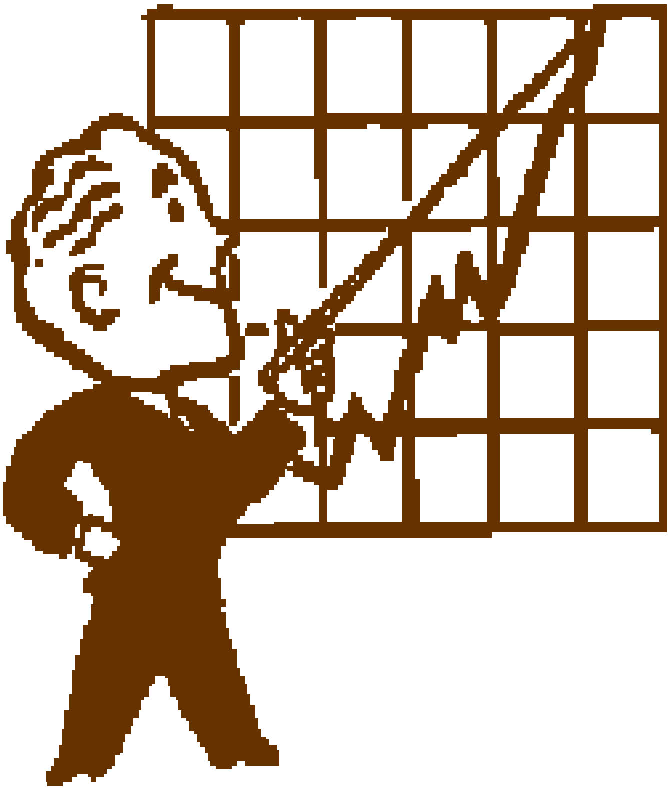
UNIDAD III

# ¿Qué es la inferencia en el modelo líneal?

# UNIDAD III

# ¿Qué es la inferencia en el modelo lineal?

**Con formato:** Fuente: Times New Roman



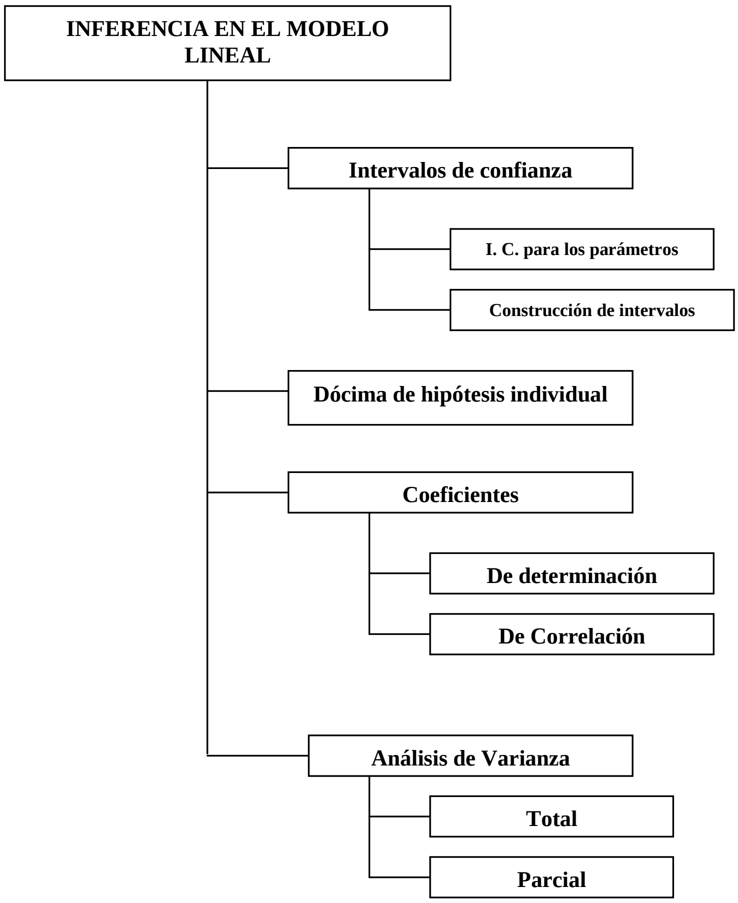
**"Toda empresa extraordinaria sólo puede lograrse mediante la aplicación de una extrema cautela en lo relacionado con sus comienzos y bases"**

**I Ching**

- **¿Qué es un intervalo de confianza, una dócima de hipótesis?**
- **¿Qué es un coeficiente de determinación, de correlación y de determinación?**
- **¿En que consiste la inferencia en el modelo lineal, el análisis de varianza total y parcial?**
- **¿Qué permite la prueba de Reset de Ramsey?**
- **¿En qué consiste el análisis de la estabilidad estructural?**

# INFERENCIA EN EL MODELO LINEAL

## ESQUEMA CONCEPTUAL




### COMPETENCIAS A LOGRAR

| CONCEPTUAL | PROCEDIMENTAL | ACTITUDINAL |
|---|---|---|
| **Explica que es una inferencia, intervalos y coeficientes utilizados.** | **Ejecuta el proceso de inferencia para explicar las variables.** | **Valora y analiza los modelos de predicción** |

**CONCEPTOS –CLAVE**

**Inferencia, intervalos, coeficientes, varianza.**

# LECCIÓN 1

## INTERVALO DE CONFIANZA

### 1. ESTIMACIÓN

- La estimación puntual.
- La estimación de intervalo

**La estimación puntual.-** Se vale de un estadístico para estimar el parámetro en un sólo valor o punto.

**Una estimación de intervalo**.- Es la que define un intervalo probable (nivel de confianza) dentro del cual puede estar el parámetro desconocido.

**Ejemplo de estimadores**
Si se desea conocer el nivel de ingresos promedio mensual de los hogares del departamento de Lima en el año 2004, es difícil calcular la media de toda la población. Es más fácil calcular la media de una muestra de hogares y a partir de ella hacer una estimación de la media poblacional.

El **ejemplo de estimación puntual** de la media poblacional sería el valor que resulta de una muestra de n = 500 hogares y hallar $\overline{X} = 437.10$ nuevos soles.

El **ejemplo de estimación de intervalo** sería asegurar con un 95% de confianza que el ingreso promedio se encuentra entre 325.60 y 548.60 nuevos soles.

Es decir una estimación puntual utiliza un número o valor único para determinar una estimación del parámetro. Un intervalo de confianza denota un rango o recorrido dentro del cual se podría encontrar el parámetro dado un nivel de confianza.

**Nivel de confianza**.
Es la probabilidad (1-$\alpha$), (generalmente expresada en porcentaje) de contener dentro de sus límites el valor verdadero del parámetro. Los más usados convencionalmente son tres: 99, 95 y 90%. Un ejemplo para este caso estaría formado por el ingreso promedio (media poblacional), de una empresa dedicada a la venta de artículos para oficina que espera obtener al 95% de probabilidad entre 35000 y 38000 nuevos soles de ingresos para el mes de junio del 2004.

**El principio del intervalo de confianza.-** Todo intervalo de confianza tiene un límite superior y un límite inferior de confianza.

### 2. INTERVALO DE CONFIANZA PARA LOS PARÁMETROS

A fin de establecer los intervalos de confianza para los coeficientes de regresión ($\beta_i$) y teniendo la varianza poblacional desconocida se construye un intervalo asumiendo que esta variable tiene una distribución estadística “t- student”.Es así que el intervalo de confianza para $\beta_1$ se define como:

$$\Pr ob(\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2} < \beta_1 < \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}) = 1 - \alpha$$

El cual puede ser interpretado como los límites de confianza donde puede estar el parámetro $\beta_1$, con una probabilidad de $1-\alpha$ :

**El nivel de confianza**: $1-\alpha$

**Límites de confianza inferior**: $\hat{\beta}_1 - \hat{\sigma}_{\beta_1} t_{\alpha/2}$

**Límites de confianza superior**: $\hat{\beta}_1 + \hat{\sigma}_{\beta_1} t_{\alpha/2}$

Por lo tanto se tiene que el intervalo de confianza para $\beta_1$ será:

$$\hat{\beta}_1 - \hat{\sigma}_{\beta_1} t_{\alpha/2}, < \beta_1 < \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2} \text{ con un nivel de confianza de } 1\text{-}\alpha$$

**Ejemplo ilustrativo:**

El cuadro siguiente, presenta los datos observados de las tasas anuales de retorno ($X_t$) sobre el fondo "Afuture" ($Y_t$), un fondo cuyo principal objetivo de inversión es obtener una ganancia máxima sobre el capital y sobre el portafolio del mercado con base en el índice de Fisher (período 1990:2001).

| **AÑO** | $Y_t$ | $X_t$ |
|---|---|---|
| 1990 | 67.50 | 19.50 |
| 1991 | 19.20 | 8.50 |
| 1992 | -35.20 | -29.40 |
| 1993 | -41.00 | -26.50 |
| 1994 | 63.70 | 61.90 |
| 1995 | 19.20 | 45.00 |
| 1996 | 3.60 | 9.50 |
| 1997 | 20.00 | 14.00 |
| 1998 | 40.30 | 35.00 |
| 1999 | 38.00 | 32.00 |
| 2000 | 21.00 | 11.50 |
| 2001 | 15.30 | 12.20 |

Construya un intervalo de confianza al 95 % para los parámetros que relacionan el fondo "Afuture" ($Y_t$ ) respecto de las tasas anuales de retorno ($X_t$), ¿Qué puede decir al respecto?

**Solución**

**Planteamiento del modelo:**

$$Y = \beta_1 + \beta_2 X + \mu$$

Donde:

$Y_t$ = Fondo Afuture

$X_t$ = Tasa anual de retorno

Aplicando el paquete E-views, se obtienen los siguientes resultados:

| | | | | |
|---|---|---|---|---|
| Dependent Variable: Y<br>Method: Least Squares<br>Date: 08/27/02 Time: 15:13<br>Sample: 1990 2001<br>Included observations: 12 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 2.209520 | 6.334633 | 0.348800 | 0.7345 |
| X | 1.061520 | 0.212218 | 5.002038 | 0.0005 |
| R-squared | 0.714452 | Mean dependent var | | 19.30000 |
| Adjusted R-squared | 0.685897 | S.D. dependent var | | 32.97040 |
| S.E. of regression | 18.47822 | Akaike info criterion | | 8.822074 |
| Sum squared resid | 3414.446 | Schwarz criterion | | 8.902892 |
| Log likelihood | -50.93244 | F-statistic | | 25.02038 |
| Durbin-Watson stat | 0.918447 | Prob(F-statistic) | | 0.000536 |

En el cual se observa que los valores estimados de los parámetros son:

$\hat{\beta}_1 = 2.209520$ ;que significa que cuando la Tasa anual de retorno es nula el Fondo Afuture es de 2.209520.

$\hat{\beta}_2 = 1.061520$ Es decir que por una unidad porcentual que aumente la tasa de retorno, el fondo Afuture aumentará en 1.06152.

Con los datos de las salidas, se calcularán los intervalos de confianza para dichos parámetros estimados con un nivel de confianza del 95%.

Por lo tanto se puede decir que el fondo Afuture mínimo ($\beta_1$) estará entre -11.904 y 16.323 con un 95% de confianza.

$$P\left(\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2}, < \beta_1 < \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}\right) = 1 - \alpha$$

$$P(2.209520 - 6.334633 * 2.228 < \beta_1 < 2.209520 + 6.334633 * 2.228) = 1 - 0.95 = 0.05$$

$$P(-11.904 < \beta_1 < 16.323) = 5\%$$

Asimismo la razón de aumento del fondo Afuture ($\beta_2$) podrá estar entre 0.589 y 1.534 por cada unidad de aumento de la tasa de retorno.

$$P\left(\hat{\beta}_2 - \hat{\sigma}_{\beta 2} t_{\alpha/2}, < \beta_2 < \hat{\beta}_2 + \hat{\sigma}_{\beta 2} t_{\alpha/2}\right) = 1 - \alpha$$

$$P(1.06152 - 0.212218 * 2.228 < \beta_2 < 1.06152 + 0.212218 * 2.228\ ) = 1 - 0.95 = 0.05$$

$$P(0.589 < \beta_2 < 1.534) = 5\%$$

## 3. DÓCIMAS DE HIPÓTESIS

Se realiza con el fin de determinar si los coeficientes son significativos, es decir diferentes de cero.

1. Planteamiento de hipótesis

   $H_0$: $\beta_i=0$ hipótesis nula
   $H_1$: $\beta_i\neq 0$ hipótesis alternativa

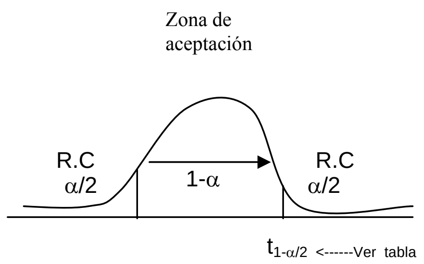
2. Estadístico de contraste
   Utilizando la distribución t-student el estadístico de contraste se define como:

$$t = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\beta_i}} = \frac{\hat{\beta}_i - 0}{\hat{\sigma}_{\beta_i}} = \frac{\hat{\beta}_i}{\hat{\sigma}_{\beta_i}}$$

3. Nivel de significancia ($\alpha$): Es la probabilidad de error al afirmar que el coeficiente es diferente de cero, también se le conoce como la probabilidad de cometer el Error de tipo I, es decir, Rechazar Ho cuando es verdadera.

4. Establecer una región crítica (RC ) o zona de rechazo de la hipótesis nula.

$$RC = (t / t_c > t_{(n-k,\ 1-\alpha/2)} \text{ ó } t_c < -t_{(n-k,\ 1-\alpha/2)})$$




   ($1-\alpha$) zona de aceptación de hipótesis nula $\beta_i = 0$

$$\text{Si: } -t_{1-\alpha/2} < t_c < t_{1-\alpha/2} \text{ no se rechaza Ho, es decir } \beta_i = 0$$

5. Comparar el t calculado con el t de tabla con un nivel de significación $\alpha$:

   **Regla de decisión*:***

   Si: $t_c > t_{(n-k,\ 1-\alpha/2)}$ ó $t_c < -t_{(n-k,\ 1-\alpha/2)}$ (t-student con n-k grados de libertad) entonces **se rechaza $H_0$**

   Donde:
   n: Es el tamaño de la muestra (número de unidades o casos)
   k: Es el número de parámetros estimados (número de variables más uno)

Si el t calculado cae dentro de la región crítica (RC), se rechaza la hipótesis nula, por lo tanto $\beta_i \neq 0$, en consecuencia $X_i$ si explica el comportamiento de la variable dependiente. Se concluye que el coeficiente de la variable es **significativo.**

## Enfoque del Intervalo de Confianza

1. El primer paso es construir un intervalo de confianza para los parámetros (coeficientes de regresión):

$$\Pr ob(\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2} < \beta_1 < \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}) = 1 - \alpha$$

2. El siguiente paso es comparar el β de la hipótesis nula con el intervalo establecido.

   **Regla de decisión:**
   Si el β de la Hipótesis nula está dentro del intervalo de confianza se acepta la hipótesis nula; contrariamente, si el β está fuera del intervalo se rechaza la hipótesis.

**Ejemplo:**
Utilizando las salidas del ejemplo anterior, verificar si la variable X (Tasa anual de retorno) explica el comportamiento de la variable Y (Fondo Afuture)

**Solución**
- Del ejercicio anterior se tiene**:**

$\hat{\beta}_1 = 2.209520$ $\hat{\sigma}_{\beta_1} = 6.334633$

$\hat{\beta}_2 = 1.061520$ $\hat{\sigma}_{\beta_2} = 0.212218$

1. Planteamiento de hipótesis
$H_0$: $\beta_2$=0 La tasa anual de retorno no explica el comportamiento del Fondo Afuture
$H_1$: $\beta_2 \neq$0 La tasa anual de retorno explica el comportamiento del Fondo Afuture

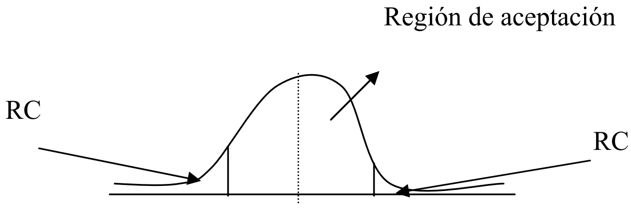
2. Estadístico de Contraste
Utilizando la distribución t-student el estadístico de contraste se define:

$$t_c = \frac{\hat{\beta}_2 - 0}{\hat{\sigma}_{\beta_2}} = \frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}} = \frac{1.061520}{0.212218} = 5.0020 \quad ,$$

Donde $t_{n-k,\alpha/2} = t_{12-2,0.05/2} = t_{10,\ 0.025} = 2.228$

3. Al igual que en el ejemplo anterior el nivel de significancia (α) escogido es de 0.05

4. Región crítica (RC)




como $|t_c| = 5.0020 > t_{10,0.025} = 2.228$ se Rechaza Ho

Por lo tanto se concluye que será rechazada la hipótesis $H_0$, es decir que el comportamiento del fondo “Afuture” (Y) es explicado por la tasa de retorno (X)

- **Por el Enfoque del Intervalo de Confianza:**

$$P\left(\hat{\beta}_2 - \hat{\sigma}_{\beta 2} t_{\alpha/2}, < \beta_2 < \hat{\beta}_2 + \hat{\sigma}_{\beta 2} t_{\alpha/2}\right) = 1 - \alpha \quad \text{, por lo anteriormente mostrado se tiene:}$$

$$P\left(1.06152 - 0.212218 * 2.228 < \beta_2 < 1.06152 + 0.212218 * 2.228\right) = 1 - 0.05 = 0.95$$

$$\text{por lo tanto:} \qquad P\left(0.589 < \beta_2 < 1.534\right) = 95\%$$

Lo que significa que $\beta_2$ estará entre 0.589 y 1.534 con un 95% de confianza.
Luego, como la hipótesis nula afirma que $\beta_2 = 0$; como no pertenece al intervalo <0.589, 1.534> se rechaza la hipótesis Ho. Es decir, la variable tasa anual de retorno explica el comportamiento de la variable Fondo Afuture.

## 4. PRUEBA DE NORMALIDAD DE JARQUE-BERA (JB)

El supuesto de normalidad de las perturbaciones es importante por cuanto dependiendo de la validez de dicho supuesto, podremos hacer inferencia estadística sobre los parámetros y cualquier prueba de hipótesis.

Ho: Las perturbaciones tienen una distribución normal
$H_1$: Las perturbaciones no tienen una distribución normal.

El test de Jarque Bera permite contrastar la normalidad de las perturbaciones. Si se cumple la hipótesis nula, la cual ha sido planteada inicialmente, indicará la presencia de una distribución normal, por lo que la representación de su histograma respectivo será en forma de una campana simétrica y con un apuntamiento similar al de una distribución normal de la cual ya se hablo anteriormente en la UNIDAD I.

La prueba de normalidad de Jarque Bera (JB) consiste en determinar como se encuentra afectado su valor por la presencia de un mayor apuntamiento (mayor a 3) o menor asimetría (cercano a cero) de las perturbaciones.

$$JB = n\left[\frac{A^2}{6} + \frac{(C-3)^2}{24}\right]$$

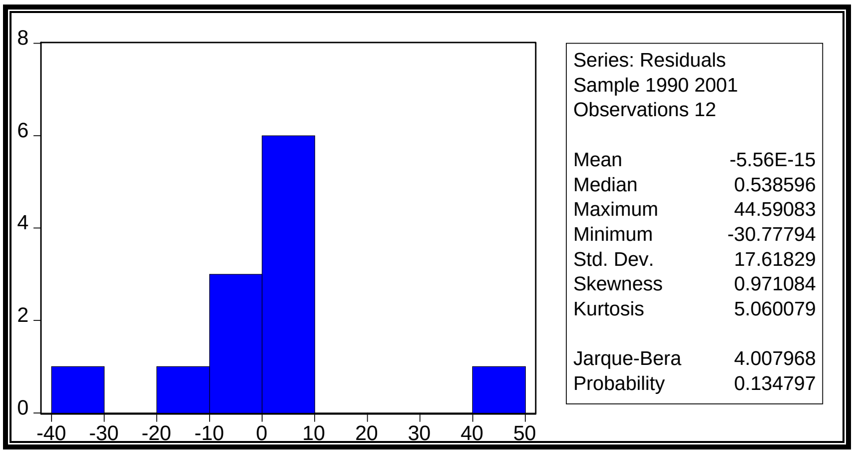
Donde “A” significa asimetría y “C” apuntamiento o curtosis.

$$A = \frac{\left[E(x-\mu)^3\right]^2}{\left[E(x-\mu)^2\right]^3} \qquad C = \frac{E(x-\mu)^4}{\left[E(x-\mu)^2\right]^4}$$

Habitualmente la probabilidad de rechazar la hipótesis nula es del 5%

**Ejemplo:**

En el gráfico adjunto (elaborado en el paquete E-views[1]), para el modelo anterior sobre el fondo “Afuture” ($Y_t$) en función a las tasas anuales de retorno ($X_t$), podemos observar una serie de indicadores relacionados con el término de perturbación a partir del análisis de los residuales en la serie 1990-2001 (e) que como se sabe son sus estimadores :




Se observa que el valor del estadístico es de 4.007; al mismo tiempo la probabilidad asociada al estadístico Jarque Bera (0.134797).Es decir la probabilidad de rechazar la hipótesis nula es de 13%; que es mayor al 5%. Por lo tanto no podemos rechazar la hipótesis nula de normalidad de los errores.

Se concluye que se cumple el supuesto de normalidad de los errores del Modelo Lineal General.

[1] Ver Anexo: “La Inferencia Estadística y el Programa E-Views”

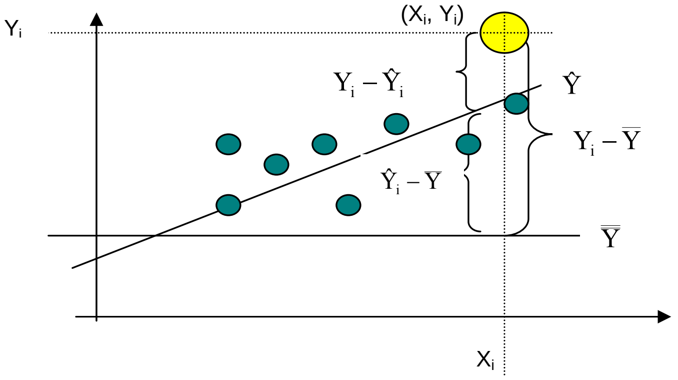
# LECCIÓN 2

## COEFICIENTE DE DETERMINACIÓN, CORRELACIÓN Y DETERMINACIÓN MÚLTIPLE

### 1. COEFICIENTE DE DETERMINACIÓN ($R^2$)

Es un indicador de la **bondad de ajuste** de la línea de regresión, que mide la proporción de la variación total en la variable dependiente Y, que "se explica" o "se debe a" la variación de la(s) variable(s) independiente(s) X.

Para obtener el $R^2$ será necesario descomponer la variación de Y, lo cual se puede ilustrar en el gráfico siguiente:




En este gráfico se observa que la variación total entre el punto $Y_i$ y el promedio $\overline{Y}$: $Y_i - \overline{Y}$ se puede dividir en la variación $Y_i - \hat{Y}_i$ y $\hat{Y}_i - \overline{Y}$

Planteada la relación inicial, ésta se mantendrá cuando se establezcan las relaciones a partir de las sumatorias de sus desviaciones cuadráticas, por lo tanto:

$$\sum(Y_i - \overline{Y})^2 = \sum\left(Y_i - \hat{Y}_i\right)^2 + \sum\left(\hat{Y}_i - \overline{Y}\right)^2$$

$$SCT = SCR + SCE$$

Donde:

**SCT** (Suma Total del Cuadrado): Variación total del $Y_i$ observado con respecto a su media muestral,

$$SCT = \sum(Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

**SCR** (Suma de los Cuadrados Residuales): Variación residual o no explicada de los valores de Y respecto a la línea de regresión.

$$SCR = \sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

**SCE** (Suma de los cuadrados explicados): Variación de los valores estimados $Y_i$ con respecto a su media.

$$SCE = \sum\left(\hat{Y}_i - \overline{Y}\right)^2 = \sum(\hat{\beta}_1 - \hat{\beta}_2 X_i - \overline{Y})^2$$

Luego $\mathbf{R^2}$ se define como:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \qquad 0 \le R^2 \le 1$$

El coeficiente de determinación se interpreta como la proporción de la variación total de Y que la regresión es capaz de explicar. Es decir, el $\mathbf{R^2}$ mide la efectividad que posee la variable independiente X para explicar la variación que la variable dependiente experimenta a lo largo de la muestra. Por lo tanto, cuando $\mathbf{R^2}$ es muy cercano a 1 se dice que el modelo de regresión es capaz de explicar un alto porcentaje de las variaciones que registra la variable explicada.

**Propiedades**

1. Es una cantidad no negativa.

2. Sus límites son $0 \le R^2 \le 1$, por lo que R variará entre cero y uno.

   $R^2$=1 cuando el ajuste es perfecto, es decir los valores observados coinciden perfectamente con la recta estimada

   $R^2 \approx 0$ es decir que no hay relación entre la variable dependiente y las variables explicativas.

Este $R^2$ no mide el grado de asociación entre x e y, para ello se acude a otro indicador que es el Coeficiente de correlación (R)

**Nota:** Para la regresión lineal múltiple la interpretación es similar, es decir el $R^2$ mide el nivel de ajuste de las variables independientes con respecto a la variable dependiente, su cálculo manual es más complejo por lo cual necesitaremos de un paquete (software) estadístico (SPSS).

## 2. COEFICIENTE DE CORRELACIÓN

Es una medida de asociación lineal entre dos variables

**Poblacional** **Muestral**

$$r = \frac{Cov(X,Y)}{\sqrt{\sigma_x^2 \sigma_y^2}} \qquad = \qquad \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1} \frac{\sum (Y_i - \overline{Y})^2}{n-1}}}$$

Donde $r = \pm\sqrt{R^2} = \pm\sqrt{\frac{SCE}{SCT}} = \pm\sqrt{1 - \frac{SCR}{SCT}}$ ; siendo el signo de la correlación el mismo al del coeficiente estimado.

**Propiedades**

Sus límites están entre: $-1 \leq r \leq 1$

- Es de naturaleza simétrica, es decir el coeficiente de correlación entre X e Y ($r_{xy}$) es igual al coeficiente de correlación entre Y y X ($r_{yx}$)

- Si X, Y son estadísticamente independientes, el coeficiente de correlación es cero; pero si r = 0 no implica necesariamente independencia entre las variables.

- Es una medida de asociación lineal, es decir mide la asociación lineal entre dos variables.

- El signo negativo indica una relación inversa es decir a medida que aumenta (o disminuye) la variable X, la variable Y disminuye (o aumenta), mientras que el signo positivo indica una relación directa, es decir a medida que aumenta (o disminuye) la variable X , la variable Y también aumenta (o disminuye)

**Nota:** Para el caso múltiple el Coeficiente de Correlación Global R varía entre 0 y 1, y se interpreta como el grado de asociación entre las variables explicativas y la variable dependiente.

## 3. COEFICIENTE DE DETERMINACIÓN MÚLTIPLE CORREGIDO

En el caso de regresión lineal múltiple, en ocasiones se desea comparar el nivel de ajuste que pueden proporcionar diferentes combinaciones de variables explicativas con respecto a una variable dependiente común, para ello el $R^2$ no es muy preciso, pues a medida que el número de variables independientes se incrementa, el $R^2$ tiende a incrementarse, es decir que el valor $R^2$ favorecería a los modelos con mayor cantidad de variables, para corregir esta deficiencia se construye un coeficiente $R^2$ corregido por los grados de libertad.

$$\overline{R}^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)/n-k}{\sum (Y_i - \overline{Y})/n-1} = 1 - \frac{SCR/n-k}{SCT/n-1}$$

**Ejercicio ilustrativo:**

Volviendo al modelo sobre el fondo "Afuture" ($Y_t$) en función a las tasas anuales de retorno (Xt), y las salidas, obtenemos el coeficiente de determinación y el Coeficiente de determinación múltiple corregido:

| | |
|---|---|
| R-squared | 0.714452 |
| Adjusted R-squared | 0.685897 |

El $\mathbf{R^2}$ es 0.714; es cercano a 1, por lo tanto se concluye que el modelo de regresión es capaz de explicar un alto porcentaje de las variaciones que registra la variable explicada (fondo "Afuture").

Evaluando el $\overline{\mathbf{R}}^2$, este también es alto 0.685.

El coeficiente de correlación se calcula mediante: $r = \pm\sqrt{R^2}$ y dado que el coeficiente $\hat{\beta}_2 = 1.061520$, es positivo, entonces el coeficiente de correlación también toma el signo positivo:

$$r = +\sqrt{R^2} = \sqrt{0.714} = 0.845$$

Se concluye que la asociación lineal entre la variable independiente e independiente, es "alta y directa". Es decir, cuando aumenta la tasa de retorno ($X_t$), aumenta el fondo "Afuture" ($Y_t$).

## LECCIÓN 3

### INFERENCIA EN EL MODELO LINEAL

El proceso de inferencia consiste en establecer la validez de determinadas afirmaciones acerca de los parámetros (desconocidos) utilizando un estimador obtenido a partir de una muestra, pero del cual se puede determinar su distribución muestral.

### 1. ANALISIS DE VARIANZA

El análisis de varianza tiene por finalidad investigar la explicación conjunta de todas las variables explicativas que participan en el modelo. Es decir, mide la significancia del modelo de regresión lineal múltiple, a partir del estudio de los componentes de la variabilidad total.

**SCT = SCR + SCE**

El análisis de varianza se interpreta mediante la siguiente tabla ANOVA (Análisis Of Variance):

| Fuentes de Variación | Grados de libertad | Suma de cuadrados | Cuadrado Medio | Estadístico F F-calculado |
|---|---|---|---|---|
| Explicada por las variables "$X_i$" que intervienen en el modelo | k-1 | SCE | $CME = \frac{SCE}{k-1}$ | $Fc = \frac{CME}{CMR}$ |
| Residual | n-k | SCR | $CMR = \frac{SCR}{n-k}$ | |
| Total | n-1 | SCT=SCE+SCR | | |

Donde:
SCT: Suma Total del Cuadrado
SCR: Suma de los Cuadrados Residuales
SCE: Suma de los Cuadrados Explicados
CME: Cuadrado Medio Explicada
CMR: Cuadrado Medio Residual

k: Número de parámetros estimados (número de variables independientes más uno).
n: Tamaño de muestra o número de unidades de análisis (personas, hogares, empresas, áreas de trabajo, etc.)

En esta tabla se distinguen las tres fuentes de variación y su finalidad es construir el estadístico de contraste F, que podrá ser usado para contrastar la hipótesis de significación conjunta de las variables independientes

**a) Planteamiento de hipótesis (significancia del modelo)**

El análisis de varianza plantea una prueba global de los parámetros (coeficientes de regresión) con la lógica de que, si al menos uno de los coeficientes de regresión es diferente de cero el modelo es significativo, por ello las hipótesis se plantean:

$H_0$: $\beta_i = 0$ para i = 1,2,…, k-1
$H_1$: Al menos un $\beta_i \neq 0$ para i = 1,2,…, k-1

**b) Estadístico de contraste (F)**

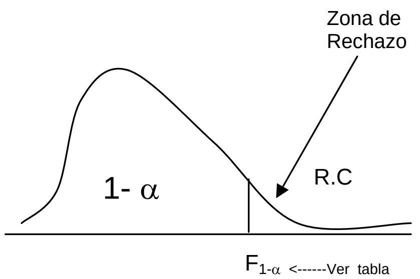
El estadístico que nos ayuda a probar la hipótesis anterior es como se menciona en la tabla ANOVA:

$Fc = \frac{CME}{CMR}$, la cual se distribuye con una F-Fisher con (k-1, n-k) grados de libertad

**c) Nivel de Significancia ($\alpha$)**

Se plantea un nivel de error al rechazar Ho, normalmente es del 5%

**d) Región Crítica (RC)**




Se acepta $H_0$ si $F \leq F_{\alpha;k-1;n-k}$
Se rechaza $H_0$ si $F > F_{\alpha;k-1;n-k}$ ,es decir, rechazamos $H_0$ para valores grandes de F.

## 2. ANÁLISIS DE VARIANZA PARCIAL

Usado para la comparación de modelos, en donde es posible determinar si la incorporación de alguna variable o grupo de ellas pueden explicar mejor el modelo.

| Fuente de Variación | Suma de Cuadrados | G.L. | Media de Cuadrados |
|---|---|---|---|
| Debido a: $X_1,....,X_r$ | $SCE_I$ | r – 1 | $SCE_I / r-1$ |
| Debido a: $X_1,....,X_r,X_{r+1},...,X_{r+s}$ | $SCE_{II}$ | (r – 1)+s | $SCE_{II} /(r-1+s)$ |
| Debido a: $X_{r+1},...,X_{r+s}$ | $SCE_s = SCE_{II} - SCE_I$ | s | $SCE_s/s$ |
| Residual del $Mod_{II}$ | $SCR_{II}$ | n – (r+s) | $SCR_{II}/n\ \ (r+s)$ |
| Total | SCT | n - 1 | |

r: Número de variables consideradas en el primer modelo
s: Número de variables consideradas en el segundo modelo

SCR: Suma de los Cuadrados Residuales
$SCE_I$: Suma de los Cuadrados Explicados del primer modelo
$SCE_{II}$: Suma de los Cuadrados Explicados del primer modelo

Entonces:

$$F_C = \frac{SCE_s / s}{SCR_{II} / n-(r+s)} \quad \text{( Valor comparado con el obtenido de tabla)}$$

### a) Planteamiento de hipótesis

$H_0$: $\beta_{r+1} = \beta_{r+2} = ..... = \beta_s = 0$

$H_1$: $\beta_{r+1} \neq \beta_{r+2} \neq ..... \neq \beta_{r+s} \neq 0$

### b) Región Crítica (RC)

Bajo el enfoque de la prueba de significancia, se construye la región crítica de la siguiente manera:

$$R.C. = \{ F_C > F \}$$

Siendo: $F = F_{s,n-(r+s)}$

Se acepta $H_0$ si $F_c \leq F$

Se rechaza $H_0$ si $F_c > F$ Es decir, rechazamos $H_0$ para valores grandes a F, es decir, la incorporación de variables mejoraría el modelo.

**Ejemplo:**

Se tiene los siguientes modelos para el consumo (C) de las familias

$$C = \beta_1 + \beta_{21} YND \qquad (I)$$
$$C = \beta_1 + \beta_2 YND + \beta_3 PR + \beta_4 IT \qquad (II)$$

Donde:
YND : Ingreso Disponible
PR : Precios
TI : Tasa de interés

Además la tabla de análisis de varianza es:

| Fuente de Variación | Suma de Cuadrados | G.L. | Media de Cuadrados |
|---|---|---|---|
| Debido a: **YND** | $SCE_I$ = 963188.02 | 2 – 1 | 963188.02 / 1 |
| Debido a: **YND, PR, TI** | $SCE_{II}$ = 1002196.15 | (2– 1)+2 | 100219.15 / 3 |
| Debido a: **PR, TI** | $SCE_s$ =1002196.15 – 963188.02 = 39008.13 | 2 | 39008.13 / 2... (A) |
| Residual del $Mod_{II}$ | SCR = 155862.6 | 19 - 4 | 155862.60 /15...(B) |

Entonces: $F_C = \frac{A}{B} = \frac{19504.07}{10390.84} = 1.877$ (que se compara con el de la tabla)

$F_{2,15;\,0.05}$ = 3.68 (valor obtenido de tabla)

Dado que: $F_C = 1.877 < F_{2,15;\,0.05} = 3.68$.
Se concluye que la incorporación de las variables precios relativos y la tasa de interés general no mejoran la explicación del modelo, estando ya incorporada la variable ingreso disponible.

# LECCIÓN 4

## CORRECTAMENTE ESPECIFICACIÓN DEL MODELO: TEST DE RESET DE RAMSEY

La prueba de Reset de Ramsey permite comprobar la correcta especificación polinómica de un modelo estimado.

El contraste se basa en la prueba de regresión aumentada tal como se indica a continuación.

- Si la regresión actual es:

$$Y = X\beta + \mu$$

- El test construirá la siguiente relación:

$$Y = X\beta + Z\varphi + \mu$$

Donde las variables en la matriz Z son regresores presumiblemente omitidos en la regresión original. La prueba de hipótesis que se llevará a cabo, intentará contrastar si $\varphi$ es nulo.

Supongamos por ejemplo que el modelo inicial es el siguiente:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$$

Si la verdadera forma del modelo es la que se muestra a continuación, se tiene una especificación incorrecta del modelo inicial, debido a la omisión de la variable $X_1$ elevada al cuadrado.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varphi X_3^2 + \mu$$

Luego, la estructura de la ecuación coincide con la planteada para el test.

Para un segundo ejemplo supóngase que la especificación correcta del modelo fuera:

$$Y = \hat{a}_1 X_2^{\hat{a}_2} X_3^{\hat{a}_3} + \mu$$

Dicha forma corresponde a un modelo no lineal, sin embargo es posible contrastar esta hipótesis mediante la estructura de la ecuación de prueba del test de Ramsey, ya que la matriz Z podría estar construida de las potencias enteras productos cruzados de todas las variables explicativas incluidas en el modelo original.

Además para realizar esta prueba no hace falta incluir explícitamente todas las potencias y variables cruzadas en el modelo de prueba, ya que para este fin puede utilizarse el término ajustado del modelo.

Obsérvese que, dado: $Y = \hat{a}_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3$

Se tiene:

$$\hat{Y}^2 = (\hat{a}_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3)^2$$
$$\hat{Y}^3 = (\hat{a}_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3)^3$$

En el desarrollo de las potencias anteriores se tendrán potencias y productos cruzados de todas las variables incluidas en el modelo original.

La implementación del Test de Ramsey, se realiza en dos etapas.

- En la primera, se estima en el modelo sujeto a análisis en su forma original:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \mu$$

- En la segunda, se toma la serie estimada por los parámetros de la regresión anterior y se anexan sus potenciales enteras a la misma regresión como parámetros auxiliares:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + (\varphi_2 \hat{Y}^2 + \varphi_3 \hat{Y}^3 + ...) + u$$

La dócima de hipótesis que se realiza en este caso será:

$$H_0 : \varphi = \varphi_2 \hat{Y}^2 + \varphi_3 \hat{Y}^3 + .... = 0$$
$$H_1 : \varphi \neq 0$$

Estadístico de prueba:

$$F = \frac{\left(R^2_{\text{nuevo}} - R^2_{\text{viejo}}\right) / \text{número de regresores nuevos}}{\left(1 - R^2_{\text{nuevo}}\right)\left(n - \text{numero de parametros en el modelo nuevo}\right)}$$

**Ejercicio ilustrativo**

Retomando el modelo sobre el fondo “Afuture” ($Y_t$) en función a las tasas anuales de retorno ($X_t$), y las salidas, obtenemos el test de Ramsey (elaborado en el paquete Eviews):

La hipótesis:

$H_0$: El modelo está correctamente especificado
$H_1$: El modelo no está correctamente especificado

El nivel de significancia ($\alpha$) o la probabilidad de rechazar la hipótesis nula cuando es verdadera es del 5%. La prueba se hace utilizando el estadístico “F”
El test de Reset Ramsey indica que añadiendo 2 términos al test “$Y^2$”, “$Y^3$“ el valor del estadístico “F” es 1.16 y la probabilidad asociada al error de rechazar la hipótesis nula cuando es verdadera es de 35.99% mayor al 5%; por lo tanto se acepta que el modelo está correctamente especificado.

| Ramsey RESET Test: | | | | | |
|---|---|---|---|---|---|
| F-statistic | 1.164495 | | Probability | | 0.359856 |
| Log likelihood ratio | 3.066156 | | Probability | | 0.215870 |

Test Equation:
Dependent Variable: Y
Method: Least Squares
Date: 08/27/02 Time: 16:20
Sample: 1990 2001
Included observations: 12

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 11.24011 | 11.44699 | 0.981928 | 0.3549 |
| X | 1.211993 | 0.393447 | 3.080453 | 0.0151 |
| FITTED^2 | -0.016729 | 0.017389 | -0.962047 | 0.3642 |
| FITTED^3 | 0.000151 | 0.000293 | 0.515680 | 0.6200 |

| | | | |
|---|---|---|---|
| R-squared | 0.778838 | Mean dependent var | 19.30000 |
| Adjusted R-squared | 0.695902 | S.D. dependent var | 32.97040 |
| S.E. of regression | 18.18156 | Akaike info criterion | 8.899894 |
| Sum squared resid | 2644.553 | Schwarz criterion | 9.061530 |
| Log likelihood | -49.39937 | F-statistic | 9.390841 |
| Durbin-Watson stat | 1.244793 | Prob(F-statistic) | 0.005340 |

# LECCIÓN 5

## ANÁLISIS DE LA ESTABILIDAD ESTRUCTURAL

Uno de los supuestos que se acepta al estimar el modelo de regresión, es que los valores de los estimadores de los parámetros se mantienen constantes para la muestra utilizada. Es decir, la relación planteada es válida para todas las observaciones, y que no hay elementos que muestren patrones de relación diferente. Para verificar estos supuestos existen 3 tipos de contrastes: A partir de la prueba de Chow, del análisis de los residuales, y del uso de las variables dicótomicas. En primer lugar haremos uso de los dos primeros

### 1. TEST DE CHOW (CONTRASTE DE CAMBIO ESTRUCTURAL)

Un contraste de mucha importancia, tanto por su interés como por la frecuencia con que aparece en aplicaciones empíricas, es la hipótesis nula donde dos submuestras son generadas por una misma estructura económica. Luego se contrasta la hipótesis nula de ausencia de cambio estructural, y el contraste suele denominarse **Test de Chow** Generalmente se produce cuando se tiene información acerca de una variación estructural que ocurrió en algún momento del periodo muestral, y se pretende contrastar si dicha variación fue suficientemente importante como para generar cambios en los coeficientes del modelo.

El modelo restringido (MR) es:

$$y_t = X_t'\beta + \mu_t \qquad t = 1,2,\ldots,T_1,T_2,\ldots T$$

mientras que el modelo sin restringir (MSR) es :

$$Y_t = X_t'\beta_1 + \mu_t \qquad t = 1,2,\ldots,T_1$$

$$Y_t = X_t'\beta_2 + \mu_t \qquad t = T_2,\ldots,T$$

La suma residual restringida es la que proviene de la estimación del modelo restringido (MR), denotada por SRR, mientras que la suma residual sin restringir es el agregado de las sumas residuales de cada una de las regresiones de las submuestras, que denotamos por $SR_1$ y $SR_2$.

El estadístico F para el contraste de la hipótesis nula de ausencia de cambio estructural es:

$$F_C = \frac{\dfrac{SRR - (SR_1 + SR_2)}{k}}{\dfrac{SR_1 + SR_2}{n-2k}} \qquad \rightarrow \qquad F_{(k,n-2k)}$$

## 2. RESIDUOS RECURSIVOS (CONTRASTE DE ESTABILIDAD)

Los Residuos Recursivos se obtienen a partir de una estimación recursiva de los parámetros β del modelo. La estimación recursiva es similar a la estimación por MCO pero realizada ésta de un modo recursivo, es decir, aumentando el tamaño de la muestra de modo paulatino. Esto es, trabajando con una muestra de tamaño r-1, el estimador que se obtiene es:

$$\hat{\beta}_{r-1} = \left(X'_{r-1}X_{r-1}\right)^{-1} X'_{r-1}Y_{r-1}$$

donde el subíndice r-1, hace referencia al número de observaciones que se utilizan para la estimación de los parámetros[2].

Una vez estimados los parámetros del modelo con las r-1 primeras observaciones, se incorpora, para la matriz de regresores, la información de la observación siguiente; es decir el vector fila $X_r$ , y con ésta se realiza una predicción para la observación r de la variable endógena $\hat{Y}_r$. A partir de ésta se calcula el error de predicción para la observación r utilizando las primeras r-1 observaciones. Este error de predicción formará parte del residuo recursivo $w_r$ que se define como una tipificación de aquél.

$$w_r = \frac{Y_r - X'_r\hat{\beta}_{r-1}}{\sqrt{1 + X'_r\left(X'_{r-1}X_{r-1}\right)^{-1}X_r}} \qquad r = k+1, k+2, ..., n$$

Este procedimiento se repite sucesivamente hasta finalizar realizando la predicción para la última observación. Así, de este modo, se obtiene una serie de n-k residuos recursivos homocedásticos e incorrelacionados.

$$w \rightarrow N\left(0,\ \sigma^2 I\right)$$

Estos nuevos residuos permiten analizar la perturbación del modelo de regresión con mayor objetividad ya que, a diferencia de los residuos MCO, verifican las hipótesis deseables para dicha perturbación puesto que su distribución sí es normal esférica.

Los contrastes basados en los residuos recursivos los podemos clasificar en dos grupos; contrastes gráficos y contrastes numéricos. Los primeros se utilizan de un modo general para detectar si existe o no estabilidad en el modelo de regresión; puesto que se basan en análisis gráficos. Son contrastes bastante generales que detectan de modo vago la presencia de problemas en el modelo.

Los contrastes numéricos permiten detectar específicamente si se cumple o no, por separado, las hipótesis de homocedasticidad e incorrelación de las perturbaciones.

Como hipótesis nula se especifica la estabilidad, tanto de los parámetros β como de las varianzas de la perturbación, a lo largo de los "n" periodos considerados.

[2] Nótese que el número mínimo de observaciones que se necesita para la estimación del modelo es r-1= k, siendo k el número de parámetros β del modelo de regresión.

$H_0$: $\beta_1 = \beta_2 = ..... = \beta_n = \beta$

$H_1$ $\sigma_1^2 = \sigma_2^2 = .... = \sigma_n^2 = \sigma$

y se especifican dos contrastes que permiten decidir si existe estabilidad o ruptura estructural.

## 3. CONTRASTE DE SUMA ACUMULADA (TEST CUSUM)

Este contraste consiste en la acumulación progresiva de los residuos recursivos que posteriormente se normalizan dividiéndolos entre la estimación insesgada de la desviación típica de la perturbación (S). De este modo se calcula el valor acumulado $W_r$, que se representa gráficamente frente al número de valores acumulados (r).

$$W_r = \frac{\sum_{j=k+1}^{r} w_j}{S} \qquad r = k+1, k+2, ... , n$$

Donde: $S = \sqrt{\frac{SCR}{n-k}}$

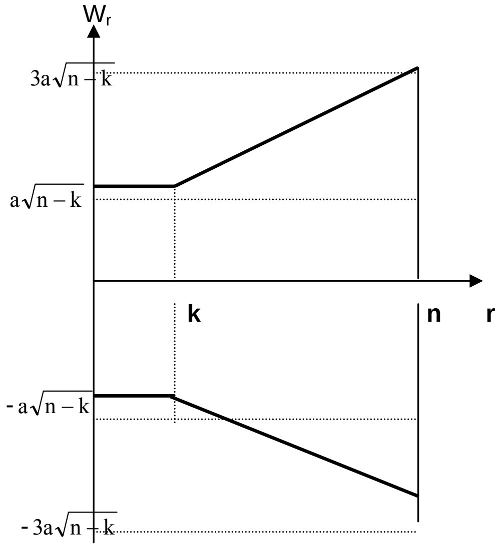
Estos valores de las sumas acumuladas, en el supuesto de estabilidad, deberían oscilar entre las líneas de significación representadas por las rectas que se definen a partir de los pares de puntos siguientes:

$$\{k, a\sqrt{n-k}\} \text{ y } \{n, 3a\sqrt{n-k}\}$$
$$\{k, -a\sqrt{n-k}\} \text{ y } \{n, -3a\sqrt{n-k}\}$$

En caso contrario, es decir cuando los valores de $W_r$ sobrepasen dichas rectas (marcadas con trazos más gruesos) se puede considerar falta de estabilidad en el modelo. Para el cálculo de estas rectas se necesita determinar los valores de “a” que se encuentran tabulados para distintos niveles de significación siendo los más usuales.

| $\alpha$ (%) | 1% | 5% | 10% |
|---|---|---|---|
| A | 1.143 | 0.948 | 0.850 |

La representación gráfica de este contraste dibujaría los residuos recursivos sobre el gráfico siguiente:




## 4. CONTRASTE DE SUMA ACUMULADA DE CUADRADOS (TEST CUSUM²)

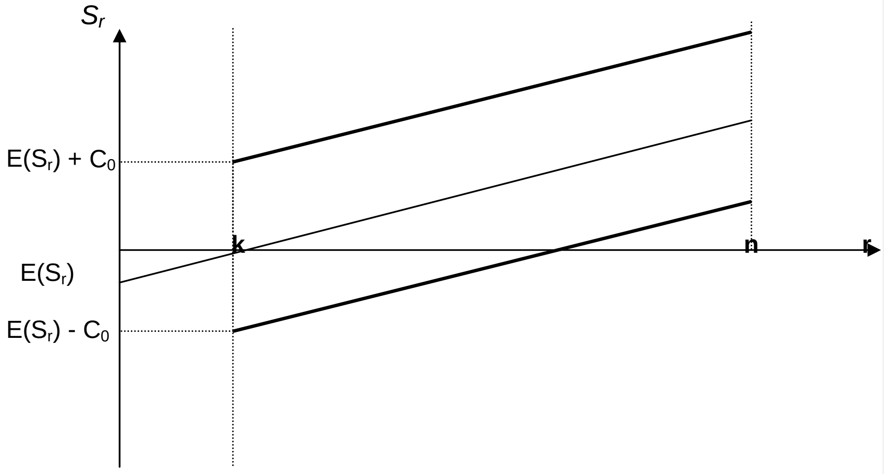
Este contraste, análogo al anterior, utiliza en el numerador la suma acumulada del cuadrado de los residuos recursivos y en el denominador el valor de la Suma de Cuadrados de la totalidad de los Residuos Recursivos.

$$S_r = \frac{\sum_{j=k+1}^{r} w_j^2}{\sum_{j=k+1}^{n} w_j^2} \qquad r = k+1, k+2, ... , n$$

Para este estadístico $S_r$ se consideran también unas rectas de significación definidas a partir del valor esperado del estadístico sumándole (y restándole) una cantidad fija dependiendo del nivel de significación elegido ($\alpha$).

Nótese que el valor esperado del estadístico oscila entre cero y uno; así, $E(S_r) = 0$ cuando $r = k$, y, cuando $r = n$, $E(S_r) = 1$.

Estos límites se dibujan frente a r, obteniéndose el siguiente gráfico conocido como CUSUM$^2$.




Al igual que en el contraste anterior, se considera que existe evidencia suficiente para rechazar la hipótesis nula de homogeneidad del modelo, cuando la representación gráfica de los valores del estadístico $S_r$ se sitúan fuera de las bandas (marcadas en trazo más grueso). Con respecto a este contraste existe cierta evidencia empírica que permite considerarlo más poderoso que el test de suma acumulada (Test CUSUM).

**Ejercicio ilustrativo**

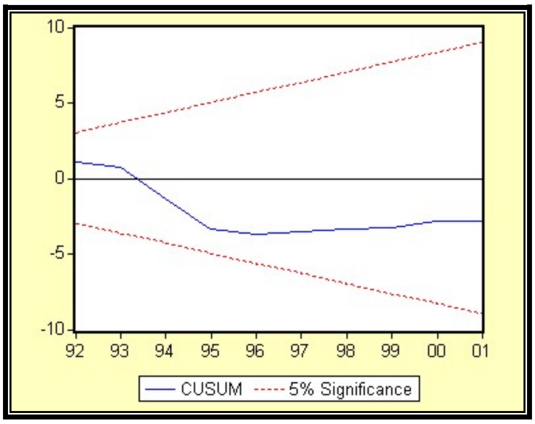
En el modelo definido anteriormente: $Yt = 2.209520 + 1.061520X_t + \mu_t$

Donde: $Y_t$ = fondo "Afuture" ; $X_t$ = tasa anual de retorno

En el paquete E-Views, se calculará el Test de CUSUM para analizar $H_o$:

$H_0$: Los parámetros son estables en el período de prueba
$H_1$: Los parámetros no son estables en el período de prueba

En el resultado se observa que el estadístico CUSUM se mantiene dentro de las bandas de confianza, con lo cual se puede afirmar que los parámetros son estables a lo largo del período de análisis en un 95% de confianza.



**ANEXO**

**LA INFERENCIA ESTADÍSTICA Y EL PROGRAMA E-VIEWS**

El E-views es un programa especializado que sirve para hacer análisis, regresiones y predicción así como simulaciones y evaluaciones de eficiencia y predicción de modelos.

Dado los siguientes datos hipotéticos (Período 1991-1995)

| AÑO | Y | X1 | X2 |
|---|---|---|---|
| 1991 | 3 | 3 | 5 |
| 1992 | 1 | 1 | 4 |
| 1993 | 8 | 5 | 6 |
| 1994 | 3 | 2 | 4 |
| 1995 | 5 | 4 | 6 |

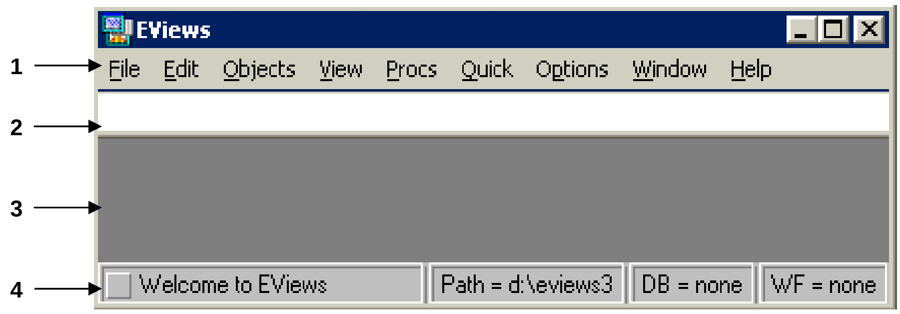
Estime el modelo $Y_t = \beta_1 + \beta_2 X_{1t} + \beta_3 X_{2t} + \mu_t$

$Y_t$: variable dependiente o endógena
$X_{1t}$: variable independiente o exógena
$X_{2t}$: variable independiente o exógena

**PROCEDIMIENTO:**

**A.** Ingresar al programa E-Views

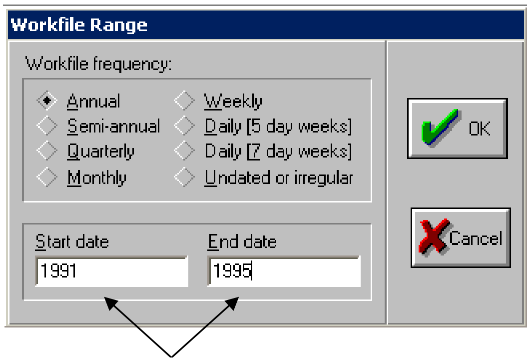
- Inicio
- Programas
- Eviews (hacer clic)
- Aparece el cuadro siguiente:




1. **Barra de Menú**
2. **Barra de comandos**
3. **Area de Trabajo**
4. **Línea de Estado**

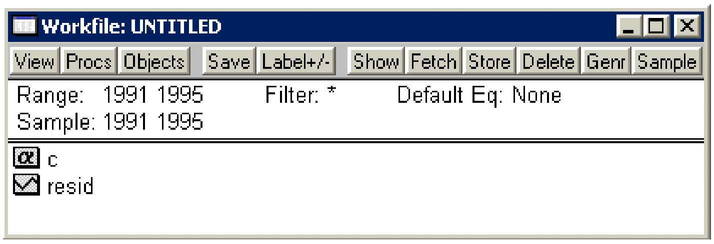
**B.** Crear un Workfile (archivo de trabajo)

- File
- New
- Workfile
- Aparece la siguiente caja de dialogo
- Se especifica periodicidad o frecuencia

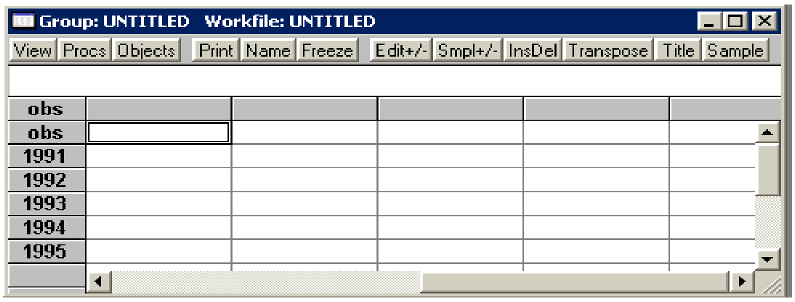



**Escribir la fecha de inicio y la fecha final de los datos**

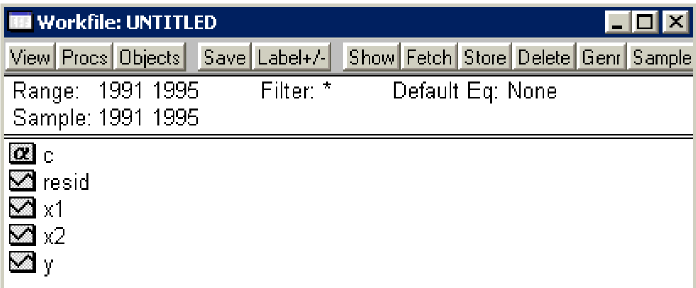
- Seleccionar OK
- Se crea vector c y serie de residuos
- Aparece el workfile:




- En la barra del menú Principal, hacer clic en Quick y seleccionar Empty Group (Edit series)
- Se crea grupo
- Aparece el siguiente cuadro:




- Se sombrea la primera columna vacía (de la izquierda)
- Se escribe el nombre de la primera variable en la primera celda vacía: Y, y ENTER.
- Se sombrea la segunda columna vacía y se escribe en la primera celda X1 y luego ENTER.
- Luego se selecciona la tercera columna y se escribe en la primera celdas X2 y Luego ENTER.
- Seguidamente digitar los datos (igualmente se puede importar la información de otro software, por ejemplo el excel), que puede ser por fila o por columna.
- Terminado el tipeado o pegado de los valores numéricos hacer clic en el botón de la esquina superior del cuadro y hacer clic en YES.
- El workfile aparecerá con las nuevas series:




## Alternativas Adicionales

Adicionalmente, también se puede editar series mediante las siguientes formas:

### Crear y editar series

- Para crear **una serie** haga clic en **Objects / New Objects/ Series.**
- En la ventana emergente escribir el nombre de la serie y **OK.**
- Para llenar o editar una serie generada o importada hacer doble clic en la serie (o clic derecho en **Open**).
- Una vez abierta la serie en el menú hacer clic en **Edit +/-** y editar la serie y para finalizar nuevamente **Edit +/- .**
- **Copy/paste:** Seleccionar las celdas a copiar desde Excel (Copy) y pegarlas en Eviews (Paste).

**Importar datos de hoja de cálculo**

- Para importar información desde archivos hojas de cálculo, hacer clic en **Procs / Import./ (Read Text-Lotus-Excel)**
- Aparecerá la ventana **Open** donde se debe buscar y seleccionar el archivo a utilizar.
- Si usted hace doble clic en le nombre del archivo, verá un segundo cuadro de diálogo pidiéndole detalles acerca del archivo **(Excel Spreadsheet Import)**
- Para un archivo de hoja de cálculo (Excel o Lotus), se debe especificar las coordenadas de la celda superior-izquierda que contiene la información. Por ejemplo, si los datos empiezan el la celda A2, entonces usted deberá especificar esta celda como celda de datos superior izquierda.
- Luego escribir el nombre o los nombres de las series.
- Si el archivo cuenta con varias hojas de trabajo, entonces se debe especificar el nombre de donde se desea importar.
- Hacer clic en OK y los datos serán incorporados en el archivo de trabajo.

**C.** Estimación de la ecuación: $Y_t = \beta_1 + \beta_2 X_{1t} + \beta_3 X_{2t} + \mu_t$

- En la barra de comandos se escribe: **LS Y C X1 X2** y luego ENTER
  Donde: LS (Last Square) (Mínimos Cuadrados)
- Aparece las siguientes salidas (ESTIMATION OUTPUT):

Dependent Variable: Y
Method: Least Squares
Date: 07/26/02 Time: 16:28
Sample: 1991 1995
Included observations: 5

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 4.000000 | 4.474930 | 0.893869 | 0.4657 |
| X1 | 2.500000 | 0.866025 | 2.886751 | 0.1020 |
| X2 | -1.500000 | 1.369306 | -1.095445 | 0.3876 |

| | | | |
|---|---|---|---|
| R-squared | 0.946429 | Mean dependent var | 4.000000 |
| Adjusted R-squared | 0.892857 | S.D. dependent var | 2.645751 |
| S.E. of regresión | 0.866025 | Akaike info criterion | 2.833904 |
| Sum squared resid | 1.500000 | Schwarz criterion | 2.599567 |
| Log likelihood | -4.084760 | F-statistic | 17.66667 |
| Durbin-Watson stat | 1.666667 | Prob(F-statistic) | 0.053571 |

**Encabezado**
Se especifica cuál es la variable dependiente, el número de observaciones, las variables explicativas y el método de estimación.

**Primera Columna**
Se refiere a qué parámetro está estimando. Es decir aquel que acompaña a la variable que se señala.

**Segunda Columna**
Se tienen los valores estimados de los parámetros.

**Tercera Columna**
Muestra la desviación estimada de los parámetros.

**Cuarta Columna**
Se presentan los valores calculados de los estadísticos t donde se tiene como hipótesis nula que cada uno de los parámetros es igual a cero. Para ello los valores t calculados para cada parámetro son la división de los respectivos valores de la segunda y tercera columna. La prueba individual de significancia estadística para un parámetro es justamente el valor del parámetro calculado dividido por la desviación estándar calculada y ello es lo que se obtiene en la cuarta columna.

**Quinta Columna**
Presenta, el valor de la probabilidad (p-value), de rechazar la hipótesis nula cuando es verdadera, (nivel de significación) con los datos estimados de la muestra que tenemos.

Al escoger el nivel de significación estadística estamos eligiendo el punto que separa la región de rechazo de la región de aceptación de la hipótesis nula cuando Ho es verdadero. Si se escoge el nivel de significación del 5%, quiere decir que si la probabilidad de que la hipótesis nula es cierta es mayor al 5% no podemos rechazar la hipótesis nula.

**R-squared** ($R^2$): Es el coeficiente de determinación.

**Adjusted R-squared** ($R^2$ ajustado): Coeficiente de determinación ajustado.

**F-statistic**: Valor del estadístico F, permite contrastar la capacidad explicativa conjunta de las variables introducidas en el modelo.

**Prob(F-statistic)**: Valor que mide la probabilidad de rechazar la hipótesis nula de significancia conjunta.

**Durbin-Watson stat**: Valor usado para contrastar la hipótesis de autocorrelación

**ANÁLISIS DE LAS SALIDAS DE LA REGRESIÓN:**

**a.** Verificación de la significancia individual de cada uno de los coeficientes a partir de la hipótesis nula, que nos dice que “la variable “$X_i$” no es significativa en el modelo (prueba t, para lo cual su valor debe ser superior a un t de tabla, que para este caso debe ser con (n-2, es decir 5-2 grados de libertad y un nivel de significación = 0.05).

Alternativamente la prueba t se docima con la observación de la última columna (Prob = Probabilidad de rechazar la hipótesis nula cuando es cierta).

Si la probabilidad asociada es mayor al 0.05, entonces se acepta la Ho de no significatividad de la variable “$X_i$”, en caso contrario se rechaza la Ho a un nivel de confianza del 95%.

En este modelo las probabilidades asociadas son superiores a 0.05, por lo tanto se acepta la Ho, es decir que todas las variables no son significativas en el modelo.

**b.** Verificación de la significancia global del modelo (Prueba F). Al igual que la prueba estadística “t” se puede analizar de 2 formas: En base a la lectura del estadístico “F” statistic, o en base a la lectura del valor de Prob(F-statistic). Cualquiera conduce a la misma decisión.
La Hipótesis Nula es: la variable dependiente no es explicada por el modelo en su conjunto.
La Hipótesis Alternativa es: la variable dependiente es explicada por el modelo en su conjunto.

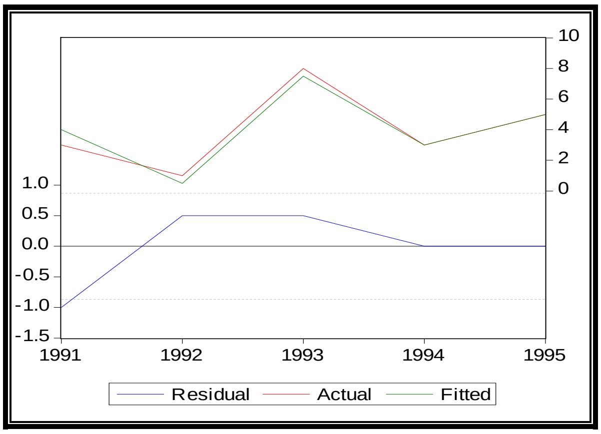
Si la probabilidad asociada (valor de Prob(F-statistic)), es superior a 0.05, entonces se acepta la Ho.
En nuestro modelo se observa que la probabilidad asociada es superior a 0.05, entonces se acepta la Ho, es decir que la variable dependiente no es explicada por el modelo en su conjunto.

**D.** Para ver la representación clásica de la regresión hacer:
- VIEW
- REPRESENTATIONS
- Aparece la siguiente ecuación: $\mathbf{Y = 4 + 2.5 * X_1 - 1.5 * X_2}$

**E.** Para ver los valores observados de la variable dependiente (Y), los valores estimados con la ecuación y los residuos, proceder de la siguiente manera:
- Estando en las salidas de la regresión (ESTIMACIÓN OUTPUT), hacer clic en VIEW que muestra una serie de alternativas.
- Escoger ACTUAL FITTED RESIDUAL para ver los residuos, si presenta o no autocorrelación.
- Luego hacer clic en: ACTUAL FITTED RESIDUAL, TABLE

| obs | Actual | Fitted | Residual | Residual Plot |
|---|---|---|---|---|
| 1991 | 3.00000 | 4.00000 | -1.00000 | |
| 1992 | 1.00000 | 0.50000 | 0.50000 | |
| 1993 | 8.00000 | 7.50000 | 0.50000 | |
| 1994 | 3.00000 | 3.00000 | -2.7E-15 | |
| 1995 | 5.00000 | 5.00000 | 2.7E-15 | |

- También se puede escoger la alternativa ACTUAL FITTED RESIDUAL GRAPH (para observar la gráfica):




**F.** Hallar la matriz de varianzas y covarianzas:

- ➢ Estando en las salidas de la regresión, hacer clic en VIEW
- ➢ Luego hacer clic en covariance matriz, aparecerá el siguiente cuadro:

| | C | X1 | X2 |
|---|---|---|---|
| **C** | 20.025 | 3.375 | -6 |
| **X1** | 3.375 | 0.75 | -1.125 |
| **X2** | -6 | -1.125 | 1.875 |

**G.** Normalidad de los Residuos

La hipótesis de normalidad de las perturbaciones es fundamental para la realización de inferencia en el modelo. En el E-views se puede calcular el estadístico Jarque Bera, el cual permite contrastar la hipótesis de normalidad de los residuos.

**Procedimiento**

- ➢ Regresionar el modelo MCO
- ➢ Ingresar a VIEW/ RESIDUAL TEST
- ➢ Seleccionar la opción HISTOGRAM-NORMALITY TEST.
- ➢ El estadístico Jarque Bera nos permite verificar la normalidad de los residuos. La Ho es que los residuos se distribuyen normalmente. Si la probabilidad asociada al estadístico es mayor al 5%, entonces no se puede rechazar la Ho de normalidad de los residuos.

**H.** Test de Punto de Quiebre de Chow

- ➢ Regresionar el modelo MCO
- ➢ Ingresar a VIEW/ STABILITY TEST / BREAKPIONT TEST
- ➢ Ingresar la Fecha de quiebre estructural.
- ➢ La Hipótesis nula es que no hay cambio estructural. Si la probabilidad asociada al test es mayor al 5%, entonces no se puede rechazar la Ho.

**I.** Test de Residuos Recursivos

- ➢ Regresionar el modelo MCO

- Ingresar a VIEW/ STABILITY TEST / RECURSIVE ESTIMATES
- Seleccionar la opción RECURSIVE RESIDUALS
- Si la gráfica de residuos recursivos sale fuera de las bandas, entonces los parámetros son inestables en el período de análisis (al 5% de significancia).

**J.** Test Cusum y Cusum Cuadrado

- Regresionar el modelo MCO
- Ingresar a VIEW/ STABILITY TEST / RECURSIVE ESTIMATES
- Seleccionar la opción CUSUM O CUSUM OF SQUARES TEST.
- El programa reporta un gráfico, conteniendo la evolución de un estadístico. Si el estadístico se encuentra dentro de las bandas, entonces los parámetros son estables en el período de análisis (al 5% de significancia). No existe suficiente evidencia para rechazar la Ho de estabilidad de los parámetros.

**K.** Test de Reset Ramsey

- En el Output de la regresión hacer clic en View/ Stability Tests/ Ramsey RESET Test/ 2
- La Ho es que el modelo está correctamente especificado.
- El programa reporta unas salidas. Si la probabilidad asociada a este estadístico es mayor que el nivel de significancia del 5%, no se debe rechazar la Ho, es decir que el modelo está correctamente especificado.

## LABORATORIO

### Ejercicio Aplicativo 1

Se presenta un modelo para la inflación en el Perú para el periodo de Enero de 1992 – Diciembre de 1998. El carácter exploratorio de este modelo, apunta a esclarecer el rol de variables adicionales a los agregados monetarios en la determinación de los niveles inflacionarios.

### 1. Estimación, inferencia e interpretación.

Al estimar el modelo en logaritmos por MCO, para el período 1992:01–1998:12, encontramos los siguientes resultados:

$$LIPC = \beta_1 + \beta_2 CIRPROM + \beta_3 LTC$$

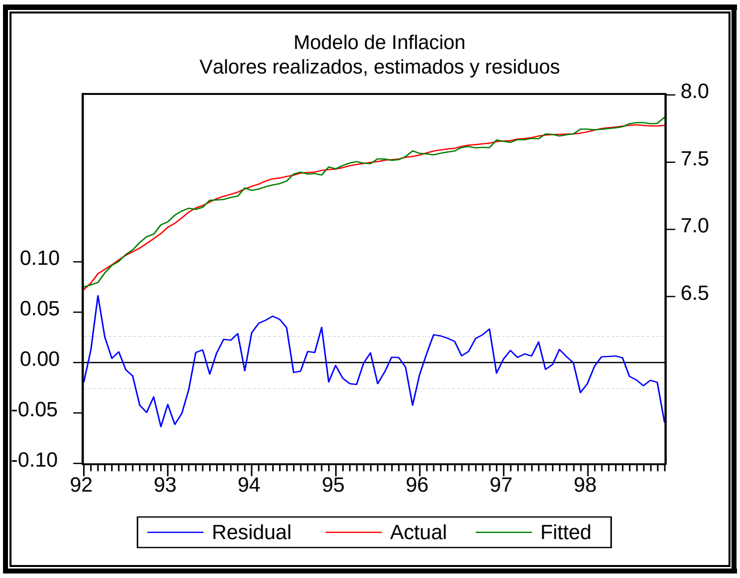
Donde:

LIPC : Logaritmo del Índice de Precios al Consumidor
CIRPROM : Circulante promedio
LTC : Logaritmo del Tipo de Cambio

Dependent Variable: LIPC
Method: Least Squares
Date: 12/13/01 Time: 21:35
Sample: 1992:01 1998:12
Included observations: 84

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 4.342678 | 0.082939 | 52.35958 | 0.0000 |
| LCIRPROM | 0.355330 | 0.013456 | 26.40709 | 0.0000 |
| LTC | 0.500555 | 0.026611 | 18.81012 | 0.0000 |
| R-squared | 0.993782 | Mean dependent var | | 7.414901 |
| Adjusted R-squared | 0.993628 | S.D. dependent var | | 0.325085 |
| S.E. of regression | 0.025949 | Akaike info criterion | | -4.430295 |
| Sum squared resid | 0.054542 | Schwarz criterion | | -4.343480 |
| Log likelihood | 189.0724 | F-statistic | | 6472.723 |
| Durbin-Watson stat | 0.628449 | Prob(F-statistic) | | 0.000000 |

Al hacer una descripción de los resultados de la regresión vemos que los test t miden apropiadamente la significancia estadística de los parámetros. Además, todas las probabilidades asignadas para cada uno de los parámetros son todas iguales a cero por lo que los parámetros estimados son significativos a un nivel de aceptación del 95%. Por otro lado, nuestro modelo tiene un nivel de ajuste del 99.37% tal como lo muestra el coeficiente de determinación $R^2$ (y su valor ajustado). Además la prueba de significancia global (F- statistic) nos permite afirmar que en conjunto ninguno de los parámetros estimados es NO significativo. El grado de ajuste, tanto de los valores realizados como de los estimados, así como los residuos de la ecuación anterior se puede apreciar de manera gráfica

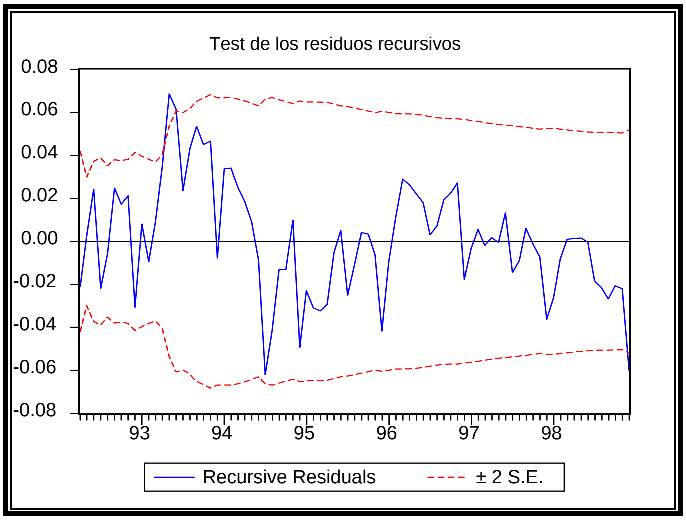



## 2. Tests de Estabilidad de Los Parámetros

Para analizar la estabilidad de los parámetros, se propone utilizar los siguientes test:

a. Test de Residuos Recursivos.
b. Test CUSUM.
c. Test CUSUM de Cuadrados.
d. Test de Punto de Quiebre de Chow:

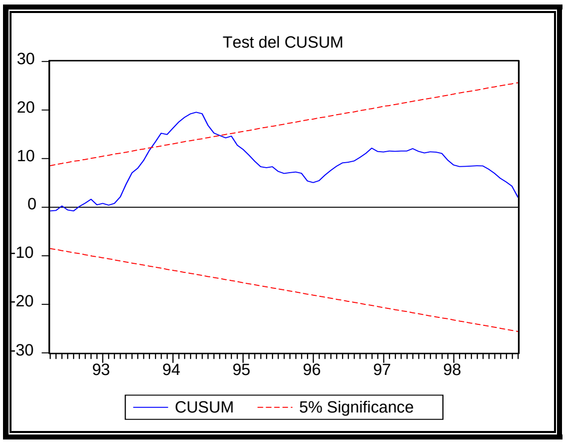
### a. Test de Residuos Recursivos

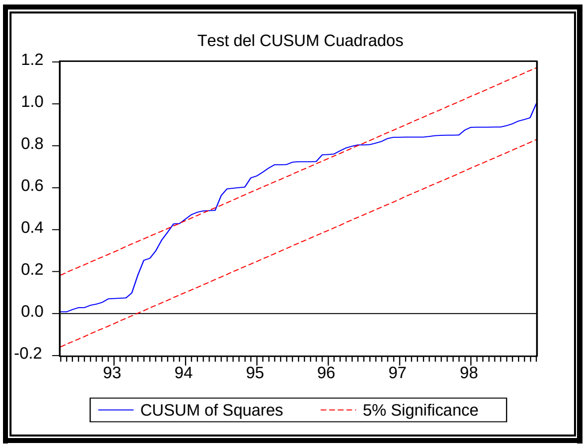


La representación de la serie de los residuos recursivos junto con sus bandas de confianza nos permite detectar la inestabilidad en los parámetros cuando uno o varios de los residuos sobrepasan sus bandas tal como ocurre para junio de 1993 hasta aproximadamente octubre del mismo año, fecha en la que "retorna" a situarse dentro de las bandas (como también ocurre para julio de 1994 y para finales de 1998) por lo que concluimos que los parámetros son inestables.

### b. Test del Cusum y Cusum Cuadrado

Dado que la curva sale fuera de las bandas, en ambos casos, concluimos que los parámetros estimados son inestables.




### c. Test de Punto de Quiebre de Chow

Probamos la posibilidad que exista un quiebre estructural en julio de 1994:

| Chow Breakpoint Test: 1993:12 | | | |
|---|---|---|---|
| F-statistic | 7.059748 | Probability | 0.000293 |
| Log likelihood ratio | 20.17847 | Probability | 0.000156 |

Rechazamos la hipótesis de que no hay cambio estructural al 99% de confiabilidad. Por lo tanto, concluimos que en julio de 1994 se produjo un cambio estructural.

### 3. Test de Error de Especificación

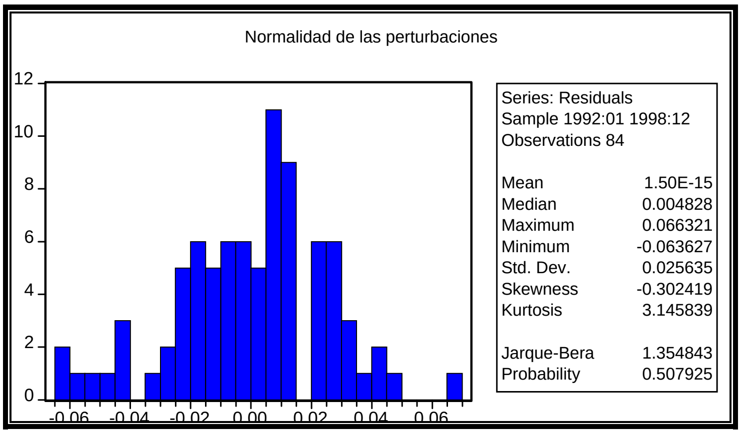
El E-views permite usar el RESET test propuesto por Ramsey (1969). La hipótesis nula es que el modelo esta bien especificado. A un nivel de 5% de significancia añadiendo 2 términos al test, la probabilidad asociada es de 43.53% mayor al 5%; por lo tanto se rechaza la hipótesis nula, concluyéndose que el modelo está correctamente especificado.

| Ramsey RESET Test: | | | | |
|---|---|---|---|---|
| F-statistic | 0.614881 | Probability | | 0.435271 |
| Log likelihood ratio | 0.643157 | Probability | | 0.422570 |

Test Equation:
Dependent Variable: LIPC
Method: Least Squares
Date: 09/22/02 Time: 09:18
Sample: 1992:01 1998:12
Included observations: 84

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 4.558636 | 0.287682 | 15.84610 | 0.0000 |
| LCIRPROM | 0.493996 | 0.177350 | 2.785424 | 0.0067 |
| LTC | 0.667595 | 0.214686 | 3.109635 | 0.0026 |
| FITTED^2 | -0.025285 | 0.032246 | -0.784144 | 0.4353 |
| R-squared | 0.993829 | Mean dependent var | | 7.414901 |
| Adjusted R-squared | 0.993598 | S.D. dependent var | | 0.325085 |
| S.E. of regression | 0.026011 | Akaike info criterion | | -4.414142 |
| Sum squared resid | 0.054126 | Schwarz criterion | | -4.298389 |
| Log likelihood | 189.3940 | F-statistic | | 4294.837 |
| Durbin-Watson stat | 0.645581 | Prob(F-statistic) | | 0.000000 |

### 4. Normalidad de las Perturbaciones



Aunque podemos observar en el histograma que las frecuencias máximas no se presentan en el centro de manera clara, como debería de ser si se sigue una distribución normal, sin embargo, el valor del coeficiente de asimetría muestral (-0.3024) es próximo a cero y el coeficiente de apuntamiento –kurtosis- es de 3.1458 no es lejano a 3. Finalmente, el estadístico Jarque-Bera no permite rechazar la hipótesis nula de normalidad, pues su valor (1.3548) genera una probabilidad (de rechazar dicha hipótesis siendo verdadera) superior a 0.05

**Ejercicio Aplicativo 2**

Suponga que Ud. desea dedicarse a la producción de Kiwi, producto que requiere gran cuidado en cuanto a las condiciones del medio ambiente como también de riego.

La siguiente información corresponde a la experiencia de una cooperativa, suponiendo que el resto de los factores que inciden en la producción de Kiwi son homogéneos a los del resto del Perú, utilice esta información para planificar su producción.

| **Año** | **Producto ton./h.** | **Cantidad de agua** | **Temperatura Promedio (C°)** |
|---|---|---|---|
| 1990 | 60 | 9 | 13.3 |
| 1991 | 50 | 10 | 8.3 |
| 1992 | 70 | 11 | 11.70 |
| 1993 | 70 | 10 | 11.70 |
| 1994 | 80 | 12 | 13.3 |
| 1995 | 50 | 9 | 8.3 |
| 1996 | 60 | 11 | 6.7 |
| 1997 | 40 | 8 | 6.7 |

1. Estime un modelo que considere solamente el agua como variable explicativa y otro que incluya al agua y la temperatura.

**Considerando Solamente al Agua**, realizando una regresión lineal simple, tendremos que la ecuación de regresión resulta:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| CANT_AGUA | 8.333333 | 2.256677 | 3.692745 | 0.0102 |
| C | -23.33333 | 22.73539 | -1.026300 | 0.3443 |
| R-squared | 0.694444 | Mean dependent var | | 60.00000 |
| Adjusted R-squared | 0.643519 | S.D. dependent var | | 13.09307 |
| S.E. of regression | 7.817360 | Akaike info criterion | | 7.162889 |
| Sum squared resid | 366.6667 | Schwarz criterion | | 7.182749 |
| Log likelihood | -26.65155 | F-statistic | | 13.63636 |
| Durban-Watson stat | 1.856061 | Prob(F-statistic) | | 0.010176 |

Se puede observar de la tabla anterior, que la variable cantidad de agua es significativa para el modelo planteado, cosa que no ocurre para la constante (tiene un p-valor superior a 0.05). Por lo tanto será necesario estimar otra vez el modelo pero sin considerarla, este resultado se muestra en la tabla siguiente:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| CANT_AGUA | 6.034483 | 0.275378 | 21.91347 | 0.0000 |
| R-squared | 0.640805 | Mean dependent var | | 60.00000 |
| Adjusted R-squared | 0.640805 | S.D. dependent var | | 13.09307 |
| S.E. of regression | 7.847060 | Akaike info criterion | | 7.074624 |
| Sum squared resid | 431.0345 | Schwarz criterion | | 7.084554 |
| Log likelihood | -27.29849 | Durbin-Watson stat | | 1.490000 |

La ecuación de regresión finalmente queda determinada por la siguiente relación:

$$\Pr oducto = 6.034 * \text{Agua}$$

**Considerando el Agua y la Temperatura:** Al igual que en el caso anterior no será necesario considerar el término constante en el modelo (p>0.05), en tal sentido la ecuación de regresión estimada final resulta:

$$\Pr oducto = 3.662 * \text{Agua} + 2.379 * \text{Temperatura}$$

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| CANT_AGUA | 3.661719 | 0.782738 | 4.678091 | 0.0034 |
| TEMP_PROM | 2.378622 | 0.762748 | 3.118492 | 0.0206 |
| R-squared | 0.862946 | Mean dependent var | | 60.00000 |
| Adjusted R-squared | 0.840104 | S.D. dependent var | | 13.09307 |
| S.E. of regression | 5.235532 | Akaike info criterion | | 6.361132 |
| Sum squared resid | 164.4647 | Schwarz criterion | | 6.380992 |
| Log likelihood | -23.44453 | F-statistic | | 37.77838 |
| Durbin-Watson stat | 1.580261 | Prob(F-statistic) | | 0.000850 |

2. Suponga que la producción se realizará en un galpón cerrado en el cual, podrá controlar la temperatura a un nivel de 10 grados Centígrados ¿Cuál será la relación relevante para planificar su producción?

Si la producción se realizará en un galpón cerrado en el cual es posible mantener la temperatura constante a un nivel de 10Cº la relación relevante para planificar la producción de kiwi será:

$$\Pr oducto = 3.662 * \text{Agua} + 2.379 * 10$$

con lo cual al realizar las operaciones respectivas se obtendrá la relación siguiente :

$$\Pr oducto = 27.452 + 3.662 * \text{Agua}$$

### Ejercicio Aplicativo 3

Dado el modelo siguiente: $I_t = \beta_1 + \beta_2 PBI + \mu_t$

Donde:
I : Inversión
PBI : Producto Bruto Interno

| Año | I | PBI |
|---|---|---|
| 1992 | 14,758 | 83,401 |
| 1993 | 16,487 | 87,375 |
| 1994 | 21,931 | 98,577 |
| 1995 | 26,374 | 107,039 |
| 1996 | 25,094 | 109,709 |
| 1997 | 28,825 | 117,110 |
| 1998 | 28,255 | 116,485 |
| 1999 | 24,455 | 117,590 |
| 2000 | 23,554 | 121,267 |
| 2001 | 21,663 | 121,513 |

Fuente: BCRP.

**a. Calcule el estimador MCO, de los parámetros e interprételos.**

Dependent Variable: I
Method: Least Squares
Date: 02/26/03 Time: 10:36
Simple: 1992 2001
Included observations: 10

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -4873.554 | 8093.987 | -0.602120 | 0.5638 |
| PBI | 0.259365 | 0.074391 | 3.486533 | 0.0082 |

| | | | |
|---|---|---|---|
| R-squared | 0.603094 | Mean dependent var | 23139.60 |
| Adjusted R-squared | 0.553481 | S.D. dependent var | 4628.697 |
| S.E. of regression | 3092.991 | Akaike info criterion | 19.08852 |
| Sum squared resid | 76532765 | Schwarz criterion | 19.14904 |
| Log likelihood | -93.44261 | F-statistic | 12.15591 |
| Durbin-Watson stat | 0.565595 | Prob(F-statistic) | 0.008240 |

I = -4873.553604 + 0.2593652018*PBI

- $\beta_1$ (-4873.554): significa que la inversión es negativa cuando el PBI es cero.
- $\beta_2$ (0.259366): significa que la inversión aumenta en 0.259366 cuando el PBI aumenta en una unidad monetaria.

**b. Matriz de Varianza y Covarianza.**

| | C | PBI |
|---|---|---|
| C | 65512622 | -597.7039 |
| PBI | -597.7039 | 0.005534 |

**c. Significancia individual de las variables.**

Para $\beta_1$:
Como se tiene un p-valor de 0.5638 el cual es mayor que 0.05 ; se acepta la hipótesis nula (no es significativo), lo que origina que en este modelo no será necesario utilizar el termino constante lo que significa que cuando el PBI sea cero no existirá inversión .

Para $\beta_2$:

$H_0$ : $\beta_2 = 0$ el PBI no explica el comportamineto de la inversión
$H_1$ : $\beta_2 \neq 0$ el PBI es capaz de explicar el comportamineto de la inversión

Se tiene un p-valor de 0.0082, el cual es menor que 0.05; lo que origina que se rechaze la hipótesis nula; por lo que se concluye que el PBI puede explicar el comportamiento de la inversión a partir del modelo inicialmente planteando.

**d. Intervalo de confianza al 95% para el segundo parámetro ($\beta_2$)**

Se sabe que: P ($\hat{\beta}_2$- t * Stad Error < $\beta_2$< $\hat{\beta}_2$+ t * Stad Error) = 0.95, reemplazando valores:

P (0.259365-2.036* 0.074391 < $\beta_2$< 0.259365 +2.036* 0.074391) = 0.95.

P (0.107904924< $\beta_2$< 0.410825076) = 0.95

Por lo tanto el valor del segundo parámetro se encuentra entre 0.107904924 y 0.410825076.

**e. Verificación de la significancia global del modelo ( Prueba F)**

$\mathrm{H}_0$ : $\beta_i = 0$ la inversión no es explicada por el uso de los regresores
$\mathrm{H}_1$ : $\beta_i \neq 0$ la inversión es explicada por el uso de los regresores

Se tiene un Prob (F-statistic) de 0.008240, el cual es mayor que cero lo que hace que se rechace la hipótesis nula, por lo tanto se concluye que el modelo en su conjunto es significativo, a un nivel de confianza del 95%.

## Ejercicio de autoconocimiento

### ¿Porqué realizar el proceso de inferencia?

| | SI | NO | NO SÉ |
|---|---|---|---|
| 1. Porque permite establecer la validez de determinadas afirmaciones acerca de los parámetros (desconocidos). | | | |
| 2. Porque mediante el análisis (de varianza), se investiga la explicación conjunta de todas las variables explicativas que intervienen en el modelo. | | | |
| 3. Para analizar el pasado y predecir el futuro de la empresa. | | | |
| 4. Porque mediante un programa especializado Eviews se podrá hacer análisis, regresiones y predicción así como simulaciones y evaluaciones de eficiencia y predicción de modelos. | | | |
| 5. Para utilizar el modelo correcto y adecuado para un pronóstico | | | |
| 6. Para establecer la importancia del estudio de las variables. | | | |
| 7. Porque se puede ingresar a las variables independientes e introducir los nuevos valores. | | | |
| 8. Para realizar aplicaciones con datos de la realidad. | | | |
| 9. Para estimar intervalos y contrastar hipótesis. | | | |
| 10. Para predecir sucesos futuros. | | | |

**CALIFICACION**

Puntuar con un punto cada respuesta “SI”.
Si obtienes de de 1 - 3 puntos tienes pocas expectativas de hacer una buena predicción empresarial.
Si tienes entre 4 - 7, tienes buenas expectativas de hacer una buena predicción empresarial.
Y si tienes entre 8 - 10, denotas excelentes expectativas de hacer una buena predicción empresarial.

## RESUMEN

Los **coeficientes de confianza** son los niveles de confianza que tenemos en que el intervalo contiene el valor desconocido del parámetro.

El **principio del intervalo de confianza**: Tiene un Límite superior de confianza y un límite inferior de confianza.

En: $\Pr ob(\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2} \leq \hat{\beta}_1 \leq \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}) = 1 - \alpha$

El coeficiente de confianza: $1 - \alpha$

Límites de confianza inferior: $\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2}$

Límites de confianza superior: $\hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}$

**Estimación de la varianza del término de perturbación:**

$$\sigma_{\mu}^2 = \frac{e'e}{(n-k)} = \frac{Y'Y - \beta' X' Y}{n-k}$$

**Docima de hipótesis,** se refiere a una distribución de frecuencias, y se plantea con el fin de comprobar si se cumple una relación.

Hipótesis propuesta o sometida a análisis.

$H_0$: $\beta_i = C$ hipótesis nula

$H_1$: $\beta_i \neq C$ hipótesis alternativas

**Coeficiente de determinación ($R^2$),** es un indicador de la bondad de ajuste de la línea de regresión que mide la proporción de la variación total en la variable dependiente Y, que "se explica" o "se debe a" la variación de la variable independiente X. ($0 \leq R^2 \leq 1$).

**Coeficiente de correlación,** es una medida de asociación lineal entre dos variables.

| COEFICIENTE | FORMULA | | |
|---|---|---|---|
| COEFICIENTE DE DETERMINACION | $R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$ | | $0 \leq R^2 \leq 1$ |
| COEFICIENTE DE CORRELACION | **Poblacional** $r = \frac{Cov(x,y)}{\sqrt{\sigma_x^2 \sigma_y^2}}$ | = | **Muestral** $\frac{\sum (x_i - \bar{x})(x_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1} \frac{\sum (y_i - \bar{y})^2}{n-1}}}$ |
| COEFICIENTE DE DETERMINACION MULTIPLE CORREGIDO | $\bar{R}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / n-k}{\sum (y_i - \bar{y}_i)^2 / n-1}$ | | |

La finalidad del **Análisis de varianza** es investigar la explicación conjunta de todas las variables explicativas que intervienen en el modelo, a partir del estudio de los componentes de la variabilidad total.

Para efectuar la dócima se construye un estadístico: $F_c = \dfrac{\text{Varianza explicada}}{\text{varianza no explicada}}$

Eviews es un programa especializado que sirve para hacer análisis, regresiones y predicción así como simulaciones y evaluaciones de eficiencia y predicción de modelos.

**EXPLORACION ON LINE**

1. Regresión lineal entre dos variables: INFERENCIA

   http://bayes.escet.urjc.es/~jmmarin/libroelec,tema 8/inferencia

2. Inferencia sobre los coeficientes del modelo

   http://www.udc.es/mate/estadisticas2/sec8.htm

3. Inferencia en el Modelo Lineal

   http://uhu.es/45132/ficheros-datos

## LECTURA

### Aplicación estadística del Análisis de Varianza

The Wall Street Journal publicó hace poco un artículo sobre Ruport Murdoch, el editor nacido en Australia que ha construido un imperio mundial de medios de difusión con préstamos concedidos por bancos de todo el mundo.

Como afirmaba el Wall Street Journal, ahora se enfrenta al período de apalancamiento al de tomar decisiones difíciles en relación con las amortizaciones.

Gran parte de la deuda se contrajo durante una época de expansión, en la compra y formación de tres nuevas empresas, TV Guide, Sky Lab y la red de televisión Fox (que atraviesa ciertas dificultades a pesar del éxito de Bart Simpson y su familia de inadaptados).

El artículo afirmaba que Murdoch tendría que comprara las situaciones deudoras relativas de cada uno de los tres nuevos aspectos de su aventura empresarial.

Quizá fuera preciso recurrir al
Análisis de Varianza para comparar los niveles medios de endeudamiento en cada una de las tres nuevas y arriesgadas empresas.

Allen L. Webster

## ACTIVIDADES

1. En esta fórmula: $\Pr ob(\hat{\beta}_1 - \hat{\sigma}_{\beta 1} t_{\alpha/2} \leq \beta_1 \leq \hat{\beta}_1 + \hat{\sigma}_{\beta 1} t_{\alpha/2}) = 1 - \alpha$

   ¿Cuál es el coeficiente de confianza?, ¿Cuál es el límite de confianza inferior y Cuál es el límite de confianza superior?

2. Menciona los puntos importantes de las Dócimas de Hipótesis

3. Escribir la fórmula del coeficiente de correlación y el significado de las siglas SCT, SCR, SCE.

4. Explicar en forma resumida el análisis de las salidas de la regresión (del programa E-Views).

5. Para los modelos lineales que se proponen a continuación:

   a. $Y = \beta 1 + \beta_2 X_2 + \mu$
      $Y$ : producción de trigo (quintales por hectárea)
      $X_2$ : Cantidad de fertilizante (kilos por hectárea)

   b. $Y = \beta_1 + \beta_2 X_2 + \mu$
      $Y$ : Ganancia de la empresa.
      $X_2$ : Gastos de Inversión

   Determinar algunas de la variables cuyo efecto están contenidas en la "$\mu$".

6. Los datos siguientes corresponden al producto bruto interno por tipo de gasto, consumo de hogares, consumo del gobierno, formación bruta del capital, exportaciones e importaciones en el periodo 1991-2002.

   a. Determinar el modelo que más se adecua a los datos. ¿Cuál se utilizaría como variable dependiente y cuales como variables independientes, justifique?
   b. Luego de realizar el inciso anterior realice la regresión más adecuada y estime los parámetros convenientes, seguidamente calcúlese sus intervalos de confianza respectivos.
   c. ¿Cómo puede asegurar que el modelo que ha seleccionado es el adecuado?
   d. El modelo que planteo inicialmente cumple con algún supuesto básico de un modelo de regresión. Realice las pruebas convenientes

   Considere:

   $Y$ : Producto Bruto Interno
   $X_1$ :Consumo de hogares
   $X_2$ :Consumo de gobierno
   $X_3$ :Formación bruta de capital
   $X_4$ :Exportaciones
   $X_5$ :Importaciones

| Año | Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|---|
| 1991 | 26686 | 20607 | 2 067 | 4613 | 3260 | 3862 |
| 1992 | 44953 | 34934 | 3566 | 7780 | 5628 | 6954 |
| 1993 | 69262 | 52996 | 5568 | 13376 | 8627 | 11304 |
| 1994 | 98577 | 71306 | 8672 | 21931 | 12590 | 15922 |
| 1995 | 120858 | 85933 | 11786 | 30013 | 15118 | 21991 |
| 1996 | 136929 | 98598 | 13827 | 31283 | 17975 | 24754 |
| 1997 | 157274 | 110782 | 15487 | 37952 | 22272 | 29219 |
| 1998 | 165949 | 118269 | 17296 | 39321 | 22076 | 31014 |
| 1999 | 173957 | 122261 | 18854 | 36894 | 25855 | 29907 |
| 2000 | 185281 | 131565 | 19717 | 37545 | 29851 | 33396 |
| 2001 | 188172 | 136040 | 20290 | 35082 | 30020 | 33260 |
| 2002 | 198437 | 142960 | 20703 | 36563 | 32612 | 34402 |

**Fuente: Instituto Nacional de Estadística e Informática-Dirección Nacional de Cuentas Nacionales.**

## AUTOEVALUACIÓN

Encierra en un círculo la letra que contenga la alternativa correcta.

1. Si el β de la Hipótesis nula está dentro del intervalo de confianza se..……….. la hipótesis nula; contrariamente, si el β está fuera del intervalo se..………. la hipótesis.

   a. Rechaza-acepta
   b. Rechaza-rechaza
   c. Acepta-rechaza
   d. Todas las anteriores

2. El rango del coeficiente de determinación $R^2$ es el siguiente:

   a. 0> R2> 1.
   b. 0≤ R2≤ 1.
   c. 0> R2≤ 1.
   d. Ninguna de las anteriores

3. Cuando $\mathbf{R^2}$ es muy…………….. se dice que el modelo de regresión es capaz de explicar un alto porcentaje de las variaciones que registra la variable explicada.

   a. Cercano a 1
   b. Cercano a 0
   c. Lejano a 1
   d. Todas las anteriores

4. El proceso de inferencia consiste en establecer la validez de determinadas afirmaciones acerca de………….. (desconocidos) utilizando un estimador obtenido a partir de una muestra, pero del cual se puede determinar su distribución muestral.

   a. Los estimadores
   b. Las muestras

c. Los coeficientes
d. Los parámetros

5. El análisis de varianza tiene por finalidad investigar la explicación conjunta de todas:

   a. Las variables explicadas intervinientes en el modelo.
   b. Las variables aleatorias en el modelo.
   c. Las variables explicativas intervinientes en el modelo
   d. Ninguna de las anteriores.

6. Mencione el Test de contraste de cambio:

   a. Test de Chow
   b. Test de Cusum
   c. Test de suma acumulada
   d. Ninguno de los anteriores

7. Los datos siguientes corresponden al gasto promedio del hogar según actividad económica del jefe del hogar y corresponden al año 2002.

   Considere:

   Y: Gasto mensual del hogar
   $X_1$: Alimentos
   $X_2$: Vestido y Calzado
   $X_3$: Alquiler de Vivienda, Combustible, Electricidad y Conservación de Vivienda
   $X_4$: Muebles y Enseres
   $X_5$: Cuidado y conservación de la salud
   $X_6$: Transportes y Comunicaciones

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | Y |
|---|---|---|---|---|---|---|
| 344.13 | 29.50 | 65.50 | 17.12 | 44.56 | 23.43 | 585.08 |
| 596.23 | 47.36 | 380.40 | 43.98 | 111.60 | 116.82 | 1 547.68 |
| 628.64 | 68.70 | 237.27 | 51.29 | 151.15 | 136.79 | 1 586.46 |
| 597.22 | 53.36 | 324.35 | 51.58 | 129.70 | 162.37 | 1 599.22 |
| 699.82 | 75.72 | 470.63 | 63.70 | 138.02 | 180.20 | 1 972.95 |
| 559.50 | 49.00 | 266.64 | 37.65 | 84.25 | 68.18 | 1 277.54 |
| 592.19 | 42.84 | 332.73 | 46.96 | 114.69 | 94.26 | 1 498.61 |
| 791.74 | 42.81 | 289.59 | 35.69 | 134.48 | 65.78 | 1 603.19 |
| 676.34 | 61.72 | 346.36 | 59.45 | 133.70 | 173.60 | 1 720.58 |
| 655.51 | 66.57 | 374.11 | 64.67 | 146.09 | 145.50 | 1 848.41 |
| 635.96 | 54.49 | 434.68 | 80.71 | 176.65 | 227.35 | 1 992.75 |
| 367.28 | 27.93 | 373.74 | 40.66 | 128.28 | 124.38 | 1 212.56 |

**Fuente: ENAHO: 2002**

Planteando un modelo de regresión lineal para estas variables, se estimó el modelo por MCO y se obtuvieron los siguientes resultados:

Dependent Variable: Y
Method: Least Squares

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| X1 | 1.0474 | 0.1202 | 8.7116 | 0.0003 |
| X2 | 2.8746 | 1.0435 | 2.7548 | 0.0401 |
| X3 | 1.0607 | 0.1404 | 7.5557 | 0.0006 |
| X4 | 6.9652 | 1.8482 | 3.7687 | 0.0130 |
| X5 | 1.7287 | 0.5802 | 2.9797 | 0.0308 |
| X6 | -0.1579 | 0.5122 | -0.3083 | 0.7703 |
| C | -118.9702 | 41.8257 | -2.8444 | 0.0361 |

| | | | |
|---|---|---|---|
| R-squared | 0.9977 | Mean dependent var | 1537.0860 |
| Adjusted R-squared | 0.9950 | S.D. dependent var | 383.6970 |
| S.E. of regresión | 27.2004 | Akaike info criterion | 9.7355 |
| Sum squared resid | 3699.3140 | Schwarz criterion | 10.0184 |
| Log likelihood | -51.4132 | F-statistic | 363.9769 |
| Durbin-Watson stat | 2.2961 | Prob(F-statistic) | 0.0000 |

¿Cuál de las siguientes alternativas es la correcta?

a. No existe ningún parámetro estimado significativo, $\mathbf{R^2}$ = 0.9950
b. Solo el parámetro de $X_1$ es significativo, $\mathbf{R^2}$ = 0.9590
c. Solo el parámetro de $X_6$ es no significativo, $\mathbf{R^2}$ = 0.9977
d. Los parámetros de $X_1$ y $X_2$ son significativos, $\mathbf{R^2}$ = 0.9850

**RESPUESTAS DE CONTROL**

1. c, 2. b, 3. a, 4. d, 5. c, 6.a 7.c