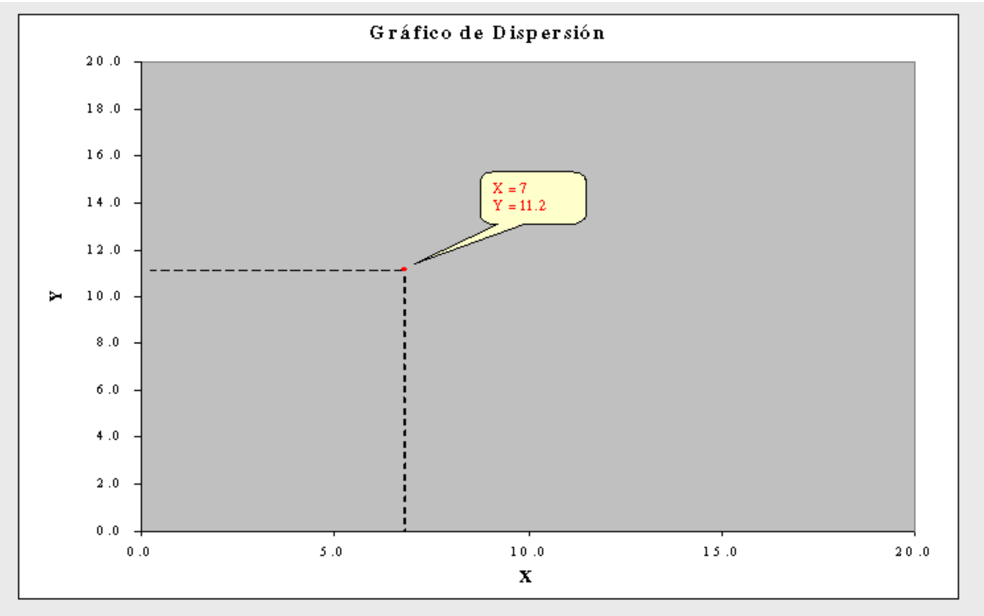
5.1. DIAGRAMAS DE DISPERSIÓN

Diagramas de Dispersión

Los Diagramas de Dispersión o Gráficos de Correlación permiten estudiar la relación entre 2 variables. Dadas 2 variables X e Y, se dice que existe una correlación entre ambas si cada vez que aumenta el valor de X aumenta proporcionalmente el valor de Y (Correlación positiva) o si cada vez que aumenta el valor de X disminuye en igual proporción el valor de Y (Correlación negativa).

En un gráfico de correlación representamos cada par X, Y como un punto donde se cortan las coordenadas de X e Y:



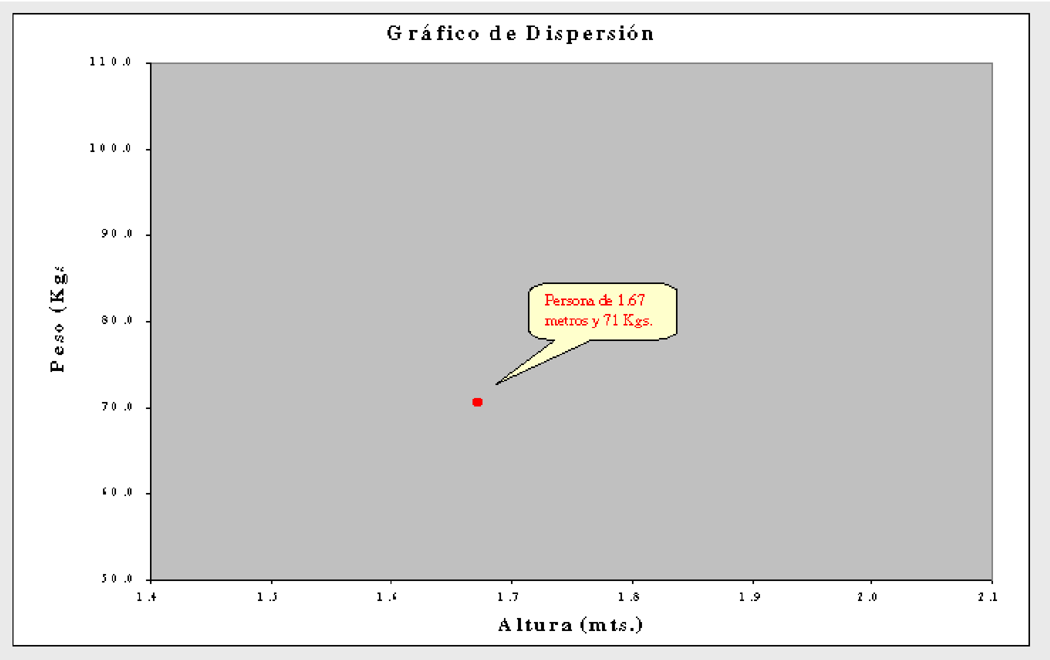
Veamos un ejemplo. Supongamos que tenemos un grupo de personas adultas de sexo masculino. Para cada persona se mide la altura en metros (Variable X) y el peso en kilogramos (Variable Y). Es decir, para cada persona tendremos un par de valores X, Y que son la altura y el peso de dicha persona:

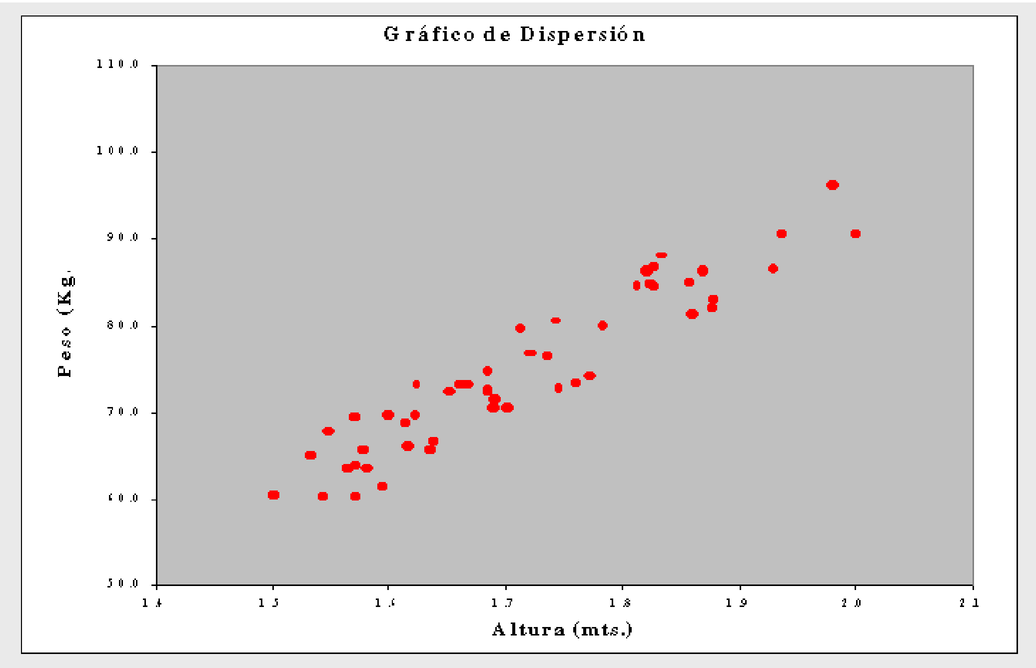
Nº Persona	***Altura (m)***	***Peso (Kg.)***	***Nº Persona***	***Altura (m)***	***Peso (Kg.)***
001	1.94	95.8	026	1.66	74.9
002	1.82	80.5	027	1.96	88.1
003	1.79	78.2	028	1.56	65.3
004	1.69	77.4	029	1.55	64.5

005	1.80	82.6	030	1.71	75.5
006	1.88	87.8	031	1.90	91.3
007	1.57	67.6	032	1.65	66.6
008	1.81	82.5	033	1.78	76.8
009	1.76	82.5	034	1.83	80.2
010	1.63	65.8	035	1.98	97.6
011	1.59	67.3	036	1.67	76.0
012	1.84	88.8	037	1.53	58.0
013	1.92	93.7	038	1.96	95.2
014	1.84	82.9	039	1.66	74.5
015	1.88	88.4	040	1.62	71.8
016	1.62	69.0	041	1.89	91.0
017	1.86	83.4	042	1.53	62.1
018	1.91	89.1	043	1.59	69.8
019	1.99	95.2	044	1.55	64.6
020	1.76	79.1	045	1.97	90.0
021	1.55	61.6	046	1.51	63.8
022	1.71	70.6	047	1.59	62.6
023	1.75	79.4	048	1.60	67.8
024	1.76	78.1	049	1.57	63.3
025	2.00	90.6	050	1.61	65.2

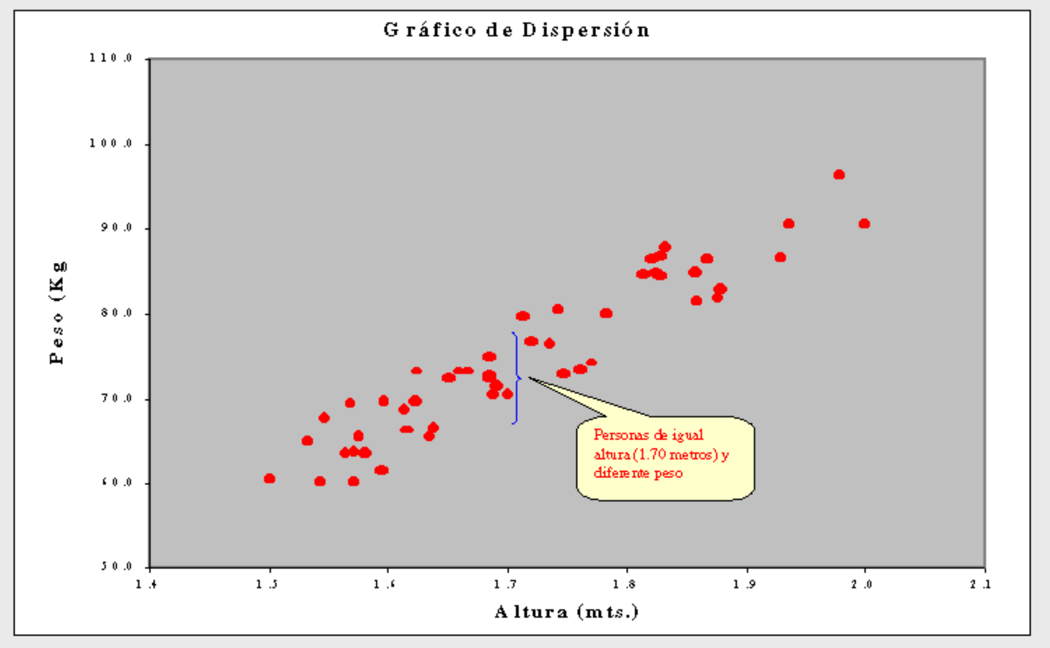
Entonces, para cada persona representamos su altura y su peso con un punto en un gráfico:

Una vez que representamos a las 50 personas quedará un gráfico como el siguiente:



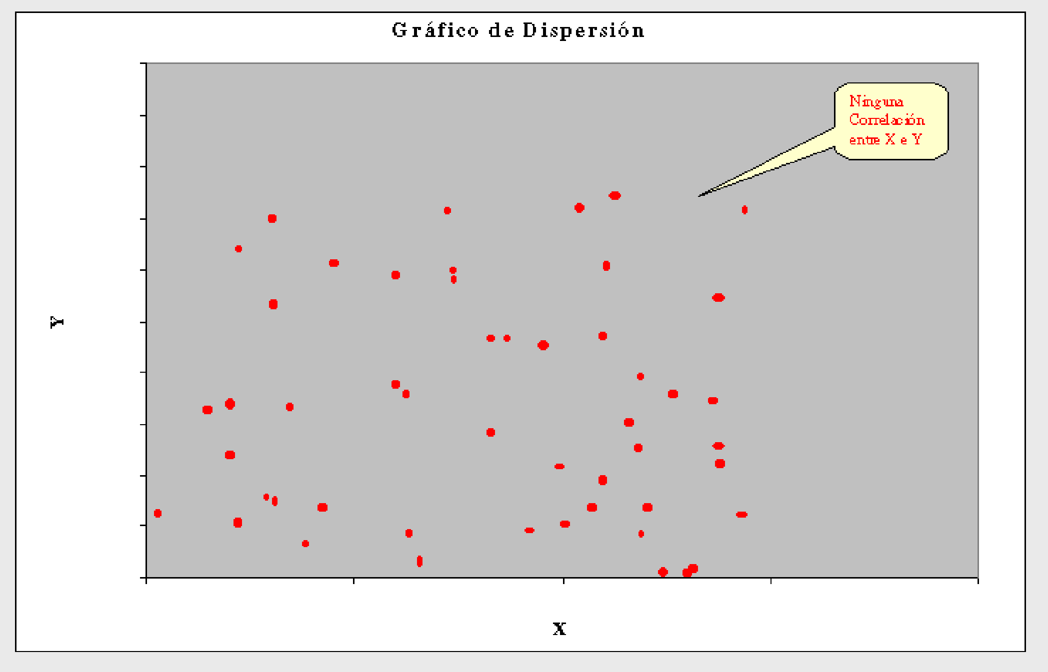


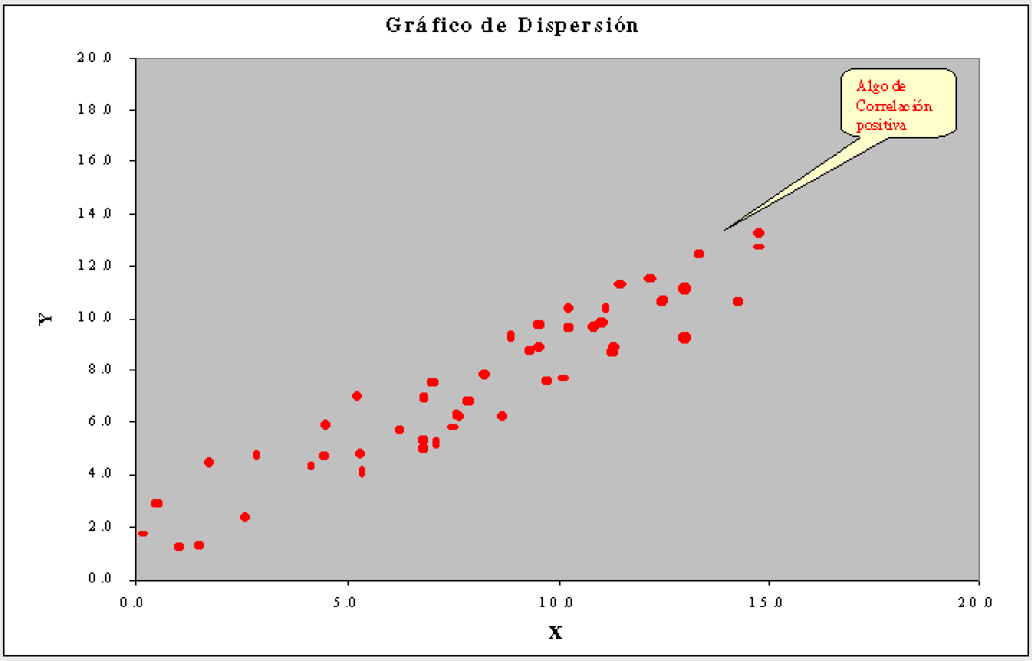
Qué nos muestra este gráfico? En primer lugar podemos observar que las personas de mayor altura tienen mayor peso, es decir parece haber una correlación positiva entre altura y peso. Pero un hombre bajito y gordo puede pesar más que otro alto y flaco. Esto es así porque no hay una correlación total y absoluta entre las variables altura y peso. Para cada altura hay personas de distinto peso:

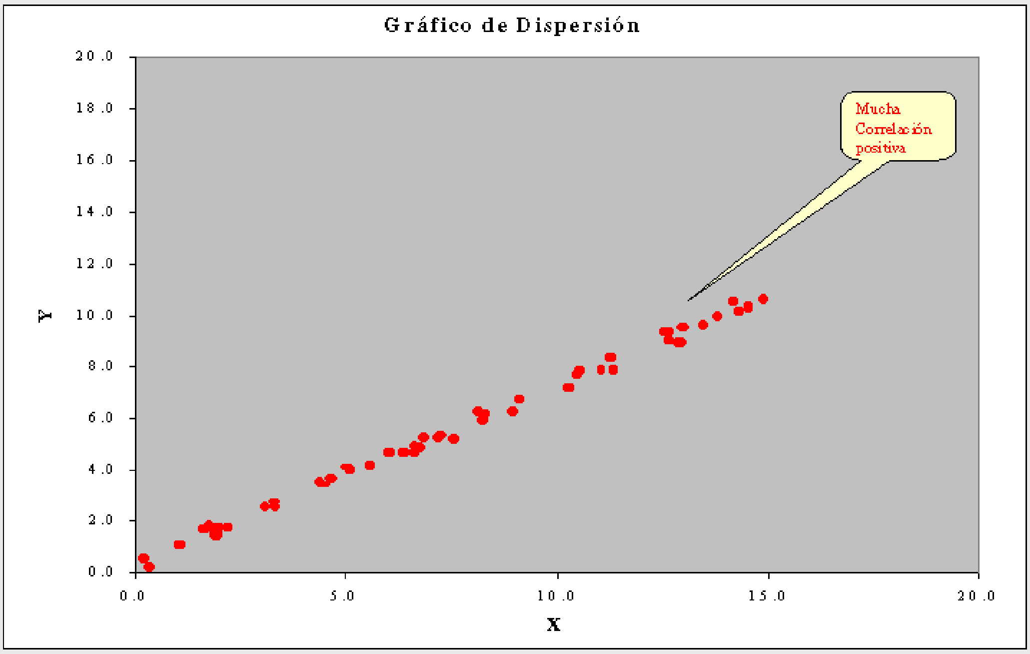


Sin embargo podemos afirmar que existe cierto grado de correlación entre la altura y el peso de las personas.

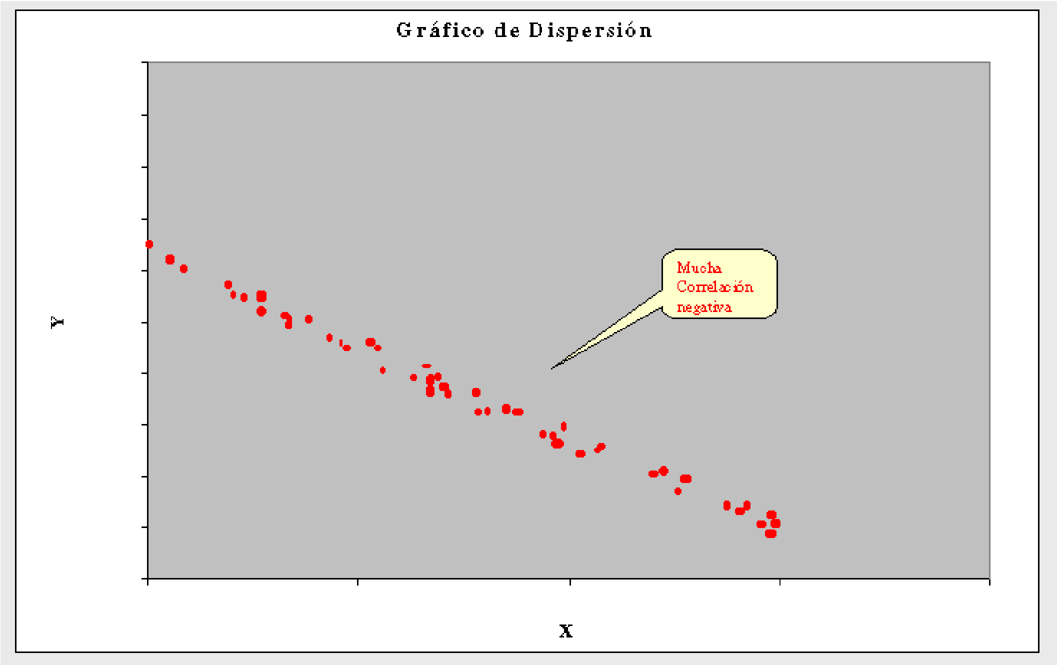
Cuando se trata de dos variables cualesquiera, puede no haber ninguna correlación o puede existir alguna correlación en mayor o menor grado, como podemos ver en los gráficos siguientes:







Por ejemplo, en el siguiente gráfico podemos ver la relación entre el contenido de Humedad de hilos de algodón y su estiramiento:



5.2. LA ASOCIACIÓN ENTRE VARIABLES EXPRESADAS EN ESCALAS DE INTERVALOS O RAZÓN.

El nivel de intervalo procede del latín interval lun (espacio entre dos paredes). Este nivel integra las variables que pueden establecer intervalos iguales entre sus valores. Las variables del nivel de intervalos permiten determinar la diferencia entre puntos a lo largo del mismo continuo. Las operaciones posibles son todas las de escalas anteriores, más la suma y la resta.

En este tipo de medida, los números asignados a los objetos tienen todas las características de las medidas ordinales, y además las diferencias entre medidas representan intervalos equivalentes. Esto es, las diferencias entre una par arbitrario de medidas puede compararse de manera significativa. Por lo tanto, operaciones tales como la adición, la sustracción tienen significado. El punto cero de la escala es arbitrario y se pueden usar valores negativos. Las diferencias se pueden expresar como razones. Las medidas de tendencia central pueden representarse mediante la moda, la mediana al promedio aritmético. EL promedio proporciona más información.

Las variables medidas al nivel de intervalo se llaman variables de intervalo o variables de escala.

Ejemplos de este tipo de variables son la fecha, temperatura.

Medida racional

El nivel de razón, cuya denominación procede del latín ratio (cálculo), integra aquellas variables con intervalos iguales pueden situar un cero absoluto. El cero absoluto supone identificar una posición de ausencia total del rasgo o fenómeno. Tiene varias características importantes: El valor cero no es arbitrario (no responde a las conveniencias de los investigadores). Un ejemplo claro es la temperatura. La existencia de un cero en la escala Celsius no supone la ausencia de temperatura, puesto que el cero grados centígrados está situado por arbitrio de los creadores de la escala. Por el contrario, la escala Kelvin sí tiene un cero absoluto, precisamente allí donde las moléculas cesan su actividad y no se produce por lo tanto roce entre los componentes moleculares. El cero absoluto de la escala Kelvin se sitúa a unos -273 grados centígrados. - La presencia de un cero absoluto permite utilizar operaciones matemáticas más complejas a las otras escalas. Hasta ahora se podía asignar, establecer la igualdad (nominal), mayor o menor que (ordinal), sumar y restar (intervalo) a las que se añade multiplicar, dividir, etc.

Los números asignados a los objetos tienen todas las características de las medidas de intervalo y además tienen razones significativas entre pares arbitrarios de números. Operaciones tales como la multiplicación y la división tienen significado.

La posición del cero no es arbitraria para este tipo de medida. Las variables para este nivel de medida se llaman variables racionales. La mayoría de las cantidades físicas, tales como la mas, longitud, energía, se miden en la escala racional, así como también la temperatura (en kelvins) relativa al cero absoluto. Las medidas de tendencia central de una variable medida a nivel racional pueden representarse por la moda, la mediana, el promedio aritmético o su promedio geométrico. Lo mismo que con la escala de intervalos, el promedio aritmético proporciona la mayor información.

Otros ejemplos de variables racionales son la edad, y otras medidas de tiempo.

5.2.1. EL COEFICIENTE DE R DE BRAVAIS- PEARSON.

COEFICIENTE DE PEARSON

En las distribuciones simétricas, la media , la mediana y la moda coinciden y conforme la distribución se separa de la simetría estos valores se separan, por lo que la más corriente de las medidas de asimetría es la diferencia entre la moda y la media que se la más sensible a los valores extremos

Sk = (X -Mo) / S

Para cuando la moda no se encuentra bien definida se puede sustituir por la mediana

Sk= 3 (X -Me) / S

Estas medidas se conocen como el primero y segundo coeficiente de Pearson y varían entre el intervalo + 3, es cero para la distribución normal.

MEDIDA CUARTIL DE ASIMETRIA

En una distribución simétrica los cuartiles quedan simétricamente colocados respecto a la mediana, pero si es asimétrica un cuartil se separa más que otro. La medida cuartil de asimetría marca esta relación

Sk =[(Q3-Me) -(Me-Q1)]/ (Q3-Q1)

Si la asimetría es a la derecha Q3 está más lejos de la mediana que Q1, si la asimetría es a la izquierda Q1 está mas alejada de la mediana que Q3.Esta medida varía siempre entre + 1, si es cero la distribuciones normal.

COEFICIENTE DE SESGO PERCENTÍLICO

Se aplica con el mismo criterio de la medida Cuartil de Asimetría

Sk = [(P90-P50) -(P50-P10)]/ (P90-P10)

MEDIDAS DE CURTOSIS

Al comparar cuán aguda es una distribución en relación con la Distribución Normal, se pueden presentar diferentes grados de apuntalamiento.

1. Mesocúrtica, Normal

2. PlarticúrtiCa, Menor apuntalamiento

3. Leptocúrtica, Mayor apuntalamiento

COEFICIENTE DE CURTOSIS PERCENTILICO

Una medida del apuntalamiento o curtosis de la distribución está basada en los cuartiles y percentiles, y está dada por el coeficiente de Curtosis Percentílico

K= (0.5 (Q3- Q1)) / (P90-P10)

Para la distribución normal K toma un valor de 0.263 y las distribuciones se definen como:

Leptocúrtica si k es mayor que 0.263

Platicúrtica si k es menor que 0.263

5.3. LA ASOCIACIÓN ENTRE VARIABLES EXPRESADAS EN UNA ESCALA ORDINAL.

Las clases en las escalas ordinales no solo se diferencian unas de otras (característica que define a las escalas nominales) sino que mantiene una especie de relación entre sí. También permite asignar un lugar específico a cada objeto de un mismo conjunto, de acuerdo con la intensidad, fuerza, etc.; presentes en el momento de la medición. Una característica importante de la escala ordinal es el hecho de que, aunque hay orden entre las categorías, la diferencia entre dos categorías adyacentes no es la misma en toda la extensión de la escala. Algunas escalas consisten en calificaciones de múltiples factores que se agregan después para llegar a un índice general.

Debe mencionarse brevemente una clase espacial de escala ordinal llamada "escala de posición", donde las observaciones se clasifican de mayor a menor (o viceversa). Al igual que en las escalas nominales, se emplean a menudo porcentajes y proporciones en escalas ordinales.

5.3. 1. EL COEFICIENTE RHO DE SPEARMAN Y TAU DE KENDALL

TAU- B DE KENDALL

Este procedimiento estadístico para medir la correlación o asociación es complementario del coeficiente de correlación parcial de Kendall; a su vez, es una segunda opción de la correlación de Spearman.

La razón por la que se expone este modelo estadístico se debe a la necesidad de comprender la mecánica aritmética y la interpretación de la prueba, pues se requiere conocerla para realizar el coeficiente parcial de Kendall.

La fórmula es la siguiente:

$$\tau \text{ (tau)} = \frac{S}{1/2\ N\ (N-1)}$$

Donde:
t (tau) = coeficiente de correlación de Kendall.
S = puntuación efectiva de los rangos.
N = tamaño de la muestra en parejas de variables.

Fórmula para determinar el nivel de significancia mediante el valor Z:

$$Z = \frac{\tau}{\sqrt{\dfrac{2\ (2N+5)}{9N\ (N-1)}}}$$

Donde:
Z = valor Z de la distribución normal.
t = coeficiente de correlación de Kendall.
N = tamaño de la muestra.

Pasos:

1. Alinear las observaciones del rango menor al mayor de la variable independiente (X), de manera que se deje el rango que corresponde a la pareja de la variable dependiente (Y).
2. Obtener la puntuación efectiva (S) en la variable dependiente, en función del orden de ocurrencia de los rangos de Y con respecto a X.
3. Contar el número de parejas y aplicar la fórmula.
4. Calcular el nivel de significancia en función del valor Z, de acuerdo con la ecuación, presentada anteriormente.
5. Una vez calculado el valor Z, se obtiene la probabilidad de su magnitud en la tabla de coeficientes de correlación en niveles de p 0.05 y 0.01.
6. Decidir si se acepta o rechaza la hipótesis.

Ejemplo:

Un investigador está interesado en saber si el desarrollo mental de un niño se asocia a la educación formal de la madre. De esta manera, obtiene la calificación de desarrollo mental en la escala de Gesell de ocho niños elegidos aleatoriamente y se informa del grado de escolaridad de las madres.

Elección de la prueba estadística.
Se desea medir asociación o correlación. Las calificaciones de la educación formal de cada madre están dadas en una medición cualitativa, pero tienen una escala ordinal, por lo cual es posible ordenarlas en rangos. *Véase:* *Flujograma 6*

Planteamiento de la hipótesis.

- Hipótesis alterna (Ha). El desarrollo mental de los hijos es una variable dependiente de la educación formal de la madre; por lo tanto, existe una correlación significativa.
- Hipótesis nula (Ho). La asociación entre las variables educación formal de la madre y desarrollo mental de los hijos no es significativa, ni hay correlación.

Nivel de significación.
Para todo valor de probabilidad igual o menor que 0.05, se acepta Ha y se rechaza Ho.

Zona de rechazo.
Para todo valor de probabilidad mayor que 0.05, se acepta Ho y se rechaza Ha.

Desarrollo mental de algunos niños y escolaridad de las madres.

Escolaridad de la madre (X)	Calificación del desarrollo mental de los niños (Y)
Primero de secundaria	90
Primero de primaria	87
Profesionista	89
Sexto de primaria	80
Tercero de secundaria	85
Tercero de primaria	84
Analfabeta	75
Preparatoria	91

Aplicación de la prueba estadística.
Inicialmente, las observaciones de las variables X y Y se ordenan en rangos.

Arreglo en rangos de las observaciones presentadas en la tabla anterior.

Rango	Escolaridad materna	Rango	Desarrollo mental del hijo (Y)
5	Primero de secundaria	7	90
2	Primero de primaria	5	87
8	Profesionista	6	89
4	Sexto de primaria	2	80
6	Tercero de secundaria	4	85
3	Tercero de primaria	3	84
1	Analfabeta	1	75
7	Preparatoria	8	91

De acuerdo con esto, se efectúa un ordenamiento natural de los rangos de las variables X y Y.

Rangos de la variable independiente X y su correspondiente de la variable dependiente.

Variable	Rangos							
(X)	1	2	3	4	5	6	7	8

El cálculo de la puntuación efectiva (S) se realiza con el ordenamiento de los rangos de la variable dependiente (Y).
El primer valor del rango de Y es 1. Respecto a los demás rangos, existen siete mayores que Y y ninguno es menor, de manera que queda:
S = (7 - 0) +

Después está el rango 5, luego se hallan tres por arriba y tres por debajo de éste y se continúa:
S = (7 - 0) + (3 - 3) +

En rango siguiente es el 3, del cual cuatro son mayores y uno menor, y queda:
S = (7 - 0) + (3 - 3) + (4 - 1) +

El rango inmediato es el 2, y los cuatro subsecuentes son mayores y ninguno menor:
S = (7 - 0) + (3 - 3) + (4 - 1) + (4 - 0) +

Después se halla el rango 7, en el que uno es mayor y dos menores:
S = (7 - 0) + (3 - 3) + (4 - 1) + (4 - 0) + (1 - 2) +

Finalmente, se encuentra el rango 8, el subsecuente es el 6, que es menor y se concluye el cálculo de S, como sigue:
S = (7 - 0) + (3 - 3) + (4 - 1) + (4 - 0) + (1 - 2) + (0 - 1)
S = 7 + 0 + 3 + 4 - 1 - 1 = 12

Aplicamos la ecuación de la prueba estadística.

$$\tau \text{ (tau)} = \frac{S}{1/2\, N\, (N - 1)} = \frac{12}{1/2\; 8(8 - 1)} = 0.43$$

Calculamos el nivel se significancia.

$$Z = \frac{\tau}{\sqrt{\frac{2\,(2N + 5)}{9N\,(N - 1)}}} = \frac{0.43}{\sqrt{\frac{2(2 \text{ X } 8 + 5)}{9 \text{ X } 8(8 - 1)}}} = \frac{0.43}{0.289} = 1.49$$

Una vez calculado el valor Z, se obtiene la probabilidad en la tabla de coeficientes de correlación en niveles de p 0.05 y 0.01; a su vez en buscamos en la tabla de probabilidades asociadas en valores extremos como los de 2 en la distribución normal.

Se localiza el valor 1.4 y en la intersección de la columna 0.09, se observa el valor 0.0681, el cual corresponde a la probabilidad de obtener un valor Z de esta magnitud, que difiere del promedio y es mayor que el nivel de significancia.

Decisión.
Como el valor Z tiene mayor probabilidad que el nivel de significancia, se acepta Ho y se rechaza Ha.

Interpretación.
La correlación entre las variables educación materna y desarrollo mental del hijo no es significativa. Esta misma conclusión se obtuvo mediante el coeficiente de correlación de Spearman.

En la sección de coeficiente de correlación de Spearman se señaló que al aumentar el tamaño de la muestra, existe mayor probabilidad de empates o ligas entre los rangos de las observaciones. Para esta condición se presenta la siguiente ecuación:

$$\tau \text{ (tau)} = \frac{S}{\sqrt{1/2\, N\, (N - 1) - Lx}\; \sqrt{1/2\, N\, (N - 1) - Ly}}$$

Donde:
t (tau) = coeficiente de correlación de Kendall.
S = puntuación efectiva de los rangos.
N = tamaño de la muestra en parejas de variables.
Lx = sumatoria de ligas o empates dados en la variable independiente (X).
Ly = sumatoria de ligas o empates dados en la variable dependiente (Y).

El nivel de significancia se obtiene de la misma manera

5.4.1 LA ECUACIÓN DE REGRESIÓN LINEAL SIMPLE: CÁLCULO E INTERPRETACIÓN.

Si sabemos que existe una relación entre una variable denominada dependiente y otras denominadas independientes (como por ejemplo las existentes entre: la experiencia profesional de los trabajadores y sus respectivos sueldos, las estaturas y pesos de personas, la producción agraria y la cantidad de fertilizantes utilizados, etc.), puede darse el problema de que la dependiente asuma múltiples valores para una combinación de valores de las independientes.

La dependencia a la que hacemos referencia es relacional matemática y no necesariamente de causalidad. Así, para un mismo número de unidades producidas, pueden existir niveles de costo, que varían empresa a empresa.

Si se da ese tipo de relaciones, se suele recurrir a los estudios de regresión en los cuales se obtiene una nueva relación pero de un tipo especial denominado función, en la cual la variable independiente se asocia con un indicador de tendencia central de la variable dependiente. Cabe recordar que en términos generales, una función es un tipo de relación en la cual para cada valorde la variable independiente le corresponde uno y sólo un valor de la variable dependiente.

2. ASPECTOS TEÓRICOS

REGRESIÓN SIMPLE Y CORRELACIÓN

La Regresión y la correlación son dos técnicas estadísticas que se pueden utilizar para solucionar problemas comunes en los negocios.

Muchos estudios se basan en la creencia de que es posible identificar y cuantificar alguna Relación Funcional entre dos o más variables, donde una variable depende de la otra variable.

Se puede decir que Y depende de X, en donde Y y X son dos variables cualquiera en un modelo de Regresión Simple.

"Y es una función de X"

Y = f(X)

Como Y depende de X,

Y es la variable dependiente, y

X es la variable independiente.

En el Modelo de Regresión es muy importante identificar cuál es la variable dependiente y cuál es la variable independiente.

En el Modelo de Regresión Simple se establece que Y es una función de sólo una variable independiente, razón por la cual se le denomina también Regresión Divariada porque sólo hay dos variables, una dependiente y otra independiente y se representa así:

Y = f (X)

"Y está regresando por X"

La variable dependiente es la variable que se desea explicar, predecir. También se le llama REGRESANDO ó VARIABLE DE RESPUESTA.

La variable Independiente X se le denomina VARIABLE EXPLICATIVA ó REGRESOR y se le utiliza para EXPLICAR Y.

ANÁLISIS ESTADÍSTICO: REGRESIÓN LINEAL SIMPLE

En el estudio de la relación funcional entre dos variables poblacionales, una variable X, llamada independiente, explicativa o de predicción y una variable Y, llamada dependiente o variable respuesta, presenta la siguiente notación:

Y = a + b X + e

Donde:

a es el valor de la ordenada donde la línea de regresión se intercepta con el eje Y.

b es el coeficiente de regresión poblacional (pendiente de la línea recta)

e es el error

SUPOSICIONES DE LA REGRESIÓN LINEAL

1. Los valores de la variable independiente X son fijos, medidos sin error.
2. La variable Y es aleatoria
3. Para cada valor de X, existe una distribución normal de valores de Y (subpoblaciones Y)
4. Las variancias de las subpoblaciones Y son todas iguales.
5. Todas las medias de las subpoblaciones de Y están sobre la recta.
6. Los valores de Y están normalmente distribuidos y son estadísticamente independientes.

ESTIMACIÓN DE LA ECUACIÓN DE REGRESIÓN MUESTRAL

Consiste en determinar los valores de "a" y "b " a partir de la muestra, es decir, encontrar los valores de a y b con los datos observados de la muestra. El método de estimación es el de **Mínimos Cuadrados**, mediante el cual se obtiene:

$a = \overline{Y} - b\ \overline{X}$

$$b = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum X_i \; Y_i - n \; \overline{X} \; \overline{Y}}{\sum X_i^2 - \overline{X}^2}$$

Luego, la ecuación de regresión muestral estimada es

$$\hat{Y} = a + b \; X$$

Que se interpreta como:

a es el estimador de a

Es el valor estimado de la variable Y cuando la variable X = 0

b es el estimador de b , es el coeficiente de regresión

Está expresado en las mismas unidades de Y por cada unidad de X. Indica el número de unidades en que varía Y cuando se produce un cambio, en una unidad, en X (pendiente de la recta de regresión).

Un valor negativo de b sería interpretado como la magnitud del decremento en Y por cada unidad de aumento en X.

3. ANTECEDENTES DEL PROBLEMA

Los datos de la siguiente tabla representan las estaturas (X, cm) y los pesos (Y, kg) de una muestra de 12 hombres adultos. Para cada estatura fijada previamente se observó el peso de una persona seleccionada de entre el grupo con dicha estatura, resultando:

X	152	155	152	155	157	152	157	165	162	178	183	178
Y	50	61.5	54.5	57.5	63.5	59	61	72	66	72	84	82

Con estos datos vamos a plantear una ecuación de regresión simple que nos permita pronosticar los pesos conociendo las tallas. Utilizaremos a = 0.05, y contrastaremos nuestra hipótesis con la prueba F.

4. DESARROLLO

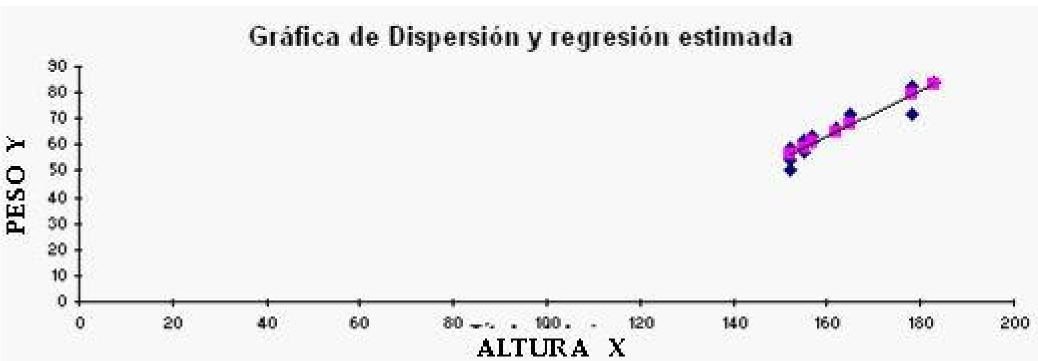
- Representación matemática y gráfica de los datos:

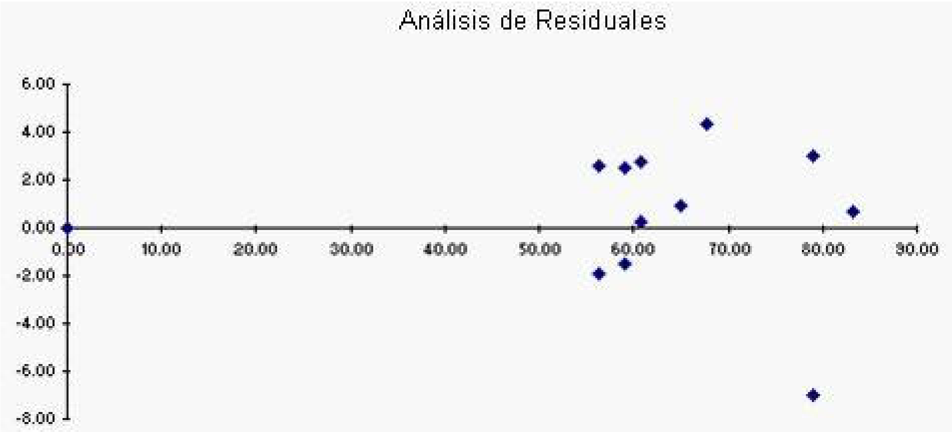
Representación Matemática

	estatura	pesos		Regresión Lineal				I.C. para la media		I. C. individual	
datos	x	y	x ^2	y ^2	xy	y est.	Residual	L. I.	L. S.	L. I.	L. S.
1	152	50	23104	2500	7600	56.43	-6.43	53.07	59.79	47.30	65.56

2	155	61.5	24025	3782.3	9532.5	59.03	2.47	56.09	61.97	50.05	68.02
3	152	54.5	23104	2970.3	8284	56.43	-1.93	53.07	59.79	47.30	65.56
4	155	57.5	24025	3306.3	8912.5	59.03	-1.53	56.09	61.97	50.05	68.02
5	157	63.5	24649	4032.3	9969.5	60.77	2.73	58.05	63.48	51.85	69.68
6	152	59	23104	3481	8968	56.43	2.57	53.07	59.79	47.30	65.56
7	157	61	24649	3721	9577	60.77	0.23	58.05	63.48	51.85	69.68
8	165	72	27225	5184	11880	67.71	4.29	65.17	70.24	58.85	76.57
9	162	66	26244	4356	10692	65.11	0.89	62.65	67.56	56.27	73.94
10	178	72	31684	5184	12816	78.99	-6.99	74.65	83.33	69.45	88.52
11	183	84	33489	7056	15372	83.32	0.68	78.01	88.64	73.31	93.34
12	178	82	31684	6724	14596	78.99	3.01	74.65	83.33	69.45	88.52

Representación Gráfica





5. HIPÓTESIS

H_O: No hay relación entre la variable peso y la variable estatura.

H_A: Hay relación entre la variable peso y la variable estatura.

Tabla de análisis de varianza								
Fuente de	Grados de		Suma de		Cuadrados			
Variación	libertad		cuadrados		medios			estadístico F
Debido a								
la regresión	1		1061.1		1061.1			73.08
error	10		145.2		14.5			
total	11		1206.3					

Se obtiene un valor F = 73.08 > 4.96, con lo cual se rechaza la hipótesis nula y aceptamos que la variable estatura está relacionada con la variable peso con un 95% de confianza.

- De acuerdo al desarrollo matemático hemos obtenido los siguientes cálculos:

$$n = 12 \quad \overline{X} = 162.167 \quad \overline{Y} = 65.25$$

$$SC(X) = \sum X^2 - n(\overline{X})^2 = 316986 - 12(162.167)^2 = 1409.667$$

$$SP(XY) = \sum XY - n(\overline{X})(\overline{Y}) = 128199.5 - 12(162.167)(65.25) = 1223$$

Lo que nos permite obtener los coeficientes a y b.

Luego,

b = 1223 / 1409.667 = 0.8676

a = 65.25 – (0.8676) (162.167) = -75.446

6. INTERPRETACIÓN

- La ecuación de regresión estimada es: $\hat{Y} = -75.446 + 0.8676\ X$

Coeficiente de correlación: R= 0.9379

Coeficiente de determinación: R^2=0.8796

El valor de b = 0.8676 indica el incremento del peso en kilogramos, en promedio, por cada centímetro de aumento en la estatura de los hombres adultos.

El valor de a, no tiene interpretación práctica en el ejemplo, se interpretaría como el valor obtenido, en promedio, para el peso Y, cuando la estatura es 0.

Utilizando la ecuación de regresión para estimar o predecir valores de la variable Y: Para una talla de 180 se obtiene un peso de 80.7 kg.

¿Cuánto se espera que pese (en promedio) una persona que mide 1.60 m?

Sustituyendo el valor de interés en la ecuación:

$$\hat{Y} = -75.446 + 0.8676\ X$$

Se obtiene:

$$\hat{Y} = -75.446 + 0.8676\ (160) = 63.37 \text{ kg.}$$

7. CONCLUSIÓN

La ecuación de Regresión Lineal estimada para las variables estatura y peso muestran, de acuerdo a la prueba F, relación.

Esta relación se ha estimado en un R = 93.7, que indica una fuerte relación positiva.

Además si consideramos el coeficiente de determinación R^2 = 87.9 podemos indicar que el 87.9% de las variaciones que ocurren en el peso se explicarían por las variaciones en la variable estatura.

BIBLIOGRAFÍA

CRISTÓBAL CRISTÓBAL, J.A.; (2003): “Lecciones de Inferencia Estadística”. Prensas Universitarias de Zaragoza.

DEGROOT, M.H.; (1988): “Probabilidad y Estadística”. Addison-Wesley Iberoamericana.

LOPEZ DE LA MANZANARA BARBERO, J.; (1989): “Problemas de Estadística”. Pirámide. 9ª ed.

LOPEZ CACHERO, M.; (1990): “Fundamentos y Métodos de Estadística”. Pirámide. 9ª ed.

MARTÍN MARTÍN, Q; (2001): “Contrastes de Hipótesis”. La Muralla.

MENDENHALL, W.; SCHEAFFER, R.L. y WACKERLY, D.D.; (1986): “Estadística Matemática con Aplicaciones”. Grupo Editorial Iberoamérica.

MEYER, P.L.; (1998): “Probabilidad y Aplicaciones Estadísticas”. Addison-Wesley.