

## Taller 2 - Econometría II - Universidad Santo Tomás

Nicolás Ronderos

En todos los casos muestre las salidas de STATA

1. Existe un resultado asintótico que establece el siguiente punto: si el error en un modelo de regresión no se distribuye normal pero el número de observaciones es lo suficientemente grande, entonces la distribución muestral de los coeficientes de regresión será normal. Por lo que la inferencia estadística puede realizarse sin inconvenientes. Sin embargo, el término suficientemente grande es subjetivo y depende de la distribución del error. Utilizando STATA simule el siguiente proceso.

$$x \sim N(0, 1) \quad y = 5 + 3x + u \quad u \sim iid$$

- a. Si  $u \sim N(0, 1)$  (rnormal en STATA) realice 2500 simulaciones del anterior proceso con  $n$  igual a 10, 25 y 75. Gráfique el histograma de frecuencias de las 2500 estimaciones de  $\beta_1$ , sus estadísticas descriptivas y realice la prueba Jarque-Bera sobre estas estimaciones. ¿con que tamaño de  $n$  es suficiente para que  $\beta_1$  se distribuya normal?
- b. Si  $u \sim \chi^2_{10}$  (rchi2(10) en STATA) realice 2500 simulaciones del anterior proceso con  $n$  igual a 10, 25 y 75. Gráfique el histograma de frecuencias de las 2500 estimaciones de  $\beta_1$ , sus estadísticas descriptivas y realice la prueba Jarque-Bera sobre estas estimaciones. ¿con que tamaño de  $n$  es suficiente para que  $\beta_1$  se distribuya normal?
- c. Si  $u \sim t_5$  (rt(5) en STATA) realice 2500 simulaciones del anterior proceso con  $n$  igual a 10, 25 y 75. Gráfique el histograma de frecuencias de las 2500 estimaciones de  $\beta_1$ , sus estadísticas descriptivas y realice la prueba Jarque-Bera sobre estas estimaciones. ¿con que tamaño de  $n$  es suficiente para que  $\beta_1$  se distribuya normal?

2. Para el desarrollo de este punto tenga en cuenta el siguiente modelo de regresión  $y_i = \beta_0 + \beta_1 x_{i1} + u_i$

- a. Con la expresión  $Var(B) = \sigma_u^2 (X'X)^{-1}$  calcule la matriz de varianza covarianza de los coeficientes del anterior modelo.
- b. ¿Que sucede con la varianza de los coeficientes si aumenta la variación de  $x_i$ ? ¿que sucede con la varianza si se incrementa el número de observaciones?

3. La base de datos *br.dta* contiene información sobre un conjunto de viviendas.
  - a. ¿Que miden las variables *price*, *sqft*, *bedrooms*, *baths*, *age*, *pool*, *fireplace*, *waterfront*, y *dom*? Calcule sus estadísticas descriptivas, histograma y comente.
  - b. Utilizando álgebra matricial en STATA calcule los estimadores del modelo  $price_i = \beta_0 + \beta_1 sqft_i + \beta_2 bedrooms_i + \beta_3 baths_i + \beta_4 age_i + \beta_5 pool_i + \beta_6 fireplace_i + \beta_7 waterfront_i + \beta_8 dom_i + u_i$  con las formula  $\hat{B} = (X'X)^{-1}X'Y$ . Interprete sus estimaciones. ¿Los resultados son los esperados?
  - c. Calcule usando álgebra matricial en STATA la matriz de varianza covarianza de los estimadores de MCO asumiendo homocedasticidad. utilice la formula  $Var(\hat{B} | X) = \sigma_u^2 (X'X)^{-1}$ . Muestre sus resultados.
  - d. Calcule usando álgebra matricial en STATA los estadísticos t para todos los regresores del modelo incluida la constante. ¿Son todos los coeficientes estadísticamente significativos?
  - e. Utilizando el comando *reg* valide los resultados encontrados anteriormente.
  
4. Utilizando alguna base de datos de su elección utilice componentes principales para estimar una variable latente. Explique detalladamente el proceso de estimación y la fuente de información.
  
5. ¿El consumo de alcohol genera más enfermedades? la base de datos *wine.dta* contiene información del consumo de alcohol y el número de muertes por enfermedades del hígado y del corazón para un grupo de 21 países.
  - a. ¿Que miden las variables *alcohol*, *heart* y *liver*? Calcule sus estadísticas descriptivas e histograma. Genere un gráfico de barras donde muestre el consumo de vino promedio por país usando STATA.
  - b. Corra las regresiones  $Heart_i = \beta_0 + \beta_1 alcohol_i + u_i$  y  $Liver_i = \beta_0 + \beta_1 alcohol_i + u_i$  usando MCO e interprete sus resultados. ¿Considera que se cumple Gauss-Markov?
  - c. Corra las regresiones  $Heart_i = \beta_0 + \beta_1 alcohol_i + u_i$  y  $Liver_i = \beta_0 + \beta_1 alcohol_i + u_i$  usando Ridge e interprete sus resultados. Instale el comando *ridgereg* y utilice *orr* para sus estimaciones. Calcule los parámetros con *c* 0, 1 y 19. Interprete sus resultados.
  
6. En la base de datos *pubexp* puede encontrar información del gasto publico en educación, PIB en millones de dolares de 1980 y población de 34 países.
  - a. Obtenga estadísticas descriptivas e histogramas de todas las variables y estime la siguiente regresión por MCO  $ee_i = \beta_0 + \beta_1 gdp_i + u_i$ . Comente sobre sus resultados.

- b. Calcule los residuales de la anterior regresión y con ellos realice un scatter plot entre los residuales elevados al cuadrado y el PIB. ¿existen indicios de heterocedasticidad? ¿por qué?
- c. Calcule las pruebas de Breusch-Pagan, White, Harvey y Gleijser. Muestre sus resultados ¿que concluye?
- d. Estime la anterior regresión usando MCG y errores robustos a la heterocedasticidad. Utilice como ponderaciones la población de los países. Muestre sus resultados y comente.

**7.** Utilizando la base de datos *fertility02* determine como la fertilidad afecta la oferta de trabajo, es decir determine la influencia de tener más hijos sobre las semanas trabajadas.

- a. Calcule estadísticas descriptivas de las variables *morekids*, *weeksm1*, y *samesex*. ¿Que miden estas variables?
- b. Estime el siguiente modelo de regresión por MCO  $weeksm1_i = \beta_0 + \beta_1 morekids_i + u_i$  ¿como interpreta este resultado? ¿es esta una estimación adecuada? Comente.
- c. ¿La variable *samesex* puede ser un candidato a instrumento de *morekids*? ¿cuantas observaciones se incluyen en la regresión?. Verifique la condición de relevancia.
- d. Estime los parámetros del modelo  $weeksm1_i = \beta_0 + \beta_1 morekids_i + u_i$  utilizando como variable instrumental *samesex* e interprete sus resultados.
- e. Realice la prueba de Hausman-Wu y concluya sobre sus resultados.

**8.** Usando STATA simule el siguiente proceso para  $n = 10,000$ .

$$x_i \sim nid(0, 1) \quad \sigma_{1i}^2 = \sigma_u^2 \mid x_i \mid$$

$$\sigma_{2i}^2 = \sigma_u^2 x_i^2 \text{ y } \sigma_{3i}^2 = \sigma_u^2 i$$

$$u_1 = e_i \sigma_{1i}$$

$$u_2 = e_i \sigma_{2i}$$

$$u_3 = e_i \sigma_{3i} \text{ donde } e_i \sim nid(0, 1)$$

$$y_{1i} = x_i + u_1$$

$$y_{2i} = x_i + u_2 \text{ y } y_{3i} = x_i + u_3$$

- a. Muestre las estadísticas descriptivas de  $y_i$  para  $i = 1, 2, 3$ . Interprete sus resultados.
- b. Corra las tres regresiones  $y_{ji} = \alpha_0 + \alpha_1 x_i + u_j$  para  $j = 1, 2, 3$  y calcule los residuales de cada una. Muestre e interprete sus resultados.

- c. Construya un scatter plot de  $u_1$  y  $u_2$  contra  $x_i$ . Y de  $u_3$  contra  $i$ . Muestre e interprete sus resultados.
  - d. Calcule las pruebas de White y Breusch-Pagan sobre los tres modelos del segundo literal. Muestre e interprete sus resultados.
- 9.** Realice una rutina de programación matricial en STATA que permita estimar los parámetros de la regresión RIDGE para un conjunto de  $k+1$  variables independientes.
- a. Simule cinco variables independientes distribuidas normal e independientemente y simule una variable dependiente como otra normal independiente. Corra la regresión entre la dependiente y las cinco independientes usando el comando `ridgereg`.
  - b. Corra la anterior regresión usando su código de programación y compare los resultados.
  - c. Simule diez variables independientes distribuidas normal e independientemente y simule una variable dependiente como otra normal independiente. Corra la regresión entre la dependiente y las cinco independientes usando el comando `ridgereg`.
  - d. Corra la anterior regresión usando su código de programación y compare los resultados.