1. Suponga el modelo de regresión lineal múltiple

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5 + \epsilon \tag{MC}$$

y considere la hipótesis conjunta $H_0 : \beta_1 = \beta_3$ y $\beta_2 = \beta_4$.

   a) Si supone 100 observaciones disponibles, construya un estadístico de prueba. Indique como se distribuye éste y cuál sería la región de rechazo.

   b) Si considera el procedimiento de *suma de cuadrados extra*, ¿qué modelo propondría como *modelo reducido* (MR) para contrastar la hipótesis anterior?

2. Considere el modelo de regresión lineal múltiple

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad i = 1, \ldots, n \tag{MC}$$

   a) Derive el estadístico de prueba t_j, para la hipótesis $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$, está dado

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{C_{jj}}} \sim t_\nu$$

   donde $s^2 = \text{CM}_{\text{Res}}$, C_{jj} es la j-ésima entrada de la diagonal de la matriz $C = (X'X)^{-1}$ y $\nu = n - q$ son los grados de libertad de los residuales. Indique cuál sería la región de rechazo.

   b) Si considera el procedimiento de *suma de cuadrados extra*, ¿qué modelo propondría como *modelo reducido* (MR) para contrastar la hipótesis anterior?

3. Considere la matriz "gorro" $H = (h_{ij})$.

   a) Encuentre la matriz de covarianza del vector de respuestas ajustadas $\hat{y}$.

   b) Muestre que si $X_{n\times q}$, la suma de los h_{ii} es q.

   c) Muestre que cada uno de los elementos de la diagonal h_{ii}, llamado algunas veces el *"apalancamiento"* de la i-ésima observación ($1 \leq i \leq n$) está entre 0 y 1.

   d) Muestre que el error estándar del i-ésimo residual es $\text{ee}(\hat{\epsilon}_i) = \sigma\sqrt{1 - h_{ii}}$.

   e) Muestre que el promedio de las varianzas de las respuestas ajustadas es $\overline{\text{var}(\hat{y})} = q\sigma^2/n$, donde n es el número de observaciones y $q = p + 1$ el número de coeficientes en el modelo.

4. Sean $X_1, \ldots, X_m$, *v.a.i.i.d.* distribuidas $\text{N}(\mu_X, \sigma^2)$, y sean $Y_1, \ldots, Y_n$, *v.a.i.i.d.* distribuidas $\text{N}(\mu_Y, \sigma^2)$ e independientes de las X's. Derive un estadístico de prueba para $H_0 : \mu_X = \mu_Y$. [*Sugerencia*: considere el problema como uno de regresión lineal. Enuncie claramente los supuestos sobre el modelo y sus componentes.]

5. Considere el modelo elemental

$$y_i = \beta_0 + \epsilon_i \quad (i = 1, \ldots, n)$$

donde los errores ϵ_i son no correlacionados, de media cero y varianza común σ^2. Muestre que $\hat{\beta}_0 = \bar{y}$ es el estimador de varianza mínima entre *todos* los estimadores lineales insesgados de β_0. Esto es, si $\tilde{\beta}_0$ es otro estimador tal que $\tilde{\beta}_0 = \sum_{i=1}^n w_i y_i$, donde $0 \leq w_i \leq 1$ y $\sum w_i = 1$, entonces

$$\text{var}(\hat{\beta}_0) \leq \text{var}(\tilde{\beta}_0)$$

6. Considere el modelo elemental

$$y_i = \beta_0 + \epsilon_i \quad (i = 1, \ldots, n)$$

donde los errores ϵ_i son no correlacionados, de media cero y varianza común σ^2. Entonces, por el *teorema de Gauss-Markov*, el mejor estimador lineal insesgado de β_0 es el de mínimos cuadrados $\hat{\beta}_0 = \bar{y}$. El estimador es "el mejor" en el sentido que tiene varianza mínima, a saber σ^2/n, entre todos los estimadores lineales insesgados.

Considere ahora el estimador $\tilde{\beta}_0 = \sum_i w_i y_i / \sum_i wi$, donde los pesos w_i son arbitrarios. Claramente se ve que $\tilde{\beta}_0$ es un promedio ponderado de las observaciones y_i y que es un estimador insesgado de β_0. Utilice la desigualdad de Cauchy-Schwarz para mostrar que efectivamente

$$\text{var}(\tilde{\beta}_0) \geq \text{var}(\hat{\beta}_0)$$

como lo garantiza el teorema de Gauss-Markov.

7. Considere el modelo de regresión lineal múltiple

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad (i = 1, \ldots, n)$$

Utilice el resultado 3.3 del documento sobre Prueba de Hipótesis

$$F = \frac{(SCE_H - SCE)/1}{SCE/(n-q)} = \frac{(c - A\hat{\beta})' \left[A(X'X)^{-1}A'\right]^{-1} (c - A\hat{\beta})}{s^2} \sim F_{1,n-q}$$

para verificar que el estadístico de prueba t_j, para la hipótesis $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$, está dado

$$t_j = \frac{\hat{\beta}_j}{\sqrt{s^2 C_{jj}}} \sim t_\nu$$

donde $s^2 = \mathrm{CM}_{\mathrm{Resid}}$, C_{jj} es la j-ésima entrada de la diagonal de la matriz $C = (X'X)^{-1}$ y $\nu = n - p - 1$ son los grados de libertad de los residuales.

8. Considere el modelo de regresión

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

y la hipótesis nula $H_0 : \beta_1 = \beta_3;\ \beta_2 = 0$.

   a) Utilice el resultado 3.3 de la nota para encontrar el estadístico de prueba F_0 correspondiente.

   b) ¿Cómo llevaría a cabo la prueba de hipótesis mediante el principio de *suma extra de cuadrados*? Esto es, ¿cuál sería el modelo completo (MC) y cual el reducido (MR)?

9. Sean $U_1, \ldots, U_m$, *v.a.i.i.d.* distribuidas $\mathrm{N}(\mu_U, \sigma^2)$, y sean $V_1, \ldots, V_n$, *v.a.i.i.d.* distribuidas $\mathrm{N}(\mu_V, \sigma^2)$ e independientes de las U's. Derive el estadístico de prueba para $H_0 : \mu_U = \mu_V$. *Sugerencia*: considere el modelo de regresión

$$\begin{bmatrix} U_1 \\ \vdots \\ U_m \\ V_1 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_U \\ \mu_V \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \\ \epsilon_{m+1} \\ \vdots \\ \epsilon_{m+n} \end{bmatrix}$$

10. Recupere la encuesta sobre servicio de televisión por cable que estudio en la Tarea 1. Considere el modelo de regresión lineal múltiple

$$y = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \tag{1}$$

donde y representa la respuesta renta, x_1 el número de niños por casa habitación, x_2 el número d adultos, x_3 el número total de horas que la familia ve televisión, y x_4 el valor catastral de la vivienda en miles de pesos.

   a) Realice la prueba de significancia del modelo de regresión (1), esto es, $H_0 : X\beta = 0$.

   b) Utilice la tabla del análisis de varianza para verificar que el estadístico F_0 obtenido en el inciso anterior está dado por

   $$F_0 = \frac{\mathrm{CM}_{\mathrm{Reg}}}{\mathrm{CM}_{\mathrm{Resid}}}$$

   donde $\mathrm{CM}_{\mathrm{Reg}} = \mathrm{SC}_{\mathrm{Reg}}/4$, y $\mathrm{SC}_{\mathrm{Reg}} = \mathrm{SC}_1 + \cdots + \mathrm{SC}_4$, con SC_{i} la suma de los cuadrados debido al regresor x_i.

   c) Verifique que $F_4 = t_4^2$, donde F_4 es el estadístico F asociado al regresor x_4 en la tabla de ANOVA y $t_4 = \hat{\beta}_4/\mathrm{ee}(\hat{\beta}_4)$, donde ee es el error estándar del estimador.

   d) Explique por qué, para el regresor niños (x_1), los valor-p correspondientes a los estadísticos F_1 y t_1 no coinciden. ¿Qué prueban cada uno de los estadísticos?

   e) Reordene el modelo (1) de manera que cada uno de los regresores quede al final de modelo. Verifique que en ese caso el estadístico F es el cuadrado del estadístico t respectivo.

11. Considere el desarrollo de un proceso en el que se estudiaron cuatro factores o variables: carga de catalizador x_1, temperatura x_2, presión x_3 y concentración de uno de los reactivos x_4. En la tabla se muestra el resultado de los ensayos, la respuesta y -el porcentaje de conversión- en cada una de las 16 condiciones de reacción.

```
n  y x1 x2 x3 x4     n  y x1 x2 x3 x4
1 70 -1 -1 -1 -1     9 60 -1 -1 -1  1
2 60  1 -1 -1 -1    10 49  1 -1 -1  1    Var:  Factor                    +1  -1
3 89 -1  1 -1 -1    11 88 -1  1 -1  1     x1: Carga de catalizador (lb) 10  15
4 81  1  1 -1 -1    12 82  1  1 -1  1     x2: Temperatura (C)          220 240
5 69 -1 -1  1 -1    13 60 -1 -1  1  1     x3: Presión (psi)             50  80
6 62  1 -1  1 -1    14 52  1 -1  1  1     x4: Concentración (%)         10  12
7 88 -1  1  1 -1    15 86 -1  1  1  1
8 81  1  1  1 -1    16 79  1  1  1  1
```

a) Ajuste el modelo lineal
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$
y verifique que, con excepción de la presión, los factores parecen afectar el porcentaje de conversión significativamente.

b) Verifique que los regresores son ortogonales entre ellos $X_i \perp X_j,\ i \neq j$. Esto es,
$$\sum_{\ell=1}^{16} x_{\ell i} x_{\ell j} = 0$$

c) Verifique que en este caso los estadísticos F_i del ANOVA son el cuadrado de los estadísticos $t_i = \hat{\beta}_i / \text{ee}(\hat{\beta}_i)$. Explique por qué.

d) Como consecuencia del inciso anterior, cuando los regresores son ortogonales entre ellos, no importa el orden en que aparecen en el modelo, las sumas de cuadrados correspondientes en el ANOVA no cambian.