Trabajo para el parcial 3 **Valor 40 puntos**

El objetivo de este trabajo es que aprendan además de realizar el ajuste de un modelo de regresión usando Excel, y demuestren su capacidad para explicar a través de un documento elaborado en Word el procedimiento que se debe seguir para realizar el ajuste de un modelo de regresión lineal simple. En este sentido deberán utilizar la aplicación para realizar los cálculos y utilizar Word para realizar el trabajo.

Los puntos a tratar ene le trabajo son:

1. Que es un modelo de regresión lineal simple.
2. Qué utilidad tiene el ajuste de un modelo de regresión lineal simple en la psicología.
3. Características del modelo de regresión lineal simple. El modelo matemático, los coeficientes del modelo y la forma correcta de interpretarlos.
4. El principio del método de mínimos cuadrados.
5. El diagrama de dispersión. Qué es y cuál es su utilidad
6. Ajustar un modelo de regresión lineal simple utilizando datos reales. En este sentido deberán buscar una base de datos real y ajustar un modelo de regresión lineal simple usando los datos de dos variables, la cual deben **garantizar** sea accesible para la consulta y verificación de los datos.

Este trabajo lo pueden hacer en grupos de dos y a continuación les presento un ejemplo de cómo pueden realizar el ajuste de un modelo de regresión lineal simple. Es importante que comprendan que el ajuste del modelo es fácil de hacer, por lo que la importancia de este trabajo no está en la estimación del modelo como tal, sino en evaluar la capacidad del estudiante para elaborar un material que permita a otros comprender desde la definición del modelo y su importancia hasta el hacer uso del modelo para predicción de valores de una variable a partir de otra. Copia y pegue y diferentes formatos y tipos de letra no se permiten, por lo cual una vez hecho léanlo y siéntanse orgullosos del trabajo que hicieron antes de enviarlo.

Ejemplo del ajuste de un modelo de regresión lineal simple

Un equipo de investigadores de la UCV estudio la relación entre la temperatura promedio y la cosecha de maíz durante la estación de crecimiento. Las medidas durante algunos años produjeron los siguientes resultados.

Años	Temperatura promedio °C
2000	296
2001	661
2002	162
2003	243
2004	263
2005	213

Calcular:
a.- Obtener la recta de regresión lineal mediante el método de lo mínimos cuadrado para predecir la temperatura

La recta de regresión estimada $\hat{y}$ esta dada por:

$$\hat{y} = b_1 x + b_0$$

donde b_1 y b_0 se conocen como los coeficientes estimados de la recta de regresión y se obtienen a partir de las siguientes expresiones cuando se utiliza el método de los mínimos cuadrados

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ y}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - b_1 \bar{x}.$$

:

y los coeficientes $b_0 \; y \; b_1$ los hallaremos haciendo los cálculos con la ayuda de Excel.

	Años (X)	Temperatura promedio °C (Y)	XY	X^2	Y^2
	2000	296	592000	4000000	87616
	2001	661	1322661	4004001	436921
	2002	162	324324	4008004	26244
	2003	243	486729	4012009	59049
	2004	263	527052	4016016	69169
	2005	213	427065	4020025	45369
Suma =	12015	1838	3679831	24060055	724368

De esta manera,

$$b_1 = \frac{n\sum_{i=1}^{n} x_i\, y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{6 * 3679831 - (12015)(1838)}{6 * 24060055 - (12015)^2} = \frac{-4584}{105} = -43.6571$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} = \frac{1838 - (-43.6571) * 12015}{6} = \frac{526378.571429}{6} = 87729.7619$$

por lo tanto, la línea de regresión está dada por:

$$\hat{y} = -43.6571x + 87729.7619$$

b.- Calcular el error estándar del modelo e interpretarlo

El modelo de regresión lineal simple está dado por

$$y = \beta_1 x + \beta_0 + \varepsilon$$

donde ε es una variable aleatoria que se supone está distribuida con $E(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2$.

El error estándar de error ε es uno de los parámetros del modelo. Al estimador de la varianza de ε se le conoce como varianza del error o varianza residual y se estima a través de la expresión:

$$s^2 = \frac{SCE}{n-2} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-2} = \frac{S_{yy} - b_1 S_{xy}}{n-2}$$

y a S^2 se le denomina error cuadrado medio.

Estimación de σ^2.

$$S^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2}$$

donde

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}{n} = \frac{6 * 724368 - (1838)^2}{6} = \frac{967964}{6} = 161327.3333$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{n\sum_{i=1}^{n} x_i\, y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n} = \frac{6 * 3679831 - (12015)(1838)}{6} = \frac{-4584}{6}$$

$$S_{xy} = -764$$

Así,

$$S^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{161327.3333 - (-43.6571) * (-764)}{6-2} = \frac{127973.276}{4} = 31993.319$$

Entonces, el error estándar de ε esta dado por $S = \sqrt{S^2} = \sqrt{31993.319} = 178.87$ y representa la variabilidad de las perturbaciones no explicadas por el modelo de regresión.

c.- Uso del modelo encontrado para a partir de él estimar que temperatura promedio tendrá el año 2006

Para $x = Año\ =\ 2006$ la temperatura se estima es

$$\hat{y} = -43.6571 * 2006 + 87729.7619 = 153.3353\ °C$$