Bienvenidos a su segundo cuatrimestre estimados Psicólogos en proceso de formación.

Materia: Estadística descriptiva

Orientador: Rosario Gómez lujano

PASIÓN POR EDUCAR

UNIDAD III. PORCENTAJES ACUMULADOS

Investigar y realizar un cuadro sinóptico de los siguientes temas: **porcentajes acumulados, puntuaciones típicas, medidas de tendencia central y representación gráfica.**

CUADRO SINOPTICO: Los cuadros sinópticos son representaciones gráficas de la información y de sus relaciones. Con ellos puedes realizar la clasificación y síntesis de datos. Los cuadros sinópticos establecen una relación entre dos conjuntos de datos, del lado izquierdo de la forma llamada “llave”, se ponen datos generales, del lado derecho datos particulares o específicos, englobados o abarcados por los primeros.

Para realizarlo debes considerar lo siguiente:

Realizar lectura e interpretación del material para el buen uso de los conceptos.

Buen uso de sistema de llaves, filas y columnas que de una fácil lectura y comprensión.

Jerarquía según la importancia de los conceptos.

Con la distribución de frecuencias que realizaste en la primera tarea construye una columna de porcentajes así como también encuentra las medidas de tendencia central para velocidades.

Investigar y realizar un ensayo de 3 cuartillas de **estadísticos de asociación entre variables.**

PASIÓN POR EDUCAR

Responde las siguientes preguntas
¿Cuántos alumnos obtuvieron una calificación mayor que 6?
¿ Que porcentaje de alumnos aprobaron la materia?
¿Qué porcentaje de alumnos obtuvieron 5?

CALIFICACION	FRECUENCIA		
4	3		
5	6		
6	10		
7	5		
8	2		
9	4		
10	5		

Distribuciones de frecuencias para las velocidades
90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118, 100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

Intervalo de clase	f	fa	fr	fra	mc	%	% acumulado
[90-95)	2	2	.07	.07	92.5	7	7
[95-100)	8	10	.27	0.34	97.5	27	34
[100-105)	5	15	0.17	0.51	102.5	17	51
[105-110)	4	19	0.13	0.64	107.5	13	64
[110-115)	6	25	0.2	0.84	112.5	20	84
[115-120)	5	30	0.16	1	117.5	16	16
Total	30		1			100	

Media para Datos No agrupados y para Datos Agrupados

$$\bar{x} = \frac{\sum fixi}{n} = \frac{\sum fimc}{n}$$

Moda para Datos Agrupados

$$\hat{x} = Li + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) i$$

Mediana para Datos Agrupados

$$\tilde{x} = Li + \left(\frac{\frac{n}{2} - FA}{fi}\right) i$$

Donde:
Li = Límite inferior de la mediana
Σ*n* = Suma total de frecuencias absolutas
f anteriores a Σ = Suma de todas las frecuencias absolutas que anteceden a la mediana

f mediana = Frecuencia de la mediana
i= Amplitud del intervalo de clase

moda
Límite inferior modal
d_a= Diferencia anterior
d_p= Diferencia posterior

Introducción y cálculo

Las puntuaciones típicas son un procedimiento alternativo para expresar la posición de las puntuaciones directas en relación al grupo, y se definen:

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

En el numerador se obtiene la diferencia de la puntuación X con la media del grupo de datos. Esta diferencia mide la distancia al punto central de la distribución. El denominador tiene la función de normalizar el resultado respecto de la variación de la distribución.

Consideremos los siguientes datos, que representan las puntuaciones obtenidas por un grupo de estudiantes en dos exámenes:

XA	XB
6	9
5	8
4	7

Observamos que un estudiante ha obtenido un 6 en la prueba A y un 9 en la prueba B. Si comparamos estos dos resultados parece mejor el resultado en el examen B que en el examen A. Ahora bien, si los expresamos como la diferencia entre la puntuación y la Media de cada examen, (en símbolos: $x=X-M_X$) tenemos que:

XA	XB
1	1
0	0
-1	-1

Es decir, las puntuaciones diferenciales muestran que el estudiante ha obtenido el mismo resultado en las dos pruebas (en relación al rendimiento promedio en cada examen). Por tanto, si interesa expresar la posición de las puntuaciones respecto del grupo, un procedimiento alternativo a los Rangos Percentiles consiste en obtener las puntuaciones diferenciales.

Sin embargo veamos los siguientes datos:

XA	XB		XA	XB
6	10		1	4
5	6		0	0
4	2		-1	-4

Las puntuaciones diferenciales del primer estudiante son diferentes. En cambio las posiciones relativas son las mismas (tanto el 6 como el 10 son las puntuaciones más grandes en cada grupo). La conclusión es que el procedimiento que consiste en obtener puntuaciones diferenciales no mide la posición relativa correctamente. Al comparar con el primer ejemplo vemos que lo único que ha cambiado es la variación entre los datos del grupo B. En consecuencia hay que corregir las puntuaciones diferenciales respeto de la variabilidad, y lo haremos dividiendo por la Desviación Típica:

Supongamos que un determinado país la estatura de la población adulta sigue una distribución N(170,12)

¿Qué porcentaje de esa población mide menos de 185 cm?

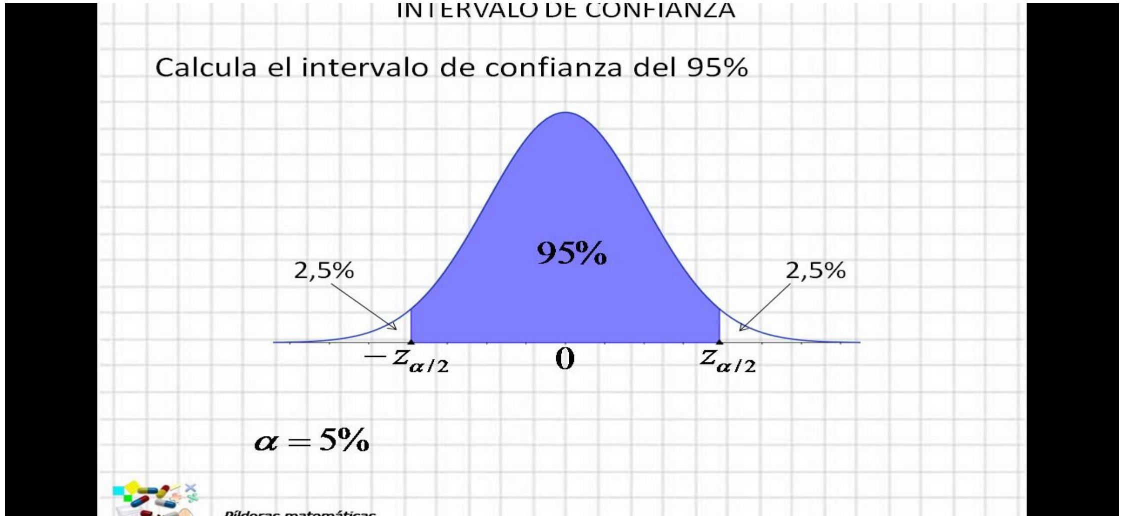
$$Z=\frac{X-M}{S}=$$

Supongamos que un determinado lugar el peso de la población adulta sigue una distribución N(65,5)
¿Qué porcentaje de esa población pesa menos de 80 kg?

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

Dada una distribución normal N(0,1) ¿Qué valor deja por debajo de si al 97.5% de la población?

UNIDAD IV. Relaciones entre variables


INTERVALO DE CONFIANZA
Calcula el intervalo de confianza del 95%
95%
2,5%
2,5%
$-z_{\alpha/2}$
0
$z_{\alpha/2}$
$\alpha = 5\%$

Correlación

En este artículo trataremos de valorar la asociación entre dos variables cuantitativas estudiando el método conocido como correlación. Dicho cálculo es el primer paso para determinar la relación entre las variables. La predicción de una variable. La predicción de una variable dado un valor determinado de la otra precisa de la regresión lineal que abordaremos en otro artículo.

La cuantificación de la fuerza de la relación lineal entre dos variables cuantitativas, se estudia por medio del cálculo del coeficiente de correlación de Pearson. Dicho coeficiente oscila entre –1 y +1. Un valor de –1 indica una relación lineal o línea recta positiva perfecta. Una correlación próxima a cero indica que no hay relación lineal entre las dos variables.

PASIÓN POR EDUCAR

El coeficiente de correlación posee las siguientes características:

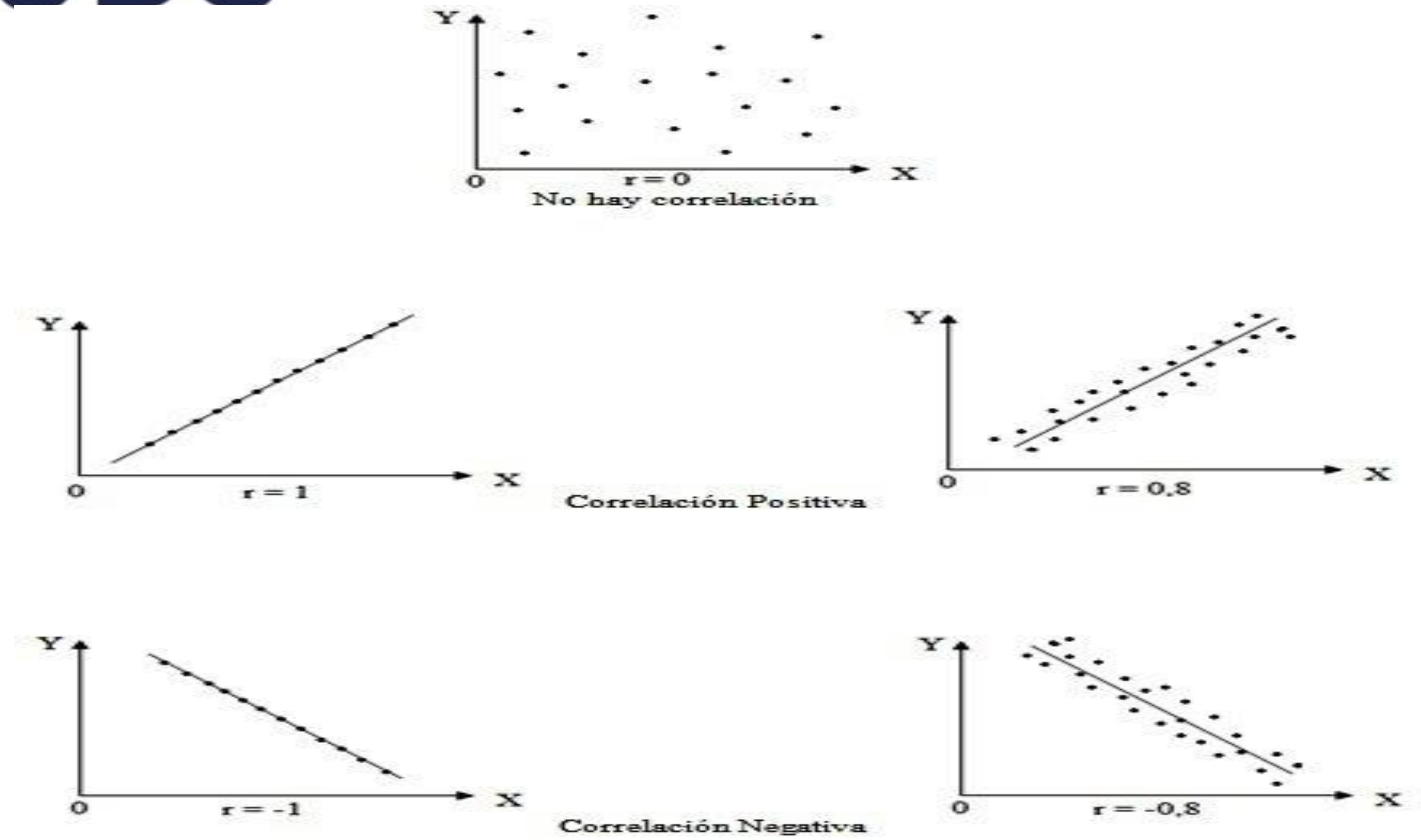
a. El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables.

b. El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica. Ante estas situaciones conviene realizar una transformación de datos que cambia la escala de medición y modera el efecto de valores extremos (como la transformación logarítmica).

c. El coeficiente de correlación mide solo la relación con una línea recta. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña. Por tanto cuando analicemos las relaciones entre dos variables debemos representarlas gráficamente y posteriormente calcular el coeficiente de correlación.

d. El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente entre X e Y puede cambiar fuera de dicho rango.

e. La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.


Y
X
0
r = 0
No hay correlación
Y
X
0
r = 1
Correlación Positiva
Y
X
0
r = 0,8
Y
X
0
r = -1
Correlación Negativa
Y
X
0
r = -0,8

Como vimos el test de **correlación de Pearson mide la relación existente entre dos variables**, su intensidad y su sentido (positivo o negativo).
A continuación veremos un **ejemplo de su aplicación**:

El Equipo Directivo de una empresa está interesado en conocer la relación que existe entre el tiempo semanal (horas) que dedican los trabajadores a formación y la productividad media de los mismos al final del año. Eligiendo 11 trabajadores al azar, han encontrado los siguientes resultados tras calcular el coeficiente de correlación de Pearson.

Media

$$X = \frac{\sum X}{n} = \frac{60}{11} = 5.46 \qquad Y = \frac{\sum Y}{n} = \frac{72}{11} = 6.55$$

PASIÓN POR EDUCAR

Como se puede ver en la tabla inferior se muestran los resultados de x e y que son puntuaciones diferenciales que se han conseguido restándoles las medias a las puntuaciones directas. El valor de la media en cada una de las variables las mostraré a continuación junto con las operaciones para calcular la covarianza de cada variable.

1)X	Y	X · Y	x	y	x · y
5	7	35	-0.46	0.45	-0.21
2	4	8	-3.46	-2.55	8.82
4	5	20	-1.46	-1.55	2.26
7	8	56	1.54	1.45	2.23
10	9	90	4.54	2.45	11.12
4	6	24	-1.46	-0.55	0.80
8	9	72	2.54	2.45	6.22
1	4	4	-4.46	-2.55	11.37
3	5	15	-2.46	-1.55	3.81
6	7	42	0.54	0.45	0.24
10	8	80	4.54	1.45	6.58
60	72	446			53.24

X^2	Y^2	x^2	y^2

Covarianza:

$$S_{xy} = \frac{\sum X_i \cdot Y_i}{n} - \bar{X} \cdot \bar{Y} = \frac{446}{11} - 5.45 \cdot 6.55 = 4.79$$

El resultado de la covarianza al ser positivo, nos indica cierta tendencia a que a un tiempo semanal de estudio por encima de la media corresponden calificaciones por encima de la media, y a un tiempo de estudio por debajo de la media corresponden calificaciones por debajo de la media.
Para calcular el coeficiente de correlación de Pearson nos ayudarán los resultados del ejercicio anterior. Pero además tendremos que calcular la desviación típica de x e y. Y para la desviación típica necesitaremos los resultados de la varianza de x e y respectivamente. Mostraré a continuación el procedimiento y los resultados del proceso necesario para obtener lo solicitado por el problema.

PASIÓN POR EDUCAR

$$S^2x = \frac{\sum x'^2 \cdot fi}{n} - X^2 = \frac{420}{11} - 29.7 = 8.48$$

$$S^2y = \frac{\sum x'^2 \cdot fi}{n} - X^2 = \frac{506}{11} - 42.9 = 3.1$$

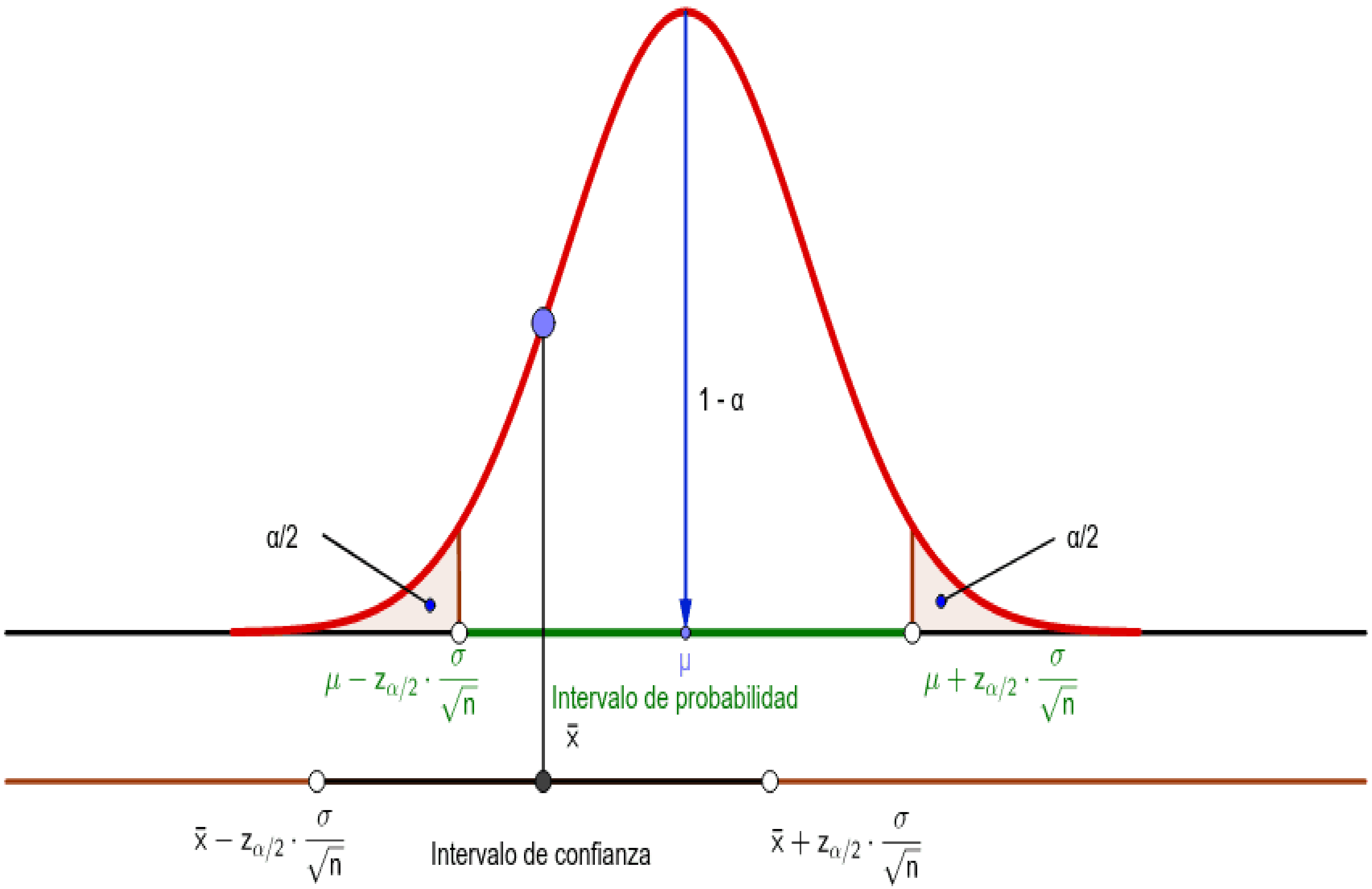
$$S\,x = \sqrt{8.48} = 2.91$$

$$S\,y = \sqrt{3.1} = 1.76$$

<u>Coeficiente de correlación de Pearson</u> $\rightarrow r\,xy = \dfrac{S\,xy}{Sx \cdot Sy}$

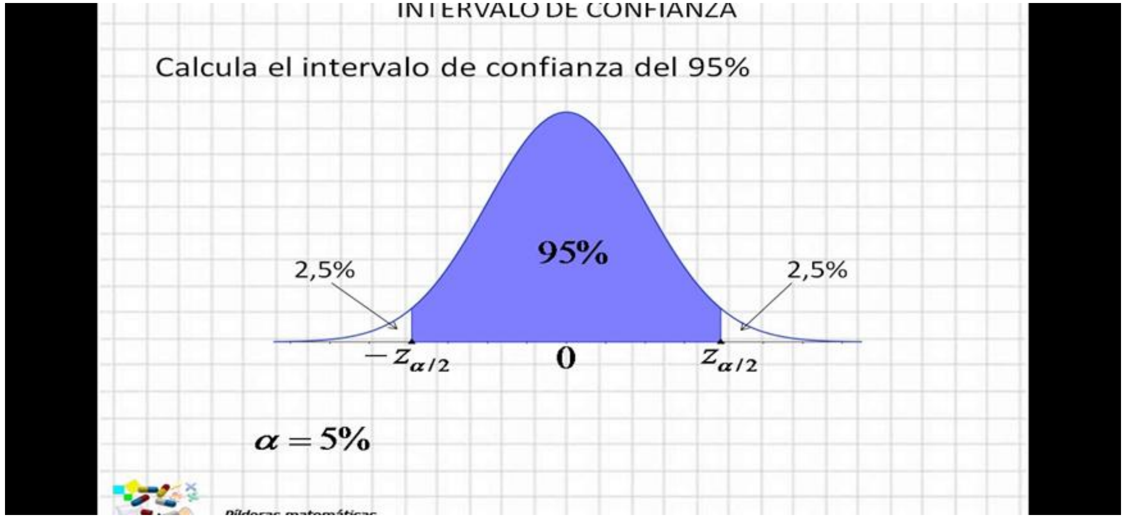
$$r\,xy = \frac{4.79}{2.91 \cdot 1.76} = 0.94$$

Viendo los resultados llegamos a la conclusión de que estamos ante una correlación muy alta, lo que quiere decir a puntuaciones altas en cuanto a notas se corresponden altas horas de trabajo y estudio semanal.


1 - α
α/2
α/2
μ
$\mu - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
Intervalo de probabilidad
$\mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
$\bar{x}$
$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
Intervalo de confianza
$\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

1.- Determina el intervalo de confianza de una muestra de 20, sabiendo que la desviación estándar de la población es 0.25 usando un nivel de confianza de 95% para una media de 80

DATOS
MEDIA=80
n=20
Varianza(σ)=0.25
Nivel de confianza
95%=0.95



$$\bar{\bar{x}} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

80+1.96