Lección 9

Marcos Bujosa

16 de octubre de 2023

Índice

1. Intervalos y regiones de confianza

Guión: EjPvivienda4.inp

Cargue los datos de precios de casas `data3-1.gdt` del libro de Ramanthan.

```
open data3-1
```

- Estime el modelo de siempre y guárdelo como icono.

  ```
  Modelo1 <- ols price const sqft --vcv
  ```

- Observe la matriz de covarianzas entre los parámetros estimados del modelo de regresión. Hay covarianza entre los estimadores, ¿con qué signo?

- Calcule los intervalos de confianza de los parámetros beta estimados: desde la ventana del modelo estimado siga los pasos *Análisis ->Intervalos de confianza para los coeficientes* o bien directamente en un guión o la consola de Gretl aplique directamente las expresiones vistas en clase:

  ```
  scalar ns = 0.05  # nivel de signifacion

  scalar i1 = $coeff(const) - critical(t, $df, ns/2) * $stderr(const)
  scalar s1 = $coeff(const) + critical(t, $df, ns/2) * $stderr(const)

  scalar i2 = $coeff(sqft)  - critical(t, $df, ns/2) * $stderr(sqft)
  scalar s2 = $coeff(sqft)  + critical(t, $df, ns/2) * $stderr(sqft)

  matrix IC = {i1, s1; i2, s2}
  print IC
  ```

- Recuerde que los intervalos de confianza al 95 % nos sirven para contrastar hipótesis al 5 % de significación. Piense qué valores están en el umbral de ser rechazados según los intervalos obtenidos.

- Dado el signo de la covarianza entre los estimadores, ¿qué relación cabe esperar entre ellos?

- Visualice la región de confianza de los parámetros: desde en la ventana del modelo estimado siga los pasos *Análisis ->Elipse de confianza* y seleccione ambos regresores para ver la elipse de confianza. ¿Confirma sus expectativas?

- Contrastemos algunas hipótesis compuestas. En particular, contraste las distintas combinaciones de valores que están en el umbral de ser hipótesis a rechazar individualmente (las correspondientes a las esquinas del cuadrado que se ve en el gráfico del apartado anterior). En la ventana del modelo *Contrastes ->Restricciones lineales* y se introducen las restriciones; o bien ejecute el código de más abajo.

  ¿Cuál es la conclusión respecto a la elipse de confianza en relación a los contrastes de hipótesis de dos parámetros?

  ```
  restrict Modelo1
   b[1] = i1
   b[2] = s2
  end restrict

  restrict Modelo1
   b[1] = s1
   b[2] = s2
  end restrict
  ```

```
restrict Modelo1
 b[1] = i1
 b[2] = i2
end restrict

restrict Modelo1
 b[1] = s1
 b[2] = i2
end restrict

/* notese que el contraste individual H0: b[2]=i2 esta en el limite de ser rechazado al 5%
   Pero si conjuntamente tambien se considera que b[1]=s1 entonces no se rechaza ni con un
   13% de sigificacion */
```

1.1. Código completo de la práctica `EjPvivienda4.inp`

```
open data3-1

Modelo1 <- ols price const sqft --vcv

scalar ns = 0.05   # nivel de signifacion

scalar i1 = $coeff(const) - critical(t, $df, ns/2) * $stderr(const)
scalar s1 = $coeff(const) + critical(t, $df, ns/2) * $stderr(const)

scalar i2 = $coeff(sqft)  - critical(t, $df, ns/2) * $stderr(sqft)
scalar s2 = $coeff(sqft)  + critical(t, $df, ns/2) * $stderr(sqft)

matrix IC = {i1, s1; i2, s2}
print IC

restrict Modelo1
 b[1] = i1
 b[2] = s2
end restrict

restrict Modelo1
 b[1] = s1
 b[2] = s2
end restrict

restrict Modelo1
 b[1] = i1
 b[2] = i2
end restrict

restrict Modelo1
 b[1] = s1
 b[2] = i2
end restrict

/* notese que el contraste individual H0: b[2]=i2 esta en el limite de ser rechazado al 5%
   Pero si conjuntamente tambien se considera que b[1]=s1 entonces no se rechaza ni con un
   13% de sigificacion */
```

2. Experimento de Montecarlo para los intervalos de confianza

Guión: samplinghouses5.inp

Cargue los datos de precios de casas `data3-1.gdt` del libro de Ramanthan. Este experimento de Montecarlo es una extensión a los ya realizados con estos mismos datos.

- Defina la serie `x` con las superficies y la serie `y` con los precios. Programe un bucle como las otras veces:

```
open data3-1
series x = sqft
series y = price
#set seed 3213798
loop 100 --progressive --quiet
    <<Estimación MCO>>
    <<Cálculo de los límites de los intervalos de confianza>>
    <<Comprobación de si los parámetros pertenecen al intervalo>>
    <<Estimación de la varianza de los errores>>
    <<Muestra de resultados y guardado en disco>>
endloop
```

La serie de cálculos son los siguientes (todos dentro del bucle).

1. El primer bloque de cálculos simula el modelo con nuevas perturbaciones, lo estima por MCO y guarda los betas estimados y sus errores estándar:

```
series U  = randgen(n, 0, 39)
series ys = 52 + 0.14*x + U
ols ys const x
scalar b1 = $coeff(const)
scalar b2 = $coeff(x)
scalar s1 = $stderr(const)
scalar s2 = $stderr(x)
```

2. Luego calculamos los intervalos de confianza al 95\

```
scalar c1L = b1 - critical(t,$df,.025)*s1
scalar c1R = b1 + critical(t,$df,.025)*s1
scalar c2L = b2 - critical(t,$df,.025)*s2
scalar c2R = b2 + critical(t,$df,.025)*s2
```

3. Verificamos si los verdaderos valores pertenecen al intervalo estimado

```
scalar p1 = (52  >c1L && 52  <c1R)
scalar p2 = (0.14>c2L && 0.14<c2R)
```

4. Al finalizar todas las cuentas, queremos que Gretl nos muestre los estadísticos de los parámetros estimados, y el porcentaje de veces que el intervalo contuvo a los parámetros, y que guarde todo lo calculado en el fichero de datos `cicoeff.gdt`.

```
print b1 b2 p1 p2
store cicoeff.gdt b1 b2 c1L c1R c2L c2R
```

Por último abrimos el fichero `cicoeff.gdt`. Generamos una serie `B1` constante e igual a valor del parámetro constante del modelo simulado (52) y otra `B2` igual a valor del parámetro de la pendiente del modelo simulado (0,14); y finalmente generamos unos gráficos que muestren los intervalos estimados en cada iteración y si incluyen o no al verdadero valor del parámetro correspondiente.

```
open "cicoeff.gdt"
series B1 = 52
series B2 = 0.14
IntervaloConstante <- gnuplot c1L c1R B1 --time-series --with-lines --output="display"
IntervaloPendiente <- gnuplot c2L c2R B2 --time-series --with-lines --output="display"
```

2.1. Código completo de la práctica `samplinghouses5.inp`

```
open data3-1
series x = sqft
series y = price
#set seed 3213798
loop 100 --progressive --quiet
    series U  = randgen(n, 0, 39)
    series ys = 52 + 0.14*x + U
    ols ys const x
    scalar b1 = $coeff(const)
    scalar b2 = $coeff(x)
    scalar s1 = $stderr(const)
    scalar s2 = $stderr(x)
    scalar c1L = b1 - critical(t,$df,.025)*s1
    scalar c1R = b1 + critical(t,$df,.025)*s1
    scalar c2L = b2 - critical(t,$df,.025)*s2
    scalar c2R = b2 + critical(t,$df,.025)*s2
    scalar p1 = (52  >c1L && 52  <c1R)
    scalar p2 = (0.14>c2L && 0.14<c2R)

    print b1 b2 p1 p2
    store cicoeff.gdt b1 b2 c1L c1R c2L c2R
endloop

open "cicoeff.gdt"
series B1 = 52
series B2 = 0.14
IntervaloConstante <- gnuplot c1L c1R B1 --time-series --with-lines --output="display"
IntervaloPendiente <- gnuplot c2L c2R B2 --time-series --with-lines --output="display"
```

3. Precio de casas unifamiliares (constante más tres regresores)

Guión: Houses.inp

Objetivos. Son tres:

- Asimilar la interpretación "ceteris paribus" de los coeficientes de un modelo de regresión.
- Eliminar variables no significativas de un modelo.
- Comparar el ajuste de dos modelos.

Los datos son los del ejemplo de clase junto con dos variables adicionales: número de dormitorios (`bedrms`) y cuartos de baño (`baths`).

Para empezar cargue los datos de la base de datos de Gretl desde la pestaña del manual de Ramanathan

```
open data4-1
```

3.1. Tareas

Actividad 1 Piense cuáles son los signos esperados de los parámetros del siguiente modelo

$$PRICE = \beta_1 + \beta_2 SQFT + \beta_3 BEDRMS + \beta_4 BATHS + U.$$

donde $PRICE$ es el precio de una vivienda, $SQFT$ su superficie, $BEDRMS$ es su número de dormitorios y $BATHS$ su número de cuartos de baño.

Actividad 2 Con los datos de la muestra Ajuste dicho modelo de regresión por MCO y guárdelo como un icono.

```
Modelo1 <- ols price 0 sqft bedrms baths
```

Actividad 3 ¿Confirman los resultados su previsión respecto a los signos de los parámetros?

Actividad 4 El modelo estimado sugiere que una casa de 1600 pies al cuadrado, con 3 habitaciones y 2 cuartos de baño tiene un precio esperado de

$$129{,}062 + 0{,}154800 \times (1600) - 21{,}5875 \times (3) - 12{,}1928 \times (2) = 287{,}59 \text{ miles de dólares.}$$

Según este modelo (y sabiendo que un pie cuadrado son $0{,}092m^2$) *ampliar* esta casa con una habitación adicional de unos 20 metros cuadrados (unos 220 pies cuadrados más de casa) arrojaría un precio esperado de ¿cuanto?

$$129{,}062 + 0{,}154800 \times (1820) - 21{,}5875 \times (4) - 12{,}1928 \times (2) = 300{,}06;$$

es decir, *ampliar* esta casa con una habitación adicional de unos 20 metros cuadrados *eleva* el precio en unos 13 mil dolares. ¿Contradice esto su idea inicial?

```
series yhat1 = $coeff(const)+$coeff(sqft)*1600+$coeff(bedrms)*3+$coeff(baths)*2
series yhat2 = $coeff(const)+$coeff(sqft)*(1600+220)+$coeff(bedrms)*(3+1)+$coeff(baths)*2
```

Actividad 5 Aunque, con la correcta interpretación de los coeficientes, el resultado parece sensato, sólo una de las variables es estadísticamente significativa.

- Esto quiere decir, que la estimación de los parámetros es muy imprecisa. Sin embargo, el estadístico $\mathcal{F}$ indica que el modelo es conjuntamente significativo (así que la previsión de precios de la actividad anterior es fiable); pero no el efecto individual de cada regresor.
- Observe además que hay un elevado grado de correlación entre los regresores (en la ventana de iconos, "pinche" en *Correlaciones*) o ejecute los siguiente

```
corr sqft bedrms baths
```

Esto sugiere que pudiera surgir un problema de multicolinealidad. Además, la interpretación caeteris paribus cuestiona la relevancia de algunos los regresores. Todo ello apunta a la posible conveniencia de excluir uno o más regresores del modelo inicial.

Actividad 6 Puesto que la variable menos significativa es `baths`, ésta será la primera variable a omitir del modelo inicial:

- En la ventana del Modelo 1, "pinche" en ***Contrastes ->omitir variables*** y seleccione la variable `baths` (y pulse ***Aceptar***). O bien ejecute el código

  ```
  omit baths
  ```

  Note como el coeficiente de determinación ha disminuido, pero el corregido ha aumentado.

Aunque ha aumentado la significatividad de los coeficientes, `bedrms` continua siendo no significativa. Así que la vamos a omitir:

- En la ventana del Modelo 1, "pinche" en ***Contrastes ->omitir variables*** y seleccione la variable `bedrms` (y pulse ***Aceptar***). O bien ejecute el código

  ```
  omit bedrms
  ```

  Note como de nuevo el coeficiente de determinación ha disminuido, pero el corregido ha aumentado.

La constante sigue sin ser significativa ¿debemos hacer como con `baths` *y* `bedrms` *y tratar de omitirla también?*

Actividad 7 Hemos excluido las variables `baths` y `bedrms` del modelo original, debido a que individualmente son no significativas.

Pero pudiera ocurrir que conjuntamente si fueran significativas. Vamos a verificar que conjuntamente tampoco son significativas:

- En la ventana del Modelo 1, "pinche" en ***Análisis ->Elipse de confianza*** y seleccione `baths` y `bedrms`.
- Compruebe que $(0, 0)$ es un punto que pertenece a la elipse de confianza, y que por tanto, la hipótesis nula $\beta_3 = 0,\ \beta_4 = 0$ no se puede rechazar al nivel de confianza de la elipse.
- Omita de golpe las dos variables: En la ventana del Modelo 1, "pinche" en ***Contrastes ->omitir variables*** y seleccione las variables `baths` y `bedrms` (y pulse ***Aceptar***)

  ```
  ols price 0 sqft bedrms baths
  Modelo2 <- omit baths bedrms
  ```

- Observe los resultados.

Actividad 8 Comparar las previsiones del modelo ampliado y el reducido.

Vamos a comprobar que las predicciones no difieren demasiado entre uno y otro modelo. Pero hay mayor precisión con el modelo reducido (no multicolinealidad).

Primero añadimos una observación adicional con los datos de superficie, número de dormitorios y cuartos de baño del ejemplo anterior.

- "Pinche" en ***Datos ->Añadir observaciones*** y añada una observación
- Marque con el ratón las variables `sqft`, `bedrms` y `baths`}, y con las tres variables marcadas, pinche en ***Datos ->Editar valores***. Teclee 1820 en la fila 15 de la columna `sqft`, teclee 4 en la fila 15 de la columna `bedrms` y 2 en la fila 15 de la columna `baths`.
- Fije la muestra con las nuevas observaciones: "pinche" en ***Muestra ->Establecer rango*** y fíjelo para las observaciones de 1 a 15. O bien teclee en la consola de comandos

```
dataset addobs 1
genr sqft[15] =1820
genr bedrms[15]=4
genr baths[15] =2
```

- Re-estime el primer modelo pero usando la muestra ampliada.

  ```
  smpl 1 15
  ols price const sqft bedrms baths
  ```

  - En la ventana del modelo re-estimado "pinche" en *Análisis ->Predicciones*. Nos aseguramos de que el dominio de predicción contiene la observación 15 y pulsamos *Aceptar*.

    ```
    fcast 15 15 --static
    fcast --plot="display"
    ```

  - Observe la previsión puntual y el intervalo de confianza

- Re-estime también el segundo modelo con la muestra ampliada

  ```
  ols price const sqft
  ```

  - En la ventana del modelo re-estimado "pinche" en *Análisis ->Predicciones*. Asegúrese de que el dominio de predicción contiene la observación 15 y pulse *Aceptar*.

    ```
    fcast 15 15 --static
    fcast --plot="display"
    ```

- Note cómo ambas previsiones puntuales están contenidas en ambos intervalos, por lo que no son estadísticamente distintas.
- Pero fíjese en cómo la falta de precisión del primer modelo se plasma en una mayor desviación típica, y en un intervalo de confianza más amplio.

3.2. Código completo de la práctica Houses.inp

```
open data4-1

Modelo1 <- ols price 0 sqft bedrms baths

series yhat1 = $coeff(const)+$coeff(sqft)*1600+$coeff(bedrms)*3+$coeff(baths)*2
series yhat2 = $coeff(const)+$coeff(sqft)*(1600+220)+$coeff(bedrms)*(3+1)+$coeff(baths)*2

corr sqft bedrms baths

omit baths

omit bedrms

ols price 0 sqft bedrms baths
Modelo2 <- omit baths bedrms

dataset addobs 1
genr sqft[15] =1820
genr bedrms[15]=4
genr baths[15] =2

smpl 1 15
ols price const sqft bedrms baths

fcast 15 15 --static
fcast --plot="display"

ols price const sqft

fcast 15 15 --static
fcast --plot="display"
```

4. Los determinantes del número de viajeros de autobús

Guión: BusTravelers.inp

Aplicación 4.6 del libro de texto [cite:@Ramanathan:IEW-98]. Para más detalles consulte el citado manual.

Objetivo Especificación de un modelo de regresión.

En general, eliminar variables no significativas incrementa la precisión de la estimación del resto de parámetros. Pero eliminar de golpe todas las variables no significativas no es recomendable. Al quitar algunas variables puede que otras se vuelvan significativas. Un procedimiento más cauteloso es ir quitando variables de una en una.

La significación estadística no es el único criterio, ni el más conveniente, para tomar una decisión. Si hay razones teóricas para mantener una variable debemos cuestionarnos seriamente el quitar dicha variable por el mero hecho de que ésta resulte no significativa estadísticamente.

Por otra parte, siempre debemos mantener el término constante.

Los datos y el modelo Ésta es una aplicación con datos reales que relaciona el número de horas viajadas en autobús (en miles) `Bustravl` (B) en 40 ciudades de los Estados Unidos con las siguientes variables explicativas:

- `Fare` = la tarifa del billete en dólares (F).
- `Gasprice` =Precio del galón de gasolina en dólares (G).
- `Income` = Renta percápita media de la ciudad en dólares (I).
- `Pop` = Población de la ciudad en miles (P).
- `Density` = Densidad de población de la ciudad (personas por milla cuadrada) (D).
- `Landarea` = Extensión del área urbana en millas cuadradas (L).

La especificación inicial del modelo es:

$$B = \beta_1 + \beta_2 F + \beta_3 G + \beta_4 I + \beta_5 P + \beta_6 D + \beta_7 L + U.$$

Se trata de decidir cuales de todas las variables disponibles es razonable mantener en un modelo final.

Para empezar cargue los datos de la base de datos de Gretl desde la pestaña del manual de Ramanathan.

```
open data4-4
```

4.1. Tareas

Actividad 1 Piense cuáles son los signos esperados de los parámetros del modelo

Actividad 2 Observe los diagramas de dispersión entre la variable dependiente y los regresores. Observe así mismo los estadísticos descriptivos de las variables y la matriz de correlaciones. Hágalo con los menus desplegables o ejecutando el siguiente código:

```
summary BUSTRAVL FARE GASPRICE INCOME POP DENSITY LANDAREA --simple
corr BUSTRAVL FARE GASPRICE INCOME POP DENSITY LANDAREA --plot="display"
fare_bustravl     <- gnuplot BUSTRAVL FARE     --output="display"
gasprice_bustravl <- gnuplot BUSTRAVL GASPRICE --output="display"
income_bustravl   <- gnuplot BUSTRAVL INCOME   --output="display"
pop_bustravl      <- gnuplot BUSTRAVL POP      --output="display"
density_bustravl  <- gnuplot BUSTRAVL DENSITY  --output="display"
landarea_bustravl <- gnuplot BUSTRAVL LANDAREA --output="display"
```

Actividad 3 Estime por MCO el modelo inicial completo y guárdelo como un icono

```
ModeloInicial <- ols BUSTRAVL 0 FARE GASPRICE INCOME POP DENSITY LANDAREA
modeltab add
```

El nombre (en la línea anterior `ModeloInicial`) puede ser el que usted elija, pero si tiene espacios en blanco debe escribirlo entre comillas, por ejemplo `"Modelo uno" <- ols Y 0 x1 x2`). Recuerde además que Gretl entiende la variable `0` como el regresor constante).

Compare los resultados con sus expectativas iniciales.

- Observe qué variables son significativas y cuáles no. ¿Cuál es la menos significativa?
- Observe también el coeficiente de determinación ajustado, así como los criterios de información para la selección de modelos "Criterio de información de Akaike (AIC)" y "Criterio de información Bayesiano de Schwarz (BIC)" para compararlos con otros alternativos más adelante.

Actividad 4 La variable con mayor p-valor ("la menos significativa") es el precio del combustible `GASPRICE`. Realice una nueva regresión omitiendo esta variable.

Puede hacerlo partiendo del modelo anterior, o definiendo un nuevo modelo (le recomiendo la primera):

- Partiendo del modelo anterior podemos omitir la variable:
  - Desde la ventana del modelo estimado "pinche" en ***contrastes -->omitir variables*** seleccione `GASPRICE`; o bien ejecute[1]

    `omit GASPRICE`
  - Alternativamente podemos definir un nuevo modelo sin `GASPRICE`. "pinchando" en ***Modelo -->Mínimos Cuadrados Ordinarios*** y seleccionando todos los regresores menos `GASPRICE`, o bien ejecutando `ols BUSTRAVL 0 FARE INCOME POP DENSITY LANDAREA`

  Analice los resultados:

  - ¿Mejora la precisión de las estimaciones? ¿Qué variable es ahora la menos significativa?
  - Observe también el coeficiente de determinación ajustado, así como los criterios de información para la selección de modelos "Criterio de información de Akaike (AIC)" y "Criterio de información Bayesiano de Schwarz (BIC)". ¿Le parece que ha mejorado el modelo?

Actividad 5 El mayor p-valor corresponde ahora a `FARE`, pero la teoría económica sugiere que el precio es una importante variable a la hora de explicar la demanda de un bien o servicio. Por ello la debemos mantener. La siguiente variable menos significativa (un inaceptable p-valor del 50 %) es la extensión de la ciudad `LANDAREA`

Repita los pasos del punto anterior, pero omitiendo `LANDAREA`.

```
omit LANDAREA
```

Actividad 6 La variable `FARE` continua teniendo un p-valor excesivo (49 %). Esto sugiere, que dadas las demás variables explicativas, la tarifa del billete no afecta demasiado a la demanda de este servicio; es decir, los viajeros no parecer ser muy sensibles al precio del billete[2]. Observe de nuevo el diagrama de dispersión entre `FARE` y `BUSTRAVL` que siguiere este comportamiento. Por tanto, veamos que pasa si finalmente omitimos la variable `FARE` (si tenemos fundadas dudas sobre este comportamiento de los consumidores, entonces *no deberíamos eliminar la variable* `FARE` del modelo).

Repita los pasos omitiendo `FARE`.

```
omit FARE
```

[1] el comando opuesto a `omit` es `add`

[2] al menos para tarifas en torno a las observadas en la muestra: alrededor de 0.8 dolares con una tarifa mínima de 0.5$ y una máxima de 1.5$.

Actividad 7 Las variables del modelo final son significativas y los criterios de selección de modelos han mejorado, pero queda un último paso.

Hemos ido “quitando” variables basándonos en los estadísticos $\mathcal{T}$. Sabemos que esto es incorrecto puesto que conjuntamente estas variables podrían ser significativas. Así pues, debemos realizar un contraste de significación de las tres variables conjuntamente; es decir, partiendo del modelo inicial debemos contrastar

$$H_0: \ \beta_2 = 0, \quad \beta_3 = 0, \quad \beta_7 = 0.$$

- *Partiendo del modelo inicial* omita de golpe las variables `FARE`, `GASPRICE` y `LANDAREA` para obtener un modelo restringido. Seleccione con el ratón la ventana del primer modelo y en *`omitir`* y marque las tres variables, o bien ejecute el código:

  ```
  ols BUSTRAVL 0 FARE GASPRICE INCOME POP DENSITY LANDAREA
  ModeloFinal <- omit FARE GASPRICE LANDAREA
  modeltab add
  ```

  ¿Son conjuntamente significativas? ¿Mejora la precisión de las estimaciones? ¿y los criterios de selección? ¿Son los signos de los parámetros los esperados? A la luz de los resultados, viajar en autobús ¿parece ser un bien normal, o un bien inferior?

Actividad 8 Toda la inferencia que hemos realizado se basa en la hipótesis de normalidad de los residuos. Contraste si se puede rechazar o no la hipótesis de normalidad en ambos modelos (el inicial y el final). El contraste Jarque-Bera es el más frecuente en Econometría.

- En la ventana de resultados de un modelo “pinche” en *`contrastes -->Normalidad de los residuos`*, o bien ejecute el código:

  ```
  series e = $uhat
  normtest --all e
  ```

  Piense que los contrastes realizados en la anteriores actividades, asumen la normalidad de las perturbaciones. Si esta hipótesis es rechazada toda la inferencia de los contrates anteriores podría quedar cuestionada.

4.2. Código completo de la práctica `BusTravelers.inp`

```
open data4-4

summary BUSTRAVL FARE GASPRICE INCOME POP DENSITY LANDAREA --simple
corr BUSTRAVL FARE GASPRICE INCOME POP DENSITY LANDAREA --plot="display"
fare_bustravl     <- gnuplot BUSTRAVL FARE     --output="display"
gasprice_bustravl <- gnuplot BUSTRAVL GASPRICE --output="display"
income_bustravl   <- gnuplot BUSTRAVL INCOME   --output="display"
pop_bustravl      <- gnuplot BUSTRAVL POP      --output="display"
density_bustravl  <- gnuplot BUSTRAVL DENSITY  --output="display"
landarea_bustravl <- gnuplot BUSTRAVL LANDAREA --output="display"

ModeloInicial <- ols BUSTRAVL 0 FARE GASPRICE INCOME POP DENSITY LANDAREA
modeltab add

omit GASPRICE

omit LANDAREA

omit FARE

ols BUSTRAVL 0 FARE GASPRICE INCOME POP DENSITY LANDAREA
ModeloFinal <- omit FARE GASPRICE LANDAREA
modeltab add

series e = $uhat
normtest --all e
```