Universidad de la República **Modelos Lineales**
Facultad de Ingeniería - IMERL **Primer semestre 2016**

Práctico 2

1. Complete la siguiente tabla:

Source	grados libertad	Suma Cuadrados	Promedio Cuadrados	F
Regresión		12.597		0.000
Error residual				
Lack of Fit	3			
Error Puro		0.157		
Total	14	15.522		

2. Justificar las fórmulas vistas en el teórico, con los supuestos hechos, para hallar los intervalos de confianza de β_0, β_1 y σ^2 de un modelo lineal simple y normal.

3. *Intervalos de confianza para la respuesta media y para la predicción*

   Leer y entender las secciones 6.3.4 y 6.3.5 de las notas de F. Carmona, *Modelos Lineales*.

4. Suponemos que tenemos n observaciones independientes $Y_1, \dots, Y_n$ con distribución normal con la misma varianza σ^2 y con esperanzas respectivas $\mu_i = \beta_0 + \beta_1 x_i$ donde β_0 y β_1 son desconocidos y $x_1, \dots, x_n$ son conocidos.

   a) Calcular los estimadores de máxima verosimilitud de β_0, β_1 (observe que coinciden con los estimadores por el método de los mínimos cuadrados) y de σ^2.

   b) Deducir de las ecuaciones normales

$$n\widehat{\beta}_0 + \left(\sum_{i=1}^{n} x_i\right)\widehat{\beta}_1 = \sum_{i=1}^{n} Y_i$$

$$\left(\sum_{i=1}^{n} x_i\right)\widehat{\beta}_0 + \left(\sum_{i=1}^{n} x_i^2\right)\widehat{\beta}_1 = \sum_{i=1}^{n} x_i Y_i$$

   que $\widehat{\beta}_0$ y $\widehat{\beta}_1$ tiene distribución normal y calcular sus esperanzas y varianzas

5. La cantidad de cierto producto químico que se disuelve en una cantidad de agua dada, a varias temperaturas está dada por la siguiente tabla:

Y=cantidad disuelta	8	12	21	31	39	58
x=temperatura	0	10	20	30	40	60

   Utilizando el modelo $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, encontrar un intervalo de confianza para β_1 al 95 %.

6. Los siguientes datos representan la altura x y el peso Y de 12 individuos. Aplicar un modelo de regresión lineal para predecir el peso a partir de la altura.

x (cm)	163	163	165	166	168	169	170	170	171	172	172	174
Y	60	61	62	63	65	67	68	69	70	72	71	70

Probar la hipótesis de que la pendiente es nula (y por lo tanto la altura no sirve para predecir el peso).

7. Se hace un estudio del efecto de la temperatura (t) en la cantidad producida por cierto material (Y), en un proceso químico. Se obtienen los resultados siguientes:

t	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y	1	5	4	7	10	8	9	13	14	13	18

   a) Ajustar un modelo de regresión lineal para predecir Y en función de t.

   b) Probar la hipótesis de que la pendiente es nula con un nivel del 95 %.

   c) Hallar un intervalo de confianza para el valor esperado de Y evaluado en $t = 3$ con un nivel del 95 %.

   d) Hallar un intervalo de confianza para una futura Y evaluado en $t = 3$con un nivel del 95 %.

8. *a*) Hallar la recta de regresión simple de la variable *velocidad* sobre la variable regresora *densidad* con los datos siguientes:

Densidad	12,7	17	66	50	87,8	81,4	75,6	66,2	81,1	62,8	77	89,6
Velocidad	62,4	50,7	17,1	25,9	12,4	13,4	13,7	17,9	13,8	17,9	15,8	12,6

Densidad	18,3	19,1	16,5	22,2	18,6	66	60,3	56	66,3	61,7	66,6	67,8
Velocidad	51,2	50,8	54,7	46,5	46,3	16,9	19,8	21,2	18,3	18	16,6	18,3

   b) Calcular la estimación de σ^2 y a partir de ella las estimaciones de $s.e(\widehat{\beta}_0)$ y $s.e(\widehat{\beta}_1)$.

   c) Escribir los intervalos de confianza para los parámetros con un nivel de confianza del 95 %.

   d) Construir la tabla para la significación de la regresión y realizar dicho contraste.

   e) Hallar el intervalo de predicción de la respuesta media cuando la densidad es de 50 vehículos por km con un nivel de confianza del 90 %.

   f) Repetir las preguntas anteriores hallando la recta de regresión de la variable *raíz cuadrada de la velocidad* sobre la variable regresora. *densidad*

9. *a*) Probar que un modelo de regresión lineal minimizar la suma de cuadrados de los residuos equivale a maximizar R^2.

   b) Probar que en regresión múltiple agregar una variable predictora aumenta el valor del R^2.

10. Se recuerda que si $X \in \mathcal{M}_{n\times(d+1)}$ es la matriz de diseño, entonces la matriz de proyección sobre $\langle X \rangle$ es $H = X(X'X)^{-1}X'$ (H por *hat matrix*, ya que "pone el gorro" a la Y pues $\widehat{Y} = HY$). Si el vector de residuos $e = Y - \widehat{Y}$, probar:

   a) H es simétrica e idempotente, es decir $H^2 = H$. Deducir que $(I - H)^2 = I - H$.

   b) $\mathbb{E}(e) = 0$

   c) $\text{Var}(e) = \sigma^2(I - H),\ \text{Cov}(e, Y) = \sigma^2(I - H)\ \ \text{Cov}(e, \widehat{Y}) = O$

   d) $\overline{e} = \frac{1}{n}\sum e_i = 0$

   e) $e'Y = Y'(I - H)Y,\ e'\widehat{Y} = 0,\ \ e'X = 0'$

   f) $\text{traza}(H) = d + 1$