ESTADÍSTICA APLICADA II
Tarea 2

1. Mostrar que la suma

$$S(\beta_o,\beta_1)=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i)^2$$

tiene un valor crítico en el punto $(\hat\beta_0,\hat\beta_1)$, en donde

$$\hat\beta_0=\overline{Y}-\hat\beta_1\overline{X}\ ,\ \hat\beta_1=\frac{S_{XY}}{S_{XX}},\ S_{XY}=\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})\ ,\ S_{XX}=\sum_{i=1}^{n}(X_i-\overline{X})^2\ .$$

Además, demostrar que en ese punto crítico la función $S(\cdot,\cdot)$ alcanza un mínimo absoluto.

2. Mostrar que en el modelo de regresión lineal simple, con el supuesto de errores normales, los estimadores de mínimos cuadrados (m.c.) para β_0 y β_1 coinciden con los estimadores de máxima verosimilitud (m.v.). Además, encuentra el estimador máximo verosímil para σ^2.

3. Demostrar que

$$\frac{\sum_{i=1}^{n}(X_i-\overline{X})Y_i}{\sum_{i=1}^{n}(X_i-\overline{X})^2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}=\frac{\sum_{i=1}^{n}X_iY_i-n\overline{X}\overline{Y}}{\sum_{i=1}^{n}X_i^2-n\overline{X}^2}$$

4. Mostrar que, bajo el modelo de regresión lineal simple, las relaciones siguientes son verdaderas,

(a) $\sum_{i=1}^{n}e_i$

(b) $\sum_{i=1}^{n}Y_i=\sum_{i=1}^{n}\hat Y_i$

(c) $\sum_{i=1}^{n}X_ie_i=0$

(d) $\sum_{i=1}^{n}\hat Y_ie_i=0$

(e) $SCR=\hat\beta_1^2S_{XX}$

(f) $Cov(\overline{Y},\hat\beta_1)=0$

(g) $Cov(\hat\beta_0,\hat\beta_1)=-\overline{X}\sigma^2/S_{XX}$

(h) $E(CME)=\sigma^2$

(i) $E(CMR)=\sigma^2+\beta_1^2S_{XX}$

5. Demostrar que

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2=\sum_{i=1}^{n}(\hat Y_i-\overline{Y})^2+\sum_{i=1}^{n}(Y_i-\hat Y_i)^2$$

6. Mostrar que bajo el modelo de regresión lineal simple, al hacer la prueba de hipótesis $H_0:\beta_1=0$ vs. $H_1:\beta_1\neq 0$, las pruebas de 'T' y 'F' son equivalentes. (Sugerencia: Mostrar que $T^2=F$).

7. Demostrar que en el modelo de regresión lineal simple se cumple que

$$r=\sqrt{R^2}$$

8. ¿Qué pasaría si de antemano se sabe que β_1 es igual a cero?. El modelo de regresión que se ajustaría es

$$Y=\beta_0+\varepsilon_i$$

Bajo las mismas suposiciones del modelo visto en clase,
a) Encuentra los estimadores de mínimos cuadrados para este modelo.
b) Prueba que el estimador de mínimos cuadrados para β_0 es insesgado.

9. ***Regresión a través del origen.***
Ocasionalmente, en un modelo se sabe a priori que el intercepto β_O es cero. En este caso el modelo estaría dado por

$$Y_i=\beta_1X_i+\varepsilon_i$$

con las mismas suposiciones sobre los errores ε_i que para el modelo visto en clase. En base a una muestra aleatoria de tamaño n,

a) Encontrar el estimador de β_1 por máxima verosimilitud. ¿Será este estimador igual al que se obtendría por mínimos cuadrados?

b) Demostrar que el estimador encontrado es insesgado y que su varianza es igual a $\sigma^2 \Big/ \sum_{i=1}^{n} X_i^2$.

c) Encontrar $\hat{\sigma}^2$ y determinar sus grados de libertad.

d) ¿Se cumple para este modelo que la suma de los residuos sea cero?

10. La pureza del oxígeno (Y) producido por separación se piensa que está relacionado con el porcentaje de hidrocarburos (X) en el condensador principal de la unidad de proceso. Veinte muestras se muestran a continuación.

Pureza(%)	86.91	89.85	90.28	86.34	92.58	87.33	86.29	91.86	95.61	89.86
H.carb.(%)	1.02	1.11	1.43	1.11	1.01	0.95	1.11	0.87	1.43	1.02

Pureza(%)	96.73	99.42	98.66	96.07	93.65	87.31	95.00	96.85	85.20	90.56
H.carb.(%)	1.46	1.55	1.55	1.55	1.40	1.15	1.01	0.99	0.95	0.98

a) Construya un diagrama de dispersión con los datos. ¿Se sugiere una asociación lineal entre las variables?

b) Ajuste por máxima verosimilitud el modelo $Y = \beta_0 + \beta_1 X + \varepsilon$, en donde los errores son independientes y $\varepsilon \sim N(0,\sigma^2)$. Interprete cada uno de los valores estimados.

c) Obtenga el coeficiente de determinación del modelo ajustado e interprete su valor.

d) Construya intervalos de confianza para cada uno de los parámetros del modelo. Usar 97% de confianza.

e) Verifica la significancia del modelo.

f) Alguien afirma que β_1 es menor que 12. Para confirmar o refutar esta afirmación, plantear las hipótesis adecuadas y realizar la prueba correspondiente.

g) Encuentra un intervalo del 95% de confianza para la pureza media cuando el porcentaje de hidrocarburo es 1.00.

h) Encuentra un intervalo de predicción para el verdadero valor de la pureza cuando el porcentaje de hidrocarburo es 1.00.

Suponiendo que la pureza y el porcentaje de hidrocarburo son conjuntamente variables aleatorias normales.

i) ¿Cuál es la correlación entre la pureza del oxígeno y el porcentaje de hidrocarburo?

j) Prueba la hipótesis H_0: $\rho = 0$. Compara este resultado con el obtenido en el inciso (e).

k) Construye un intervalo aproximado de 95% de confianza para ρ.

11. Se ha creado un nuevo tipo de báscula que está diseñada para dar los pesos con una exactitud de milésimas de libra. Se tienen a la mano cinco objetos cuyos pesos verdaderos son: 0.1, 0.2, 0.3, 0.4 y 0.5 libras respectivamente. Se colocan dos veces cada uno de estos cinco objetos en la báscula y se registra la lectura obtenida. Se supone que el peso marcado en la báscula es una variable aleatoria normal Y con media ($\beta_1 X$) y varianza constante σ^2, en donde X es el peso verdadero.

Los datos son:

X	0.1	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.5	0.5
Y	0.098	0.099	0.208	0.200	0.302	0.298	0.405	0.401	0.502	0.495

a) Grafique Y vs X. ¿Qué modelo propone para explicar Y en función de X?.

b) Ajuste la recta de regresión por mínimos cuadrados e interprete cada valor estimado.

c) Encuentre el estimador de máxima verosimilitud de β_1.

d) Dé una estimación del peso promedio que registrará la báscula, de un objeto cuyo peso verdadero es de 0.8 libras. Comente, que tan buena cree que es la estimación.

e) Dé una estimación para la varianza del error.

f) Dé una estimación de la varianza del estimador de β_1 .

g) Encuentre un intervalo para β_1 con 95% de confianza.

h) Use los diez pesos obtenidos para probar las hipótesis

H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$

Comente, por qué sería importante realizar esta prueba de hipótesis.