

# Problem Set 1: Repaso y primeros pasos en Stata

---

**Fecha de entrega:** 23 de agosto de 2024

Paola Llamas y Tomás Pacheco

## Instrucciones

- Agreguen los nombres y apellidos de los integrantes en el encabezado del archivo que entregan.
- El título de los archivos entregados debe tener la siguiente estructura:  
`PS1_Garcia_Gomez_Gonzalez_Perez.pdf`
- **UN ÚNICO** miembro del grupo debe subir el PDF y *do-file* al Campus Virtual en la sección "Problem Set 1"
- El código lo tendrán que entregar de dos formas distintas. Primero deberán exportar el *do-file* como PDF y unirlo al informe ([ayuda](#)). Luego, deberán entregar el código en formato `.do` en el lugar indicado en el Campus Virtual.
- El directorio de trabajo debe tener tres carpetas: `programs`, `input` y `output`. El *do-file* debe ser entregado de tal manera que los tutores podamos correr el código **solamente** modificando el global con el *main directory* (**esto es muy importante**).
- Recuerden que siempre deben presentar la información en tablas y gráficos estéticos. Nunca presenten una tabla o figura sin dedicarle unas líneas a comentar lo que se observa.

## Repaso teórico

En este primer ejercicio se plantearán distintas oraciones que ustedes deberán comentar:

1. Bajo ciertos supuestos sobre el método, el modelo de MCO es el mejor estimador lineal e insesgado.
2. ¿Cuál es el supuesto que creen que es necesario hacer para que el estimador de Mínimos Cuadrados Ordinarios sea insesgado? ¿Es el mismo supuesto que necesitamos para garantizar la existencia del estimador?
3. ¿Cuál(es) es(son) los problemas de levantar supuestos tales como el de homocedasticidad o no correlación serial?
4. ¿Por qué decimos que  $\hat{\beta}$  tiene distribución normal cuando suponemos  $\mu_i \sim N(0, \sigma^2)$ ?
5. Para hacer inferencia siempre se necesita suponer la distribución del término de error.
6. Un estimador no puede ser consistente y asintóticamente normal al mismo tiempo.
7. Consistencia e insesgadez son dos propiedades muy similares.

## Primeros pasos en Stata

Esta sección del trabajo práctico tiene como objetivo que empiecen a familiarizarse con el uso de Stata. La base de datos que tendrán que usar se llama `data_russia.dta`.

1. Como podrán observar, la base de datos está muy sucia. Idealmente, todas las variables deberían ser numéricas. Utilizando los comandos `replace`, `split`, `destring` y `encode`, emprolijen la base. Se les recomienda que utilicen loops para hacer los cambios. Pueden ahorrar mucho tiempo!
2. ¿Hay alguna variable que tenga más del 5 % de valores faltantes?
3. Cuando uno trabaja con datos, es normal encontrarse variables o observaciones que pueden considerarse “irregulares” (por ejemplo, ingreso con valores negativos o gastos mayores a ingresos). Procedan a buscar si hay observaciones de este tipo. Expliquen cómo lo hacen. Si encuentran algún valor irregular, reemplácenlo con *missing*.

4. Ahora el objetivo es que ordenen los datos. La primera variable que aparezca en la base debería ser el *id* del individuo, la segunda el sitio (*site*) donde se encuentra y la tercera el sexo (*sex*). Luego, ordenen las filas de mayor a menor según *totexpr*.
5. Siempre que trabajamos con datos, tenemos que conocer la base. Hagan un cuadro con estadísticas descriptivas del sexo, la edad **en años**, la satisfacción con la vida, la circunferencia de la cintura, la circunferencia de la cadera y el gasto real (*totexpr*). Asegúrense de que las variables tengan etiquetas. **Recuerden siempre comentar los resultados de una tabla.**
6. Supongan que hay un determinado *paper* que se llama “Hips don’t lie” (Shakira, 2005) que dice que las caderas de los hombres son mayores que la de las mujeres.
  - a. Presenten un gráfico comparando la distribución de *hipsiz* para los hombres y para las mujeres.
  - b. Hagan un test de medias usando las variables *hipsiz* y *sex*. El comando que tienen que usar es `ttest`. Presenten y comenten los resultados.
7. Finalmente, supongan que quieren correr una regresión para explicar la felicidad de las personas (*satlif*) en función de variables explicativas como sexo, edad, altura, etc. Una vez que decidan qué covariables utilizar:
  - a. Hagan dos gráficos que los ayuden a pensar en cómo será el resultado de su estimación. Asegúrense de que esté autocontenido, es decir, que se entienda de qué trata solo con verlo. Comenten.
  - b. Estimen dos especificaciones distintas del modelo. **Interpreten detalladamente los coeficientes y la significatividad estadística.**