

GUIA N°1 CONTROL 3

P1)

Se busca estudiar la relación entre dos variables X e Y.

- a) X e Y variables cuantitativas. ¿Qué herramientas estadísticas y gráficas utilizaría? ¿Qué tipo de relaciones podría detectar? Describa.
- b) X variable cualitativa e Y cuantitativa. ¿Qué herramientas estadísticas y gráficas utilizaría? ¿Qué tipo de relaciones podría detectar? Describa.
- c) Al realizar una regresión de la variable Y en función de una variable X se obtiene el gráfico (1) entre y e \hat{y} . El gráfico (2), corresponde a la gráfica de y v/s $\hat{\varepsilon}$. Según su intuición, ¿Es posible mejorar el modelo? Explique claramente que es lo que haría para mejorarlo.

Gráfico 1: y v/s \hat{y}

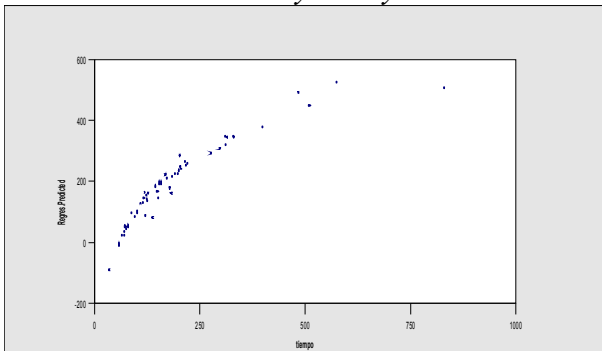
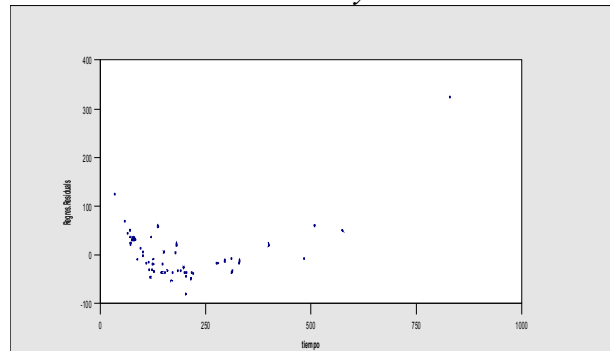


Gráfico 2: y v/s $\hat{\varepsilon}$



P2)

Investigadores de la conducta de compra de las personas desean saber si cuando un producto tiene varias promociones en un mes, los consumidores esperan que el precio del producto sea aún mas bajo en el futuro. Para ello, en cuatro supermercados distintos se realizaron 1, 3, 5 y 7 promociones respectivamente y 40 personas por supermercado (elegidas al azar), observaron el precio durante 5 semanas. Al término de las 5 semanas cada persona hizo una estimación del precio del producto para el día siguiente (ver tabla 8). Se desea testear si el número de promociones influye en el precio esperado.

Tabla 8

N° Promociones	Precio esperado por los 160 encuestados									
1	3.78	4.13	3.89	4.54	4.36	4.56	4.31	4.46	4.18	3.82
	3.97	4.73	3.96	4.32	4.22	4.24	4.34	3.91	3.98	4.38
	3.62	4.23	4.37	4.36	4.4	4.18	4.46	4.58	4.79	4.27
	4.06	4	4.5	4.6	3.97	3.94	4.1	4.33	4.26	3.86
3	4.12	3.91	3.96	4.22	3.88	4.14	4.17	4.07	4.16	4.12
	3.84	4.01	4.42	4.01	3.84	3.95	4.26	3.95	4.3	4.33
	4.17	3.97	4.32	3.87	3.91	4.21	3.86	4.14	3.93	4.08
	4.07	4.08	3.95	3.92	4.36	4.05	3.96	4.29	3.6	4.11
5	3.32	3.86	4.15	3.71	3.78	3.93	3.73	3.71	4.1	3.65
	3.69	3.83	3.58	3.99	3.72	4.41	4.12	3.73	3.56	4.08
	3.25	3.76	3.56	3.47	3.58	3.76	3.57	3.87	3.92	3.48
	3.39	3.54	3.86	4.37	3.77	3.81	3.71	3.58	3.69	3.77
7	3.45	3.64	3.37	3.27	3.58	4.01	3.67	3.74	3.5	3.6
	3.97	3.57	3.5	3.81	3.55	3.08	3.78	3.86	3.29	3.77
	3.25	3.07	3.21	3.55	3.23	2.97	3.86	3.14	3.43	3.84
	3.65	3.45	3.73	3.12	3.82	3.7	3.46	3.73	3.79	3.94

- a) Realice un resumen de los datos encontrando el tamaño de la muestra, la media y la desviación estándar de cada grupo. Construya la tabla ANOVA.
- b) ¿El supuesto de igualdad de desviación estándar es razonable?
- c) Realice el test ANOVA. Escriba claramente las hipótesis testeadas, el estadístico utilizado con sus grados de libertad y el P-valor. Concluya.

P3)

Sea $Y_i = \beta_0 + \beta_1 x + \varepsilon_i$ ($i=1 \dots n$) un modelo lineal simple, donde $\varepsilon_i \sim N(0, \sigma^2)$, y sean $\hat{\beta}_0$ y $\hat{\beta}_1$ los estimadores MCO para β_0 y β_1 respectivamente. Muestre que:

$$a) \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i \cdot y_i - n \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b) \text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$$

$$c) \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$d) \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$e) \tilde{\sigma}^2 = \frac{n S_y^2 (1 - R^2)}{n - 2}$$

P4)

Una persona posee una balanza con la cual quiere determinar su peso corporal "Y". La balanza no es perfecta, y puede devolver lecturas para el peso con un error que se distribuye según una Normal de media 0 y varianza σ^2 . Proponga un modelo lineal para analizar el fenómeno y encuentre el estimador MCO correspondiente al modelo propuesto. Resuelva el problema desde el punto de vista de la estimación clásica y compare con la solución anterior. Comente.

P5)

Se obtiene el siguiente resultado de correr un modelo lineal simple $Y_i = \beta_0 + \beta_1 x + \varepsilon_i$ ($i = 1 \dots n$) con los supuestos usuales de normalidad de los errores:

Output del Modelo

Variable	Coficiente	Desv.est.	T	P-Valor
Cte	3,830	???	2,17	0,038
X	???	0,05012	18,03	???

R-cuadrado: ???

Análisis de Varianza de la Regresión

Fuente	Grados de libertad	Suma de Cuadrados	Cuadrados Medios	F	P-Valor
Regresión	1	???	???	???	???
Residuos	???	323,3	???		
Total	32	3713,9			

a) Complete los datos faltantes de las tablas anteriores y el valor de R-cuadrado. Comente los resultados.

b) Encuentre una estimación insesgada para σ^2 .

c) Encuentre un intervalo de confianza al 95% para β_1 .

d) Resuelva el test hipótesis

$$H_0: \beta_1 = 1$$

$$H_1: \beta_1 < 1$$

e) Encuentre una predicción para $x=20$

P6)

En un taller, se reúne un conjunto de datos para determinar si la proporción de artículos defectuosos producidos por los trabajadores es la misma para el turno matutino, vespertino o nocturno. Se obtuvo lo siguiente:

Turno:	Matutino	Vespertino	Nocturno
Defectuosos	45	55	70
No defectuosos	905	890	870

Utilizando un nivel de significación de 0,025 determine si la proporción de defectos depende del turno.