

## Examen Parcial (13 de septiembre de 2020)

Dr. Irving Arturo De Lira Salvatierra

Puntos totales 100

- El parcial está dividido en dos secciones, una parte teórica y una parte de programación con datos.
- **La entrega del examen será a más tardar el jueves 17 de septiembre a las 9:00 am.**
- El último día para dudas es el lunes 14 de septiembre.
- Se deben enviar 3 archivos (al correo [irvingdelir@gmail.com](mailto:irvingdelir@gmail.com)):
  - El primero contendrá las respuestas de la Sección 1 (enviar en formato PDF). Usar el siguiente nombre: CU\_Nombre\_S1\_Parcial.
  - El segundo y tercer archivo corresponden a la Sección 2, enviarán un archivo do y un log con los siguientes nombres: CU\_Nombre\_S2\_Parcial.do y CU\_Nombre\_S2\_Parcial.log. Nota: algunas preguntas de la Sección 2 requieren contestar no solo código sino interpretación, en cuyo caso deberán incluir las respuestas como texto comentado en el archivo do. En esta sección todos los pasos necesarios se deberán hacer directamente en Stata, ningún cálculo o limpieza de información se hará en otro software. Se debe comentar el código mostrando las secciones y los pasos explícitos de los distintos ejercicios.
- **Código de ética: los alumnos tienen prohibido utilizar el mismo código o compartir respuestas, cualquier plagio se penalizará de acuerdo con los lineamientos institucionales del ITAM.**
- El puntaje de cada pregunta se encuentra explícito en el documento

### Sección 1 (70 puntos)

#### 1. Introducción (10 pts)

- a. (2.5 pts) Describir la utilidad y diferencias entre un archivo log y un archivo do.
- b. (2.5 pts) ¿Cuál es la sintaxis correcta para ejecutar un archivo do directamente en la ventana "command" (suponer que el archivo do se llama "Intro.do")?
- c. (2.5 pts) En general, ¿qué se debe hacer previo a ejecutar un archivo do? Escribir la sintaxis correcta.
- d. (2.5 pts) Supongamos que se desea comentar el procedimiento en un archivo do, y se quiere poner el siguiente texto "se supone que la función de producción del mercado minorista es una Cobb-Douglas". Escribir la sintaxis correcta del comentario en el archivo do.

#### 2. Descripción de bases de datos (10 pts)

- a. (2.5 pts) ¿Cuál es el comando utilizado para generar estadísticas descriptivas?

- b. (2.5 pts) Utilizar el comando del inciso anterior y escribir la sintaxis correcta para describir la variable “salario” entre los años 1995 y 2005 para el estado de Nuevo León y Jalisco. Nota: los años son variables numéricas (nombre de la variable “años”) y los estados son una variable de texto (nombre la variable “estados”).
- c. (2.5 pts) Escribir la sintaxis correcta para crear una variable “salario\_promedio”, que es igual al promedio de la variable “salario”.
- d. (2.5 pts) Escribir la sintaxis correcta para crear un escalar “salario\_promedio\_escalas”, que es igual al promedio de la variable “salario”. Generar su visualización.

### 3. Regresión simple (10 pts)

- a. (1 pts) ¿Cuál es el comando utilizado para ejecutar una regresión lineal?
- b. (1 pts) Supongamos que se quiere ejecutar una regresión lineal con la variable independiente mpg (combustible en millas por galón) y como regresores el precio (precio del automóvil en dólares) y el peso (peso del automóvil en kilogramos). Escribir la sintaxis correcta de la regresión.
- c. (1 pts) Si se quiere suprimir el resultado visual de la regresión del inciso anterior, ¿cuál sería la sintaxis adecuada?
- d. (3.5 pts) Escribir la sintaxis correcta para crear una variable “residOLS”, que es igual a los residuos de la regresión del inciso (b).
- e. (3.5 pts) Supongamos que se quiere calcular la  $R^2$  de manera manual después de ejecutar la regresión del inciso (b). Escribir la sintaxis correcta para calcular el escalar de  $R^2$  (nombre “r2”) y que ésta se visualice. Recordar que la  $R^2$  es igual al cociente de la suma de los cuadrados del modelo entre (i) la suma de los cuadrados de los residuos y (ii) la suma de los cuadrados del modelo.

### 4. Análisis post-regresión (10 pts)

- a. (1 pts) Supongamos que se ejecuta el siguiente comando:  
`quietly regress ltotexp suppins phylim actlim totchr age female, robust`  
 Describir que se está ejecutando.
- b. (3 pts) Escribir la sintaxis correcta para realizar probar de hipótesis de que los coeficientes de los regresores phylim y actlim son iguales (suponer errores no robustos y la misma especificación que en (a)).
- c. (1.5 pts) ¿Cuándo se rechaza la hipótesis nula del inciso (b)?
- d. (3 pts) Escribir la sintaxis correcta para realizar probar de hipótesis de que los coeficientes de los regresores phylim, actlim y totchr son (conjuntamente) iguales a cero.
- e. (1.5 pts) ¿Cuándo se rechaza la hipótesis nula del inciso (d)?

5. Gráficos (10 pts)

- a. (3 pts) Escribir la sintaxis correcta para generar una gráfica de dispersión entre las variables  $x_1$  y  $x_2$ .
- b. (4 pts) Escribir la sintaxis correcta para generar una gráfica de densidad kernel con un bandwidth definido (.2) y con la superposición de la densidad normal (considerar como variable objetivo  $x_1$ ).
- c. (4 pts) Escribir la sintaxis correcta para generar un histograma superpuesto por una estimación de la densidad del kernel para  $x_2$ .

6. Bucles (loops) (20 pts)

- a. (5 pts) En esta sección se debe ir a la documentación del comando "foreach". Supongamos que se tienen 4 variables constantes  $x_1$ ,  $x_2$ ,  $x_3$  y  $x_4$  (ya definidas en la base de datos), y se requiere generar una nueva variable que sea la suma de las  $x$ 's. Para ello escribir la sintaxis correcta para crear una variable "suma" que guarde el valor de la suma usando el comando "foreach".
- b. (3 pts) Escribir la sintaxis para generar 100 observaciones y generar las variables  $x_i=i$  para  $i=1, \dots, 4$ . Utilizar la sintaxis del inciso (a) y dar el resultado exacto de la variable "suma" (explorar con el comando sum y ver el promedio).
- c. (5 pts) En esta sección se debe ir a la documentación del comando "while". Supongamos que se tienen 4 variables constantes  $x_1$ ,  $x_2$ ,  $x_3$  y  $x_4$ , y se requiere generar una nueva variable que sea la suma de las  $x$ 's (ya definidas en la base de datos). Para ello escribir la sintaxis correcta para crear una variable "suma2" que guarde el valor de la suma usando el comando "while".
- d. (3 pts) Escribir la sintaxis para generar 100 observaciones y generar las variables  $x_i=i$  para  $i=1, \dots, 4$ . Utilizar la sintaxis del inciso (c) y dar el resultado exacto de la variable "suma2".
- e. (4 pts) ¿Qué diferencia encuentran entre sum y sum2?

**Sección 2 (30 puntos)**

1. Descripción de bases de datos y regresión (10 pts)

- a. (2.5 pts) Utilizar la base de datos auto.dta y obtener los principales estadísticos que resumen las variables mpg (combustible en millas por galón) y weight (peso del automóvil en kilogramos) de acuerdo con la variable foreign (indica si el automóvil es producción nacional o extranjera) (Hint: utilizar el prefijo by foreign:).
- b. (2.5 pts) Comentar sobre las diferencias entre los automóviles nacionales y extranjeros.
- c. (2.5 pts) Ejecutar una regresión con mpg como variable dependiente y regresores weight y foreign.
- d. (2.5 pts) Comentar sobre cualquier diferencia para automóviles extranjeros.

2. Regresión y valores generados (10 pts)

- a. (2.5 pts) Usar la base de datos auto.dta, ejecutar una regresión con variable dependiente mpg y regresores price y weight.
- b. (2.5 pts) Utilizar los resultados almacenados de e() del inciso (a) para calcular un escalar de la  $R^2$  ajustada,  $r^2_{adj}$ . Recordar que la  $R^2$  ajustada es igual a  $R^2 - (1 - R^2) (k - 1) / (N - k)$ , donde N es el número de observaciones y k es el número de regresores, incluido el intercepto ¿Cuál es el valor?
- c. (2.5 pts) Utilizar los resultados almacenados de e() del inciso (a) para calcular un escalar tweight, igual al estadístico t para probar que el coeficiente de weight es cero.
- d. (2.5 pts) ¿El coeficiente de weight es cero? Justificar

3. Bucles (10 pts)

- a. (5 pts) Usando la base de datos auto. dta y el comando foreach crear una variable "total", igual a la suma de headroom y length.
- b. (5 pts) Confirmar el resultado utilizando el comando summarize, de tal forma que total tiene un promedio igual a la suma de los promedios de headroom y length.

4. (OPCIONAL) Variables aleatorias (10 pts)

- a. (5 pts) Cree un conjunto de datos simulado con 100 observaciones sobre dos variables aleatorias (x1, x2) que se obtienen cada una a través de una muestra con distribución uniforme. Utilizar seed igual a 12345.
- b. (5 pts) En teoría, estas variables aleatorias tienen una media de 0.5 y una varianza de 1/12, ¿esta conclusión se mantiene para el conjunto de datos creados?