Taller 3: Regresión Lineal Múltiple
Econometría 06216

17-01-2011

Profesores: Julio César Alonso.

Monitoras: Sasha Magyaroff - Carolina Restrepo.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller puede subirse en la plataforma de Moodle hasta las 7:10 del 7 de febrero de 2011. **Sólo se recibirán talleres en formato pdf**. Cualquier otro formato no será calificado

Instrucciones:

- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo grupal. Sólo se admiten grupos de dos personas, y por lo tanto debe reflejar tan sólo el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.
- Este taller debe ser escrito en computador.

Pregunta 1

Una vendedora de casas está interesada en saber, en cuánto puede vender una determinada casa si elige venderla. Algunos analistas creen que el valor de las casas está determinado por las siguientes variables. Sea V el valor de una casa; VM el valor promedio de la demás casas del vecindario; L el tamaño del lote donde fue construida la casa, y H el tamaño de la casa.

1.1. Escriba el modelo que la vendedora de casas debe estimar para conocer el precio de las mismas. Interprete los coeficientes a priori y diga su signo esperado.

1.2. Estime el modelo del literal anterior. Haga los cálculos matriciales en Excel y muestre claramente todo el procedimiento y resultados parciales, esto es, muestre claramente la matriz X^TX y X^Ty (no emplee EasyReg).

Pregunta 2

Siguiendo con el ejercicio anterior:

2.1 Interprete los coeficientes estimados.

2.2 Estime σ^2, y la matriz de varianzas y covarianzas. Haga los cálculos en Excel y muestre todo su procedimiento.

Pregunta 3

Con la misma información de los dos puntos anteriores:
3.1 Suponga que se quiere predecir el valor de una casa usando la predicción del modelo que se expresó en 1.1. Si el error de predicción se define de la siguiente forma: $f = V - \textit{la predicción}$, Cuál es la media poblacional y la varianza poblacional de la predicción del error?

3.2 Halle la significancia individual de los coeficientes estimados.Interprete los coeficientes estimados teniendo en cuenta su significancia.

Pregunta 4

4.1 Si V y VH están medidos en dólares, interprete el significado de β_1

4.2 Si V está medido en dólares y VH está medido en millones de dólares, interprete el significado de β_1

4.3 Cuál es la elasticidad de V con respecto a L? Si L aumenta en uno por ciento, qué porcentaje se espera que cambie V?

Pregunta 5

5.1 Construya la tabla ANOVA. Halle el R^2 y el R^2 ajustado. Comente sus resultados.

Pregunta 6

Realice una prueba de significancia global, muestre las hipótesis, el procedimiento para calcular el estadístico y concluya.

Taller 3: Regresión Lineal Múltiple
Econometría 06216
Respuestas sugeridas

17-01-2011

Profesores: Julio César Alonso.

Monitoras: Sasha Magyaroff - Carolina Restrepo.

Notas:

- Recuerde que únicamente tres preguntas, seleccionadas al azar, serán calificadas.
- Este taller puede subirse en la plataforma de Moodle hasta las 7:30 del 7 de febrero de 2011. **Sólo se recibirán talleres en formato pdf**. Cualquier otro formato no será calificado

Instrucciones:

- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.
- Este taller es un trabajo grupal. Sólo se admiten grupos de dos personas, y por lo tanto debe reflejar tan sólo el trabajo de la pareja.
- Si bien no es necesario reportar todos los números decimales, sí lo es hacer los cálculos con todos ellos.
- Este taller debe ser escrito en computador.

Pregunta 1

Una vendedora de casas está interesada en saber, en cuánto puede vender una determinada casa si elige venderla. Algunos analistas creen que el valor de las casas está determinado por las siguientes variables. Sea V el valor de una casa; VM el valor promedio de la demás casas del vecindario; L el tamaño del lote donde fue construida la casa, y H el tamaño de la casa.
1.1. Escriba el modelo que la vendedora de casas debe estimar para conocer el precio de las mismas. Interprete los coeficientes a priori y diga su signo esperado.

El modelo que la vendedora debe estimar es:

$$V_i = \beta_0 + \beta_1 VM_i + \beta_2 L_i + \beta_3 H_i + \varepsilon_i$$

Interpretación a priori de los coeficientes:
β_0 : Valor de la casa que no depende ni del valor promedio de las demas casas del vecindario, ni del tamaño del lote donde fue construida la casa, ni del tamaño de la casa. Se espera que tenga signo positivo.
β_1 : Ante un aumento de una unidad en el valor promedio de las demás casas del vecindario, se espera que el valor de una de las casas varie en β_1 unidades. Se espera que tenga signo positivo.

β_2 : Ante un aumento de una unidad en el tamaño del lote donde fue construída la casa, se espera que el valor de dicha casa varie en β_2 unidades. Se espera que tengo signo positivo.
β_3 : Ante un aumento de una unidad en el tamaño de la casa, se espera que su valor varie en β_3 unidades. Se espera que tenga signo positivo.
1.2. Estime el modelo del literal anterior. Haga los cálculos matriciales en Excel y muestre claramente todo el procedimiento y resultados parciales, esto es, muestre claramente la matriz X^TX y X^Ty (no emplee EasyReg).

$$X^TX = \begin{bmatrix} 10 & 293.4 & 2625 & 2130 \\ 293.4 & 8724.36 & 77956.5 & 63445 \\ 2625 & 77956.5 & 720525 & 585450 \\ 2130 & 63445 & 585450 & 492100 \end{bmatrix}$$

$$X^Ty = \begin{bmatrix} 286.8 \\ 8506.75 \\ 77501.5 \\ 63197 \end{bmatrix}$$

$$\widehat{\beta} = \left(X^TX\right)^{-1} X^Ty$$

De acuerdo con lo anterior, es necesario encontrar la inversa de X^TX, esta es:

$$\left(X^TX\right)^{-1} = \begin{bmatrix} 7.68437855 & -0.24992781 & -0.00330757 & 0.00292407 \\ -0.2499278 & 0.011585449 & -0.000248067 & -0.00011677 \\ -0.00333076 & -0.00024807 & 0.000079826 & -0.00004857 \\ 0.00292407 & -0.00011677 & -0.000048569 & 0.000062214 \end{bmatrix}$$

Finalmente, se encuentra:

$$\widehat{\beta} = \begin{bmatrix} 4.46021635 \\ 0.27024306 \\ 0.05168914 \\ 0.01278147 \end{bmatrix}$$

Pregunta 2

Siguiendo con el ejercicio anterior:

2.1 Interprete los coeficientes estimados.
β_0 : El valor de la casa que no depende ni del valor promedio de las demas casas del vecindario, ni del tamaño del lote donde fue construida la casa, ni del tamaño de la casa, es 4.4602 unidades.
β_1 : Ante un aumento de una unidad en el valor promedio de las demás casas del vecindario, el valor de una de las casas aumenta en 0.2702 unidades.
β_2 : Ante un aumento de una unidad en el tamaño del lote donde fue construída la casa, el valor de dicha casa aumenta en 0.517 unidades.
β_3 : Ante un aumento de una unidad en el tamaño de la casa, su valor aumenta en 0.0127 unidades.

2.2 Estime σ^2, y la matriz de varianzas y covarianzas. Haga los cálculos en Excel y muestre todo su procedimiento.
Para encontrar σ^2 se utiliza la fórmula: $s^2 = \frac{Y^TY - \widehat{\beta}^T X^TY}{n-k}$. Por tanto son necesarios los siguientes valores:

$$Y^TY = 8394.06$$

$$\widehat{\beta}^T X^TY = \begin{bmatrix} 4.460216 & 0.270243 & 0.051689 & 0.012781 \end{bmatrix} \begin{bmatrix} 286.8 \\ 8506.75 \\ 77501.5 \\ 63197 \end{bmatrix} = 8391.816618$$

$$n-k = 10-4 = 6$$

Finalmente, se encuentra: $s^2 = 0.373897$
A continuación, para encontrar la matriz de varianzas y covarianzas se emplea la siguiente fórmula:

$$s^2\left(X^TX\right)^{-1} = \begin{bmatrix} Var\left(\widehat{\beta}_o\right) & Cov\left(\widehat{\beta}_o, \widehat{\beta}_1\right) & Cov\left(\widehat{\beta}_o, \widehat{\beta}_2\right) \\ & Var\left(\widehat{\beta}_1\right) & Cov\left(\widehat{\beta}_1, \widehat{\beta}_2\right) \\ & & Var\left(\widehat{\beta}_2\right) \end{bmatrix}$$

De donde se obtiene que la matriz de varianzas y covarianzas es:

$$\begin{bmatrix} 2.873166 & -0.093447 & -0.001245 & 0.001093 \\ -0.093447 & 0.004332 & -0.000093 & -0.000044 \\ -0.001245 & -0.000093 & 0.000029 & -0.000018 \\ 0.001093 & -0.000044 & -0.000018 & 0.000023 \end{bmatrix}$$

Pregunta 3

Con la misma información de los dos puntos anteriores:
3.1 Suponga que se quiere predecir el valor de una casa usando la predicción del modelo que se expresó en 1.1. Si el error de predicción se define de la siguiente forma: $f = V - la\ predicción$, Cuál es la media poblacional y la varianza poblacional de la predicción del error?
Sea la predicción:

$$\beta_0 + \beta_1 VM + \beta_2 L + \beta_3 H$$

Además se tiene que el modelo 1.1 está dado por:

$$V = \beta_0 + \beta_1 VM + \beta_2 L + \beta_3 H + \varepsilon$$

Por tanto, el error estimado corresponde a:

$$f = \beta_0 + \beta_1 VM + \beta_2 L + \beta_3 H + \varepsilon - (\beta_0 + \beta_1 VM + \beta_2 L + \beta_3 H)$$
$$f = \varepsilon$$

Por lo tanto, la media poblacional del error estimado es igual a:

$$E(f) = E(\varepsilon) = 0$$

Además la varianza poblacional del error estimado corresponde a:

$$Var(f) = Var(\varepsilon) = \sigma^2$$

3.2 Halle la significancia individual de los coeficientes estimados.Interprete los coeficientes estimados teniendo en cuenta su significancia.
Para encontrar la significancia individual de los coeficientes estimados es necesario calcular, en primer lugar, los

respectivos estadísticos t utilizando la fórmula: $t = \frac{\widehat{\beta}_i - c}{s_{\widehat{\beta}_i}}$ En este caso se obtiene:

$$t = \frac{\widehat{\beta}_0 - 0}{s_{\widehat{\beta}_0}} \rightarrow t = \frac{4.46021635}{1.69504177} \rightarrow t = 2.631331$$

$$t = \frac{\widehat{\beta}_1 - 0}{s_{\widehat{\beta}_1}} \rightarrow t = \frac{0.27024306}{0.06581615} \rightarrow t = 4.10603$$

$$t = \frac{\widehat{\beta}_2 - 0}{s_{\widehat{\beta}_2}} \rightarrow t = \frac{0.05168914}{0.00546322} \rightarrow t = 9.461291$$

$$t = \frac{\widehat{\beta}_3 - 0}{s_{\widehat{\beta}_3}} \rightarrow t = \frac{0.01278147}{0.00482301} \rightarrow t = 2.6501$$

En segundo lugar, se deben contrastar las siguientes hipótesis: $\begin{array}{l} H_o : \beta_i = 0 \\ H_A : \beta_i \neq 0 \end{array}$

Ahora, para dicho fin, con n-k grados de libertad, que en este caso n=10 y k=4 se encuentran los t críticos:

$$\alpha = 0.01 \rightarrow t_c = 3.7074$$
$$\alpha = 0.05 \rightarrow t_c = 2.4469$$
$$\alpha = 0.1 \rightarrow t_c = 1.9432$$

Al hacer el contraste se encuentra que $\widehat{\beta}_0 \wedge \widehat{\beta}_3$ son significativos al 95% de confianza; mientras que $\widehat{\beta}_1 \wedge \widehat{\beta}_2$ lo son al 1% de significancia.
De acuerdo con lo anterior, es posible hacer las siguientes interpretaciones:

β_0 : Con un 95% de confianza es posible inferir que el valor de la casa que no depende ni del valor promedio de las demas casas del vecindario, ni del tamaño del lote donde fue construida la casa, ni del tamaño de la casa, es 4.4602 unidades.
β_1 : Ante un aumento de una unidad en el valor promedio de las demás casas del vecindario, el valor de una de las casas aumenta en 0.2702 unidades; esto con un 99% de confianza.
β_2 : Ante un aumento de una unidad en el tamaño del lote donde fue construída la casa, el valor de dicha casa aumenta en 0.517 unidades; esto con un 99% de confianza.
β_3 : Ante un aumento de una unidad en el tamaño de la casa, su valor aumenta en 0.0127 unidades; esto con un 95% de confianza.

Pregunta 4

4.1 Si V y VH están medidos en dólares, interprete el significado de β_1
β_1 : Ante un aumento de un dólar en el valor promedio de las demás casas del vecindario, se espera que el valor de una de las casas cambie en β_1 dólares. Esto es, el cambio esperado en dólares en el valor de una de las casas, cuando aumenta en un dólar VM.
4.2 Si V está medido en dólares y VH está medido en millones de dólares, interprete el significado de β_1
β_1 : Ante un aumento de un millón de dólares en el valor promedio de las demás casas del vecindario, se espera que el valor de una de las casas cambie en β_1 dólares. Esto es, el cambio esperado en dólares en el valor de una de las casas, cuando VM aumenta en un millón de dólares.

Pregunta 5

5.1 Construya la tabla ANOVA. Halle el R^2. Comente sus resultados.

Fuente de la variación	SS	Grados de libertad	MS
Regresión	SSR	k-1	MSR
Error	SSE	n-k	MSE
Total	SST	n-1	

Donde,

$$SSR = \hat{\beta}^T X^T y - n\bar{Y}^2$$

$$SSE = y^T y - \hat{\beta}^T X^T y$$

$$SST = y^T y - n\bar{Y}^2$$

$$MSR = \frac{SSR}{k-1}$$

$$MSE = s^2 = \frac{SSE}{n-k}$$

Con lo anterior se obtiene:

Fuente de la variación	SS	Grados de libertad	MS
Regresión	166.3926	3	55.4642
Error	2.243382	6	0.373897
Total	168.636	9	

Para encontrar el R^2 es necesario utilizar la fórmula: $R^2 = \frac{SSR}{SST}$ por lo que obtenemos:

$$R^2 = \frac{166.392618}{168.636} \rightarrow R^2 = 0.9867$$

De acuerdo con este resultado es posible afirmar que el 98.67% de las variaciones en la variable dependiente son explicadas por las variables independendientes incluídas en la regresión. Esto es, el 98.67% de las variaciones en el valor de una casa son explicadas por el valor promedio de las demás casas del vecindario; el tamaño del lote donde fue construida la casa, y el tamaño de la casa.

Pregunta 6

Realice una prueba de significancia global, muestre las hipótesis, el procedimiento para calcular el estadístico y concluya.

Debido a que lo que se pide es una prueba de significancia global, no debe incluirse el intercepto en las hipótesis a contrastar, por lo que estas son:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_A : No\ H_o$$

Para saber si se rechaza o no H_o se pueden utilizar los datos reportados en la tabla ANOVA de la pregunta anterior. El F calculado es posible encontrarlo utilizando la fórmula: $F = \frac{MSR}{MSE}$. Con lo que se obtiene que $F = 148.34$; este valor se compara con un F crítico, con k-1 grados de libertad en el numerados y con n-k en el denominador; en este caso, k-1=3 y n-k=6, por lo que $F_c = 9.780$ con un nivel de significancia del 1%.

De acuerdo con lo anterior, es posible afirmar que existe evidencia suficiente para afirmar que al menos alguno de los coeficientes de la regresión es estadísticamente diferente de cero.