Taller #3
Econometría 06169
Grupo 3

Profesor: Julio César Alonso

Nota: Recuerden que únicamente dos preguntas seleccionadas al azar serán calificadas. Admás recuerden que en todas las preguntas ustedes deben mostrar todo su trabajo.

1. En el taller pasado estimamos por MCO el el siguiente model $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ usando las siguientes observaciones recogidas por un investigador.

$$y := \begin{pmatrix} 3 \\ 2 \\ 8 \\ 3 \\ 5 \end{pmatrix} \qquad X := \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix}$$

Ahora, (no neceistan hallar de nuevo los estimadores, pueden usar los resultados encontrados anteriormente)

a) Contraste la hipótesis nula $H_o: \beta_2 + \beta_3 = 0$. Explique las implicaciones de su decisión

b) Construya la Tabla ANOVA para el modelo estimado.

2. Usando la misma información de la pregunta 1,

a) Calcule el R^2 del modelo estimado. Explique en sus propias palabras el significado de su respuesta numérica.

b) Contraste la hipótesis nula $H_o: \beta_2 = \beta_3 = 0$. Explique las implicaciones de su decisión.

3. El jefe de la división de planeación del departamento de mercadeo de una cádena de almacenes espera que las ventas diarias en millones de pesos, Y_t, tengan un comportamiento descrito por el siguiente modelo estadístico: $Y_t = \beta_1 X_{1t} + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + \varepsilon_t$, donde X_{1t} representa el número de reclamos en el dia t, X_{2t} representa el número de ofertas en el día t y X_{3t} representa el número de anuncios en el periodico en el dia t. ε_t representa una variable estocastica que esta normalemte e idependientemente distribuida con media cero y varianza constante.

Los economistas de la división de planeación ya han recogido 400 observaciones y han efectuado los siguientes cáculos:

$$Y^T Y = 47797 \qquad Y^T X = (5200 \;\; 13600 \;\; 22800) \qquad X^T X = \begin{pmatrix} 800 & 1600 & 2400 \\ 1600 & 4000 & 6400 \\ 2400 & 6400 & 11200 \end{pmatrix}$$

a) Calcule los mejores estimadores lineales insesgados del vector $\beta = (\beta_1 \ \ \beta_2 \ \ \beta_3)^T$.

b) Estime σ^2 y la matriz de varianzas y covarianzas de los estimadores de β.

4. Usando la misma información de la pregunta 3, considere las siguientes preguntas:

a) ¿Es β_2 significativo? Explique las implicaciones de su respuesta.

b) En la última reunión del departamento de planeación uno de los asesores afirmaba: "Si logramos que en un día no existán reclamos, hacemos 3 promociones y publicamos un anuncio de prensa, entonces nuestros ingresos diarios serán de 6.5 millones". Pruebe la hipotesis de dicho asesor y explique sus conclusiones.

c) El departamento de finanzas de esta cadena de almacenes desea conocer un rango para el valor promedio de las ventas diarias para poder efectuar la programación financiera del próximo trimestre. Si sabemos que existirán 5 ofertas diarias, 40 anuncios de prensa diarios y ningún reclamo, calcule el rango requerido por el departamento de finanzas (con un 95% de confianza).

Respuestas Sugeridas
Taller #3
Econometría 06169
Grupo 3

Profesor: Julio César Alonso

Nota: Recuerden que únicamente dos preguntas seleccionadas al azar serán calificadas.
Admás recuerden que en todas las preguntas ustedes deben mostrar todo su trabajo.

1. En el taller pasado estimamos por MCO el el siguiente model $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ usando las siguientes observaciones recogidas por un investigador.

$$y := \begin{pmatrix} 3 \\ 2 \\ 8 \\ 3 \\ 5 \end{pmatrix} \qquad X := \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{pmatrix}$$

Ahora, (no neceistan hallar de nuevo los estimadores, pueden usar los resultados encontrados anteriormente)

a) Contraste la hipótesis nula $H_o: \beta_2 + \beta_3 = 0$. Explique las implicaciones de su decisión

Esta hipótesis nula puede ser escrita en la forma $R \cdot \beta = C$. Como sólo existen una restricción y tenemos tres parametros estimados, entonces la matriz R tendrá dimensiones 1x3. En este caso $R = (0 \;\; 1 \;\; 1)$ y $C = 0$. Así, el Fcalculado esta dado por

$$F_{calculado} = \frac{\dfrac{((C - R\cdot \beta hat))^T \cdot \left[R \cdot \left((X)^T \cdot X\right)^{-1} \cdot R^T\right]^{-1} \cdot (C - R\cdot \beta hat)}{r}}{s^2}$$

Del taller anterior sabemos que:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{pmatrix} \frac{16}{5} \\ 2 \\ -1 \end{pmatrix} \qquad \left(X^T X\right)^{-1} = \begin{pmatrix} \frac{267}{10} & \frac{9}{2} & -8 \\ \frac{9}{2} & 1 & -\frac{3}{2} \\ -8 & -\frac{3}{2} & \frac{5}{2} \end{pmatrix} \qquad s^2 = \frac{7}{5}$$

Entonces tenemos que

$$F_{calculado} = \frac{\dfrac{\left[(0\ \ 1\ \ 1)\cdot\begin{pmatrix}\frac{16}{5}\\2\\-1\end{pmatrix}\right]^{T}\cdot\left[(0\ \ 1\ \ 1)\cdot\begin{pmatrix}\frac{267}{10} & \frac{9}{2} & -8\\ \frac{9}{2} & 1 & -\frac{3}{2}\\ -8 & -\frac{3}{2} & \frac{5}{2}\end{pmatrix}\cdot\begin{pmatrix}0\\1\\1\end{pmatrix}\right]^{-1}\cdot\left[(0\ \ 1\ \ 1)\cdot\begin{pmatrix}\frac{16}{5}\\2\\-1\end{pmatrix}\right]}{1}}{\frac{7}{5}}$$

$$F_{calculado} = \frac{(1)^{T}\cdot\left[(0\ \ 1\ \ 1)\cdot\begin{pmatrix}\frac{267}{10} & \frac{9}{2} & -8\\ \frac{9}{2} & 1 & -\frac{3}{2}\\ -8 & -\frac{3}{2} & \frac{5}{2}\end{pmatrix}\cdot\begin{pmatrix}0\\1\\1\end{pmatrix}\right]^{-1}\cdot 1}{\frac{7}{5}} = \frac{\left(\frac{1}{2}\right)^{-1}}{\frac{7}{5}} = \frac{10}{7}$$

Este $F_{calculado}$ lo tenemos que comparar con el $F_{0.05,(1,2)} = 18.513$. Claramente el $F_{calculado}$ es menor que el F de la tabla y por tanto no hay suficiente evidencia para rechazar la hipótesis nula que $\beta_2 + \beta_3 = 0$. Noten que este resulta implica que $\beta_2 = -\beta_3$, así nuestra conclusión implica que las pendientes asociadas con X_{2i} y X_{3i} son iguales en valores absolutos.

b) Construya la Tabla ANOVA para el modelo estimado.

Sabemos que la tabla ANOVA viene dada por:

Tabla ANOVA

Fuente de variación	**SS**	**G de L**	**MS**
Regresión	$SSR = \hat{\beta}^T X^T y - n\overline{Y}^2$	$k - 1$	$\frac{SSR}{k-1}$
Error (Residuos)	$SSE = y^T y - \hat{\beta}^T X^T y$	$n - k$	$\frac{SSE}{n-k}$
Total	$y^T y - n\overline{Y}^2$	$n - 1$	

En este caso tenemos que $k := 3,\ n := 5,\ y^T y = (111)$, $\beta hat \cdot X^T y = \frac{541}{5}$ y $Ybar = \frac{21}{5}$. Asi, la tabla Anova es

Tabla ANOVA

Fuente de variación	**SS**	**G de L**	**MS**
Regresión	$\frac{541}{5} - \left[5\cdot\left(\frac{21}{5}\right)^2\right] = 20$	$3 - 1 = 2$	$\frac{20}{2} = 10$
Error (Residuos)	$111 - \frac{541}{5} = \frac{14}{5} = 2.8$	$5 - 3 = 2$	$\frac{\frac{14}{5}}{2} = \frac{7}{5}$
Total	$20 + \frac{14}{5} = \frac{114}{5} = 22.8$	4	

2. Usando la misma información de la pregunta 1,

a) Calcule el R^2 del modelo estimado. Explique en sus propias palabras el significado de su respuesta numérica.

Sabemos que $R^2 = \frac{SSR}{SST}$. Entonces, en este caso tenemos que $R^2 = \frac{20}{\frac{114}{5}} = \frac{50}{57} = 0.877$. Esto implica que el 87.7% de la variación de Y es explicada por nuestro modelo.

b) Contraste la hipótesis nula $H_o: \beta_2 = \beta_3 = 0$. Explique las implicaciones de su decisión.

Existen varias formas de hacer esta pregunta. Una de las formas es usando la información de la tabla ANOVA. Sabemos que el F calculado para pobar dicha Ho esta dado por

$$F_{calculado} = \frac{MSR}{MSE} = \frac{10}{\frac{7}{5}} = \frac{50}{7} = 7.143$$

El F de la tabla par este caso es $F_{0.05,(2,2)} = 19$. Como el $F_{calculado} < F_{0.05,(2,2)}$, entonces no hay suficiente evicencia para rechazar la Ho. Así podemos decir que un modelo con sólo un intercepto explicaría mejor la variabilidad de la variable Y.

3. El jefe de la división de planeación del departamento de mercadeo de una cádena de almacenes espera que las ventas diarias en millones de pesos, Y_t, tengan un comportamiento descrito por el siguiente modelo estadístico: $Y_t = \beta_1 X_{1t} + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + \varepsilon_t$, donde X_{1t} representa el número de reclamos en el dia t, X_{2t} representa el número de ofertas en el día t y X_{3t} representa el número de anuncios en el periodico en el dia t. ε_t representa una variable estocastica que esta normalemte e idependientemente distribuida con media cero y varianza constante.

Los economistas de la división de planeación ya han recogido 400 observaciones y han efectuado los siguientes cáculos:

$$Y^T Y = 47797 \qquad Y^T X = (5200 \quad 13600 \quad 22800) \qquad X^T X = \begin{pmatrix} 800 & 1600 & 2400 \\ 1600 & 4000 & 6400 \\ 2400 & 6400 & 11200 \end{pmatrix}$$

a) Calcule los mejores estimadores lineales insesgados del vector $\beta = (\beta_1 \quad \beta_2 \quad \beta_3)^T$.

Sabemos que $\hat{\beta} = (X^T X)^{-1} X^T Y$ es MELI si se sumplen los 3 supuestos del teorema de Gauss-Markov. Así, necesitamos encontrar $(X^T X)^{-1}$. Noten que

$$X^T X = \begin{pmatrix} 800 & 1600 & 2400 \\ 1600 & 4000 & 6400 \\ 2400 & 6400 & 11200 \end{pmatrix} = 800 \cdot \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 8 & 14 \end{pmatrix}$$

Entonces,

$$\left|X^T X\right| = 800^3 \cdot \left|\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 8 & 14 \end{pmatrix}\right|$$

Recuerden: $|a \cdot Z| = a^k \cdot |Z|$, donde a es un escalar y Z es una matriz cuadrada de kXk.

Como $\left|\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 8 & 14 \end{pmatrix}\right| = 1$, entonces $\left|X^T X\right| = 800^3 = 2^9 \cdot 10^6$

La matriz de cofactores esta dada por

$$800^2 \cdot \begin{bmatrix} 5\cdot 14 - 8\cdot 8 & (2\cdot 14 - 8\cdot 3)\cdot(-1)^{2+1} & (2\cdot 8 - 5\cdot 3)\cdot(-1)^{1+3} \\ (2\cdot 14 - 3\cdot 8)\cdot(-1)^{2+1} & (1\cdot 14 - 3\cdot 3)\cdot(-1)^{2+2} & (1\cdot 8 - 2\cdot 3)\cdot(-1)^{3+2} \\ (2\cdot 8 - 5\cdot 3)\cdot(-1)^{1+3} & (1\cdot 8 - 2\cdot 3)\cdot(-1)^{3+2} & (1\cdot 5 - 2\cdot 2)\cdot(-1)^{3+3} \end{bmatrix} = 800^2 \cdot \begin{pmatrix} 6 & -4 & 1 \\ -4 & 5 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

Entonces

$$\left(X^T X\right)^{-1} = \frac{1}{800^3} \cdot \left[800^2 \cdot \begin{pmatrix} 6 & -4 & 1 \\ -4 & 5 & -2 \\ 1 & -2 & 1 \end{pmatrix}\right] = \frac{1}{800} \cdot \begin{pmatrix} 6 & -4 & 1 \\ -4 & 5 & -2 \\ 1 & -2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{400} & \frac{-1}{200} & \frac{1}{800} \\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400} \\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800} \end{pmatrix}$$

Ahora sí podemos encontrar los estimadores para $\beta = \left(\beta_1 \;\; \beta_2 \;\; \beta_3\right)^T$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \left(X^T X\right)^{-1} X^T y = \begin{pmatrix} \frac{3}{400} & \frac{-1}{200} & \frac{1}{800} \\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400} \\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800} \end{pmatrix} \cdot (5200 \;\; 13600 \;\; 22800)^T = \begin{pmatrix} -\frac{1}{2} \\ 2 \\ 1 \end{pmatrix}$$

b) Estime σ^2 y la matriz de varianzas y covarianzas de los estimadores de β.

Para encontrar el estimador para σ^2 empleamos la siguiente fórmula.

$$s^2 = \frac{y^T y - \beta hat^T \cdot X^T y}{n - k} = \frac{47797 - \left[\begin{pmatrix} -\frac{1}{2} & 2 & 1 \end{pmatrix} \cdot (5200 \quad 13600 \quad 22800)^T\right]}{400 - 3}$$

$$s^2 = \frac{47797 - (47400)}{400 - 3} = \frac{397}{397} = 1$$

Entonces la matriz estimada de varianzas y covarianzas de los estimadores de β esta dada por

$$\widehat{Var(\hat{\beta})} = s^2 (X^T X)^{-1} = 1 \cdot \begin{pmatrix} \frac{3}{400} & \frac{-1}{200} & \frac{1}{800} \\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400} \\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800} \end{pmatrix} = \begin{pmatrix} \frac{3}{400} & \frac{-1}{200} & \frac{1}{800} \\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400} \\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800} \end{pmatrix}$$

4. Usando la misma información de la pregunta 3, considere las siguientes preguntas:

c) ¿Es β_2 significativo? Explique las implicaciones de su respuesta.

En este caso debemos realizar la siguiente prueba de hipótesis: $H_o: \beta_2 = 0$ versus $H_A: \beta_2 \neq 0$, como sabemos el tcalculado esta dado por $t_c = \frac{\beta hat_2}{s_{(\beta hat_2)}}$. En este caso tenemos

$$t_c = \frac{2}{\sqrt{\frac{1}{160}}} = \frac{2}{\sqrt{\frac{1}{4^2 \cdot 10}}} = 8 \cdot \sqrt{10}$$

$\boxed{\sqrt{10} = 3.162}$

Este t_c lo tenemos que comparar con el $t_{0.025,397} = 2.588$ de la tabla (Aquí ustedes podían escoger cualquier nivel de significancia y la decisión sería la misma). Note que $|t_c| = 8\cdot\sqrt{10} > t_{0.025,397} = 3.588$, por lo tanto hay suficiente evidencia para decir que el coeficiente asociado con el regresor X_3 no es igual a cero. En otras palabras, podemos decir con un 95% de confianza que X_2 ayuda ha explicar la variabilidad en la variable dependiente.

b) En la última reunión del departamento de planeación uno de los asesores afirmaba: "Si logramos que en un día no existán reclamos, hacemos 3 promociones y publicamos un anuncio de prensa, entonces nuestros ingresos diarios serán de 6.5 millones". Pruebe la hipotesis de dicho asesor y explique sus conclusiones.

Esta afirmación es equivalente a probar la siguiente $H_o: \beta_1\cdot 0 + \beta_2\cdot 3 + \beta_3\cdot 1 = 6.5$. Como vimos en clase, esta hipótesis la podemos reescribir en la forma $R\cdot\beta = C$. En este caso únicamentetenemos una rtestricción y por tantola martriz R tendrá dimensiones 1x3. En este caso tenemos $R := (0\ \ 3\ \ 1)$ y $C := 7.5$. Asi el F calculado esta dado por:

$$F_{calculado} = \frac{\dfrac{((C-R\cdot\beta hat))^T\cdot\left[R\cdot\left((X)^T\cdot X\right)^{-1}\cdot R^T\right]^{-1}\cdot(C-R\cdot\beta hat)}{r}}{s^2}$$

Del taller anterior sabemos que:

$$\hat{\beta} = \begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{pmatrix}-\frac{1}{2}\\ 2\\ 1\end{pmatrix} \qquad \left(X^T X\right)^{-1} = \begin{pmatrix}\frac{3}{400} & \frac{-1}{200} & \frac{1}{800}\\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400}\\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800}\end{pmatrix} \qquad s^2 = 1$$

Entonces tenemos que

$$F_{calculado} = \frac{\dfrac{\left[6.5-\left[(0\ \ 3\ \ 1)\cdot\begin{pmatrix}-\frac{1}{2}\\ 2\\ 1\end{pmatrix}\right]\right]^T\cdot\left[(0\ \ 3\ \ 1)\cdot\begin{pmatrix}\frac{3}{400} & \frac{-1}{200} & \frac{1}{800}\\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400}\\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800}\end{pmatrix}\cdot\begin{pmatrix}0\\ 3\\ 1\end{pmatrix}\right]^{-1}\cdot\left[6.5-\left[(0\ \ 3\ \ 1)\cdot\begin{pmatrix}-\frac{1}{2}\\ 2\\ 1\end{pmatrix}\right]\right]}{1}}{1}$$

$$F_c = [\,6.5 \cdot -(7)\,]^T \cdot \left[(0\ \ 3\ \ 1) \cdot \begin{pmatrix} \frac{3}{400} & \frac{-1}{200} & \frac{1}{800} \\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400} \\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \right]^{-1} \cdot [\,6.5 \cdot -(7)\,] = \frac{-1}{2} \cdot \left(\frac{17}{400} \right)^{-1} \cdot \frac{-1}{2} = \frac{100}{17} = 5.$$

Este $F_{calculado}$ lo tenemos que comparar con el $F_{0.05,(1,397)} = 3.865$. Claramente el $F_{calculado}$ es mayor que el F de la tabla y por tanto hay suficiente evidencia para rechazar la hipótesis nula que $\beta_1 \cdot 0 + \beta_2 \cdot 3 + \beta_3 \cdot 1 = 6.5$. Noten que este resulta implica que el asesor no tenia la razón.

c) El departamento de finanzas de esta cadena de almacenes desea conocer un rango para el valor promedio de las ventas diarias para poder efectuar la programación financiera del próximo trimestre. Si sabemos que existirán 5 ofertas diarias, 40 anuncios de prensa diarios y ningún reclamo, calcule el rango requerido por el departamento de finanzas (con un 95% de confianza).

El primer paso es encontrar la predicción para el valor promedio. Sabemos que dicha prediccción esta dada por

$$\hat{y}_p = x_p^T \hat{\beta}$$

Donde x_p es el vector de valores de las variables independienetes para los cuales queremos la predicción. Asi nuestra predicción para $x_p^T = (0\ \ 5\ \ 40)$ es

$$(0\ \ 5\ \ 40) \cdot \begin{pmatrix} -\frac{1}{2} \\ 2 \\ 1 \end{pmatrix} = (50)$$

Además sabemos que la varianza de la predicción media esta dada por

$$Var\left(\hat{y}_p\right) = \sigma^2 x_p^T \left(X^T X\right)^{-1} x_p$$

Es decir,

$$1\cdot\left[(0\quad 5\quad 40)\cdot\begin{pmatrix}\frac{3}{400} & \frac{-1}{200} & \frac{1}{800}\\ \frac{-1}{200} & \frac{1}{160} & \frac{-1}{400}\\ \frac{1}{800} & \frac{-1}{400} & \frac{1}{800}\end{pmatrix}\cdot\begin{pmatrix}0\\5\\40\end{pmatrix}\right] = \frac{37}{32}$$

$$S^2_{\hat{y}_p} = \frac{37}{32}$$

Así el un intervalo del 95% de confianza para nuestra predicción esta dada por

$$\hat{y}_p \pm t_{\frac{\alpha}{2},n-k}\sqrt{S^2_{\hat{y}_p}} \qquad \text{con } t_{0.025,\,397} = 1.966$$

$$\left(50 - 1.966\cdot\sqrt{\frac{37}{32}},\, 50 + 1.966\cdot\sqrt{\frac{37}{32}}\right)$$

$$\left[50 - 1.966\cdot\left(\frac{1}{8}\cdot\sqrt{74}\right),\, 50 + 1.966\cdot\left(\frac{1}{8}\cdot\sqrt{74}\right)\right]$$

Nota: Sino tenían una calculadora a la mano, esto era suficiente

$$(47.886\,,52.114)$$

Entonces, el un intervalo del 95% de confianza para nuestra predicción es $(47.886\,,52.114)$.