Trabajo de Económetria

capítulo 5

(Regresión con 2 variables: Estimacion de Intervalos y Prueba de Hipótesis)

Integrantes:

Astrid Uribe O.

Sebastian Sotomayor

Fabian Muñoz I.

Johan Celedon S

Nicolas Blatazar O.

Nicolas Araya V.

Ignacio Isla A.

Tarea de Económetria.

5.1. Establezca si las siguientes afirmación son ciertas, falsas o inciertas. Sea preciso.

1. La prueba *t* de significancia estudiada en este capítulo requiere que las distribuciones muestrales de los estimadores β1 y β2 (muestrales) sigan una distribución normal.

**VERDADERO**, la prueba t se basa en las distribuciones normales, lo cual nos ayuda a determinar que los estimadores β1 y β2 se distribuyen normalmente bajo MCRL.

1. Aunque el término de perturbaciones en el MRCL no esté normalmente distribuido, los estimadores MCO continúan siendo insesgado.

**VERDADERO**, basándonos el teorema de Gauss-Markov además de ser lineal, insesgado y tener mínima varianza se adiciona que independientemente de las perturbaciones de MRCL no estén distribuidos normalmente los estimadores seguirán siendo insesgado.

1. Si no hay intersección en el modelo de regresión, las Ui (=Ui)(muestral) estimadas no sumaran cero.

**VERDADERO**, Cuando vemos en el gráfico las perturbaciones muéstrales y las estimadas existe una distancia entre ambas que al llevarlas de modo numérico se cancelaran teniendo origen en el punto muestral del error. Es así como las perturbaciones muestral (arriba) se cancelan con las perturbaciones estimada (abajo), y dan cero, en contraste se dice que existe intersección entonces las distancias de las perturbaciones no serán equitativas y la diferencia no resultaría igual a cero.

1. El valor *p* y el tamaño de un estadístico de prueba tienen el mismo significado.

**VERDADERO**, estos están estrechamente relacionados por el valor de p donde nos indica el nivel más bajo de importancia en el cual podemos determinar que la hipótesis nula pueda ser rechazada.

1. En un modelo de regresión que contenga la intersección, la suma de los residuos es siempre cero.

**FALSO**, la suma de las perturbaciones al existir una intersección en el modelo no siempre dará un valor igual a cero sino que también puede entregar un valor mayor o menor a cero.

1. Si una hipótesis nula no es rechazada, es verdadera.

**FALSO**, si una hipótesis nula no es rechazada no se está diciendo que sea verdadera con absoluta certeza sino que esta se acepta.

1. Entre mayor sea la el valor de f2, mayor será la varianza de β2 (muestral) dada en (3.3.1).

**FALSO**,a mayor ѳ² podrá ser contrarestado por un mayor β2 (estimado). Pero si se mantiene constante la ѳ² se podrá decir que la frase es verdadera.

1. Las medias condicional e incondicional de una variable aleatoria significan lo mismo.

**FALSO**, las medias condicionales están sujetas al comportamiento de la variable que las condiciona, en cambio, las medias incondicionales dependen de sí mismas.

1. En una FRP de dos variables, si el coeficiente de de la pendiente de β2 es cero, la intersección de β1 es estimada por la media muestral Y.

**VERDADERO**, al obtener los β (muestrales) los resultados se guiaran por la media de Y por ende al eliminar el β2 se elimina la variable X lo que provoca que la intersección de β1 sea estimada en su totalidad por la media muestral Y.

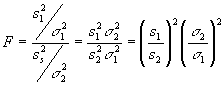
1. La varianza condicional, var(Yi/Xi)=f2 y la varianza incondicional de Y, var(Y)=f2y, serian la misma si X no tuviera influencia en Y.

5.2. Constrúyase la tabla ANOVA a la manera de la tabla 5.4 para el modelo de regresión dado en (3.7.2) y pruébese la hipótesis de que no existe relacion entre el gasto en alimentos y el gasto total en India.

|  |  |  |  |
| --- | --- | --- | --- |
| Fuente de variacion | SC | G de l | SPC |
| Debido a la regresion ( SEC) | 139023 | 1 | 139023 |
| Debido a los residuos (SRC) | 236894 | 53 | 4470 |

Total 375916

si consideramos como hipotesis que no existe relacion entre las variables planteadas que seria el gasto de alimentacion versus el gasto total de India, el valor obtenido de P resultado de la ecuacion F es casi cero, lo que sugiere que uno en primera instancia puede Rechazar la hipotesis nula planteada, y ademas con el valor obtenido podemos decir que el rechazo es bastante fuerte.



5.3. De los datos proporcionados en la tabla 2.6 respecto a los ingresos y la escolaridad, se obtuvo la siguiente regresión [véase la EC. (3.7.3)]:

Salmedio = 0.7437 + 0.6416 Escolaridad

ee = (0.8355) (……)

t = (…… ) (9.6536) r ⃰ 2 = 0.8944 n = 13

1. Complete los números faltantes:

, b1= 0 =0,8797



=0,0664



1. ¿Como se interpretaría el coeficiente 0,6417 ?

Cuando varia la en un año de escolaridad el salario medio aumenta en un 0,6417

1. ¿Se rechazaría la hipótesis de que la educación no tiene efecto alguno sobre los salarios?¿cual prueba se usaría?, ¿Por qué?¿Cual es el valor p del estadístico de prueba

α=95%= 0,05

α/2= 0,025

β2 - tα/2 \* eeβ1 < b2 < β2 + tα/2 \* eeβ1

0,6416-2,201\*0,06664 < b2 < 0,6416+2,201\*0,06664

0,6416-0,1463 < b2 < 0,6416 +0,1463

0,4953 < b2 < 0,7878

Se rechaza la hipótesis ya que β2 se encuentra dentro del rango esto quiere decir que si tiene influencia.

1. Constrúyase la tabla ANOVA para este ejemplo y pruébese la hipótesis de que el coeficiente de la pendiente es cero. ¿Cuál prueba se utilizaría y por que?

SCE=74,9389; SCR=8,8454, numerador DF=1 denominador DF=11. F=93,1929.

El valor p de los mismos por F bajo la hipótesis nula de la que no hay relación entre las dos variables es 0,000001 cual es extremadamente pequeña. Así podemos rechazar la hipótesis nula con gran confianza.

1. Supóngase que en la regresión que se acaba de dar, el valor r2 no se proporciona. ¿Se podría haber obtenido con base en otra información dada en la regresión?

En el caso de dos variables, dada H0: β2 = 0 no hay siguiente relación entre el valor de t y r2.

5.4. Sea *p²* el verdadero coeficiente de correlación poblacional. Supóngase que se desea probar que *p² = 0.* Explíquese verbalmente como se probaría esta hipótesis. Pista utilícese la ecuación (3.5.11), véase también el ejercicio 5.7



Si se sostiene que tenemos un coeficiente de correlaicon cuadratico de por si este resultara dentro del conjunto de numeros relaes positivos y el cero, ademas sera = a cero. Por interpretacion no existiria una relacion lineal entre estas dos variables a probar, pero este hecho no quieta que estas variables sean independientes independientes ya que puede haber una relacion no lineal entre ambas variables.

5.5. Lo que se conoce como la **recta característica** del análisis de inversión moderno, es sencillamente la recta de regresión obtenida del siguiente modelo.

rit = αi + βirmt + ut

Donde rit = la tasa de rendimiento del i- ésimo valor en el tiempo t.

rmt = la tasa de rendimiento del portafolio del Mercado al tiempo t.

ut = termino de perturbación estocástica.

En este modelo, βi es conocida como **coeficiente beta** del i- ésimo valor, una medida del riesgo del mercado ( o sistemático) de un valor.

Con base en 240 tasas de rendimiento mensuales para el periodo 1956-1976, Fogler y Ganapathy obtuvieron la siguiente recta característica desarrollado en la Universidad de Chicago:

rit = 0.7264 + 1.0598 rmt r´2 = 0.4710

ee = (0.3001) (0.0728) g de l =238

F1.238 = 211.896

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 tiempo bajo estudio?

Utilizar la prueba de la t para la hipótesis de que el coeficiente de la pendiente es una verdadera.

Que es obtener:

rit = 0.7264 + 1.0598rmt r2 = 0.4710

ee = (0.3001) (0.0728) g de l = 238 F1.238 = 211.896

Ho = β > 1 β = 0.7264 100(1 – α) = 95   
H1 = β < 1 1 – α = 0.95

α = 0.05

β +- tα/2ee (β1) α/2 = 0.025 t238, 0,25 = 1,960  
1,0598 +- 1,960 x 0,0728   
0,917112 < β < 1,202488 95%  
t = 1,0598 – 1/0,0728 = 0,821

Por lo tanto de una muestra de 238 df este valor t no es significativo incluso a α = 10%, la conclusión es que durante el período de la muestra, **ibm** no era volátil.

1. ¿Es el coeficiente de la intersección significativamente diferente de cero? Si lo es, ¿Cuál es su significado práctico?

|  |  |
| --- | --- |
| T= 0.7264 = | 24.205 |
| 0.3001 |  |

,

Lo cual es importante en los dos niveles de significación por ciento. pero tiene poco sentido económico interpretado literalmente, el valor de intercepción de unos 0,73

significa que incluso si la cartera de mercado tiene retorno a cero, el retorno es seguro 0,73 por ciento

5.6. La ecuación (5.3.5) puede ser escrita también como



Es decir la desigualdad debil (˂) puede ser remplazada por la desigualdad fuerte (˂). ¿Porque?

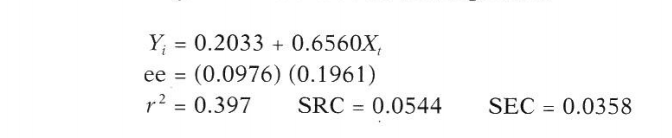
Como se señala en los supuestos de normalidad, el beta coeficiente 2 tiene una distribución normal. Dicho esto, se sabe que una variable normal que toma un valor cuando la probabilidad de que una variable aleatoria tome un valor, su resultado es cero. Según esto, se puede señalar que no existe diferencia si la igualdad es fuerte o débil.

5.7. R.A Fisher ha derivado la distribución muestral del coeficiente de correlación definido en (3.5.13). si se supone que las variables X y Y tienen una distribución normal conjunta, es decir, si provienen de una distribución normal bivariada (véase el apéndice 4A, ejercicio 4.1), entonces, bajo el supuesto de que el coeficiente de correlación poblacional p es cero, es posible demostrar que t = r√(n-2) / √(1-r2) sigue la distribución t de Student con n-2 g de l. Demuéstrese que este valor t es idéntico al valor t dado en (5.3.2) bajo la hipótesis nula de que *β2 = 0. Por tanto, estableciéndose que bajo la misma hipótesis nula F= r2 (véase la sección 5.9.).*

X e Y tiene una distribución conjunta, es decir si provienen de una distribución normal bivariada, entonces, bajo el supuesto de que el coeficiente de correlación poblacional p es cero, demostrar que t=/ sigue la distribución t student con n-2 g de l demuestre que este valor t es idéntico al valor t dado bajo la hipótesis nula F=t^2



5.8. Considere el siguiente resultado de una regresión:



Donde y =tasa de participación de la fuerza de trabajo TPFT) de las mujeres en 1972 y X= TPFT de las mujeres en las mujeres en 1968. Los resultados de la regresión se obtuvieron de una muestra de 19 ciudades de estados unidos.

1. ¿Como se interpretaría esta regresión?

Existe relación positiva en la tasa de participación de en 1972 y 1968, lo cual no resulta sorprendente a vista del hecho desde la Segunda Guerra Mundial ha habido un aumento constante de la tasa de participación de las mujeres

1. Pruébese la hipótesis Ho:β2=1, contra H1:β2≥1 ¿Que prueba se utilizaría? ¿Por qué? ¿Cuáles son las suposiciones subyacentes de la prueba que se utilizan?

Utilizar la prueba y de Una cola

t = 0,6560 - 1 = -1.7542. de 17 df, el valor t de una cola

0.1961

a = 5% es 1.740. ya que el valor t estimado es significativo, a este nivel de significativo, podemos rechazar la hipótesis de que el coeficiente de la pendiente es cierto lo mayor

1. Suponga que la TPFT para 1968 fue de 0.58 (58%) .con base en los resultados de la regresión dado antes , ¿cuál es la TPFT media en 1972? .establézcase un intervalo de covarianza de 95% para la predicción de la media

La TASA de Participación de los medios de comunicación es: 0,2033 + 0,6560 (0,58) = 0,5838. Establecer sin INTERVALO de confianza del 95% previsto de Este valor, utilice la fórmula: 0.5838 2.11 (sí del valor Medio Previsto), Donde 2,11 es el valor t critica 5% párrafo 17 df. Obtener el párrafo pronóstico error estándar del valor. Pero Tenga en Cuenta Que, Dado Que los Autores no dan el valor Medio de la TASA de Participación de las Mujeres en 1968, cuan no Florerías Calcular Este error Estándar

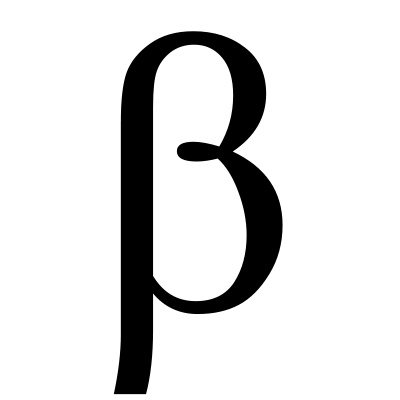
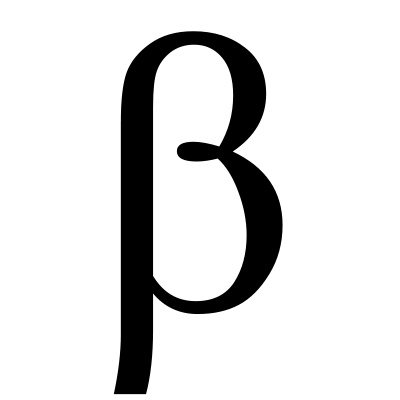
1. ¿como e probaría la hipótesis de que el termino de error en la regresión sobre la población esta normalmente distribuido? Muéstrense los cálculos necesarios?

sin los datos reales, no vamos a ser capaces de responder a esta pregunta, porque necesitamos el valor de los residuos para trazar ellos y obtener el gráfico de probabilidad normal o para calcular el valor de la jaruque - test bera

* 1. La tabla 5.5 proporciona datos sobre el salrio promedio de un maestro de escuela punblica (el sueldo anual en dolares) y el gasto en educacion (dolares) para 1985 en los 50 estados y Distritos de Columbia en Estados unidos

A fin de Averiguar si existe alguna relacion entre salario del maestro y el gasto por alumno en las escuelas publica, se sugirio el siguiente modelo:

Sueldo= 1 + 2 Gastoi + ui , donde la variable Sueldo es el salario del maestro y la variable Gasto significa gasto por alumno



a) Grafíquense los datos y Trácese la recta de regresión



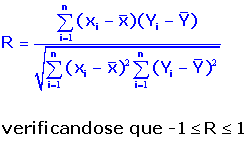
b) Supóngase, con base en el inciso), QUE SE DECIDE ESTIMAR EL MODELO DE REGRESION DADO ANTRES. Obténgase los estimados de los parámetros, sus errores estándar, R2, la SRC y la SEC.

La ecuación que representa a las variables y a los respectivos resultados de estimadores es: Y= 12129,37 + 3.3076 X. con sus respectivos errores:

ee = (1197.351) (0.3117)

Los resultados obtenidos de esta formula son:

SCT = SCE + SCR



R\*2= 0.6968 SCR = 2.65E +0.8

C) Obteniendo el modelo y sus estimadores se puede realizar un analisis entre las variables, en este caso seroa que si bien el gasto alumno aumenta en 1 dólar, la pendiente o estimador beta2 nos indica que el sueldo promedio tendria que aumentar en 3.31 aprox.

El consepto de intersecto en relacion al analisis de este caso y hablando economicamente no tendria ningun signidicado relevante por el momento.

d) El intervalo de confianza al 95% para el beta 2 seria según:



Seria: 3.3076 + 2(0.3117) = (2.6842 y 3.931)

Considerando este intervalo de confianza al 95% y sus resultados se puede concluir y decir que la hipotesos nula no sera Rechazada ya que la verdadera pendiente es cercana y es 0.3.

e) con la tabla y los datos obtenidos anteriormente se p(uede sacar la media y el valor individual promedio los cuales como resultados son considerados iguales ya que se logra obtener 12129.37 + 3.3076 (5000) = 28.667.

Pero se debe obtener datos como el error estandar del valor individual promedio que es a base de dolares con la ecuacion 2382.337.

El error estandar del valor promedio es 520.517 dolares.

Usando la misma formula de intervalos de confianza se obtiene que los de prediccion media son :

28.667+ 2(520.5117) = 27.626

28667 – 2(520.5117)= 29.708

Y los de prediccion individual serian:

28.66752 + 2(520.5117) = 33432 dolares

28.667 – 2(2382.337) = 23902 dolares

f) Una forma de probar la suposicion de la normalidad del termino error se que es el histograma residuall aproximarlo a la curvca normal.

Se obtiene con el estadistico JARQUE-BERA que es el valor de P pero siendo honesto no sabemos como obtener esos resultados por lo que no podemos como grupo asumir si rechazamos o no la hipotesis, sabemos que si hay que considerar que la cantidad de muestras que plantea el estudio debemos fijarla como una muestra grande sino no permite el razonamiento.

5.10 Refiérase al ejercicio 3.20 para construir las tablas ANOVA y probar la hipótesis de que no existe ninguna relación entre la productividad y la compensación salarial. ¿Es válido lo anterior para el comercio y para el no agrícola?

La tabla ANOVA para el comercio es:

|  |  |  |  |
| --- | --- | --- | --- |
| Las fuentes de variación | Suma de cuadrados(SC) | Grados de libertad(g de l) | SPC( división de SC y g de l) |
| Debido a la regresión(SEC) | 38685.997 | 1 | 38685.997 |
| Debido a los residuos(SRC) | 4934.138 | 37 | 133.355 |

Total 43620.135

F= 38685.997 / 133.355 = 290.0978

Basándonos en la hipótesis nula de que no existirá algún tipo de relación entre salarios y productividad en el comercio, observamos que la probabilidad de obtener un valor tal que f sea 0,000 es prácticamente cero. Por lo cual, se puede rechazar la hipótesis nula.

La tabla ANOVA para el no agrícola es:

|  |  |  |  |
| --- | --- | --- | --- |
| Las fuentes de variación | Suma de cuadrados(SC) | Grados de libertad(g de l) | SPC( división de SC y g de l) |
| Debido a la regresión(SEC) | 37887.455 | 1 | 37887.455 |
| Debido a los residuos(SRC) | 5221.585 | 37 | 141.129 |

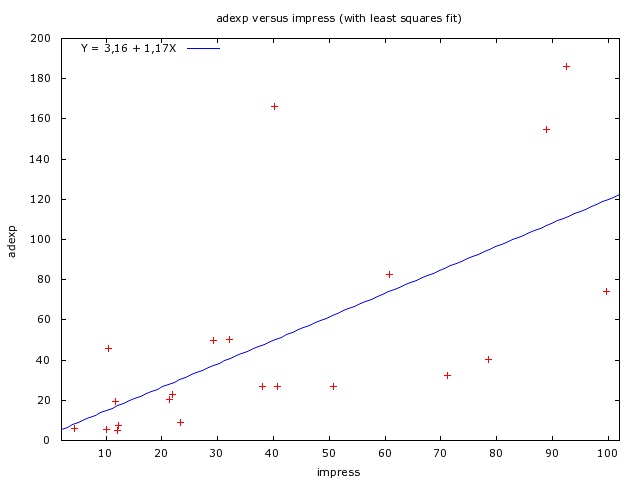
Total 43109.04

F= 37887.455 / 141.129 = 268.459

La hipótesis nula es verdadera, por la probabilidad de obtener un valor f es prácticamente cero, por lo cual, se rechaza la hipótesis.

5.11. Refiérase al ejercicio 1.7.

**a)** Grafique los datos sobre impresiones en el eje vertical y el gasto publicitario en el horizontal. ¿Que tipo de relaciones se observan?



Las dos variables son no lineales

Inicialmente, los gastos de publicidad aumentan el número de impresiones retenidas aumenta pero poco a poco disminuyen .

b) ¿seria apropiado ajustar un modelo de regresión de dos variables a los datos? ¿por qué si o porque no?¿si la respuesta es negativa, ¿qué tipo de modelo de regresión se ajustaría a los datos?¿ se cuenta con las herramientas necesarias para ajustar dicho modelo?

No es apropiado para adaptarse a un modelo de regresión lineal bivariado los datos y además no contamos con las herramientas para adaptarse a un modelo apropiado el cual podría ser apropiado, en caso que se use Y = impresiones retenidas y X2 es el gasto publicitario.

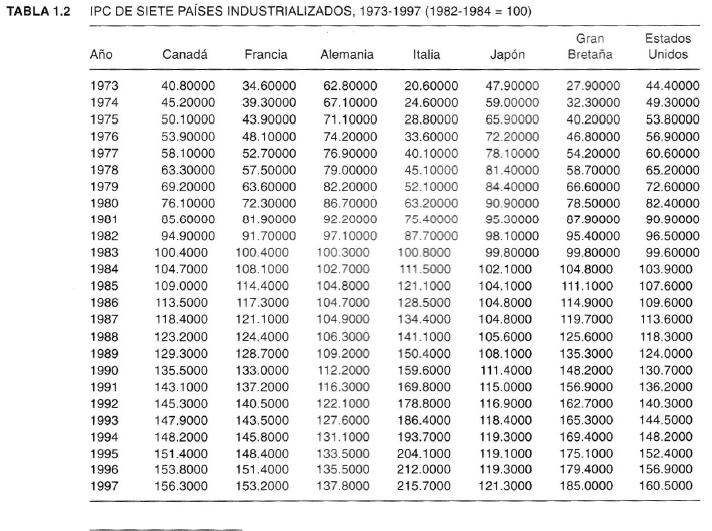
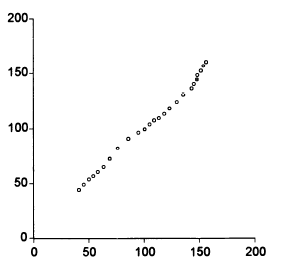
c) Supóngase que no se grafican los datos y que simplemente se ajustan el modelo de regresión con dos variables a los datos. Obténgase los resultados usuales de la regresión.

Yi= 22.163 + 0.3631 Xi

ee (7.089) (0.0971) = 0.424

5.12.- Refiérase al ejercicio 1.1

a). Grafíquese el índice de precios al consumidor (IPC) estadounidense en relación con el canadiense. ¿Qué revela la gráfica?



El grafico muestra que la inflación de los dos países se mueve en conjunto, además se observa una tendencia lineal, quiere decir un comportamiento similar de las variables, son dependientes una de la otra

IPC EEUU

b) Supóngase que se desea predecir el IPC de EU con base en el canadiense. Desarróllese un modelo apropiado.

c) Pruébese la hipótesis de que no existe relación entre ambos IPC. Utilícese [α =5%]. Si se rechaza la hipótesis nula. ¿Significa que el IPC canadiense “condiciona” al IPC estadounidense? ¿Por qué sí o por qué no?

B) y C) como se muestra el comportamiento en el grafico la relación entre las dos variables es positiva, Uno puede facilitar el rechazo de la hipótesis nula que no está relacionada entre las dos variables, como el valor t obtenido bajo la hipótesis 53.33 y el p Valium que se obtuvo con un valor t prácticamente cero.  
Aunque las dos tasas estén posiblemente relacionadas por algún factor en común, no podemos inferir que hay causalidad, ya que la regresión no necesariamente implica causalidad

5.13. Refiérase al ejercicio 3.22.

a) Estímese las dos regresiones dadas en dichos ejercicios y obténgase los errores estándar, así como los otros resultados usuales.

Las 2 regresiones son las siguientes:

Precio del orot = 186.183 + 1.842 IPCt  
ee = (125.403) (1.215)

t= (1.848) (1.515) r2 = 0.150

NYSEt = 102.060 + 2.129 IPCt

ee= (23.767) (0.230)

t= (-4.294) (9.247) r2 = 0.868

b) Pruébese la hipótesis de que las perturbación en los dos modelos de regresión están normalmente distribuidas.

La estadística Jarq-bera para la ecuación del precio del oro es de 4.751, con un valor de p 0,093, la estadística de JB para la ecuación NYSE Index es 1,218 con un valor de

P 0,544 al 5% de significación, en ambos casos, no rechazamos la hipótesis de la normalidad.

c) En la regresión del oro, pruébese la hipótesis de que β2 =1; es decir, que existe una relación uno a uno entre los precios del oro y el IPC(es decir, el oro es una perfecta barrera). ¿Cual es el valor p del estadístico estimado?

Dado que el coeficiente de la pendiente en la regresión del precio del oro no es estadísticamente diferente de cero, no tiene sentido para encontrar si es diferente de 1.

d) Repítase el paso c) para la regresión del índice de la bolsa de Nueva YORK (NYSE). ¿La inversión en el mercado de valores representa una barrera perfecta contra la inflación? ¿Cuál es la hipótesis nula que se esta probando? ¿Cuál es el valor P?

e) Entre el oro y el mercado de valores ¿Cuál inversión se elegiría? ¿En que se basaría la decisión?

d) y e) Si realizamos los cálculos usando el procedimiento de prueba de la t habitual, se obtiene:

t = 2.129-1/0.0230 = 4,91

Dado que este valor t supera el valor t crítico de 2.160, se rechaza la hipótesis nula. El coeficiente estimado es en realidad mayor que 1. Para este período de la muestra, la inversión en el mercado de valores fue probablemente una cobertura contra la inflación. Sin duda fue una mejor protección contra la inflación que la inversión en oro

5.14. La tabla 5.6 presenta los datos observados sobre el producto nacional bruto y cuadro definiciones de la oferta monetaria de estado unidos durante el periodo de 1970 – 1983. Al hacer regresiones del PNB con respecto a las diversas definiciones de dinero, obtenemos los resultados que parecen en la tabla 5.7:

Los monetaristas o partidarios de la teoría cuantitativa sostienen que el ingreso nominal (es decir, el PNB nominal ) esta determinado en gran medida por cambios en la cantidad o en la reserva de dinero, auque no hay consenso sobre la definición “correcta “ de dinero . Dados los resultados de la tabla anterior, considérense las siguientes preguntas.

a) ¿Cuál definición de oferta monetaria parece estar estrechamente relacionada con el PNB nominal?

No parece ser mejor que los demás. Todos los resultados estadísticos son muy similares. Cada coeficiente de la pendiente es estadísticamente significativo al nivel de 99% de confianza

b) Puesto que los términosson uniformemente elevados. ¿significa este hecho que la elección de la definición de dinero no tiene importancia?



Siempre una gran . No se pueden utilizar para decidir qué agregado monetario es mejor. Sin embargo, esto no quiere decir que no importa lo que la ecuación de usar

