Ejercicios teóricos (Optativos)

.

1. Sean A y B matrices de $m \times n$ de constantes y X e Y vectores aleatorios de dimensión n.
   a. Probar que $E(AX + BY) = AE(X) + BE(Y)$.
   b. Probar que si a es un vector de constantes de dimensión n entonces $\Sigma_{Y-a} = \Sigma_Y$
2. Llamemos Σ_{XY} a la matriz de covarianza entre los vectores X e Y, es decir $(\Sigma_{XY})_{ij} = cov(X_i, Y_j)$.
   a. Probar que $\Sigma_{XY} = E[(X - E(X))(Y - E(Y))']$.
   b. Probar que si X es un v.a. de dimensión m, Y un v.a. de dimensión n y A y B matrices de constantes de $l \times m$ y $p \times n$ respectivamente, entonces $\Sigma_{AX,BY} = A\Sigma_{XY}B'$.
   c. Probar que si Y es un vector aleatorio y a es un vector de constantes, entonces $E((Y - a)(Y - a)') = \Sigma_Y + (E(Y - a))(E(Y - a))'$.
3. Probar que $Var(\hat{Y}_i) \leq Var(Y_i)$. Interpretar este resultado.
4. (*) Con el fin de estimar dos parámetros θ y ϕ es posible tomar 3 tipos de observaciones:
   a. con esperanza θ,
   b. con esperanza $\theta + \phi$,
   c. con esperanza $\theta - 2\phi$.

   Todas las observaciones están sujetas a errores no correlacionados con esperanza 0 y varianza constante. Si se toman m observaciones de tipo a), m del tipo b) y n del tipo c), hallar los estimadores de mínimos cuadrados de θ y ϕ y probar que son no correlacionados si $m = 2n$.
5. Consideremos el modelo
   $$Y_i = \beta_o + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$
   donde $E(\epsilon) = 0$ y además $\Sigma_\epsilon = \sigma^2 I$.
   a. ¿Qué condición deben cumplir las X_i para que los estimadores de β_o y β_1 sean no correlacionados?
   b. Si la condición hallada en a) se cumple, ¿cuál es el estimador de β_o?
6. Probar que
   a. si el modelo incluye una constante, entonces $\sum_{i=1}^n (Y_i - \hat{Y}_i) = 0$
   b. si la matriz de diseño es de rango completo, $\sum_{i=1}^n \hat{Y}_i(Y_i - \hat{Y}_i) = 0$

7. Si X e Y son dos vectores aleatorios n–dimensionales independientes con distribución normal multivariada, y si a y b son dos constantes, probar que $U = aX + bY$ también tiene distribución normal multivariada.

8. Sea $Y \sim N_3(0, I_3)$. Hallar la esperanza de $(Y_1 - Y_2)^2 + (Y_2 - Y_3)^2$.

Ejercicios prácticos

1. Implementar una función que dado un vector y de valores de respuesta y una matriz X de valores observados, mediante las ecuaciones normales, calcule el estimador de cuadrados mínimos $\hat{\beta}$.

2. Se tiene en el archivo `girasol.txt` el rinde de diversas parcelas de girasol (en toneladas) según la cantidad de dinero invertida en fertilizantes (en miles de pesos).

   a. Graficar en un diagrama de dispersión inversión vs rinde.

   b. Plantear un modelo de regresión lineal simple obtener el estimador de mínimos cuadrados.

   c. Graficar la recta de regresión obtenida, ¿detecta algo sospechoso?

3. Considerarx el archivo `abalone.txt` que contiene información sobre distintas muestras de abalones. Los atributos están separados por coma, con los siguientes campos:

   Sexo (categórica): M (masculino), F (femenino) o I (infante).
   Longitud (continua), en milimetros.
   Diametro (continua), en milimetros.
   Altura (continua), en milimetros.
   Peso completo del abalone (continua), en gramos.
   Peso de la carne (continua), en gramos.
   Peso de las visceras (continua), en gramos.
   Peso del caparazon (continua), en gramos.
   Anillos (discreta).

   a. Plantear un modelo de regresión lineal simple para predecir el diámetro en función de la longitud.

   b. Observar que el conjunto de datos tiene información del peso total de cada espécimen junto con un desagregado por partes. Ajustar un modelo de regresión múltiple que explique el peso total en función del peso del caparazón, las vísceras y la carne.

   c. Se trata ahora de establecer una relación entre el peso total y el diámetro del espécimen. Empezar dibujando en un scatter plot ambas variables. Si definimos como P al peso total y D al diámetro, se consideran los siguientes modelos:

      - Modelo lineal simple, $P = b + aD + \epsilon$.
      - Modelo cuadrático, $P = c + bD + aD^2 + \epsilon$.
      - Modelo cubico sin términos de orden inferior, $P = aD^3 + \epsilon$.

Efectuar en cada caso una regresión y graficar las curvas superpuestas sobre el scatter plot.

4. En este ejercicio se crearán datos simulados y se ajustará un modelo de regresión lineal simple.

   a) Utilizando la función `rnorm`, crear un vector x que contenga 100 observaciones provenientes de una distribución $\mathcal{N}(0, 1)$.

   b) Utilizando la función `rnorm`, crear un vector *epsilon* que contenga 100 observaciones provenientes de una distribución $\mathcal{N}(0, 0{,}025)$.

   c) Usando x y *epsilon*, generar un vector acorde al modelo:

   $$y = -1 + 0{,}5x + epsilon$$

   ¿Cuál es la longitud del vector y? ¿Cuáles son los valores de β_0 y β_1 en el modelo?

   d) Realizar un scatterplot y observar la relación entre x e y.

   e) Ajustar un modelo lineal para predecir y en función de x utilizando el método de cuadrados mínimos. Comparar los valores exactos de β_0 y β_1 con sus estimaciones.

   f) Graficar la recta de cuadrados mínimos sobre el gráfico realizado en (d). En otro color graficar la recta $Y = -1 + 0{,}5X$.

   g) Ajustar un modelo polinomial que prediga y usando x y x^2. ¿Encuentra alguna evidencia de que el término cuadrático mejora el ajuste del modelo?

   h) Repetir los ítems (a) a (f) modificando los datos generados de manera que haya menos ruido en los datos. Una forma de hacerlo es disminuyendo el valor de la varianza de la distribución normal usada para general el término del error *epsilon*.

   i) Repetir los ítems (a) a (f) modificando los datos generados de manera que haya más ruido en los datos. Una forma de hacerlo es aumentando el valor de la varianza de la distribución normal usada para general el término del error *epsilon*.

   j) En ambos escenarios, hallar una estimación de la varianza.

5. a) Generar el siguiente modelo: Crear dos vectores de datos de tamaño 100 x_1 y x_2 a partir de una distribución uniforme en el intervalo $(0, 1)$. Crear el vector $y = 2 + 2 * x_1 + 0{,}3 * x_2 + \epsilon$, con ϵ que contenga 100 observaciones provenientes de una distribución $\mathcal{N}(0, 1)$. ¿Cuales son los coeficientes de regresión?. Estimar la correlación entre x_1 y x_2. Realizar un scatterplot en el que pueda observarse la relación entre x_1 y x_2. Utilizando los datos generados, ajustar a un modelo lineal para predecir y en función de x_1 y x_2, utilizando el método de cuadrados mínimos y comparar los valores exactos de β con sus valores estimados.

   b) Repetir el inciso *a* pero con el siguiente modelo: Crear dos vectores de datos de tamaño 100 x_1 a partir de una distribución uniforme en el intervalo $(0, 1)$, y $x_2 = 0{,}5 * x_1 + rnorm(100)/10$. Crear el vector $y = 2 + 2 * x_1 + 0{,}3 * x_2 + rnorm(100)$. Comparar los resultados obtenidos con los del ítem *a*.