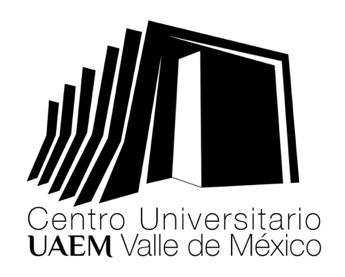
****

**UNIVERSIDAD AUTONOMA DEL ESTADO DE MEXICO**

**CENTRO UNIVERSITARIO UAEM VALLE DE MÉXICO**

**LICENCIATURA EN ACTUARÍA**

**CUADERNO DE EJERCICIOS DE ANÁLISIS DE REGRESIÓN**

**SEPTIMO SEMESTRE**

**AUTOR: D. en E. EDUARDO ROSAS ROJAS**

**OCTUBRE DE 2016**

INDICE

[1. MODELOS DE REGRESIÓN LINEAL SIMPLE 4](#_Toc463533927)

[2. MODELO DE REGRESIÓN LINEAL MULTIPLE 9](#_Toc463533928)

[3. MODELOS DE REGRESIÓN CON VARIABLES DICÓTOMAS 20](#_Toc463533929)

[4. MODELOS DE REGRESIÓN CON VARIABLES DICÓTOMAS 23](#_Toc463533930)

[EJERCICIO 1 Salarios de los maestros de escuelas públicas por región geográfica. 23](#_Toc463533931)

[EJERCICIO 2. Salario de los maestros respecto a la región y el gasto en escuelas públicas por alumno. 24](#_Toc463533932)

[EJERCICIO 3. Diferencias estructurales en la regresión ahorros-ingreso para estados unidos: el método de la variable dicótoma. 25](#_Toc463533933)

[EJERCICIO 4. Temporada en la venta de refrigeradores. 27](#_Toc463533934)

[EJERICICIO 5. Costo total con relación a la producción. 29](#_Toc463533935)

[EJERCICIO 6. Logaritmo de salarios por hora respecto al sexo. 31](#_Toc463533936)

# 

# Introducción

**El material incluido en este cuaderno de ejercicios de Análisis de Regresión ha sido diseñado de acuerdo al Programa de Estudios por Competencias de la materia de Análisis de Regresión y con base en las necesidades de los estudiantes de Actuaría, quiénes deben adquirir conocimientos sobre: Modelos de Regresión Uniecuacionales; especificamente sobre la Naturaleza del análisis de regresión con dos variables, los problemas de estimación, el Modelo Clásico de Regresión Lineal Multiple (MCRLM), estimación por intervalos, pruebas de hipótesis, el análisis de varianza, coeficentes de correlación simple y parcial, también se presenta una introducción a los modelo de regresión con variables Dicotomas. Con esto se busca cubrir la unidad de competencia I,II y la parte introductoria de la V.**

**Criterios de selección.**

**El cuaderno de ejercicios ha sido diseñado para reforzar la teoría aprendida y ejercitar de forma empírica cada concepto. Además se busca introducir al alumno en el uso del software libre “R-Project”, para lo cuál se ha incluido el codigo y los comandos necesarios para realizar los calculos indicados.**

**Se busca que el alumno entienda que el análisis de regresión lineal sienta las bases para el tratamiento de técnicas de análisis estadísticos que se deben cumplir para comprobar que el modelo planteado es confiable y que describe adecuadamente el comportamiento de los datos. Dentro de las pruebas que se solicitaran para esta técnica son las pruebas de análisis de varianza para validar el modelo con la prueba F y las prueba t para validar los parámetros del modelo planteado; además de los coeficiente de correlación y determinación y su interpretación, las intervalos e confianza del modelo, pruebas de residuales, predicción de nuevas observaciones. Es en este sentido que los ejercicios y problemas buscan reafirmar mediante la comprobación empirica cada uno de los conceptos teóricos.**

# CUESTIONARIO DE MODELOS DE REGRESIÓN LINEAL SIMPLE

**Instrucciones. Con los conocimientos adquiridos en el aula, y la realización de las tareas y lecturas, se le pide que consteste el siguiente cuestionario y realice los calculos en las preguntas que lo indiquen.**

1.- ¿Cuáles son los lineamientos de la metodología tradicional o clásica?

1. Planteamiento de la teoría o de la hipótesis

2.-Especificación del modelo matemático de la teoría

3.- Especificación del modelo econométrico o estadístico de la teoría

4.- Obtención de datos

5.- Estimación de los parámetros del modelo econométrico

6.- Pruebas de Hipótesis

7.- Pronóstico o predicción

8.- Utilización del modelo para fines de control o de política.

2.- ¿Cuál es la interpretación de Análisis de Regresión?

El análisis de regresión trata del estudio de la dependencia de la variable dependiente, respecto a una o más variables (las variables explicativas), con el objetivo de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las ultimas.

3.- ¿Cuál es el objetivo del análisis de correlación? Medir la fuerza o el grado de asociación lineal entre dos variables, medido a través del coeficiente de correlación

4.- ¿Cuáles son los tres tipos de datos disponibles para el análisis empírico?

1.- Series de Tiempo

2.- Corte transversal

3. Información combinada (Datos de Panel).

5.- ¿Qué es una función de regresión poblacional (FRP)? Función que Denota únicamente que el valor esperado de la distribución de Y dada Xi esta relacionada funcionalemente con XI. Nos dice como la media o respuesta promedio de Y varia con X. E(Y/X) =B1 +B2X.

6.- ¿Qué es una función de regresión Muestral (FRM)? Aquella que utiliza una muestra de observaciones de la población para estimar los parámetro de la FRP, mediante estimadores.

7.-¿Qué papel desempeña el término de error estocástico Ui en el análisis de regresión? Un modelo de regresión no puede ser completamente una descripción exacta de la realidad. Por lo tanto, existirán algunas diferencias entre los valores reales de las regresadas y sus valores estimados del modelo elegido.

|  |  |  |
| --- | --- | --- |
| **obs** | **Y=consumo** | **X=ingreso** |
| **1** | **80** | **90** |
| **2** | **75** | **110** |
| **3** | **100** | **120** |
| **4** | **105** | **150** |
| **5** | **120** | **170** |
| **6** | **125** | **190** |
| **7** | **130** | **215** |
| **8** | **150** | **224** |
| **9** | **165** | **247** |
| **10** | **160** | **270** |

8.- Ejercicio empírico. Dada la siguiente tabla, determine y , R y R-cuadrada. Considere que y de igual forma: . En caso de realizarlo en Excel mostrar la tabla. Si lo realiza a mano, mostrar todos los cálculos.



9.- Sabemos que el modelo de Gauss, modelo clásico o estándar de regresión lineal (MCRL) es el cimiento de la mayor parte de la teoría econométrica. Por tanto se pide que enuncie los 10 supuestos en que se encuentra sustentada.

1. Modelo de regresión lineal. El modelo de regresión es lineal en los parámetros.
2. Los valores de X son fijos en muestreo repetido.
3. El valor medio de la pertubación Ui es igual a cero
4. Homoscedasticidad o igual varianza de Ui.
5. No existe autocorrelación entre las perturbaciones.
6. La covarianza entre Ui y Xi es cero.
7. El número de observaciones “n” debe ser mayor que el número de parámetros por estimar.
8. Variabilidad en los valores de X.
9. El modelo de regresión está correctamente especificado.
10. No hay multicolinealidad perfecta.

10.. Cuales son las propiedades de los estimadores de Mínimos Cuadrados Ordinarios. Teorema de Gauss- Markov.

1.- Lineal; 2.- Insesgado; 3.-Estimador Eficiente.

Teorema de Gauss-Markov. Dados los supuestos del modelo clásico de regresión lineal, los estimadores de mínimos cuadrados, dentro de la clase de estimadores lineales insesgados, tienen varianza mínima, es decir, son MELI.

11.- ¿Qué es un coeficiente de determinación sus propiedades y su representación en términos del análisis de varianza?

El coeficiente de determinación R^2 (caso de 2 o más variables) es una medida comprendida que nos dice que tan bien se ajusta la recta de regresión muestral a los datos.

Propiedades: 1. Es una cantidad no negativa

2.- Sus limites son 0 y 1

R^2 = SEC/STC

12.- ¿Qué es un coeficiente de determinación sus propiedades y su representación en términos del análisis de varianza?

Es una medida del grado de asociación lineal entre dos variables.

Propiedades:

1. Puede tener signo positivo o negativo.
2. Cae entre los límites -1 y 1
3. Es simétrico por naturaleza rxy=ryx
4. Es independiente del origen y de la escala
5. Si Y y X son son estadísticamente independientes, e coeficiente de correlación entre ellos es cero.
6. Es una medida de asociación lienal o dependencia lienal
7. No implica necesariamente alguna relación causa efecto.

13.- ¿Cómo se clasifica la teoría clásica de la inferencia estadística?

En la estimación y la prueba de hipótesis.

# CUESTIONARIO DE MODELO DE REGRESIÓN LINEAL MULTIPLE

**Instrucciones. Con los conocimientos adquiridos en el aula, y la realización de las tareas y lecturas, se le pide que consteste el siguiente cuestionario y realice los calculos en las preguntas que lo indiquen.**

**Modelo clásico de regresión lineal normal**

**1.- ¿Cuáles son los supuestos de normalidad del termino de perturbación estoicastica?**

****

**2.- ¿Cómo se distribuyen los estimadores de la regresión, y como se distribuye la variable dependiente?**

**Dónde: **

**Y. **

**3.- ¿Cuál es la Función de Densidad de Probabilidad de una distribución normal?**

****

****

**4.-¿Cuál es la transformación para que una variable se distribuya como una Z normal estandarizada? Resuelva: Suponga que X se distribuye N(0,4). ¿Cuál es la probabilidad de que X tome un valor entre X1=-2 y X2=2?.**

****

****

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**1-pnorm(-2)\*2**

**[1] 0.9544997**

**(1-(1-pnorm(2))\*2)**

**[1] 0.9544997**

**Para X=-2**

**1-pnorm(-1)\*2**

**[1] 0.6826895**

**Para X=2**

**> (1-(1-pnorm(1))\*2)**

**[1] 0.6826895**

**5.- Cuál es la probabilidad de que en el ejercicio anterior, X exceda 2?**

|  |  |
| --- | --- |
| **EJEMPLO 19** | ¿Cuál es la probabilidad de que, en el ejemplo anterior, *X* exceda 12? Esta probabilidad es la misma de que *Z* exceda 2. De la tabla D.1, es obvio que esta probabilidad es (0.5 − 0.4772) o 0.0228. |

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**1-pnorm(2)**

**[1] 0.02275013**

**1-pnorm(1)**

**[1] 0.1586553**

**6. Suponga que X se distribuye N(0,1). ¿Cuál es la probabilidad de que X tome un valor entre X1=-1 y X2=1, después X1=-2 y X2=2; y finalmente X1=-3 y X2=3?.**

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**1-pnorm(-1)\*2**

**[1] 0.6826895**

**(1-(1-pnorm(1))\*2)**

**[1] 0.6826895**

**1-pnorm(-2)\*2**

**[1] 0.9544997**

**1-pnorm(-3)\*2**

**[1] 0.9973002**

**7.- ¿Cuál es el valor del tercero y cuarto momento estadistico de una distribución normal?**

asimetría y curtosis de 0 y 3 respectivamente.



**8.-¿A partir de que número de grados de libertad la distribución Xi-cuadrada puede tratarse como una variable normal estandarizada?**

**A partir de los 100 grados de libertad**

**9.- ¿Cuál es la media y varianza de una distribución Xi-cuadrada?**

**La media de la distribución ji cuadrada es *k* y su varianza es 2*k*, donde *k* son los grados de libertad.**

**10.- ¿Cuál es la probabilidad de obtener un valor xi-cuadrada de 19 o superior, si los g. de l. son 30?**

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**1-pchisq(19,30)**

**[1] 0.940008**

**11.- ¿Cuál es la probabilidad de obtener un valor χ2 de 40 o superior, si los gl son 20? Como muestra la tabla D.4, la probabilidad de obtener un valor χ2 de 39.9968 o mayor (20 gl) es 0.005. Por consiguiente, la probabilidad de obtener un valor χ2 de 40 es menor que 0.005, probabilidad un tanto baja.**

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**1-pchisq(40,20)**

**[1] 0.004995412**

**12.- ¿Cuál es la media y varianza de una distribución t-student?**

**La media de la distribución *t* es cero y su varianza es *k*/(*k* − 2).**

**13.- ¿Cuál es la probabilidad de obtener 1.645, 1.96 y 2.575 en una distribución t-student con 1000 grados de libertad?**

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**(1-pt(1.645,1000))\*2**

**[1] 0.1002841**

**> (1-pt(1.96,1000))\*2**

**[1] 0.05027318**

**> (1-pt(2.575,1000))\*2**

**[1] 0.01016678**

**14.- ¿Cuál es la media y varianza de una distribución t-student?**

**El valor de la media de una variable con distribución *F* es *k*2/(*k*2 − 2), el cual está definido para *k*2 > 2 y su varianza es :**



**14.- Si k1= 10 y k2= 8¿cuál es la probabilidad de obtener un valor F (a) de 2.53, (b) de 3.347 y (c) de 5.81?**

**> (1-pf(2.53,10,8))**

**[1] 0.1007467**

**> (1-pf(3.347,10,8))**

**[1] 0.05000647**

**> (1-pf(5.81,10,8))**

**[1] 0.01002326**

**15.- ¿Qué es un error tipo I y tipo II?**

**También conocida como probabilidad de cometer un error tipo I. Un error tipo I consiste en rechazar una hipótesis verdadera, mientras que el error tipo II consiste en aceptar una hipótesis falsa.**

**16.- ¿Cómo esta definido el intervalo de confianza para los estimadores del modelo de regresión ?**

**Pr[β2 −*t*α/2 ee(βˆ2)≤β2 ≤βˆ2 +*t*α/2 ee(βˆ2)]= 1 – alfa.**

**17.- ¿Qué establece la regla práctica 2t?**

**Si el número de grados de libertad es 20 o más, y si α, el nivel de significancia, se fija en 0.05, se rechaza la hipótesis nula β2 = 0 si el valor de *(B2/ee(B2))*] calculado a partir de la prueba t-student es superior a 2 en valor absoluto.**

**18. ¿Mencione y explique al menos 3 pruebas de normalidad?**

***Histograma de residuos.* Es un simple dispositivo gráfico para saber algo sobre la forma de la función de densidad pobla- cional (FDP) de una variable aleatoria. En el eje horizontal se dividen los valores de la variable de interés (por ejemplo, los residuos de MCO) en intervalos convenientes, y sobre cada intervalo de clase se construyen rectángulos cuya altura sea igual al número de observaciones (es decir, la frecuencia) para ese intervalo de clase. Si mentalmente se coloca la curva de distribución normal en forma de campana sobre el histograma, se tendrá cierta idea sobre la pertinencia o no de la aproximación normal (FDP).**

***Gráfica de probabilidad normal.* Un dispositivo gráfico relativamente sencillo para estudiar la forma de la función de densidad de probabilidad (FDP) de una variable aleatoria es la gráfica de probabilidad normal (GPN), la cual utiliza el *papel de probabilidad normal*, especialmente diseñado para gráficas. Sobre el eje horizontal, o eje *X*, se grafican los valores de la variable de interés (por ejemplo, los residuos de MCO, *u*ˆ*i*), y sobre el eje vertical, o eje *Y*, el valor esperado de esta variable si estuviera nor- malmente distribuida. Por tanto, si la variable fuese de la población normal, la GPN sería más o menos una línea recta**

***Prueba de normalidad de Jarque-Bera (JB).* La prueba de normalidad JB es una prueba *asintótica*, o de muestras grandes. También se basa en los residuos de MCO. Esta prueba calcula primero la asimetría y la curtosis (analizadas en el apéndice A) de los residuos de MCO, con el siguiente estadístico de prueba. En este caso, se espera que el valor del estadístico JB sea igual a cero.**



**19. Dada la siguiente tabla, calcule el valor de la prueba Jarque-Bera para N= 51 observaciones.**

|  |  |
| --- | --- |
| **Media** | **0** |
| **Desviación Estándar** | **66.23382** |
| **Asimetría** | **0.119816** |
| **Curtosis** | **3.234473** |
| **Prueba J-B** |  |

**Jarque–Bera 0.257585**

**20.- En la siguiente tabla se muestran los resultados del análisis de regresión de de la Natalidad Infantil en función del ingreso per cápita. Se pide responda las siguientes cuestiones.**

|  |  |  |
| --- | --- | --- |
| **Natalidad Infantilt =** | **94.2087** | **0.4368 Ingreso per cápita** |
| Error Estandar | (50.8563) | (0.0783) |
| t-student | (1.8524) | (5.5770) |
| p-value | (0.0695) | (0.0000)\* |
| R2 |  |  |
| F(1,53) | 31.1034 |  |

1. **¿Se encuentra evidencia de que el ingreso per cápita determina la natalidad infantil? Si, existe evidencia estadística.**
2. **¿Calcule el p-value de la prueba F con 1 y 53 grados de libertad respectivamente y determine si se puede concluir que ambos estimadores son estadísticamente significativos al 1%, 5% y 10%?**

**(1-pf(31.1034,1,53))**

**[1] 8.451474e-07**

**c) Suponga que la hipótesis nula fuese que B2=0.5. ¿Existe evidencia para poder rechazar la hipótesis nula, recuerde que se tienen 53 grados de libertad?**

**(1-(1-pt(-0.8071,53)))**

**[1] 0.2116081**

**La probabilidad de obtener una |*t*| de 0.8071 es mayor que 20%. Por tanto, no se rechaza la hipótesis de que el verdadero valor de β2 sea 0.5.**

**d) ¿Cual es el valor del coeficiente de determinación R2?**

Recordemos que en el caso bivariado la relació entre t yR2

 = 0.3698

**21. De acuerdo con la siguiente tabla ANOVA, determine el valor del estadístico F.**

****

|  |  |
| --- | --- |
| **F(1,53)** | **31.1034** |

**22.** **Con base en 240 tasas de rendimiento mensuales para el periodo 1986-2006, se obtuvieron la siguiente regresión para las acciones de IBM en relación con el índice de portafolio del mercado elaborado en la Universidad de Chicago:**

***RIBMt = 0.*7264 + 1.0598*IPt***

**ee (0.3001) (0.0728)**

**R2=0.4710**

**g.de l. =238**

**F(1,238) =211.89**

1. **Se dice que un valor cuyo coeficiente beta es mayor que uno es un valor volátil o agresivo. ¿Fueron las acciones de IBM valores volátiles en el periodo que se estudia a un nivel de significancia del ?**
2. **¿Es el coeficiente del intercepto significativamente diferente de cero? Si lo es, ¿cuál es su interpretación práctica?**

**a)**

**(1-pt(0.821,238))**

**[1] 0.2062339**

**b)**

**(1-pt(2.4205,238))**

**[1] 0.008124405**

**23.- El archivo a: salario profesores.xls proporciona datos sobre el salario promedio de un maestro de escuela pública (el sueldo anual está en dólares) y el gasto en educación pública por alumno (dólares) para 2005 en los 50 estados en Estados Unidos.**

**A fin de averiguar si existe alguna relación entre el salario del maestro y el gasto por alumno en las escuelas públicas, se sugirió el siguiente modelo: Sueldo*i* = β1 + β2 Gasto*i* + *ui*, donde la variable Sueldo es el salario del maestro y la variable Gasto significa gasto por alumno.**

1. **Grafique los datos y trace la línea de regresión.**
2. **Suponga, con base en el inciso *a*), que decide estimar el modelo de regresión dado antes. Obtenga las estimaciones de los parámetros, sus errores estándar, *R* 2, la SRC y la SEC.**
3. **Interprete la regresión. ¿Tiene sentido económico?.**
4. **Establezca un intervalo de confianza de 95% para β2. ¿Rechazaría la hipótesis de que el verdadero coeficiente de la pendiente es 3.0?**
5. **Obtenga el valor individual pronosticado y la media del sueldo, si el gasto por alumno es de 5,000 dólares.**
6. **¿Cómo probaría la suposición de la normalidad del término de error? Muestre la(s) prueba(s) utilizada(s).**

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

**a)**

**plot(SUELDO~GASTO)**

**abline(lm(SUELDO~GASTO))**

**b,c,d)**

**reg1<-lm(SUELDO~GASTO)**

**reg1**

**summary(reg1)**

**confint(reg1)**

**e)**

**nuevo<-data.frame(GASTO=5000)**

**predict(lm(SUELDO~GASTO),nuevo,se.fit=T)**

**$fit**

**28667.3**

**f)**

**resid<-resid(reg1)**

**shapiro.test(resid)**

**Shapiro-Wilk normality test**

**data: resid**

**W = 0.9676, p-value = 0.1754**

**jarque.bera.test(resid)**

**Jarque Bera Test**

**data: resid**

**X-squared = 2.1963, df = 2, p-value = 0.3335**

**library(moments)**

**> kurtosis(resid)**

**[1] 2.807557**

**> skewness(resid)**

**[1] 0.4991257**

**library(ggplot2)**

**datasim <- data.frame(resid)**

**ggplot(datasim, aes(x = resid), binwidth = 2) +**

**geom\_histogram(aes(y = ..density..), fill = 'red', alpha = 0.5) +**

**geom\_density(colour = 'blue') + xlab(expression(bold('Simulated Samples'))) +**

**ylab(expression(bold('Density')))**

**24.- Calcule el coeficiente de correlación (R) de los siguientes datos.**

|  |  |  |
| --- | --- | --- |
| **Obs** | **X** | **Y** |
| **1** | **42** | **75** |
| **2** | **61** | **49** |
| **3** | **12** | **95** |
| **4** | **71** | **64** |
| **5** | **52** | **83** |
| **6** | **48** | **84** |
| **7** | **74** | **38** |
| **8** | **65** | **58** |
| **9** | **53** | **81** |
| **10** | **63** | **47** |
| **11** | **55** | **78** |
| **12** | **94** | **51** |
| **13** | **19** | **93** |

**A continuación se presenta el codigo de programación en “R-Project” (Software Libre) para resolver el problema correspondiente.**

Code

library(BSDA)

attach(Correlat)

plot(X,Y,col="blue",main="Scatterplot")

m.x <- mean(X)

m.y <- mean(Y)

s.x <- sd(X)

s.y <- sd(Y)

Z.x <- (X-m.x)/s.x

Z.y <- (Y-m.y)/s.y

ZxZy <- Z.x\*Z.y

r <- (1/(length(X)-1))\*sum(ZxZy)

r

cor(X,Y)

Otra Forma es:

Cov(X,Y)

R<- cov(X,Y)/(s.x\*s.y)

Corr(X,Y)

# CUESTIONARIO DE MODELOS DE REGRESIÓN CON VARIABLES DICÓTOMAS

**Instrucciones. Con los conocimientos adquiridos en el aula, y la realización de las tareas y lecturas, se le pide que consteste el siguiente cuestionario y realice los calculos en las preguntas que lo indiquen.**

**1.- ¿qué se entiende por variable dicotoma?**

***Variables que toman dos valores 0 y 1. Tales variables son, por tanto, en esencia, un recurso para clasificar datos en categorías mutuamente excluyentes, como masculino o femenino.***

***2.-¿Qué es un modelo de análisis de varianza (ANOVA)?***

***Las variables dicótomas pueden utilizarse en los modelos de regresión en forma tan fácil como las variables cuantitativas. De hecho, un modelo de regresión puede contener variables explicativas exclusivamente dicótomas o cualitativas, por naturaleza. Tales modelos se denominan modelos de análisis de varianza (ANOVA).***

***3.¿En los modelos ANOVA, que tipo de variable es la dependiente y que tipo de variable es la o las independientes?***

**Los modelos ANOVA se utilizan para evaluar la significancia estadística de la relación entre una regresada cuantitativa y regresoras cualitativas o dicótomas. A menudo se emplean para comparar las diferencias entre los valores medios de dos o más grupos o categorías y, por tanto, son más generales que la prueba *t*, con la cual se comparan las medias de sólo dos grupos o categorías.**

**4.¿Como se llaman los modelos en los que la variable dependiente es cualitativa y las variables independientes son cualitativas o cuantitativas? Modelos LOGIT, PROBIT y TOBIT.**

**5.-¿Qué es la trampa de la variable dicótoma?**

**Una situación de perfecta colinealidad o perfecta multicolinealidad, si hay más de una relación exacta entre las variables. si una variable cualitativa tiene *m* categorías, sólo hay que agregar (*m* − 1) variables dicótomas.**

**6.-¿Para que sirve la categoría base?**

**Es la categoría a la cual no se asigna variable dicótoma. Y sirve para que las comparaciones se hagan respecto de esta categoría.**

**7.-¿Qué indican los coeficientes de las variables dicótomas?**

**Indican la medida en que el valor de la categoría que recibe el valor de 1 difiere del coeficiente de intercepto correspondiente a la categoría de comparación.**

**8.-¿Si se cambia la categoría base los resultados se modician? No.**

**9.-¿En que caso pueden permanecer en el modelo el mismo número de categorías y de variables dicótomas?**

**Para no caer en la trampa de la variable dicótoma, *se debe asegurar de que, cuando haga esa regresión, utilice la opción “no intercepto” en el paquete de software. (no colocar el coeficiente de intersección).***

**10.-¿Cómo es la interpretación de los coeficientes de las variables dicótomas en los modelos en que se omite el coeficiente de intersección?**

**La interpretación de estos coeficientes es que los resultados promedio ya están acumulados. *Obtenemos de manera directa los valores medios de las distintas categorías*.**

**11.-¿Cuál de los siguientes métodos es el mejor para introducir una variable dicótoma: 1) agregar una variable dicótoma para cada categoría y omitir el término del intercepto o 2) incluir el término del intercepto y añadir sólo (*m* − 1) variables, donde *m* es el número de categorías de la variable dicótoma?**

**Como señala Kennedy:**

**La mayoría de los investigadores piensan que es más conveniente la ecuación con intercepto porque les permite enfrentar de manera más sencilla las interrogantes que a menudo les interesan más; a saber, si la categorización genera una diferencia o no; y si lo hace, en qué medida. Si la categorización genera una diferencia, el grado de esta diferencia se mide directamente por las estimaciones de los coeficientes de las variables dicótomas. Probar si la categorización es o no es relevante se lleva a cabo mediante la prueba *t* del coeficiente de una variable dicótoma, respecto de cero (o, de forma más general, una prueba *F* sobre el conjunto apropiado de los coeficientes estimados de las variables dicótomas).**

**12.-¿En caso de que existan dos variables cualitativas se coloca el mismo número de coeficientes de intersección o solamente se coloca uno para ambas?**

**Solamente se coloca uno para ambas y se comparan respecto a éste.**

**13.-¿ *Qué es un modelo de análisis de covarianza (ANCOVA)?***

**Por lo general, en la mayor parte de la investigación económica, un modelo de regresión contiene diversas variables explicativas cuantitativas y otras cualitativas. Los modelos de regresión que muestran una mezcla de variables cuantitativas y cualitativas se llaman modelos de análisis de covarianza (ANCOVA). Tales modelos representan una generalización de los modelos ANOVA en el sentido de que proporcionan un método para controlar estadísticamente los efectos de las regresoras cuantitativas (llamadas covariantes o variables de control) en un modelo con regresoras cuantitativas y cualitativas (o dicótomas).**

**14.-¿Qué es un componente estacional?**

**Muchas series de tiempo económicas que se basan en datos mensuales o trimestrales presentan pautas estacionales (movimiento oscilatorio regular); por ejemplo, las ventas de las tiendas de departamentos en la época de Navidad y otras festividades importantes, la demanda de dinero (saldos de efectivo) por parte de las familias en épocas de vacaciones, la demanda de helado y bebidas gaseosas durante el verano y los precios de los cultivos justo después de la época de cosecha, la demanda de viajes en avión, etc. A menudo es útil eliminar el factor o *componente* estacional de las series de tiempo con el fin de concentrarse en los demás componentes, como la tendencia. El proceso de eliminar el componente estacional de una serie de tiempo se conoce como desestacionalización o ajuste estacional, y la serie de tiempo así obtenida se denomina serie de tiempo desestacionalizada o ajustada por estacionalidad**

**15.-¿Cuáles son los cuatro componentes de una serie de tiempo?**

**Una serie de tiempo puede tener cuatro componentes: (1) estacional, (2) cíclico, (3) tendencia y (4) estrictamente aleatorio.**

# EJERCICIOS DE MODELOS DE REGRESIÓN CON VARIABLES DICÓTOMAS

**Instrucciones. Con los conocimientos adquiridos en el aula, y la realización de las tareas y lecturas, se le pide que realice los calculos que se indiquen.**

**EJERCICIO 1 Salarios de los maestros de escuelas públicas por región geográfica.**

Donde:

Y=al salario (promedio) de los maestros para el estado de México

D2=1 si el estado está al noroeste o norte-centro

=0 para otra región

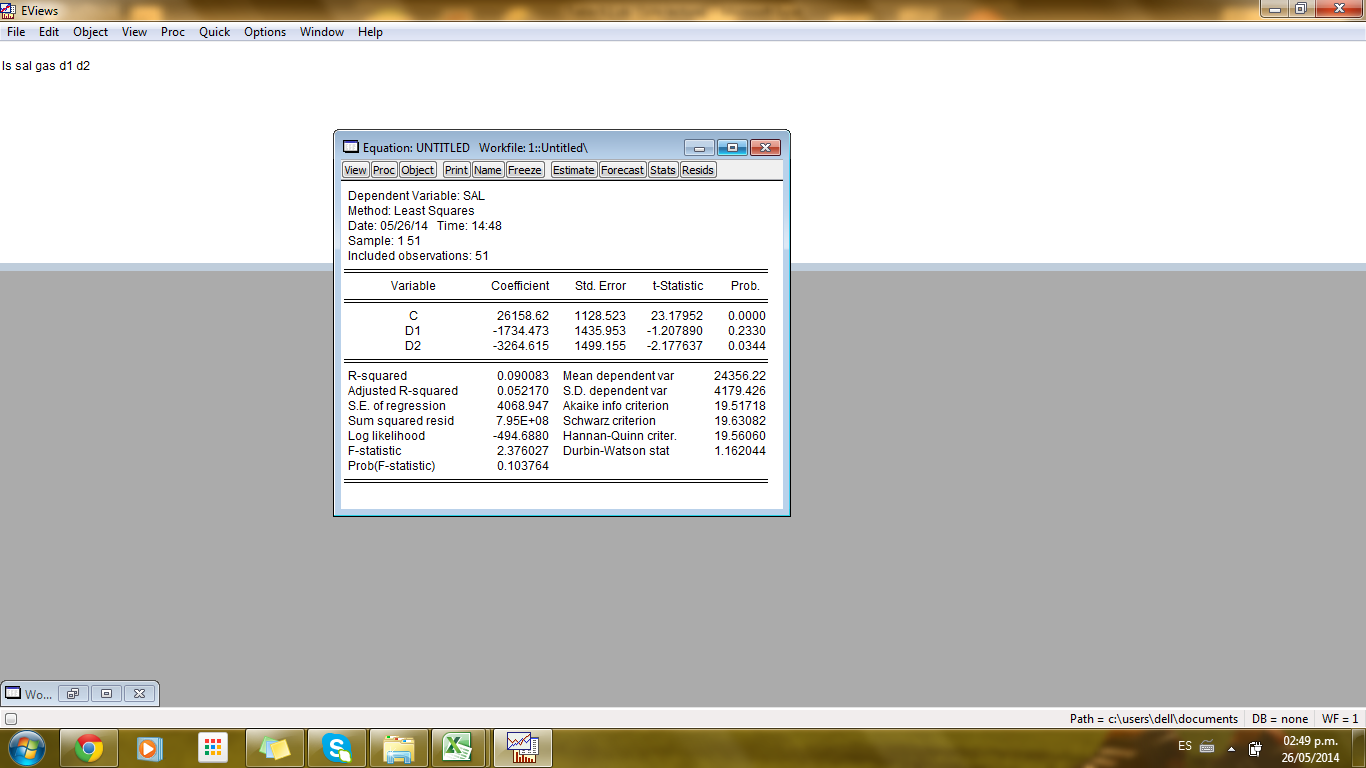
D3=1 si el estado es del sur

=0 para otra región del país

Como cualquier modelo de regresión múltiple que se haya estudiado antes, excepto que dé en vez de regreso ras cuantitativas se tienen solo variables cualitativas o dicótomas las cuales toman el valor de 1 si la observación pertenece a una categoría particular y 0 si no pertenece a esa categoría.

Salario medio de los maestros de escuelas públicas en el oeste esta dado por la intersección en la regresión múltiple además los coeficientes de la pendiente y indican la cantidad por la que los salarios promedio de los maestros del noroeste y norte –centro así como los del sur difieren respecto a los salarios medios de los profesores.

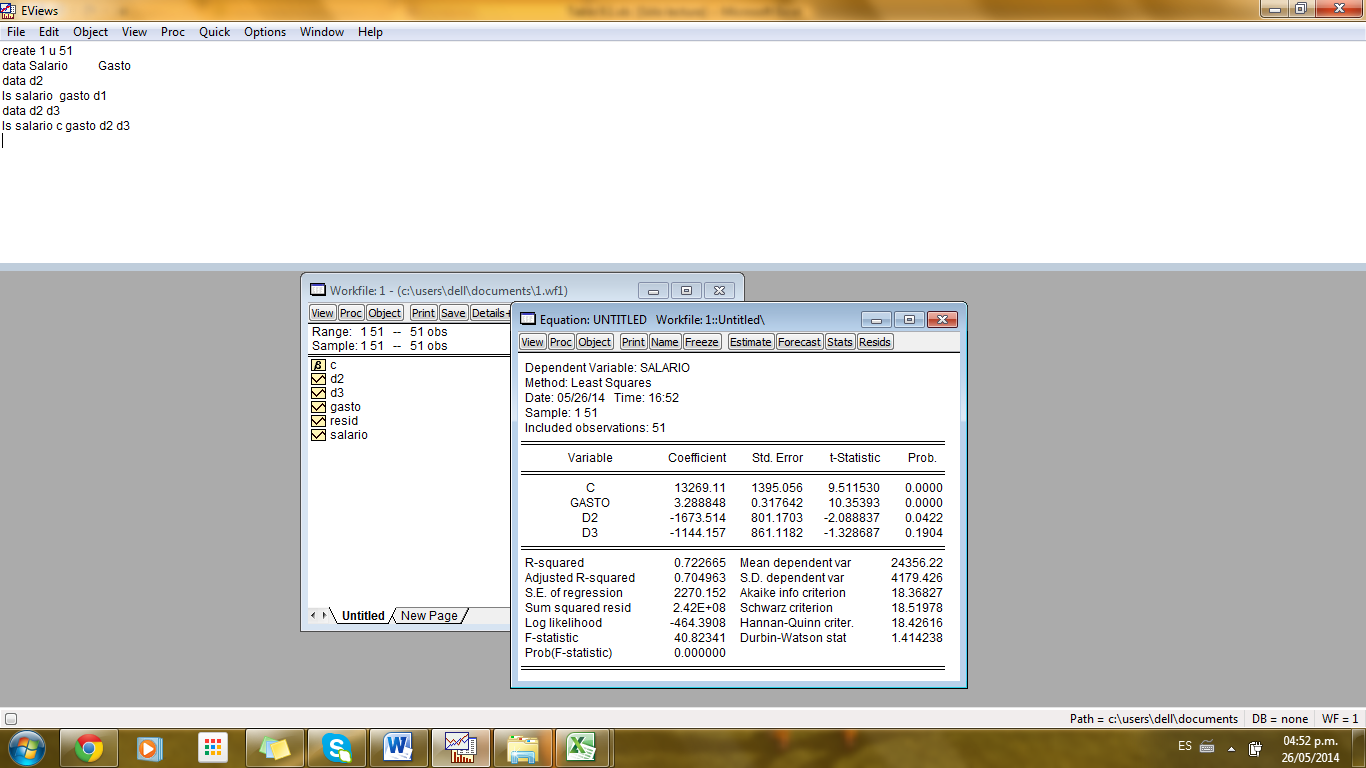
Utilizando los datos de la tabla 9.1 se obtienen los siguientes resultados



Tal como se muestran los resultados de esta regresión el salario medio de los profesores del este de casi $2 6158, el de los maestros del noreste y del norte –centro es menor por casi $1 734 y respecto a los del sur su salario es menor por casi $3 265. Los salarios medios reales en las últimas dos regiones se pueden obtener con facilidad al añadir estos salarios diferenciales al salario medio de los maestros del oeste.

**EJERCICIO 2. Salario de los maestros respecto a la región y el gasto en escuelas públicas por alumno.**

Datos de la tabla 9.1

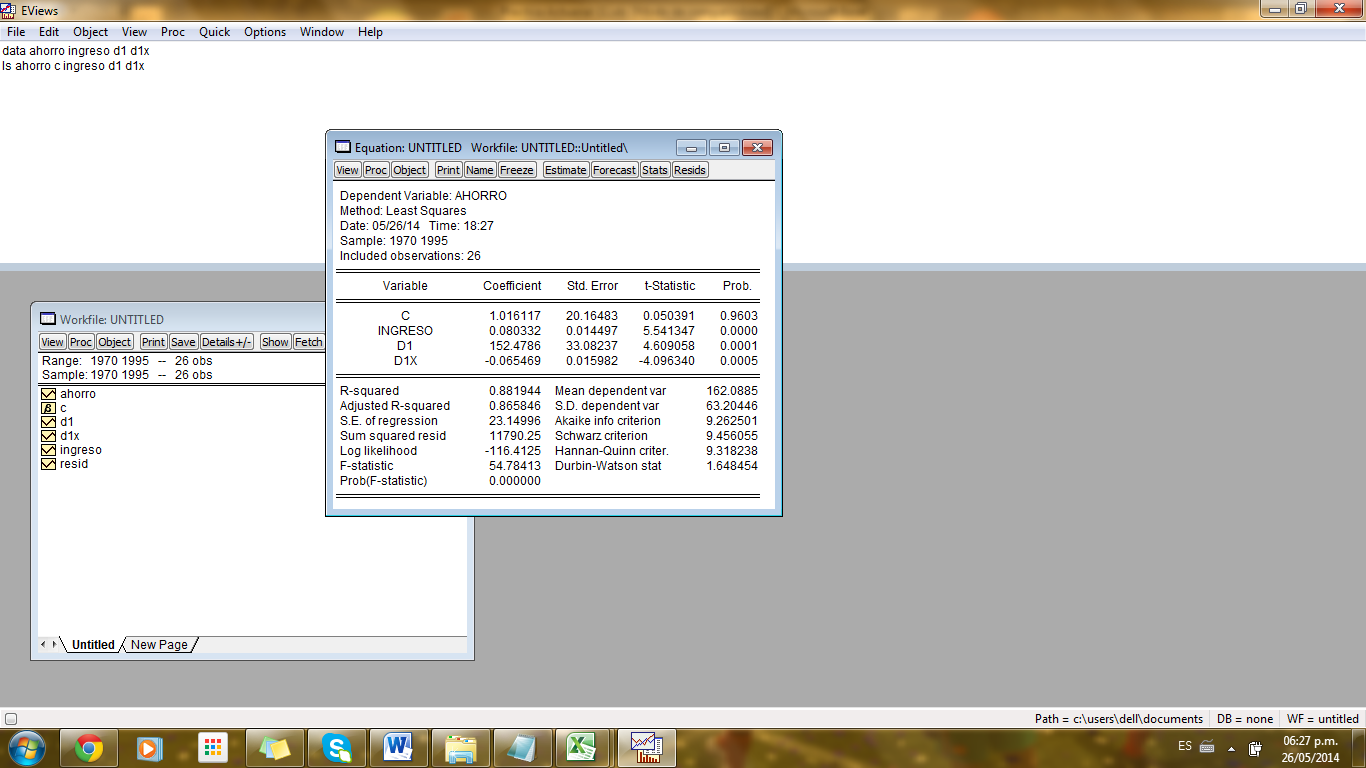


Donde indican los valores de p menores al 5% y señala los valores p mayores que 5%

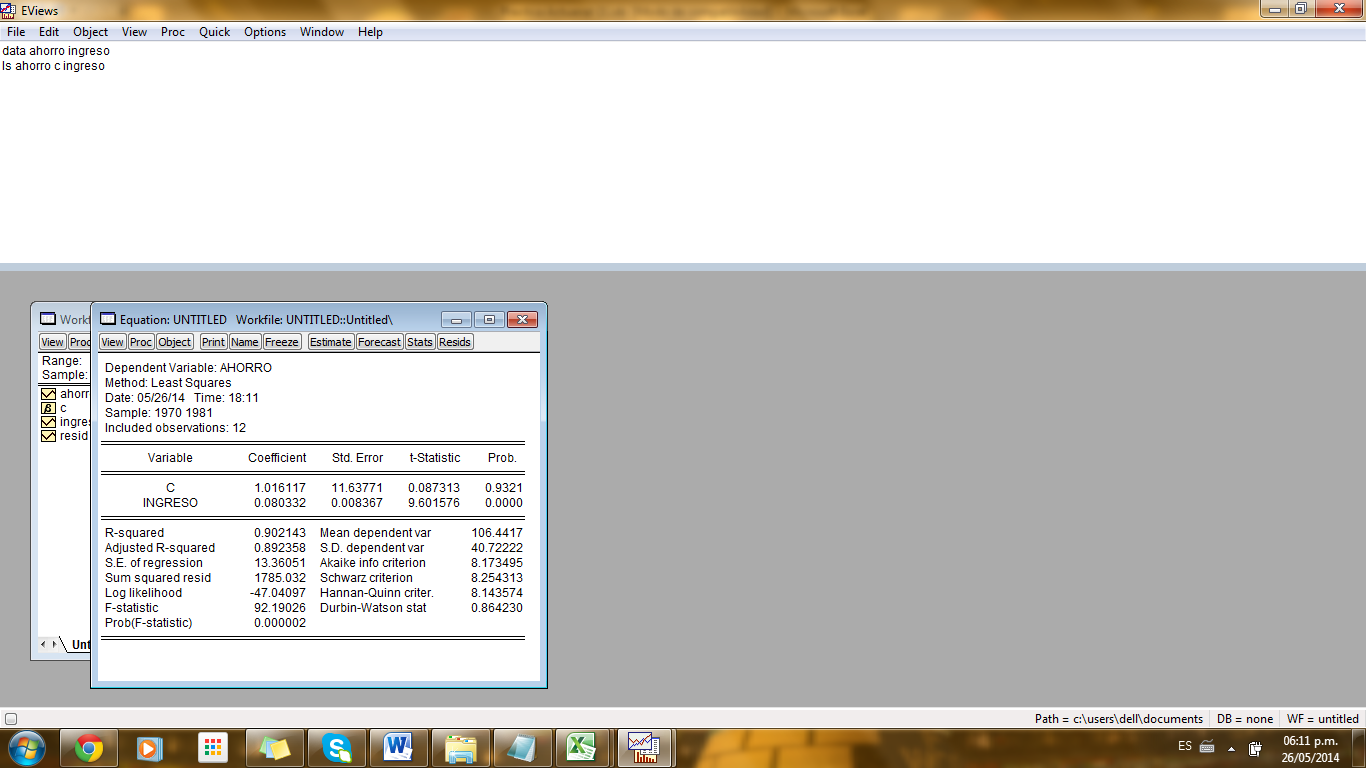
Como los resultados lo sugieren *ceteris paribus* conforme al gasto público aumenta un dólar el salario de los maestros de la escuela pública se incremente aproximadamente 3.29 si se controla el gasto en educación ahora se observa que el coeficientes diferencial de la intersección es significativo para la región noroeste y norte centro, pero no para sur.

**EJERCICIO 3. Diferencias estructurales en la regresión ahorros-ingreso para estados unidos: el método de la variable dicótoma.**

Antes de seguir adelante se presenta primero los resultados de la regresion del modelo para los satos ahorros-ingreso de Estados Unidos



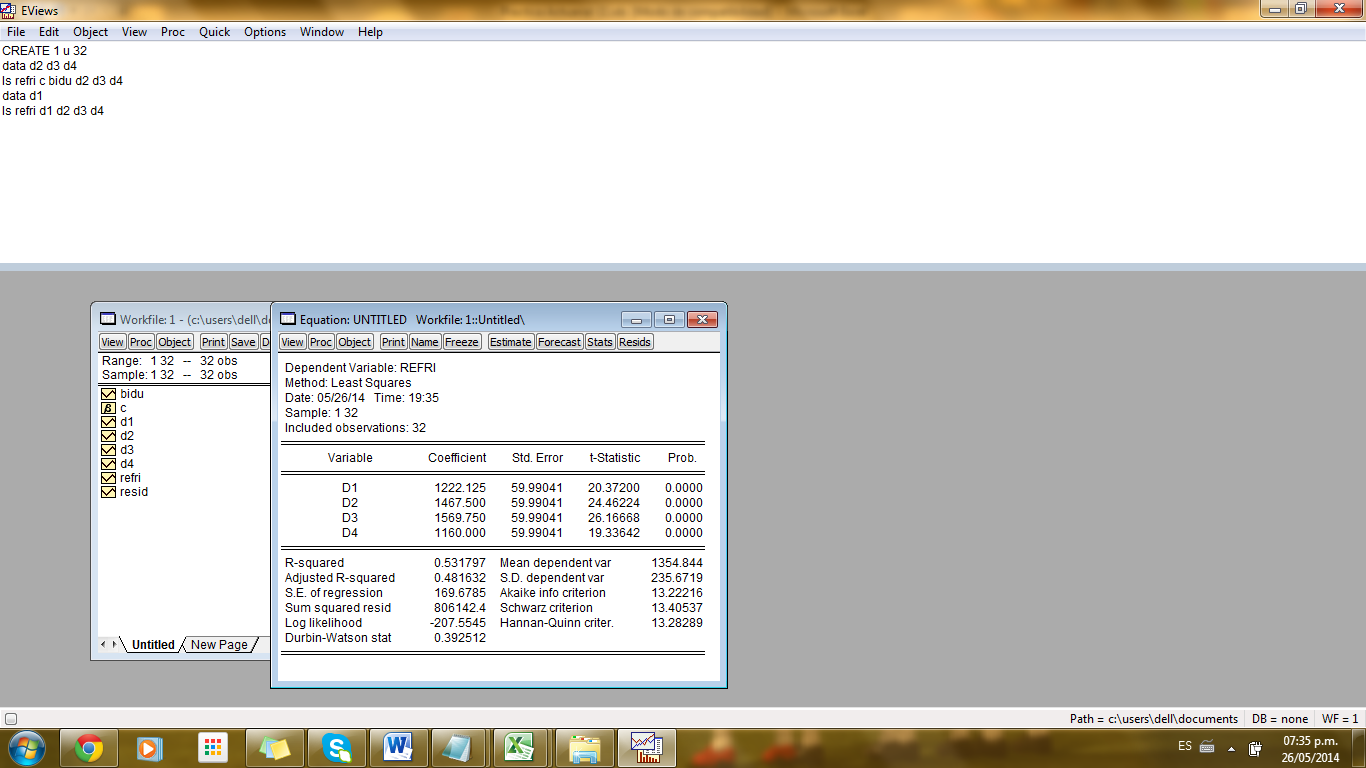
Donde indica los valores p menores al 5% y señala los valores p mayores que 5% tal y como muestra los resultados de esta regresión la intersección deferencial y el coeficiente de la pendiente son estadísticamente significativos.

Regresión ahorros-ingreso para 1982-1995

**EJERCICIO 4. Temporada en la venta de refrigeradores.**

De los datos respecto a la venta de refrigeradores datos de la tabla 9.3 se obtienen los siguientes resultados de la regresión

Omitiendo la constante



Regri = a ventas de refrigeradores

Vidu = A gasto de bienes durables

D2 = Uno en segundo trimestre

D3 = 1 en tercer trimestre

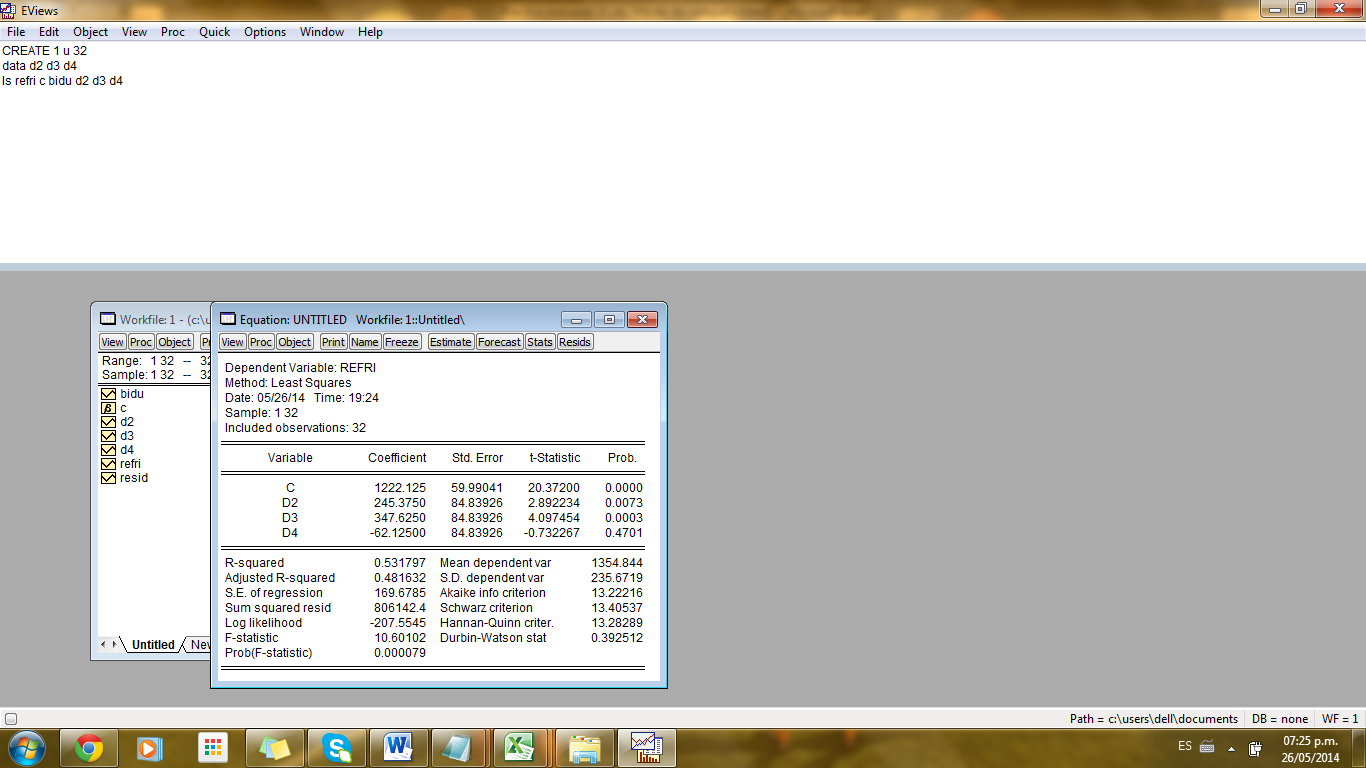
D4= 1 en cuarto trimestre

Día de refrigeradores en el primer trimestre en miles de unidades es de casi 1222.

En el segundo trimestre fue de casi 1468. Las de tercer trimestre fueron de 1570 aproximadamente y de las del último trimestre fueron de casi 1160.

Por cierto en vez de asignar una variable dicótoma a cada trimestre y suprimir el término de intersección a fin de evitar la trampa de la variable dicótoma se podrá asignar sólo 3 variables dicótomas e incluir el término de intersección. Supóngase que se considera el primer trimestre como el trimestre de referencia y se asignan variables dicótomas al tercero y cuarto. Lo anterior da los siguientes resultados de regresión

Metiendo la constante pero quitando la d1



Donde indica los valores p menores al 5% y señala los valores p mayores al 5% puesto que se está considerando al primer trimestre como el punto de referencia los coeficientes relacionados con las distintas variables dicótomas ahora son intersección diferencial que muestran en qué medida el valor promedio de Y en el trimestre que recibe un valor de 1 para la variable dicótoma difiere del trimestre que es punto de referencia expresado de manera distinta los coeficientes de las variables estacionales indicaran el incremento o decremento estacional del valor promedio de Y con relación a la temporada base si se añaden distintos valores de la referencia de 1222.125, se tendrán los valores promedio para los distintos trimestres. Al llevar a cabo lo anterior se producirán exactamente la ecuación 9.7.2 salvo errores de redondeo.

Ahora se aprecia el valor de considerar a un trimestre como punto de referencia ya que 9.7.3 muestra el valor promedio de y para el cuarto trimestre. No es estadísticamente distinto del valor promedio para el primer trimestre puesto que el coeficiente de la variable dicótoma para el cuarto trimestre no es estadísticamente significativo

**EJERICICIO 5. Costo total con relación a la producción.**

Como ejemplo de aplicación de la regresión lineal por secciones, considerese la información hipotetica sobre costo total-producción total dada en la tabla 9.6. Se dice que el costo total puede cambiar su pendiente al alcanzar un nivel de producción de 5500 unidades.

*(Tabla 9.6)*

|  |  |
| --- | --- |
| *INFORMACIÓN HIPOTÉTICA SOBRE LA PRODUCCIÓN Y EL COSTO TOTAL* | |
| *COSTO TOTAL, DÓLARES* | *UNIDADES DE PRODUCCIÓN* |
| 256 | 1 000 |
| 414 | 2 000 |
| 634 | 3 000 |
| 778 | 4 000 |
| 1 003 | 5 000 |
| 1 839 | 6 000 |
| 2 081 | 7 000 |
| 2 423 | 8 000 |
| 2 734 | 9 000 |
| 2 914 | 10 000 |

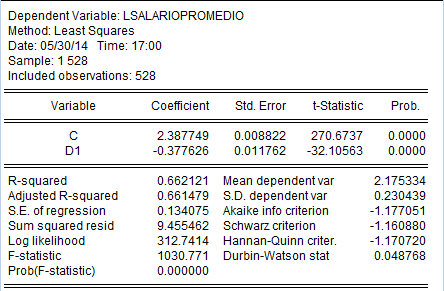
Si se permite que en (9.8.4) represente el costo total y la producción total, se obtienen los siguientes resultados:



Como lo muestran estos resultados, el costo marginal de producción es de cerca de 28 centavos de dólar por unidad y aunque éste es cerca de 37 centavos (28 + 9) para la producción por encima de 5 500 unidades, la diferencia entre los dos no es estadísticamente significativa puesto que la variable dicótoma no es significativa, por ejemplo, al nivel del 5%. Para todos los fines prácticos, entonces se puede efectuar la regresión del costo total sobre la producción total, eliminando la variable dicótoma.

**EJERCICIO 6. Logaritmo de salarios por hora respecto al sexo.**

Para ilustrar (9.10.1), se utilizan los datos que están implícitos en el ejemplo 9.2. Los resultados de la regresión basada en 582 observaciones son los siguientes:



Donde \* indica los valores p que son prácticamente cero.

Tomando el antilogaritmo de 2.1763, se tiene 8.8136 ($), que es la mediana de los ingresos por hora de los trabajadores, y si se toma el antilogaritmo de [(2.1763 – 0.2437)= 1.92857], se obtiene 6.8796 ($), que es la mediana de los ingresos por hora de las trabajadoras. Por tanto, la mediana de los ingresos por hora de las trabajadoras es menor por casi 21.94%, en comparación con sus contrapartes masculinos [(8.8136 – 6.8796)/8.8136].

Resulta interesante que se pueda obtener la semielasticidad para una regresora dicótoma de manera directa, mediante el proceso sugerido por Halvorsen y Palmquist. Tómese el antilogaritmo (de base e) del coeficiente estimado de la variable dicótoma, réstele 1 y multiplique la diferencia por 100. En consecuencia, si se toma el antilogaritmo de -0.2437, se obtendrá 0.78366. Al restar 1 de lo anterior, se tiene -0.2163, y después de multiplicar esta cifra por 100 se tiene -21.63%, lo cual sugiere que la mediana del salario de una trabajadora (D=1) es menor que la de su contraparte masculina por aproximadamente 21.63%, que es lo mismo que se obtuvo antes, salvo errores de redondeo.

**Bibliografía**

**Los cuestionarios así como los ejercicios propuestos son un compendio tanto de ejercicios propios como de ejercicos comprobados y calculados en el libro “Gujarati, Damodar N. “Econometría”. McGrawHill. Cuarta Edición, 2007”. Además de los siguientes libros:**

1. Jhonston, J. **Econometric Methods**, Edit. Mc Graw Hill, 3a. Edicion.
2. Pindyck R. Y Rubinfeld, L. (1991) **Econometric Models And Econometric Forecast.** Mc Graw Hill.

La Bibliografía complementaria recomendada es:

1. Charemza, W Y Derek F. Deadman (1992). **New Directions In Econometric Pratice: General To Specific Modelling, Cointegration And Vector Autogregresive.**
2. Greene, W. (1999) **Analisis Ecometrico**. Prentice Hall, Tercera Edicion.