



Laboratorio 1: Regresión lineal y MCO¹

Usted cuenta con la base de datos **datos lab1 E2EMI1 2024II ENAHO.dta**, la cual fue construida a partir de los módulos 100, 500, 612 y la sumaria de la ENAHO 2019.

1. Importe la base de datos a utilizar y familiarícese con las observaciones y las variables disponibles, i.e., observe el tamaño de la muestra, los nombres y etiquetas de las variables disponibles.
2. Considere que la población en edad de trabajar para el caso peruano comprende a las personas entre 14 y 64 años: haga los cambios pertinentes.
3. Observe las características estadísticas de la variable de ingreso mensual. Haga lo mismo con ayuda visual: Muestre y guarde en formato *Stata Graph* (gph) un histograma de esta variable. ¿Observa algún potencial problema? ¿Qué es un *outlier*?
4. *Trimming*/Acotando: Acote la variable excluyendo a individuos con ingresos mensuales muy altos o bajos. Muestre las características estadísticas de la nueva versión. ¿Los ingresos se distribuyen normalmente? Muestre un histograma de esta nueva versión, pero ahora mejore su gráfico y guárdelo en *Portable Document Format* (PDF) y en formato *Joint Photographic Experts Group* (JPG). Concepto nuevo: [winsorize](#) (E2TEE9-2025 I).
5. Plantee una regresión lineal en la que muestre la relación entre el ingreso laboral de la persona y su edad. Luego, plantee la misma relación, pero usando como variable dependiente el logaritmo del salario. ¿Cómo se interpretan los resultados?
6. Plantee el modelo anterior, pero ahora incluya como variable de control la edad al cuadrado de la persona ($edad^2$). ¿Qué puede decir de la nueva especificación? ¿Cómo se interpretan los resultados?
7. Muestre gráficamente lo que halló en el apartado anterior utilizando el comando *reganat*.

¹Dictado el martes 27 de agosto de 2024.

Variables factoriales

El uso de variables factoriales tiene varias ventajas: permite incluir de manera compacta un número grande de variables relacionadas al modelo, automatizan la creación de interacciones entre variables y evita errores en su creación.

Para indicar variables categóricas se usa el operador prefijo **i.**, para indicar variables continuas se usa el operador prefijo **c.** y para interacciones se usa **#**, para este último tenemos dos casos: una sola vez especifica interacciones de la variables y dos veces especifica interacciones factoriales (incluye la variable sola y elevada al cuadrado, por ejemplo).

8. Estime el mismo modelo reemplazando *edad* y *edad*² con su contraparte factorial y muestre que los resultados son los mismos. Muestre la información guardada luego de correr la regresión.
9. Pruebe que la edad como polinomio tiene significancia sobre los ingresos.

Comparando correctamente: estandarización

El uso de variables medidas en distintas unidades nos impide generalizar los efectos de cambios en las regresoras en la dependiente. Por ejemplo, no es lo mismo el cambio de un año más de educación a un hijo más, a pesar de que el incremento sea «uno». Sin embargo, al estandarizar medimos todas las variables en desviaciones estándar de sus respectivas medias.

A continuación, cree un **indicador de riqueza** del hogar estandarizado, para ello, siga los siguientes pasos:

10. Estandarice las variables material de los pisos, las paredes, el acceso a servicios básicos y el equipamiento del hogar.
11. Cree el indicador de riqueza como la agregación (suma) de estas variables estandarizadas.
12. Estandarice el indicador.

Ahora, ya con el indicador estandarizado, realice el siguiente análisis:

13. Corra tres regresiones: el modelo univariado, el bivariado y uno en el que controle por riqueza, todos contra los ingresos estandarizados. Muestre una tabla en **STATA** comparando los betas estimados obtenidos en cada regresión. Muestre significancia con asteriscos.
14. Haga lo mismo que en el apartado anterior, pero esta vez muestre los resultados como tablas en una cuadrícula de **STATA** comparando los betas estimados obtenidos en cada regresión. Esto ayuda a copiar y pegar como Excel. Muestre significancia con asteriscos.
15. Haga lo mismo que en el apartado anterior, pero esta vez mejore la tabla y expórtela en formato Excel (xls). Muestre significancia con asteriscos, muestre las etiquetas de las variables en vez del nombre, muestre solo tres decimales, agregue un título y una nota al pie.