Examen de Introducción a la Econometría
Universidad Carlos III de Madrid
2^a Convocatoria
Curso 2004/205
Conteste las preguntas siguientes en 2 horas y media

1. Sean dos variables aleatorias discretas X e Y. La siguiente tabla describe la función de masa de probabilidad conjunta de estas variables:

	$X=-1$	$X=0$	$X=1$
$Y=1$	0.24	0.08	0.12
$Y=2$	0.16	0.12	0.28

a. (0,75 puntos) Obtenga la función de esperanza condicional $E(Y|X)$, es decir, la esperanza de la variable aleatoria Y dado cada valor que toma la variable X.

b. (1,25 puntos) Obtenga el predictor lineal óptimo de Y dado X. Es decir, calcule el valor de los parámetros a y b del $PLO(Y|X)=a+bX$ y ofrezca el valor predicho para Y en función del valor de X según esta formula. ¿Son las funciones $E(Y|X)$ y $PLO(Y|X)$ idénticas? ¿Existe alguna diferencia entre $E[E(Y|X)]$ y $E[PLO(Y|X)]$? Justifique su respuesta.

SOLUCIÓN

a. A partir de la información disponible en el cuadro, distribución conjunta $\Pr(Y=y_i, X=x_j)$, se calculan las distribuciones marginales de Y y de X, para luego obtener la distribución condicional de $Y|X$.

Las distribuciones marginales $\Pr(Y=y_i)=\sum_{j=1}^{3}\Pr(Y=y_i, X=x_j)$ y $\Pr(X=x_j)=\sum_{i=1}^{2}\Pr(Y=y_i, X=x_j)$ son:

	$X=-1$	$X=0$	$X=1$	$\Pr(Y=y_i)$
$Y=1$	0.24	0.08	0.12	0.44
$Y=2$	0.16	0.12	0.28	0.56
$\Pr(X=x_j)$	0.40	0.20	0.40	

La distribución condicional $\Pr(Y=y_i|X=x_j)=\Pr(Y=y_i, X=x_j)/\Pr(X=x_j)$

	$\Pr(Y=y_i\|X=-1)$	$\Pr(Y=y_i\|X=0)$	$\Pr(Y=y_i\|X=1)$
$Y=1$	$0.60=\frac{0.24}{0.40}$	$0.40=\frac{0.08}{0.20}$	$0.30=\frac{0.12}{0.40}$
$Y=2$	$0.40=\frac{0.16}{0.40}$	$0.60=\frac{0.12}{0.20}$	$0.70=\frac{0.28}{0.40}$

Por tanto, las esperanzas condicionales $E(Y|X=x_j)=\sum_{i=1}^{2}y_i\Pr(Y=y_i|X=x_j)$

$E(Y\|X=-1)$	$1.40=0.60*1+0.40*2$
$E(Y\|X=0)$	$1.60=0.40*1+0.60*2$
$E(Y\|X=1)$	$1.70=0.30*1+0.70*2$

b. Puesto que $PLO(Y|X)=a+bX$, se tiene que calcular $a=E(Y)-bE(X)$ y $b=\frac{Cov(Y,X)}{Var(X)}$. Para esto se necesitan obtener las esperanzas, varianzas y la covarianza:

$$
\begin{aligned}
E(Y) &= \sum_{j=1}^{2} y_j \Pr(Y = y_j) = 1 * 0.44 + 2 * 0.56 = 1.56 \\
E(X) &= \sum_{i=1}^{3} x_i \Pr(X = x_i) = -1 * 0.40 + 0 * 0.20 + 1 * 0.40 = 0 \\
Var(X) &= E\left(X^2\right) - (E(X))^2 = \sum_{i=1}^{3} x_i^2 \Pr(X = x_i) - 0^2 \\
&= (-1)^2 * 0.40 + 0^2 * 0.20 + 1^2 * 0.40 = 0.80 \\
Cov(Y.X) &= E(XY) - E(X)E(Y) = \sum_{i=1}^{3}\sum_{j=1}^{2} x_i y_j \Pr(Y = y_j, X = x_i) \\
&= -1 * 1 * 0.24 + (-1) * 2 * 0.16 \\
&+ 0 * 1 * 0.08 + 0 * 2 * 0.12 \\
&+ 1 * 1 * 0.12 + 1 * 2 * 0.28 \\
&= 0.12
\end{aligned}
$$

Por tanto,

$$
\begin{aligned}
b &= \frac{0.12}{0.80} = \frac{3}{20} = 0.15 \\
a &= 1.56 - 0.15 * 0 = 1.56
\end{aligned}
$$

y el $PLO(Y|X) = 1.56 + 0.15 * X$

$PLO(Y\|X = -1)$	$1.41 = 1.56 + 0.15 * -1$
$PLO(Y\|X = 0)$	$1.56 = 1.56 + 0.15 * 0$
$PLO(Y\|X = 1)$	$1.71 = 1.56 + 0.15 * 1$

Concluimos que $E(Y|X) \neq PLO(Y|X)$. Sin embargo, $E(Y) = E[E(Y|X)] = E[PLO(Y|X)]$.

Reglas de Corrección

Hay bastantes cálculos, pero fáciles; aunque no deberían confundirse en los resultados numéricos, debe ser fundamental que sepan los conceptos.

En el primer apartado, por obtener las distribuciones marginales, las condicionales y la esperanza condicional $0,25$ cada una.

En el segundo apartado, por todos los cálculos intermedios para obtener los coeficientes a y b (esperanzas, varianzas, covarianza) $0,4$ puntos; por las expresiones de a y b otros $0,4$. Finalmente, por obtener las predicciones según el $PLO(Y|X)$ y decir que $E(Y|X)$ y $PLO(Y|X)$ no son idénticos: $0,25$ puntos. Por concluir que $E(Y) = E[E(Y|X)] = E[PLO(Y|X)]$: $0,20$ puntos.

---

2. Considere dos variables aleatorias X e Y. La siguiente tabla recoge información de la esperanza condicional de Y y de Y^2 dados los posibles valores de X y de la función de masa de probabilidad de la variable X ($\Pr(X = x_j)$).

	$X = -1$	$X = 0$	$X = 1$
$E(Y\|X = x_j)$	-0.20	0.00	-0.10
$E\left(Y^2\|X = x_j\right)$	0.60	0.40	0.70
$\Pr(X = x_j)$	0.40	0.20	0.40

a. (1,25 puntos) Calcule $E(Y)$ y $V(Y)$

b. (0,75 puntos) Considere la variable aleatoria $Z = 2X^3(2000 + Y)$; calcule $Var(Z|X = -1)$.

SOLUCIÓN

a. Según la ley de las esperanzas iteradas

$$\begin{aligned} E(Y) &= E(E(Y|X)) = \sum_{j=1}^{3} E[Y|X = x_j] \Pr(X = x_j) \\ &= -0.20 * 0.40 + 0.00 * 0.20 - 0.10 * 0.40 = -0.12 \end{aligned}$$

Por otro lado, $V(Y) = E\left(Y^2\right) - E(Y)^2$; además, se puede aplicar la ley de las esperanzas iteradas a la nueva variable Y^2

$$\begin{aligned} E\left(Y^2\right) &= E\left(E\left(Y^2|X\right)\right) = \sum_{j=1}^{3} E\left[Y^2|X = x_j\right] \Pr(X = x_j) \\ &= 0.60 * 0.40 + 0.40 * 0.20 + 0.70 * 0.40 = 0.60 \end{aligned}$$

Por lo que, $V(Y) = 0.60 - (-0.12)^2 = 0.5856$.
También se podría utilizar la descomposición de la varianza $V(Y) = V(E(Y|X)) + E(V(Y|X))$.

b. Por las propiedades de la varianza condicional, se sabe que

$$\begin{aligned} Var(Z|X = -1) &= \left(2 * (-1)^3\right)^2 * Var(2000 + Y|X = -1) \\ &= 4 * Var(Y|X = -1) \end{aligned}$$

Por tanto, sólo se necesita calcular $Var(Y|X = -1) = E\left(Y^2|X = -1\right) - [E(Y|X = -1)]^2 = 0.60 - (-0.20)^2 = 0.56$
Luego, finalmente

$$\begin{aligned} Var(Z|X = -1) &= 4 * Var(Y|X = -1) \\ &= 4 * 0.56 = 2.24 \end{aligned}$$

Reglas de Corrección

En el apartado a), repartir los puntos a la mitad para el cálculo de ambos
En el apartado b), $0,5$ puntos por utilizar las propiedadesde la varianza condicional y $0,25$ por calcular la variazna de Y dado $X = -1$

---

3. Considere la siguiente función de costes

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$$

donde Y representa el coste total de producción para una empresa y X representa la cantidad producida por esa empresa. Este modelo cumple todos los supuestos del modelo de regresión clásico. Con una muestra de 28 empresas se han ajustado una serie de modelos lineales por MCO cuyas salidas de E-Views aparecen a continuación:

Salida 1

Dependent Variable: Y
Method: Least Squares
Date: 06/09/05 Time: 16:32
Sample: 1 28
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	57.97021	29.97024	1.934259	0.0650
X^2	-11.02893	5.764614	-1.913213	0.0677
X^3	1.143118	0.335912	3.403031	0.0023
C	134.6560	44.80010	3.005707	0.0061

R-squared	0.979961	Mean dependent var	345.1071
Adjusted R-squared	0.977456	S.D. dependentvar	146.0272
SE of regression	21.92549	Akaike info criterion	9.144740
Sum squared resid	11537.45	Schwarz criterion	9.335055
Log likelihood	-124.0264	F-statistic	391.2195
Durbin-Watson stat	1.517423	Prob(F -statistic)	0.000000

Salida 2

Dependent Variable: Y
Method: Least Squares
Date: 06/09/05 Time: 16:42
Sample: 1 28
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	52.00983	3.831727	13.57347	0.0000
C	77.79520	22.03740	3.530144	0.0016

R-squared	0.876331	Mean dependent var	345.1071
Adjusted R-squared	0.871575	S.D. dependentvar	146.0272
SE of regression	52.33100	Akaike info criterion	10.82180
Sum squared resid	71201.88	Schwarz criterion	10.91696
Log likelihood	-149.5053	F-statistic	184.2390
Durbin-Watson stat	1.845223	Prob(F -statistic)	0.000000

Salida 3

Dependent Variable: Y
Method: Least Squares
Date: 06/09/05 Time: 16:44
Sample: 1 28
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	-39.54598	10.47298	-3.776001	0.0009
X^2	8.402173	0.944908	8.892049	0.0000
C	270.4372	24.30398	11.12728	0.0000

R-squared	0.970292	Mean dependent var	345.1071
Adjusted R-squared	0.967915	S.D. dependentvar	146.0272
SE of regression	26.15688	Akaike info criterion	9.467059
Sum squared resid	17104.56	Schwarz criterion	9.609795
Log likelihood	-129.5388	F-statistic	408.2552
Durbin-Watson stat	1.343745	Prob(F -statistic)	0.000000

a. (0,25 puntos) Contraste si los datos sugieren que la función de costes es lineal frente a que es una función cúbica.

b. (0,25 puntos) Contraste si los datos sugieren que la función de costes es cuadrática frente a que es una función cúbica.

c. (0,50 puntos) Contraste si los datos sugieren que la función de coste marginal es lineal frente a que es cuadrática. ¿Para qué valor de la producción se alcanza el mínimo de la función de coste marginal?

SOLUCIÓN

a. Esta hipótesis en nuestro modelo se formula como $H_0 : \beta_2 = \beta_3 = 0$ *vs.* $H_1 : \beta_2 \neq 0, \beta_3 \neq 0$

Se puede contrastar con un test$-F$ utilizando las dos primeras estimaciones

$$F = \frac{R^2_{SR} - R^2_R}{\left(1 - R^2_{SR}\right)/n} = \frac{0.9799 - 0.8763}{1 - 0.9799}\,(28) = 144.32$$

Este valor del estadístico se compara con el valor crítico de una chi-cuadrado con dos grados de libertad. Puesto que $\Pr\left(\chi^2_{(2)} > 0,01\right) = 9.21$, se rechaza H_0 a un nivel de significación del 1%, es decir la función de costes no es lineal.

b. Con un estadístico$-t$ podemos contrastar esta hipótesis en la primera estimación. La hipótesis es $H_0 : \beta_3 = 0$ vs $H_0 : \beta_3 \neq 0$. El estadístico viene en el output de E-Views y es $t = 3.403$. Este valor se compara con el valor crítico de una normal estandard. Puesto que $\Pr\left(N\left(0,1\right) > 0,005\right) = 2,576$, también rechazamos la hipótesis al 1% y los datos sugieren que con una función cuadrática no es suficiente.

También se puede obtener el resultado con el estadístico F mediante los R^2 de las Salidas 1 y 3.

c. La función de coste marginal viene dada por la expresión:

$$\text{coste marginal} = \partial Y/\partial X = \beta_1 + 2\beta_2 X + 3\beta_3 X^2$$

La hipótesis que debemos contrastar es $H_0 : \beta_3 = 0$ vs $H_0 : \beta_3 \neq 0$ que ya vimos en el apartado b) y la rechazamos. Es decir, los datos sugieren que el coste marginal es cuadrático.

El mínimo se alcanza en

$$\frac{\partial \text{ coste marginal}}{\partial X} = 2\beta_2 + 6\beta_3 X = 0 \Rightarrow X^* = \frac{-\beta_2}{3\beta_3}$$

Nuestras estimaciones señalan que

$$X^* = \frac{-\widehat{\beta_2}}{3\widehat{\beta_3}} = \frac{-(-11.0289)}{3 * 1.1432} = 3.2158 \text{ (}X^* \text{ es un mínimo porque la segunda derivada resulta } 6\widehat{\beta_3} = 6.8592 > 0\text{)}$$

---

4. Considere el siguiente modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Suponga que se dispone de una muestra de 20 observaciones y al estimar por MCO se obtienen los siguientes resultados:

$$\begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \widehat{\beta_2} \end{pmatrix} = \begin{pmatrix} 0.96587 \\ 0.69914 \\ 1.7769 \end{pmatrix}$$

y

$$\widehat{Var}\begin{pmatrix}\widehat{\beta_0}\\ \widehat{\beta_1}\\ \widehat{\beta_2}\end{pmatrix}=\begin{pmatrix}0.21812 & 0.019195 & -0.050301\\ & 0.048526 & -0.031223\\ & & 0.037120\end{pmatrix}$$

$$\widehat{\sigma^2}=2.5193 \qquad\qquad R^2=0.9466$$

a. (0,40 puntos) Halle la suma cuadrática total (SCT), la suma cuadrática explicada por la regresión (SCE) y la suma cuadrática de los residuos (SCR).

b. (0,40 puntos) Constraste la hipótesis $H_0:\beta_1+\beta_2=2$

c. (0,20 puntos) Constraste las hipótesis $H_0:\beta_1=\beta_2=0$

SOLUCIÓN

a. Se puede utilizar que $SCR=\sum\widehat{\varepsilon_i^2}=(n-k-1)*\widehat{\sigma^2}$ o bien considerar la aproximación $\sum\widehat{\varepsilon_i^2}\simeq n*\widehat{\sigma^2}$

$$\sum\widehat{\varepsilon_i^2}=(20-3)*2.5193=42.8281$$
$$\sum\widehat{\varepsilon_i^2}\simeq n*\widehat{\sigma^2}=50.386.$$

Por otro lado, $R^2=1-\frac{SCR}{SCT}\Longrightarrow SCT=\frac{SCR}{1-R^2}$

$$SCT=\frac{42.8281}{1-0.9466}=802.0243$$
$$\simeq\frac{50.386}{1-0.9466}=943.56$$

Finalmente, $SCT=SCE+SCR\Longrightarrow SCE=SCT-SCR$

$$SCT-SCR=802.0243-42.8281=759.1962$$
$$\simeq 943.56-50.386=893.17$$

b. Con un estadístico$-t$

$$t=\frac{\widehat{\beta_1}+\widehat{\beta_2}-2}{se(\widehat{\beta_1}+\widehat{\beta_2})}=\frac{0.69914+1.7769-2}{0.1523}=3.1257$$

donde

$$se(\widehat{\beta_1}+\widehat{\beta_2})=\sqrt{\widehat{V}(\widehat{\beta_1}+\widehat{\beta_2})}=\sqrt{\widehat{V}(\widehat{\beta_1})+\widehat{V}(\widehat{\beta_2})+2*\widehat{Cov}(\widehat{\beta_1},\widehat{\beta_2})}=$$
$$=\sqrt{0.048526+0.037120+2*(-0.031223)}=0.1523$$

que comparado con el valor crítico de una normal estándar, nos lleva a rechazar la hipótesis al 1% de significatividad.

c. El estadístico de contraste es

$$F=\frac{R_{SR}^2-R_R^2}{\left(1-R_{SR}^2\right)/n}=\frac{0.9466-0}{1-0.9466}*20=354.53$$

que se compara con el valor crítico de una chi-cuadrado con dos grados de libertad. Puesto que $\Pr\left(\chi^2_{(2)}>0,01\right)=9,21$, se rechaza H_0 a un nivel de significación del 1%, lo que nos lleva a rechazar la hipótesis nula de no significatividad conjunta.

5. Los distribuidores de manzanas ecológicas (cultivadas sin pesticidas ni abonos químicos) han realizado una muestra entre sus clientes en la que han obtenido 660 observaciones de las cantidades en kilos (*ecolbs*) y precios en euros (*ecoprc*) a los que se han vendido las manzanas. Además, han conseguido información sobre los precios de su competidores (los distribuidores de manzanas normales no ecológicas) en los centros comerciales más cercanos a cada cliente. Por tanto, también disponen de una variable adicional: precios de venta de las manzanas normales en euros (*regprc*). También, se han construido una variable que refleja la diferencia entre el precio de las manzanas ecológicas y el precio de las manzanas normales: $difprc = (ecoprc - regprc)$. Con estos datos se han estimado por mínimos cuadrados ordinarios las siguientes funciones de demanda lineal:

$$\widehat{ecolbs} = \underset{(0.372)}{2.388} - \underset{(0.332)}{0.845}\, ecoprc \tag{1}$$
$$n = 660 \quad R^2 = 0.0098 \quad \hat{\sigma} = 2.5153$$

$$\widehat{ecolbs} = \underset{(0.380)}{1.965} - \underset{(0.588)}{2.927}\, ecoprc + \underset{(0.711)}{3.029}\, regprc \tag{2}$$
$$n = 660 \quad R^2 = 0.0364 \quad \hat{\sigma} = 2.4831$$

$$\widehat{ecolbs} = \underset{(0.152)}{2.056} - \underset{(0.588)}{2.926}\, difprc \tag{3}$$
$$n = 660 \quad R^2 = 0.0363 \quad \hat{\sigma} = 2.4814$$

a. (1 punto) En este apartado, fíjese en la estimación del parámetro asociado con la variable *ecoprc* en las ecuaciones (1) y (2). Utilizando el concepto de covarianza, analice cómo deben estar relacionados los precios de las manzanas normales y de las manzanas ecológicas en la muestra.

b. (1 punto) Formule en términos de los parámetros de la ecuación (2) la siguiente hipótesis: si los distribuidores de manzanas normales reducen el precio en un euro, los distribuidores de manzanas ecológicas mantendrán constante su demanda si también reducen su precio en un euro. Con la información disponible, constrate esta hipótesis con un nivel de significación del 5%.

SOLUCIÓN

a. La relación entre el parámetro estimado en la ecuación (1) ("regresión corta") y la (2) ("regresión larga") es:

$$\widehat{\delta_1} = \widehat{\beta_1} + \widehat{\beta_2}\frac{\widehat{C}\,(ecoprc, regprc)}{\widehat{V}\,(ecoprc)}$$

Como $\widehat{V}\,(ecoprc) \geq 0$ y $\widehat{\beta_2} > 0$, la covarianza muestral $\widehat{C}\,(ecoprc, regprc)$ también debe ser positiva. Esto significa que los precios de ambas variedades guardan una relación positiva: cuando aumenta el precio de las manzanas ecológicas, se espera, de acuerdo con la muestra observada, un mayor precio de las manzanas normales.

b. La hipótesis a contrastar consiste en que $\beta_1 = -\beta_2$, es decir,

$$H_0 : \beta_1 + \beta_2 = 0 \quad \text{frente a} \quad H_1 : \beta_1 + \beta_2 \neq 0$$

Se puede utilizar un estadístico F basado en la ecuación (2) (modelo no restringido) y la (3) (modelo restringido). El valor del estadístico será:

$$F = \frac{R^2_{NR} - R^2_R}{\left(1 - R^2_{NR}\right)/n} = \frac{0.0364 - 0.0363}{(1 - 0.0364)\,/660} \approx 0.068$$

Sabemos que este estadísticose compara con el valor crítico de una ji-cuadrado con un grado de libertad. La regla de decisión consiste en rechazar H_0 al 5% cuando el valor del estadístico es mayor que el valor crítico $c_{0,05} = 3,841$, donde $\Pr\left(\chi^2_{(1)} < c_{0,05}\right) = 0,95$. Por tanto, NO hay evidencia empírica para rechazar la hipótesis nula.

Reglas de Corrección

En el apartado a), por discutir que el parámetro en la ecuación (1) está "sesgado" respecto (2) por la omisión de una variable $0,25$; por presentar la formula del sesgo $0,25$. Por razonar, que la covarianza debe ser positiva $0,4$ puntos. Por interpretar la covarianza $0,1$.

En el apartado b), por formular adecuadamente la hipótesis nula $0,50$ puntos; por plantear el estadístico de contraste $0,25$ puntos y por la toma de decisión sobre la hipótesis $0,25$ puntos.

---

6. La Oficina de Estadísticas Laborales de Estados Unidos realiza todos los años una encuesta, conocida como CPS, para estudiar la situación del mercado de trabajo norteamericano. Entre las variables que se pueden construir a partir de la encuesta se encuentran el logaritmo neperiano del salario nominal a la hora ($lwage$), los años potenciales de experiencia laboral ($exper$), los años efectivos de educación ($educa$), una variable artificial (dummy) que toma valor 1 si el encuestado es mujer ($female$), una dummy que toma valor 1 si el encuestado es hispano ($hispanic$), una dummy que toma valor 1 si el encuestado es negro ($black$), y una dummy que toma valor 1 si el encuestado no nació en EEUU ($alien$). Con los datos disponibles se estiman los siguientes modelos lineales por MCO:

Modelo 1:

$$\begin{aligned}\widehat{lwage} = &\underset{(.005)}{.83} + \underset{(.0002)}{.02}\, exper - \underset{(4x10^{-6})}{.0003}\, exper^2 + \underset{(.0002)}{.09}\, educa \\ &- \underset{(.002)}{.03}\, hispanic - \underset{(.002)}{.13}\, black - \underset{(.001)}{.04}\, alien \\ &n = 1174705; \quad RSS = 315338.7323\end{aligned}$$

Modelo 2:

$$\begin{aligned}\widehat{lwage} = &\underset{(.004)}{.93} + \underset{(.0002)}{.02}\, exper - \underset{(4x10^{-6})}{.0003}\, exper^2 + \underset{(.0002)}{.09}\, educa - \underset{(.0009)}{.25}\, female \\ &- \underset{(.002)}{.04}\, hispanic - \underset{(.002)}{.11}\, black - \underset{(.001)}{.04}\, alien \\ &n = 1174705; \quad RSS = 297451.9783\end{aligned}$$

Modelo 3: Se generan nuevas dummies interactuando $female$ y $hispanic$:
$nohisfem = female*(1-hispanic)$; $hisfem = female*hispanic$; $hismale = (1-female)*hispanic$

$$\begin{aligned}\widehat{lwage} = &\underset{(.005)}{.93} + \underset{(.0002)}{.02}\, exper - \underset{(4x10^{-6})}{.0003}\, exper^2 + \underset{(.0002)}{.09}\, educa \\ &- \underset{(.001)}{.25}\, nohisfem - \underset{(.003)}{.04}\, hismale - \underset{(.003)}{.29}\, hisfem - \underset{(.002)}{.11}\, black - \underset{(.001)}{.04}\, alien \\ &n = 1174705; \quad RSS = 297451.8690\end{aligned}$$

a. (0,75 puntos) Calcule el diferencial salarial medio exacto, en tantos por ciento, entre los trabajadores negros y blancos no-hispanos nacidos en EEUU, ceteris paribus, utilizando los resultados de la estimación del Modelo 1.

b. (0,75 puntos) De nuevo, utilizando los resultados de la estimación del Modelo 1, calcule el diferencial salarial medio exacto, en tantos por ciento, entre un trabajador negro con 16 años de educación (es decir, con un título universitario) y un trabajador blanco con 12 años de educación (es decir, solo con educación secundaria) ambos no-hispanos, nacidos en EEUU y permaneciendo constantes el resto de factores que pudieran afectar.

c. (0,50 puntos) Contraste $H_0 : \beta_{nhisfem}+\beta_{hismal} = \beta_{hisfem}$; es decir, que *ceteris paribus* los diferenciales salariales por sexo no difieren entre trabajadores de origen hispano y el resto.

SOLUCIÓN

a. Para calcular el diferencial salarial medio:

$$\frac{wage_{black} - wage_{white}}{wage_{white}} x100$$

tenemos en cuenta que si $\widehat{\beta}_b$ es la estimación de la dummy de raza negra para la regresión de los logaritmos de los salarios entonces:

$$\widehat{\ln(wage_{black})} - \widehat{\ln(wage_{white})} = \widehat{\beta}_b$$

Tomando antilogaritmos y sustrayendo por uno:

$$\begin{aligned} \frac{wage_{black} - wage_{white}}{wage_{white}} x100 &= \left(\exp(\widehat{\beta}_b) - 1\right) x100 \\ &= (\exp(-0.13) - 1)\, x100 \\ &= -12.1904569\% \end{aligned}$$

b. Sea $\widehat{\ln(wage_{b,16})}$ el logaritmo del salario esperado para una persona negra con 16 años de educación y $\widehat{\ln(wage_{w,12})}$ el logaritmo del salario esperado para una persona blanca con 12 años de educación. En este caso:

$$\begin{aligned} \widehat{\ln(wage_{b,16})} - \widehat{\ln(wage_{w,12})} &= \widehat{\beta}_b + \widehat{\beta}_e(16-12) \\ &= \widehat{\beta}_b + 4\widehat{\beta}_e \end{aligned}$$

Por tanto:

$$\begin{aligned} \frac{wage_{b,16} - wage_{w,12}}{wage_{w,12}} x100 &= \left(\exp(\widehat{\beta}_b + 4\widehat{\beta}_e) - 1\right) x100 \\ &= (1.2586 - 1)\, x100 \\ &= 25.86\% \end{aligned}$$

c. La hipótesis nula formulada en el modelo 3 es:

$H_0 : \beta_{nhisfem} + \beta_{hismal} = \beta_{hisfem}$

Se contrasta utilizando la suma de residuos al cuadrado (RSS) de los modelos 2 (modelo restringido) y del modelo 3 (modelo no restringido):

$$\begin{aligned} F &= \frac{RSS_R - RSS_{NR}}{(RSS_{NR})/n} = \frac{RSS_R - RSS_{NR}}{(RSS_{NR})} * n \\ &= \frac{297451.9783 - 297451.8690}{297451.8690} * (1174705) = 0.431\,65 \end{aligned}$$

que se compara con el valor crítico de una chi-cuadrado con un grado de libertad. Puesto que $\Pr\left(\chi^2_{(1)} > 0,01\right) = 6,63$, el estadístico F no es mayor que el valor crítico: NO hay evidencia empírica para rechazar la hipótesis nula con un nivel de significación del 1%.

Reglas de Corrección

En los apartados a) y b), $0,50$ puntos en cada uno si en lugar del diferencial exacto ofrecen el aproximado (utilizando diferencias de logaritmos).

VALORES CRÍTICOS:

$N(0,1)$
$\Pr(N(0,1) > 0,005) = 2,576$
$\Pr(N(0,1) > 0,01) = 2,326$
$\Pr(N(0,1) > 0,025) = 1,960$
$\Pr(N(0,1) > 0,05) = 1,645$
$\Pr(N(0,1) > 0,10) = 1,282$

$\chi^2_{(1)}$	$\chi^2_{(2)}$	$\chi^2_{(3)}$
$\Pr\left(\chi^2_{(1)} > 0,01\right) = 6,63$	$\Pr\left(\chi^2_{(2)} > 0,01\right) = 9,21$	$\Pr\left(\chi^2_{(3)} > 0,01\right) = 11,34$
$\Pr\left(\chi^2_{(1)} > 0,05\right) = 3,84$	$\Pr\left(\chi^2_{(2)} > 0,05\right) = 5,99$	$\Pr\left(\chi^2_{(3)} > 0,05\right) = 7,81$
$\Pr\left(\chi^2_{(1)} > 0,10\right) = 2,71$	$\Pr\left(\chi^2_{(2)} > 0,10\right) = 4,61$	$\Pr\left(\chi^2_{(3)} > 0,10\right) = 6,25$

Recordamos que una t de Student con n grados de libertad se comporta como un $N(0,1)$ para n razonablemente grande ($n \geq 20$). Por otro lado, una F de $Fisher$ con q grados de libertad en el numerador y n grados de libertad en el denominador se comporta como una $\chi^2_{(q)}/q$.