Lección 7

Marcos Bujosa

10 de noviembre de 2023

Índice

1. Contrastes de hipótesis simples

Guión: HtestingHouses.inp

Nota 1 la función `pvalue(t,gl,Valor)` calcula la probabilidad a la derecha de `Valor` (Por tanto, puede calcular la probabilidad por la izquierda así:

`1-pvalue(t,gl,Valor)`, o bien así: `pvalue(t,gl,-Valor)`

Nota 2 Es posible que necesite ejecutar todas las órdenes desde la línea de comandos (es decir, sin menús ni ratón).

Cargue los datos `data3-1.gdt` del libro de Ramanathan.

```
open data3-1
```

1. Ajuste por MCO el precio en función de la superficie y guarde el modelo como icono

```
Modelo   <- ols price const sqft
```

2. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0:\ \beta_2 = 0{,}1$ frente a $H_1:\ \beta_2 > 0{,}1$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b2=0.1)    Halt (b2>0.1)
scalar t1 = ($coeff(sqft)-0.1)/$stderr(sqft)
scalar c1 = critical(t,$df,.05)
scalar p1 = pvalue(t,$df,t1)
```

3. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0:\ \beta_2 = 0{,}15$ frente a $H_1:\ \beta_2 < 0{,}15$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b2=0.15) Halt (b2<0.15)
scalar t2 = ($coeff(sqft)-0.15)/$stderr(sqft)
scalar c2 = -1*critical(t,$df,.05)          # alternativa A
scalar c2 =    critical(t,$df,.95)          # alternativa B
scalar p2 = 1-pvalue(t,$df, t2)             # Alternativa 1
scalar p2 =   pvalue(t,$df,-t2)             # Alternativa 2
```

4. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0:\ \beta_1 = 0$ frente a $H_1:\ \beta_1 < 0$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b1=0)    Halt (b1<0)
scalar t3 = ($coeff(const)-0)/$stderr(const)
scalar c3 =   critical(t,$df,.95)
scalar p3 = 1-pvalue(t,$df,t3)
```

5. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0:\ \beta_1 = 0$ frente a $H_1:\ \beta_1 > 0$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b1=0)    Halt (b1>0)
scalar t3 = ($coeff(const)-0)/$stderr(const)
scalar c3 = critical(t,$df,.05)
scalar p3 = pvalue(t,$df,t3)
```

6. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0:\ \beta_2 = 0{,}15$ frente a $H_1:\ \beta_2 \neq 0{,}15$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b2=0.15)  Halt (b1~=0.15) Bilateral
scalar t4 = ($coeff(sqft)-0.15)/$stderr(sqft)
scalar c4 = critical(t,$df,.025)
scalar p4 = 2*pvalue(t,$df,abs(t4))
scalar f4 = t4^2
scalar pf4 = pvalue(f,1,$df,f4)
```

7. Efectúe los cálculos necesarios para obtener el estadístico t para contrastar $H_0: \beta_1 + \beta_2 = 10$ frente a $H_1: \beta_1 + \beta_2 \neq 10$ con una significación del 5 %. Calcule el p-valor del estadístico.

```
# Hnula (b1+b2=10)  Halt (b1+b2~=10) Bilateral
scalar t5 = ($coeff(sqft)+$coeff(const)-10)/sqrt($vcv[1,1]+$vcv[2,2]+2*$vcv[1,2])
scalar c5 = critical(t,$df,.025)
scalar p5 = 2*pvalue(t,$df,abs(t5))
scalar f5 = t5^2
scalar pf5 = pvalue(f,1,$df,f5)
```

8. Estos dos últimos contrastes son bilaterales. Los contrastes bilaterales se pueden realizar fácilmente desde los menús de Gretl. Compruebe que obtiene los mismos resultados abriendo la ventana del modelo estimado y siguiendo los pasos *Contrastes ->Restricciones lineales* y tecleando en la ventana

```
b[1] + b[2] = 10
```

Y pulse *Aceptar*. Observe que Gretl usa el contraste F. Calcule el cuadrado del contraste t y compruebe que da exactamente el mismo resultado.

1.1. Código completo de la práctica `HtestingHouses.inp`

```
open data3-1

Modelo   <- ols price const sqft

# Hnula (b2=0.1) Halt (b2>0.1)
scalar t1 = ($coeff(sqft)-0.1)/$stderr(sqft)
scalar c1 = critical(t,$df,.05)
scalar p1 = pvalue(t,$df,t1)

# Hnula (b2=0.15) Halt (b2<0.15)
scalar t2 = ($coeff(sqft)-0.15)/$stderr(sqft)
scalar c2 = -1*critical(t,$df,.05)          # alternativa A
scalar c2 =    critical(t,$df,.95)          # alternativa B
scalar p2 = 1-pvalue(t,$df, t2)             # Alternativa 1
scalar p2 =   pvalue(t,$df,-t2)             # Alternativa 2

# Hnula (b1=0) Halt (b1<0)
scalar t3 = ($coeff(const)-0)/$stderr(const)
scalar c3 =   critical(t,$df,.95)
scalar p3 = 1-pvalue(t,$df,t3)

# Hnula (b1=0) Halt (b1>0)
scalar t3 = ($coeff(const)-0)/$stderr(const)
scalar c3 = critical(t,$df,.05)
scalar p3 = pvalue(t,$df,t3)

# Hnula (b2=0.15) Halt (b1 =0.15) Bilateral
scalar t4 = ($coeff(sqft)-0.15)/$stderr(sqft)
scalar c4 = critical(t,$df,.025)
scalar p4 = 2*pvalue(t,$df,abs(t4))
scalar f4 = t4^2
scalar pf4 = pvalue(f,1,$df,f4)

# Hnula (b1+b2=10) Halt (b1+b2 =10) Bilateral
scalar t5 = ($coeff(sqft)+$coeff(const)-10)/sqrt($vcv[1,1]+$vcv[2,2]+2*$vcv[1,2])
scalar c5 = critical(t,$df,.025)
scalar p5 = 2*pvalue(t,$df,abs(t5))
scalar f5 = t5^2
scalar pf5 = pvalue(f,1,$df,f5)
```

2. p-valor, potencia del contraste y otras distribuciones para las perturbaciones

Guión: samplinghouses4.inp

Este ejercicio es una modificación del guión `samplinghouses1.inp` de la lección anterior.

Nota 1 En este ejercicio todos los contrastes son bilaterales

Nota 2 Le recomiendo abrir directamente el guión `samplinghouses4.inp` y modificar lo que sea necesario para realizar cada apartado.

Comencemos la práctica...

A) En lugar de almacenar los valores estimados para los parámetros, este guión almacena los estadísticos t para el contraste $H_0 : b_1 = 52$ frente a $H_1 : b_1 \neq 52$, así como los p valores de dichos estadísticos. Nótese que la hipótesis nula es cierta en este ejemplo simulado (¡esa es la ventaja de simular!).

B) Recupere esos datos almacenados y compruebe qué porcentaje de veces los p-valores son mayores que 0.05 (cuantas veces hubiéramos rechazado H_0 pese a ser cierta con una significación del 5 %. ¿Le sorprende el resultado?)

C) Repita desde el principio el ejercicio, pero simulando perturbaciones con una distribución muy alejada de la normal (por ejemplo empleando una distribución χ^2 con un grado de libertad y restando 1 para que su esperanza sea nula: `gl=1` y `series U = randgen(X, gl) - gl`. ¿Cambian mucho los resultados?

D) Lo visto en el apartado anterior puede ser debido a que la muestra es muy pequeña. Repita el ejercicio pero simulando superficies de pisos.

- Comente la línea `open data3-1.gdt` y añada debajo `nulldata 150`.
- Comente `series x = sqft` y añada debajo `series x = randgen(U,1000,3000)`.

Es decir, simule (con distribución uniforme) tamaños de 150 pisos con un rango igual al de la verdadera muestra (entre 1000 y 3000 pies cuadrados).

Repita el ejercicio, con los datos simulados: primero con distribución normal, y luego con una distribución alejada de la Normal.

- ¿Cambian los resultados?
- ¿Y si aumenta más aún el tamaño muestral? (por ejemplo `nulldata 500`

E) **(Función potencia)** Vuelva a usar tamaños muestrales de 14 datos (bien empleando los datos originales, o bien simulando 14 superficies), y simule perturbaciones con distribución normal.

Repita el ejercicio pero esta vez para contrastar hipótesis falsas, por ejemplo $H_0 : b_1 = 50$, ó $H_0 : b_1 = 30$, ó $H_0 : b_1 = 100$ ó $H_0 : b_1 = 0$.

- ¿Puede encontrar una pauta en los los resultados?
- ¿Sabe lo que es la potencia de un contraste?
- ¿Depende de algún modo el comportamiento del test respecto del tamaño muestral? Por ejemplo, contraste $H_0 : b_1 = 30$ con muestras de 14, 150 y 500 datos. ¿Qué observa?
- Si, siendo el verdadero parámetro 52, contrasta al 5 % la hipótesis $H_0 : b_1 = 51{,}7$ ¿Se rechaza con mucha frecuencia H_0?

F) En los apartados (A) y (D) (*donde contrastábamos una hipótesis nula* $H_0 : b_1 = 52$ *que era cierta*) hemos visto que los resultados no parecían muy dependientes de la distribución de las perturbaciones.

Emplee una muestra de tamaño 500 y simule perturbaciones con distribución χ^2 con un grado de libertad (y reste los grados de libertad para que su esperanza sea nula).

Contraste al 5 % la falsa hipótesis $H_0 : b_1 = 51{,}7$.

El porcentaje de rechazos cuando empleamos distribución normal era approx. el 5 %

- ¿qué pasa cuando simulamos una distribución muy alejada de la normal? (χ^2_1)
- ¿Y si aumenta el número de grados de libertad?
- Pruebe con distribuciones χ^2_{10}, χ^2_{25}, χ^2_{50} y χ^2_{100}.

2.1. Código completo de la práctica `samplinghouses4.inp`

```
open data3-1.gdt
# nulldata 150

series x = sqft
# series x = randgen(U,1000,3000)

# Generamos parte sistematica del modelo
series ys = 52 + 0.14*x

# Desviacion tipica de las perturbaciones
scalar s  = 39
scalar gl = 1
loop 100000 --progressive --quiet
    series U = randgen(n, 0, s)
    #series U = randgen(X, gl) - gl
    scalar m = mean(U)
    series y = ys + U

    ols y const x
    # Hnula (b1=52)
    scalar t = ($coeff(const)-52)/$stderr(const)
    scalar p = 2*pvalue(t,$df,abs(t))

    print p m
    store "@workdir\pvalue.gdt" p m
endloop

open "@workdir\pvalue.gdt"
series ns = p<0.05
summary ns
freq ns --plot="display"
```