipótesis cierta (Probabilidad de no equivocarse al rechazar la HIPOTESIS NULA)
RIESGO DE ERROR TIPO 2 : β	Probabilidad de aceptar una hipótesis falsa (Probabilidad de equivocarse al aceptar la HIPOTESIS NULA)
POTENCIA DEL CONTRASTE 1 – β	Probabilidad de no aceptar una hipótesis falsa (Probabilidad de no equivocarse al aceptar la HIPOTESIS NULA)
NIVEL DE SIGNIFICACION α	Probabilidad de rechazar una hipótesis cierta (probabilidad de equivocarse el rechazar la HIPOTESIS NULA)

CONTINUACIÓN

Como puede notarse cualquiera que sea la decisión adoptada existe un riesgo de error, y como estos son aproximadamente complementarios, no podemos disminuir uno de ellos de forma exagerada ya que aumentaría mucho el otro: Siendo el riesgo de error tipo 1 , el que puede fijarse de antemano, se fija normalmente en 5% o en 1%, y no en valores más bajos por cuanto aumentaría el riesgo de error tipo 2, el cual no puede fijarse de antemano, por depender de la alternativa que sea realmente cierta.

TIPOS DE CONTRASTES SEGÚN EL PLANTEAMIENTO DE HIPÓTESIS

De acuerdo con el planteamiento de la HIPOTESIS ALTERNATIVA (HIPOTESIS DE INVESTIGACION), se pueden presentar dos tipos de contraste:

a.- UNILATERAL (DE UNA COLA)

b.- BILATERAL (DE DOS COLAS) HACIA LA DERECHA O HACIA LA IZQUIERDA

Veamos los valores correspondientes al nivel de significación dependiendo del tipo de contraste

NIVEL DE SIGNIFICACIÓN α 5% = 0,05	CONTRASTE UNILATERAL ± 1,64 EL SIGNO DEPENDE DE CUAL SEA LA ALTERNATIVA > 0 <
	CONTRASTE BILATERAL 1,96
NIVEL DE SIGNIFICACIÓN α 1% = 0,01	CONTRASTE UNILATERAL ± 2,33 EL SIGNO DEPENDE DE CUAL SEA LA ALTERNATIVA > 0 <
	CONTRASTE BILATERAL 2,58

CONTRASTE DE HIPÓTESIS ACERCA DE LA MEDIA ARITMÉTICA MUESTRAS GRANDES

EJERCICIO: En una prueba de estadística presentada por todos los estudiantes de la carrera de educación de la UNEG de Ciudad Bolívar, se obtuvo una calificación media de 6,4 puntos con una desviación típica de 2,4 puntos respectivamente.

Recientemente la misma prueba fue aplicada a una muestra de 75 nuevos estudiantes y se obtuvo una puntuación media de 5,2 puntos. Determine con un 1% de significación si existe diferencia significativa entre la media de puntos obtenida por los nuevos estudiantes con respecto a la del resto de estudiantes de la carrera en dicha universidad.

CONTINUACIÓN

FORMULA A UTILIZAR

$$\text{Zcal} = \frac{\mu - \mu o}{\frac{S}{\sqrt{n}}}$$

SOLUCIÓN DEL EJERCICIO:

PASO 1: SE SELECCIONAN LOS DATOS:

μo = 6,4 Puntos

S = 2,4 Puntos

n = 75 Estudiantes

μ = 5,2 Puntos

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\mu o = \mu$

HI: $\mu o \neq \mu$

PASO 3: SE DETERMINA EL VALOR CRITICO DE ACUERDO AL NIVEL DE SIGNIFICACION Y TIPO DE CONTRASTE

Z Crítico con 1% de significación para un contraste bilateral (DOS COLAS) = (±2,58)

CONTINUACIÓN

PASO 4: SE REALIZAN LOS CÁLCULOS

$$Zcal = \frac{\mu - \mu o}{\frac{S}{\sqrt{n}}}$$

$$Zcal = \frac{5{,}2 - 6{,}4}{\frac{2{,}4}{\sqrt{75}}} = -4{,}33$$

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN

Si el valor de Z calculado en valor absoluto es mayor al valor de Z crítico, se rechaza la hipótesis nula.

PASO 6: SE DECIDE

En este caso como el valor de Z calculado = (- 4,33) , es mayor en valor absoluto al valor de Z crítico (-2,58) , se rechaza la hipótesis nula. Lo cual equivale a aceptar con un 1% de significación, que existe diferencia entre la media de puntuaciones de ambos grupos.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA

Como se puede apreciar el valor calculado (- 4,28) cae en la región de rechazo de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE LA MEDIA ARITMÉTICA MUESTRAS PEQUEÑAS:

EJERCICIO: En un estudio acerca de del rendimiento académico de los todos estudiantes de una institución educativa, se obtuvo una media de 13,0 puntos. En una muestra de 10 estudiantes tomada posteriormente se obtuvo una media de 14,3 puntos con una desviación típica de 3,0 puntos respectivamente. El director del instituto afirma que el rendimiento promedio de todos los estudiantes de la institución es mayor que el resultado obtenido en la muestra estudiada. Determine con un 5% de significación si se puede aceptar la afirmación hecha por el director del instituto educativo.

CONTINUACIÓN

FORMULA A UTILIZAR

$$Tcal = \frac{\mu - \mu o}{S} * \sqrt{n-1}$$

CONTINUACIÓN

PASO 1: SE SELECCIONAN LOS DATOS

μo = 13 Puntos

S = 3,0 Puntos

n = 10 Estudiantes

μ = 14,3 Puntos

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\mu o = \mu$

HI: $\mu o > \mu$

PASO 3: SE DETERMINA EL VALOR CRÍTICO DE ACUERDO AL NIVEL DE SIGNIFICACIÓN Y SE REALIZA EL CÁLCULO:

T Crítico con 5% de significación para un contraste unilateral, con 10-1=9 grados de libertad es = (1,83)

$$\text{Tcal} = \frac{14,3 - 13,0}{3,0} * \sqrt{10 - 1} = (1,3)$$

CONTINUACIÓN

PASO 4: SE ESTABLECE LA REGLA DE DECISIÓN:

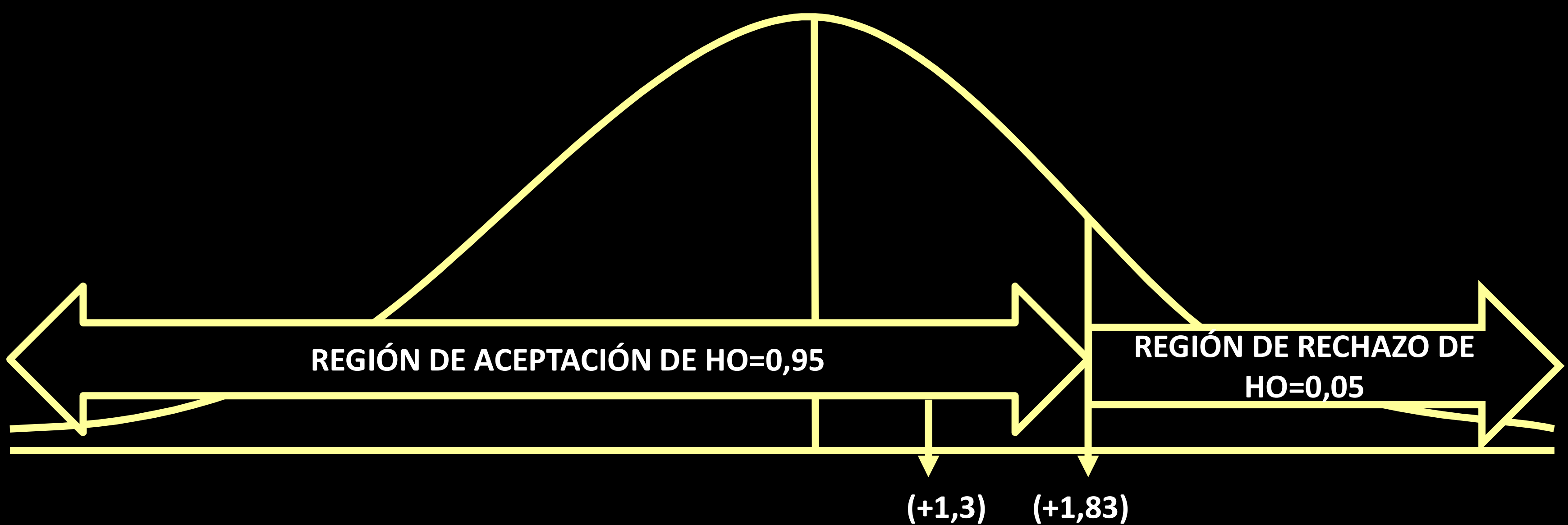
Si el valor de T calculado es mayor al valor de T crítico, se rechaza la hipótesis nula.

PASO 6: SE DECIDE:

En este caso como el valor T calculado = (1,3) , es menor al valor T crítico (1,83) , se acepta la hipótesis nula. Lo cual equivale a rechazar con un 5% de significación la afirmación hecha por el director del instituto que el rendimiento promedio de todos los estudiantes es mayor al obtenido en por los estudiantes de la muestra.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA



Como se puede apreciar el valor calculado (1,3) cae en la región de aceptación de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE UNA PROPORCIÓN O PORCENTAJE:

EJERCICIO: En un estudio de mercado para la utilización de un nuevo libro para el estudio de la estadística. se observó que de un total de 750 estudiantes encuestados, 510 manifestaron su disposición de adquirirlo . ¿La empresa editora afirma que existe una disposición de más del 50% de la población de estudiantes por adquirir el nuevo libro?. Determine si puede aceptarse tal afirmación . Utilice un nivel de significación del 0.01.

CONTINUACIÓN

FORMULA A UTILIZAR

$$Zcal = \frac{P - Po}{\sqrt{\frac{Po * Qo}{n}}}$$

SOLUCIÓN DEL EJERCICIO:

PASO 1: SE SELECCIONAN LOS DATOS

n = 750 Estudiantes

P = 510/750 = 0,68

Po = 50%/100 = 0,5

Qo = 1 – 0,5 = 0,5

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\mu p = 50\%$

HI: $\mu p > 50\%$

PASO 3: SE DETERMINA EL VALOR CRÍTICO DE ACUERDO AL NIVEL DE SIGNIFICACIÓN

Z Crítico con 1% de significación para un contraste unilateral, es = (2,33)

PASO 4: SE REALIZAN LOS CÁLCULOS

$$Zcal = \frac{0{,}68 - 0{,}50}{\sqrt{\frac{0{,}50 * 0{,}50}{750}}} = \frac{0{,}18}{0{,}018} = 10{,}0$$

CONTINUACIÓN

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN:

Si el valor calculado de Z es mayor al valor crítico de Z , se rechaza la hipótesis nula.

PASO 6: SE DECIDE:

En este caso como el valor Z calculado = (10,00) , es mayor al valor Z crítico (2,33) , se rechaza la hipótesis nula. Lo cual equivale a aceptar con un 1% de significación, que existe una disposición de más del 50% de la población de estudiantes por adquirir el nuevo libro.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA

Como se puede apreciar el valor calculado (10,00) cae en la región de rechazo de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE UNA DIFERENCIA DE MEDIAS MUESTRAS GRANDES:

EJERCICIO: La aplicación de una prueba de estadística a dos muestras estudiantes de la UCV y de la UC arrojó los siguientes resultados:

UCV	UC
n1=100 Est	n2 =100 Est
μ1=13 Ptos	μ2=11 Ptos
S1 = 2,6	S1 = 2,1

El Jefe de Control de Estudios de la UCV, afirma que existe una diferencia significativa entre los promedios de calificaciones obtenidas por los estudiantes de ambas universidades. Determine si se puede aceptar tal afirmación con un 5% de significación.

CONTINUACIÓN

FORMULA A UTILIZAR

$$Zcal = \frac{\mu 1 - \mu 2}{\sqrt{\frac{S1^2}{n1} + \frac{S2^2}{n2}}}$$

PASO 1: SE SELECCIONAN LOS DATOS:

n1=100 Est	n2 =100 Est
μ1=13 Ptos	μ2=11 Ptos
S1 = 2,6 Ptos	S2 = 2,1 Ptos
Ns = 5%	

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\mu 1 = \mu 2$

HI: $\mu 1 \neq \mu 2$

PASO 3: SE DETERMINA EL VALOR CRÍTICO DE ACUERDO AL NIVEL DE SIGNIFICACIÓN

Crítico con 5% de significación para un contraste bilateral, es = (±2,58)

PASO 4: SE REALIZAN LOS CÁLCULOS:

$$Zcal = \frac{13 - 11}{\sqrt{\frac{2{,}6^2}{100} + \frac{2{,}1^2}{100}}} = \frac{2}{0{,}33} = (9{,}09)$$

CONTINUACIÓN

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN

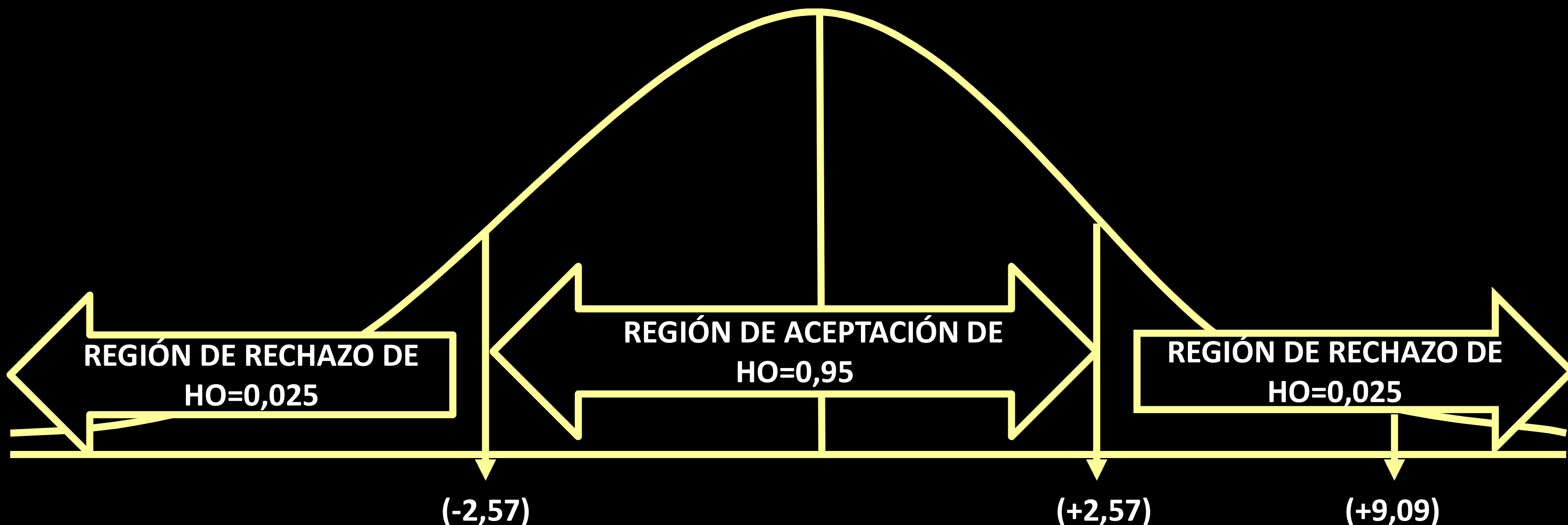
Si el valor calculado de Z es mayor al valor crítico de Z , se rechaza la hipótesis nula.

PASO 6: SE DECIDE

En este caso como el valor Z calculado = (9,09) , es mayor al valor Z crítico (2,58) , se rechaza la hipótesis nula. Lo cual equivale a aceptar con un 5% de significación, la afirmación hecha por el Jefe de Control de Estudios de la UCV, de que existe una diferencia significativa entre los promedios obtenidos por los estudiantes en ambas universidades.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA



Como se puede apreciar el valor calculado (9,09) cae en la región de rechazo de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE UNA DIFERENCIA DE MEDIAS MUESTRAS PEQUEÑAS:

EJEMPLO: La aplicación de una prueba de estadística a dos muestras estudiantes de la UNEG y de la UNEG arrojó los siguientes resultados:

ULA	UNEG
n1=10 Est	n2 =15 Est
μ1=13 Ptos	μ2=11 Ptos
S1 = 1,8 Ptos	S2 = 2,4 Ptos

¿Puede aceptarse con un 1% de significación que la media de los estudiantes de la ULA es mayor a la media de los estudiantes de la UNEG?

CONTINUACIÓN

CONTRASTE DE HIPÓTESIS ACERCA DE UNA DIFERENCIA DE MEDIAS MUESTRAS PEQUEÑAS

$$\mathrm{Tcal} = \frac{\mu 1 - \mu 2}{\sqrt{\frac{(n1 * S1^{2}) + (n2 + S2^{2})}{(n1 + n2) - 2} * \sqrt{\frac{1}{n1}}} + \frac{1}{n2}}$$

PASO 1: SE SELECCIONAN LOS DATOS:

n1=10 Est	n2 =15 Est
μ1=15 Ptos	μ2=10 Ptos
S1 = 3,8	S1 = 2,4
Ns = 1%	

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\mu 1 = \mu 2$

HI: $\mu 1 > \mu 2$

PASO 3: SE DETERMINA EL VALOR CRÍTICO DE ACUERDO AL NIVEL DE SIGNIFICACIÓN

T Crítico con 1% de significación para un contraste unilateral, con 10+15 – 2 GL es = (2,50)

PASO 4: SE REALIZAN LOS CÁLCULOS:

$$Tcal = \frac{15 - 10}{\sqrt{\frac{(10 * 3{,}8^2) + (15 * 2{,}4^2)}{(10 + 15) - 2}} * \sqrt{\frac{1}{15} + \frac{1}{10}}} = \frac{5}{3{,}16 * 0{,}40} = \frac{5}{1{,}26} = (3{,}96)$$

CONTINUACIÓN

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN

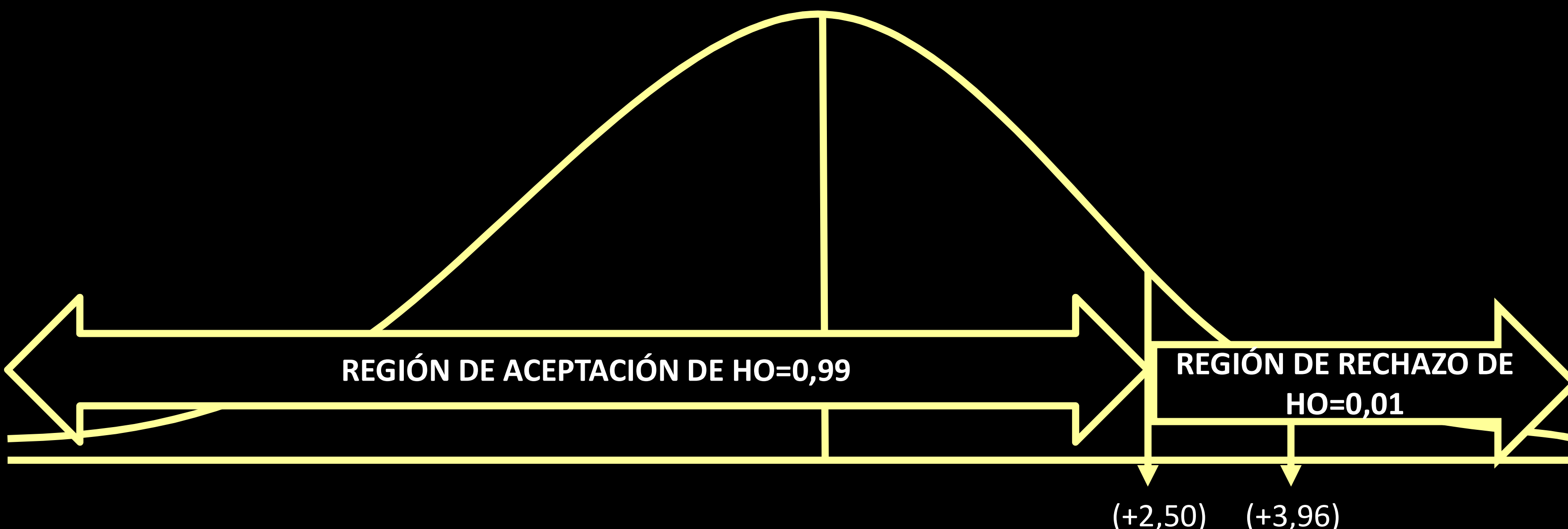
Si el valor calculado de Z es mayor al valor crítico de Z , se rechaza la hipótesis nula.

PASO 6: SE DECIDE

En este caso como el valor T calculado = (3,96) , es mayor al valor Z crítico (2,50) , se rechaza la hipótesis nula. Lo cual equivale a aceptar con un 1% de significación, que la calificación media de los estudiantes de la UCV, es mayor a la calificación media de los estudiantes de la UNEG.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA



Como se puede apreciar el valor calculado (3,96) cae en la región de rechazo de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE UNA DIFERENCIA DE PROPORCIONES O PORCENTAJES:

EJERCICIO: **Para determinar el rendimiento en matemática de los estudiantes de ingeniería de mantenimiento de una universidad, se aplicó una prueba a dos grupos de estudiantes. De la primera muestra conformada por 100 estudiantes regulares 65 salieron reprobados, y de la segunda muestra conformada por 120 estudiantes nuevos 90 salieron reprobados. Determine con un 1% de significación si se puede aceptar la hipótesis de que existe una diferencia significativa entre la proporción de estudiantes reprobados para las muestras consideradas.**

CONTINUACIÓN

CONTRASTE DE HIPOTESIS ACERCA DE UNA DIFERENCIA DE PROPORCIONES O PORCENTAJES

FORMULA A UTILIZAR

$$Zcal = \frac{P1 - P2}{\sqrt{\frac{P1 * Q1}{n1} + \frac{P2 * Q2}{n2}}}$$

PASO 1: SE SELECCIONAN LOS DATOS:

n1=100 Est	n2 =120 Est
P1= 65/100 = 0,65	P2= 90/120 = 0,75
Q1 = 0,35	Q2 = 0,25
Ns = 1%	

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: P1 = P2

HI: P1 ≠ P2

PASO 3: SE DETERMINA EL VALOR CRÍTICO DE ACUERDO AL NIVEL DE SIGNIFICACIÓN:

Z Crítico con 1% de significación para un contraste bilateral, es = (±2,57)

PASO 4: SE REALIZAN LOS CÁLCULOS:

$$Zcal = \frac{0,65 - 0,75}{\sqrt{\frac{0,35*0,65}{100} + \frac{0,25*0,75}{120}}} = \frac{-0,1}{0,06} = (-1,66)$$

CONTINUACIÓN

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN:

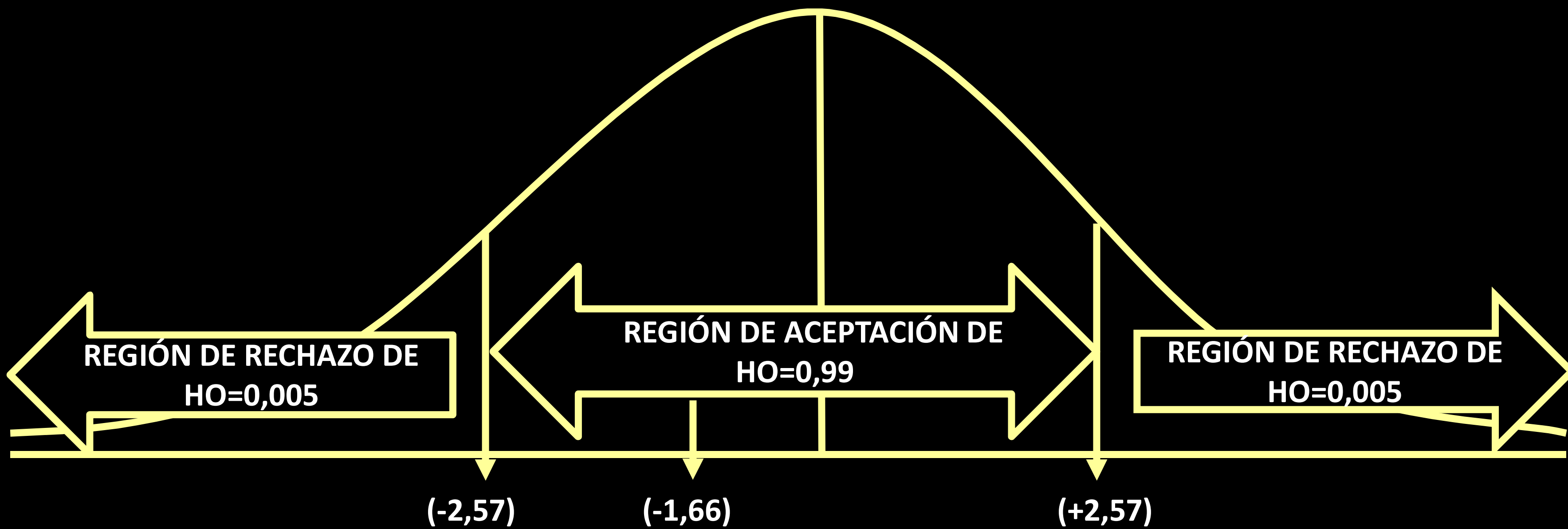
Si el valor calculado de Z es mayor al valor crítico de Z , se rechaza la hipótesis nula.

PASO 6: SE DECIDE:

En este caso como el valor Z calculado = (-1,66) , es menor en valor absoluto al valor Z crítico (-2,57) , se acepta la hipótesis nula. Lo cual equivale a aceptar con un 1% de significación, no existe una diferencia significativa entre la proporción de reprobados de ambas muestras en matemática.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA



Como se puede apreciar el valor calculado (- 1,66) cae en la región de aceptación de la hipótesis nula.

CONTRASTE DE HIPÓTESIS ACERCA DE DESVIACIONES TÍPICAS:

Antes de entrar en los cálculos propiamente dichos es necesario señalar, que la distribución muestral de la desviación típica no es normal, por lo cual estudiamos la estimación de este parámetro sólo para muestras mayores de 100 datos en la cual la distribución es aproximadamente normal.

En este caso se introduce un nuevo estadístico: $\chi^2 = n.S^2 / \sigma^2$. Este estadístico recibe el nombre de χ^2 (CHI CUADRADO) de Pearson (suponiendo que la distribución poblacional es normal), y su distribución depende exclusivamente del tamaño de las muestras consideradas (n) o con más precisión del número de grados de libertad, en que se basa su determinación (v = n-1).

Para realizar un contraste de hipótesis acerca de una desviación típica, se sigue un procedimiento similar al que hemos visto acerca de la media aritmética, sólo que en lugar de utilizar el estadístico Z o T se utilizará en estadístico χ^2

TABLA DSITRIBUCIÓN CHI χ^2 PERCENTILES								
GL	χ^2 0,995	χ^2 0,99	χ^2 0,975	χ^2 0,95	χ^2 0,90	χ^2 0,75	χ^2 0,50	χ^2 0,15
1	7,88	6,63	5,02	3,84	2,71	1,32	0,46	0,102
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6
20	40,0	37,6	34,2	31,4	28,4	28,8	19,3	15,5

CONTINUACIÓN

TABLA DSITRIBUCIÓN CHI χ^2 PERCENTILES								
GL	χ^2 0,995	χ^2 0,99	χ^2 0,975	χ^2 0,95	χ^2 0,90	χ^2 0,75	χ^2 0,50	χ^2 0,15
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8
27	49,6	47,0	43,2	40,1	36,7	31,5	16,3	21,7
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3

CONTRASTE DE HIPÓTESIS ACERCA DE LA DESVIACIÓN TÍPICA:

EJERCICIO: Se desea contrastar la hipótesis de que la desviación típica poblacional $\sigma = 15$ contra la alternativa $\sigma > 15$ con un nivel de significación de 1%. Si en una muestra al azar de $n = 150$ estudiantes se encontró una desviación típica de 6,8 (se asume que la distribución poblacional es normal)

CONTINUACIÓN

FORMULA A UTILIZAR

$$\chi^2 = \frac{n * S^2}{\sigma^2}$$

SOLUCIÓN DEL EJERCICIO

PASO 1: SE SELECCIONAN LOS DATOS

$n = 150$

$S = 6{,}8$

$\sigma = 15$

PASO 2: SE PLANTEAN LAS HIPÓTESIS

H0: $\sigma = 15$

HI: $\sigma > 15$

PASO 3: SE DETERMINA EL VALOR CRITICO DE ACUERDO AL NIVEL DE SIGNIFICACION:

χ^2 Crítico con α = 1% de significación para un contraste unilateral, y con 25 – 1 GL es = (43,0)

PASO 4: SE REALIZAN LOS CÁLCULOS

$$\chi^2 = \frac{n * S^2}{\sigma^2}$$

$$\chi^2 = \frac{150 * 6{,}8^2}{15^2} = \frac{1156}{225} = 30{,}82$$

CONTINUACIÓN

PASO 5: SE ESTABLECE LA REGLA DE DECISIÓN

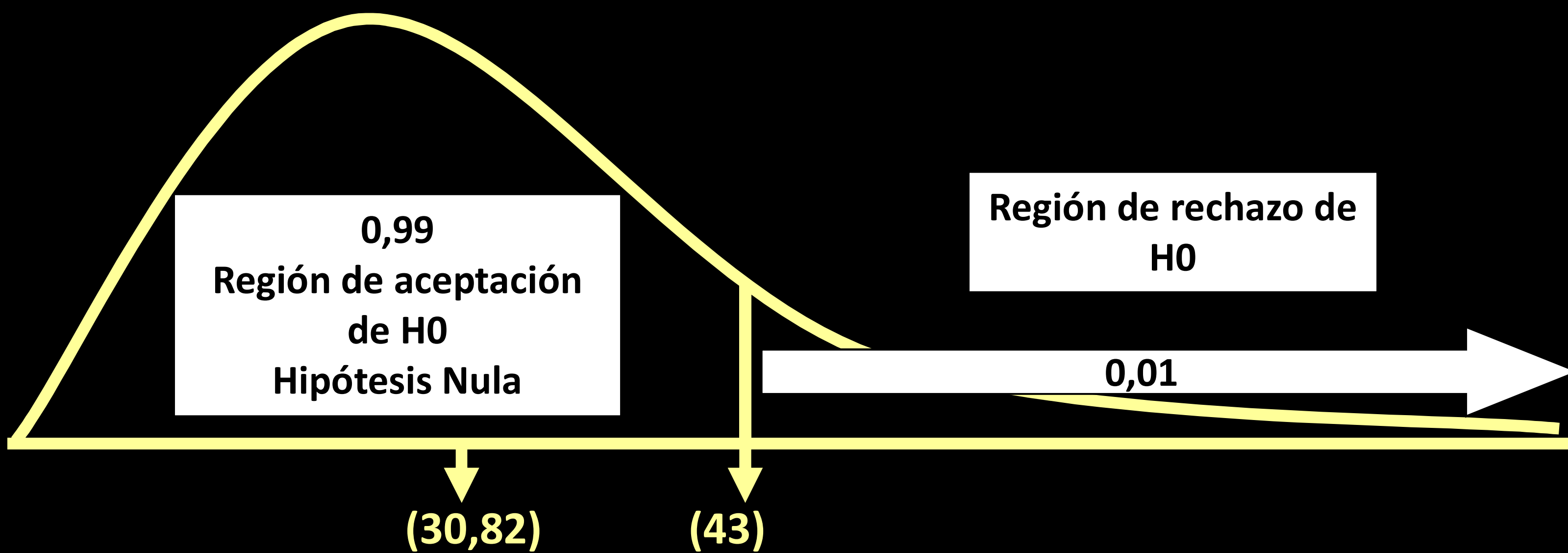
Si el valor de χ^2 calculado es mayor al valor de χ^2 crítico , se rechaza la hipótesis nula.

PASO 6: SE DECIDE

En este caso como el valor calculado χ^2 calculado = (30,82) , es menor al valor χ^2 crítico (43) , se acepta la hipótesis nula, es decir que la muestra puede considerarse proveniente de una población normal en la cual la σ = 15.

CONTINUACIÓN

REPRESENTACIÓN GRÁFICA



Como se puede apreciar el valor calculado (30,82) cae en la región de aceptación de la hipótesis nula.