Econometría Avanzada - Trabajo Práctico N°1

Supuestos clásicos

Fecha de entrega: martes 22 de marzo de 2022

Profesor: Walter Sosa Escudero
Asistente: Gastón García Zavaleta

Modalidad: enviar el informe en formato pdf por mail a Gastón García Zavaleta (`gzavaleta@udesa.edu.ar`) con el asunto "TP1 Econometría Avanzada - APELLIDOS". Además, adjuntar el script de R con el código utilizado. Esperar mail confirmando la recepción.

Reglas de formato y presentación

- El objetivo de este trabajo práctico es revisar demostraciones y conceptos ligados a los supuestos clásicos.
- Incluyendo tablas y gráficos, no debe exceder las cinco carillas A4. Se espera una redacción cuidada y profesional, prestando atención a aspectos estéticos en el diseño de tablas y en la presentación elegante de los resultados, como si se tratase de un verdadero trabajo académico o de consultoría profesional.
- El trabajo debe ser hecho en grupos de **3 personas**.

Ejercicio 1

El objetivo de este ejercicio es crear muchas bases de datos usando R de modo tal de estimar múltiples veces el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \tag{1}$$

sabiendo de antemano que los verdaderos coeficientes son:

$$\beta_0 = 400$$
$$\beta_1 = 2$$
$$\beta_2 = 2$$
$$\beta_3 = -2$$

1. Cree 50 bases de datos distintas entre sí con 100 observaciones cada una usando la siguiente información:
   - Las variables x_1, x_2 y x_3 siguen una distribución uniforme [0,100].
   - El término de error sigue una distribución normal con $E(u) = 0$ y $\sigma^2 = 1400$.
   - Una vez que tenga definidas las variables independientes y el error, defina y teniendo en cuenta la ecuación 1.

   Use como semillas los números del 1 al 50 inclusive. Es decir, cada base de datos generada tiene que tener 100 observaciones de x_1, x_2, x_3, u_i e y_i.
2. Muestre la correlación entre las variables independientes para una de las bases de datos creadas. Comente.
3. Estime β_0, β_1, β_2 y β_3 para las 50 muestras y guarde los datos en un Excel llamado `primera_estimacion.xlsx`. Tiene que quedarle una planilla con cuatro columnas y 50 estimaciones distintas.
4. Cargue la base `primera_estimacion.xlsx` a R y haga dos gráficos: uno de $\hat{\beta}_1$ contra $\hat{\beta}_2$, y otro de $\hat{\beta}_1$ contra $\hat{\beta}_3$. Comente el resultado. ¿Son buenas las estimaciones?
5. El siguiente código sirve para lograr que x_2 esté altamente correlacionada con x_1:

```
x2 <- scale(matrix( rnorm(100), ncol=1 ))
xs <- cbind(scale(x1),x2)
c1 <- var(xs)
chol1 <- solve(chol(c1))
newx <- xs
newc <- matrix(
  c(1 , 0.987,
  0.987, 1 ), ncol=2)
eigen(newc)
chol2 <- chol(newc)
xm2 <- newx%*% chol2 * sd(x1) + mean(x1)
x2 <- xm2[, 2]
```

   Úselo de modo tal de generar 50 bases de datos nuevas donde x_1 y x_2 están altamente correlacionadas. Use las mismas semillas.
6. Estime β_0, β_1, β_2 y β_3 para las 50 nuevas bases de datos y guarde los datos en un Excel llamado `segunda_estimacion.xlsx`.
7. Cargue la base `segunda_estimacion.xlsx` a R y grafique $\hat{\beta}_1$ contra $\hat{\beta}_2$. Comente el resultado. ¿Qué pasó? ¿Dejaron los estimadores de β_1 y β_2 de ser los MELI?
8. Nombre con claridad el problema con el que está lidiando. ¿Qué consecuencias tiene?

Explique.

Ejercicio 2

En este ejercicio usaremos las mismas bases de datos que creó en el ejercicio anterior.

1. Usando las 50 bases de datos donde x_1 y x_2 no están altamente correlacionadas, estime el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{3i} + u_i \tag{2}$$

   Guarde las 50 estimaciones en un Excel llamado `tercera_estimacion.xlsx`, cárguelo a R y grafique $\hat{\beta}_1$ contra $\hat{\beta}_3$. ¿Hay problemas de sesgo? ¿Y de varianza?

2. Usando las 50 bases de datos donde x_1 y x_2 están altamente correlacionadas, estime el modelo de la ecuación (2). Guarde las 50 estimaciones en un Excel llamado `cuarta_estimacion.xlsx`, cárguelo a R y grafique $\hat{\beta}_1$ contra $\hat{\beta}_3$. ¿Hay problemas de sesgo? ¿Y de varianza?

Ejercicio 3

Indique con precisión qué propiedades matemáticas o estadísticas se aplican en cada paso de la siguiente demostración. Por último, explique claramente qué se está demostrando. (Recuerde que $x_i = X_i - \overline{X}$ y que $y_i = Y_i - \overline{Y}$.)

$$\begin{aligned}
\hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} \\
E(\hat{\beta}) &= E\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) \\
E(\hat{\beta}) &= \sum w_i E(y_i) \\
E(\hat{\beta}) &= \sum w_i E(x_i\beta + \mu_i - \overline{\mu}) \\
E(\hat{\beta}) &= \sum w_i [E(x_i\beta) + E(\mu_i) - E(\overline{\mu})] \\
E(\hat{\beta}) &= \sum w_i \left[x_i\beta + E(\mu_i) - E\left(\frac{\sum \mu_i}{n}\right)\right] \\
E(\hat{\beta}) &= \sum w_i \left(x_i\beta + k - \frac{nk}{n}\right) \\
E(\hat{\beta}) &= \sum w_i (x_i\beta + k - k) \\
E(\hat{\beta}) &= \sum w_i x_i \beta \\
E(\hat{\beta}) &= \beta \sum w_i x_i \\
E(\hat{\beta}) &= \beta \frac{\sum x_i^2}{\sum x_i^2} \\
E(\hat{\beta}) &= \beta
\end{aligned}$$