SESIÓN 8

DESCRIPCIONES DE UNA RELACIÓN

I. CONTENIDOS:

1. Regresión lineal simple.
2. Interpretación de gráficas de regresión.
3. Cálculo de coeficiente de correlación.
4. Interpretación del coeficiente de correlación.

II. OBJETIVOS:

Al término de la Sesión, el alumno:

- Aprenderá a realizar cálculos de regresión lineal.
- Realizará predicciones a partir de la ecuación obtenida en la regresión lineal.
- Aprenderá a calcular e interpretar el coeficiente de correlación.

III. PROBLEMATIZACIÓN:

Comenta las preguntas con tu Asesor y selecciona las ideas más significativas

- Si los datos estadísticos dispersos en una gráfica, pueden servir para predecir el comportamiento de una variable.
- Si los datos estadísticos se distribuyen a lo largo de una recta ¿sería más fácil predecir algún comportamiento?
- ¿Qué utilidad pueden tener el hacer predicciones en el comportamiento de una variable?

IV. TEXTO INFORMATIVO-FORMATIVO:

1.1. Regresión lineal simple

Cuando se tiene la hipótesis de que dos variables se encuentran vinculadas en una relación de dependencia, es posible encontrar una función matemática que determine dicha relación. En esta clase veremos la regresión lineal simple que permite encontrar la recta de mejor ajuste entre una serie de puntos en el plano cartesiano. Cada punto correlaciona a las variables pues corresponde a observaciones hechas sobre un conjunto de elementos con el fin de conocer el grado de dependencia de las variables.

2.1. Interpretación de gráficas de regresión

La variable independiente se asocia con las abscisas del plano cartesiano, las “x”, y la variable dependiente con las ordenadas la “y”. El conjunto de puntos, que correlacionan a las variables, graficados en el plano cartesiano reciben el nombre de diagrama de dispersión, para algunos casos se puede observar una marcada tendencia lineal. La recta de mejor ajuste se calcula por el método de los mínimos cuadrados. Este método permite encontrar un conjunto de puntos que forman la recta de mejor ajuste, tales que los cuadrados de las diferencias entre los valores observados y los puntos de la recta sean los mínimos posibles.

Para encontrar la ecuación de la recta de mejor ajuste, que describe cómo se relacionan las variables, se procede de la siguiente manera:

1. Se analizan las variables para determinar cuál es la variable independiente y cuál es la variable dependiente.
2. Se disponen en columnas como “X” los datos de la variable independiente, y como “Y” los datos de la variable dependiente.
3. Se elevan al cuadrado los datos de cada columna para formar otras dos: “X^2” y “Y^2”.
4. Se multiplican los datos de las columnas “X” e “Y” para formar la columna “XY”.
5. Se suman los datos de cada columna.
6. Se aplican las siguientes fórmulas:

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$a = \frac{\Sigma y}{n} - b\frac{\Sigma x}{n}$$

“b” representa la pendiente de la recta; ésta corresponde a un número que permite indicar cómo cambia la variable dependiente al cambiar una unidad la variable independiente.

“a” representa la posición del eje de las ordenadas las “y” donde cruza la recta de mejor ajuste.
La función lineal es:

$$\hat{y} = bx + a$$

Con esta función es posible estimar los valores que tendría la variable dependiente, para algunos valores propuestos de la variable independiente.

Ejemplo 1 Los siguientes datos representan los resultados de un examen médico aplicado a 10 personas donde se midió el nivel de colesterol en la sangre (mg) y la presión sanguínea diastólica (mm Hg)

COLESTEROL (mg)	PRRESION SANGUINEA (mm Hg)
522	78
149	68
316	83
139	44
416	79
638	65
230	66
972	80
754	58
850	65

a) *Determinar la ecuación de la recta de mejor ajuste.*
b) *Elaborar el diagrama de dispersión.*
c) *¿Qué presión sanguínea se estima que se medirá en una persona que presenta 1250 mg de colesterol?*

Se inicia averiguando cuál es la variable independiente, se hace la pregunta: ¿La presión sanguínea depende del nivel de colesterol o el nivel de colesterol depende de la presión sanguínea?

La respuesta a la pregunta es que la presión sanguínea depende del nivel de colesterol en la sangre, por lo que la variable dependiente es la presión sanguínea y la independiente el nivel de colesterol.

Luego se elabora la tabla que se ha descrito:

COLESTEROL	PRESION SANGUÍNEA	XY	X^2	Y^2
522	78	40716	272484	6084
149	68	10132	22201	4624
316	83	26228	99856	6889
139	44	6116	19321	1936
416	79	32864	173056	6241
638	65	41470	407044	4225
230	66	15180	52900	4356
972	80	77760	944784	6400
754	58	43732	568516	3364
850	65	55250	722500	4225
4986	**686**	**349448**	**3282662**	**48344**

Se sustituye en las fórmulas:

$$b = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$b = \frac{349448 - \frac{(4986)(686)}{10}}{3282662 - \frac{(4986)^2}{10}}$$

$$b = \frac{349448 - 342039.6}{3282662 - 2486019.6}$$

$$b = \frac{7408.4}{796642.4}$$

$$b = 0.0093$$

$$a = \frac{\Sigma y}{n} - b\frac{\Sigma x}{n}$$

$$a = \frac{686}{10} - 0.0093\frac{4986}{10}$$

$$a = 68.6 - 4.636$$

$$a = 63.96$$

Entonces la ecuación de la recta de mejor ajuste es:

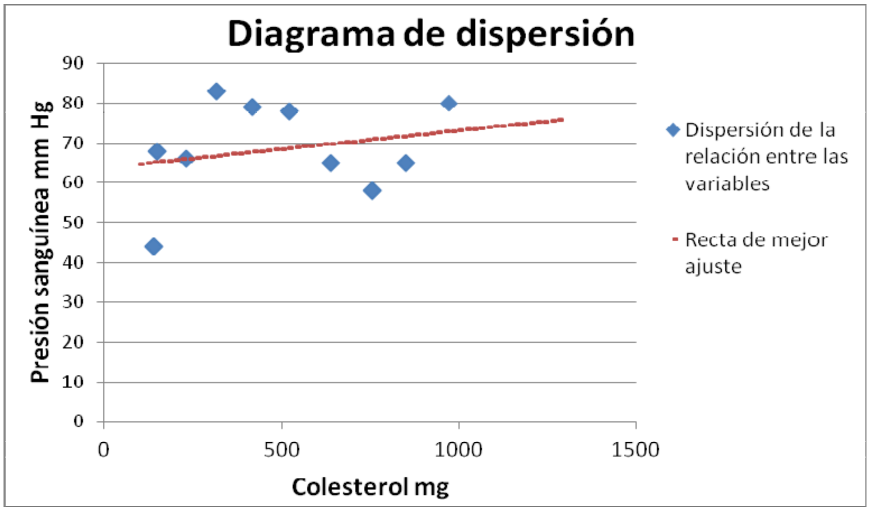
$\hat{y} = 0.0093x + 63.93$

La estimación de la presión sanguínea de una persona que tenga 1250 mg de colesterol se obtiene sustituyendo la “x” de la ecuación anterior por 1250

$\hat{y} = 0.0093(1250) + 63.93$ = 75.55

Entonces, esta persona tendría una presión sanguínea de 75.55 mm Hg

Ahora construiremos el diagrama de dispersión utilizando las valores de “x” e “y” para formar coordenadas y representarlas como puntos en el plano cartesiano.



3.1. Cálculo de coeficiente de correlación

La medición de qué tan intensa es la correlación entre las variables para integrar una función lineal, se da por el llamado coeficiente de correlación o de Pearson. Este se calcula con la fórmula:

$$r = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

Para el ejemplo anterior, el cálculo del coeficiente de correlación es:

$$r = \frac{349448 - \frac{(4986)(686)}{10}}{\sqrt{3282662 - \frac{(4986)^2}{10}}\sqrt{48344 - \frac{(686)^2}{10}}}$$

$$r = \frac{7408.4}{\sqrt{796642.4}\sqrt{1284.4}}$$

$$r = \frac{7408.4}{(892.548)(35.838)}$$

$$r = 0.23$$

A continuación se presenta la interpretación de este valor.

4.1. Interpretación del coeficiente de correlación

Un resultado cercano a la unidad tanto positiva como negativa, indica una marcada correlación lineal. Esto significa que hay una función lineal que vincula a las variables o que los puntos formados con las observaciones hechas se ajustan a una recta.

Un valor del coeficiente de correlación igual cero, indica que las variables no se ajustan a una función lineal. A continuación se muestra un cuadro para la interpretación del coeficiente de correlación.

r	Intensidad de la relación	Dispersión de los puntos en torno a la recta de regresión
$r = -1.0$	Negativa perfecta	Ninguna
$-1.0 < r \le -0.7$	Negativa intensa	Pequeña
$-0.7 < r \le -0.3$	Negativa moderada	Moderada
$-0.3 < r \le 0.0$	Negativa débil	Grande
$r = 0.0$	No hay relación lineal	Muy grande
$0.0 < r \le 0.3$	Positiva débil	Grande
$0.3 < r \le 0.7$	Positiva moderada	Moderada
$0.7 < r \le 1.0$	Positiva intensa	Pequeña
$r = 1.0$	Positiva perfecta	Ninguna

De acuerdo a la tabla, el coeficiente de correlación del ejemplo anterior indica que la relación entre las variables es positiva débil.

V. ESTRATEGIAS CENTRADAS EN EL APRENDIZAJE:

A. Realiza el siguiente ejercicio:

Los datos siguientes representan la relación entre el peso (libras) y el tamaño del pecho (pulgadas) de 12 osos. Calcula cuál es el peso estimado de un oso que tiene 50 pulgadas de pecho. Calcula e interpreta también el coeficiente de correlación lineal y realiza el diagrama de dispersión.

Pecho.	Peso.
26	90
47	351
54	420
49	363
40	262
49	377
46	353
40	253
19	40
30	152
56	419
45	357