Universidad Nacional Autónoma de México
Estadística II
Prof. Jimmy Hernández

1. La idea de este ejercicio es poner en practica la prueba $H_0 : \mathbf{H}\beta = \mathbf{h}$ vs $H_a : \mathbf{H}\beta \neq \mathbf{h}$, vamos a asumir que los datos satisfacen un modelo lineal, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

x_i	550	200	280	340	410	475	160	380	510	510
y_i	200	50	60	140	130	180	20	120	190	160

Encuentre quien es $\hat{\beta}_0$,$\hat{\beta}_1$, $\hat{\sigma}^2$. Suponga que queremos hacer una prueba de hipótesis con $\alpha = 0{,}05$,

$$H_0 : \beta_0 = 0, \beta_1 = 0 \qquad vs \qquad H_a : \beta_0 \neq 0 \text{ o } \beta_1 \neq 0$$

Usando una función monótona del cociente de verosimilitud generalizado

$$W = \left(\frac{n-p}{q}\right) \frac{\hat{\sigma}^2_\omega - \hat{\sigma}^2_\Omega}{\hat{\sigma}^2_\Omega}$$

donde,$\hat{\sigma}^2_\Omega$ es el estimador máximo verosímil de σ^2 en el modelo completo, y $\hat{\sigma}^2_\omega$ es el estimador máximo verosímil de σ^2 en el modelo reducido $\mathbf{Z} = \mathbf{B}\gamma + \varepsilon$ por la hipotesis H_0

$$\hat{\sigma}^2_\Omega = \frac{1}{n}\mathbf{Y'Y} - \hat{\beta}'\mathbf{X'Y}$$
$$\hat{\sigma}^2_\omega = \frac{1}{n}\mathbf{Z'Z} - \hat{\gamma}'\mathbf{B'Z}$$

¿Rechazaría H_0? recuerde que el estadístico $W \sim \mathscr{F}_{(q,n-p)}$

2. La tabla FootballLeague.csv contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976.

   a) Ajuste un modelo lineal múltiple que relaciona el numero de juegos ganados con
   Yardas por aire del equipo (x_2)
   El porcentaje de Yardas por Tierra (x_7)

   $$y_i = \beta_0 + \beta_{i1}x_2 + \beta_{i2}x_7 + \beta_{i3}x_8 + \varepsilon_i \tag{1}$$

   Las Yardas por tierra del contrario (x_8)

   b) Constrasta la hipótesis $H_0 : \beta_1 = \beta_3 = 0$ vs $H_a : \beta_1 \neq 0$ o $\beta_3 \neq 0$

   c) Valide el modelo
   - Normalidad de los residuales (Q-Q Plot, Histogramas)
   - Homocedasticidad de los residuales (Gráfi
     co, Prueba Levene,Prueba Barttlet)
   - Independencia de los Residuales (Durbin-Watson, Prueba de Rachas)

   d) Encuentre los intervalos al 98 % de confianza para cada uno de los parámetros β_0, β_1, β_2 y β_3

e) Encuentre el intervalo al 92 % de confianza para la respuesta media de numero juegos ganados cuando $x_2 = 2300$, $x_7 = 56$ y $x_8 = 2100$

3. Suponga que se tiene el modelo lineal con intercepto:

$$y_i = \beta_0 + \beta_{i1}x_1 + ... + \beta_{ik}x_k + \varepsilon_i \tag{2}$$

El cual al ser escrito en forma matricial toma la forma:

$$Y = \mathbf{X}\beta + \epsilon \qquad \varepsilon_i \sim N_n(0, \sigma^2 I)$$

Suponiendo que el vector de parámetros

y la matriz de diseño es particionados de la siguiente forma:

$$\underline{\beta} = \begin{pmatrix} \underline{\beta_1} \\ \underline{\beta_2} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}_{n \times p}$$

Donde

$\underline{\beta}_1$ es un vector con $p\times$ entradas y $\underline{\beta}_2$ es un vector con r ($r < p$) entradas mientras que la partición de la matriz $\mathbf{X}$ hace que$\mathbf{X}_1$ sea una sub-matriz de $n(p \times r)$ mientras que $\mathbf{X}_2$ es una sub-matriz de $n \times r$. Con esto el modelo lineal general se expreso como:

$$\mathbf{Y} = \mathbf{X_1}\underline{\beta_1} + \mathbf{X_2}\underline{\beta_2} + \varepsilon$$

El objetivo de este ejercicio es encontrar la estadística de prueba para el contraste:

$$H_0 : \underline{\beta_2} = 0 \qquad vs \qquad H_a : \underline{\beta_2} \neq 0$$

Lo cual es equivalente en terminos del modelo ajustado a:

$$H_a : \mathbf{Y} = \mathbf{X_1}\underline{\beta_1} + \varepsilon \qquad vs \qquad H_1 : \mathbf{Y} = \mathbf{X_1}\underline{\beta_1} + \mathbf{X_2}\underline{\beta_2} + \varepsilon$$

Con lo anterior defina las matrices sombreros del modelo reducido y completo como:

$$\mathbf{H}_{H_0} = \mathbf{X_1}(\mathbf{X_1'}\mathbf{X_1})^{-\mathbf{1}}\mathbf{X_1'} \qquad vs \qquad \mathbf{H}_{H_1} = \mathbf{X}(\mathbf{X'}\mathbf{X})^{-\mathbf{1}}\mathbf{X'}$$

- Demuestre que $\mathbf{H}_{H_1} - \mathbf{H}_{H_0}$ es idempotente y encuentre su rango
- Demuestre que bajo H_0
$$\frac{SCH_{H_0} - SCE_{H_1}}{\sigma^2} \sim \chi^2_r$$
  Donde SCE_{H_0} representa la suma de cuadrados del error bajo el modelo reducido, y SCE_{H_1} la suma de cuadrados del error bajo el modelo completo.
- Demuestre que $SCH_{H_0} - SCE_{H_1}$ y SCE_{H_1} son independientes ($Hint$: Utilizar el teorema para independencia de formas cuadráticas)
- Demuestre entonces que bajo H_0
$$\frac{SCH_{H_0} - SCE_{H_1}}{r\sigma^2} \sim \mathscr{F}_{(r, n-p)}$$