

Trabajo práctico evaluable

Inferencia Estadística - GCID

19 de abril de 2022

Introducción

La presente práctica debe entregarse con fecha límite el martes 3 de mayo a las 22:00 horas. La entrega se realizará mediante envío por correo electrónico a *julian.costa@udc.es*, poniendo a los coautores del trabajo en *copia carbón* (CC).

El trabajo se realizará en grupos de 5 personas. Ante la imposibilidad práctica de que todos los grupos tengan 5 integrantes, también se aceptaran grupos de 4 y de 6 personas. El nombre de todos los autores del trabajo debe estar en las primeras líneas del fichero (o ficheros) que se entregue. La práctica se calificará sin tener en cuenta el número de integrantes del grupo, y esa nota será la misma para todos los autores del trabajo.

Debe entregarse un fichero con el código R (.R) y además se valorará positivamente que también se entregue un fichero **html** o **pdf** con el código y los resultados (debe obtenerse compilando el fichero con el código). El fichero (o ficheros, si también se entrega el documento compilado), se identificará mediante **Apellidos**, **Nombre** de la persona que lo entregue en nombre del grupo.

Para la realización de este trabajo no se puede hacer uso de paquetes adicionales de R, por ejemplo de muestreo (es decir, solo se puede utilizar lo que está disponible al iniciar sesión con las opciones por defecto).

La práctica se calificará entre 0 y 2 puntos, y estos puntos se consolidarán en la nota final de la asignatura. La parte que no se obtenga de estos 2 puntos podrá recuperarse el día del examen final (tal como se explica en la guía docente).

Ejercicio 1

El fichero `Zumo.Rdata` contiene datos ficticios de unidades familiares, detallándose algunas características como por ejemplo el porcentaje de los ingresos destinado a los gastos en alimentación, el nivel económico de la familia o la región en la que vive.

| ## | Variables | Descripcion |
|------|-----------|--|
| ## 1 | Id | Id |
| ## 2 | Ingresos | Ingresos hogar |
| ## 3 | Gastos | Gastos en alimentacion |
| ## 4 | PcGastos | Pct. gastos alimentacion respecto a ingresos |
| ## 5 | UnidF | Numero de miembros de la unidad familiar |
| ## 6 | NivelEcon | Nivel economico |
| ## 7 | Region | Region |

1. Carga el fichero de datos utilizando `load('nombre.Rdata')` (se puede utilizar la ruta de acceso si se considera oportuno) y visualiza las 10 primeras filas del data frame `Zumo`. Indica cuántas familias hay en la base de datos.

[Observación: Si cargas el fichero de datos haciendo clic en el nombre del fichero (dentro de la pestaña *files*), la instrucción que se ejecuta aparecerá en la consola. En ese caso, no te olvides de copiar y pegar esa instrucción en el fichero .R que contiene el código (fichero R script).]

2. Calcula la media de **Gastos**.
3. Calcula las medias de **Gastos** de las familias para cada uno de los valores de **UnidF**.
4. Realiza un muestreo polietápico considerando como estratos tanto **UnidF** como **NivelEcon**. Debe hacerse con afijación proporcional al 1%.

[Observación: Debes obtener muestras aleatorias de familias *para cada uno de los valores del número de miembros de la unidad familiar de cada uno de los 3 niveles económicos*. Esto se puede hacer de muchas formas: utilizando **for**, **split** o **tapply**, por ejemplo. Si utilizas **tapply** debes hacerlo así **tapply(..., list(x, y), ...)**, donde **x** e **y** deben ser sustituidos por las variables que definen el número de miembros y los niveles económicos.]

5. Calcula la media de **Gastos** de las familias de la muestra del apartado anterior.

Ejercicio 2

Siguiendo con el ejercicio anterior, nos interesa estudiar la proporción p de familias de la **Región 2** en las que la unidad familiar consta de 1 o 2 miembros.

1. Calcula el valor de p (este cálculo se puede realizar al disponer del censo con toda la información necesaria).
2. Utilizando las fórmulas de la teoría, calcula el tamaño muestral n necesario para estimar la proporción con un error máximo de 0.035 y una confianza del 95 %. Por estudios previos, sospechas que $p \approx 0.3$.
3. Obtén una muestra aleatoria simple (con reemplazamiento) de tamaño n de familias de la **Región 2**.
4. Calcula una estimación de p utilizando la muestra del apartado anterior.
5. Calcula un intervalo de confianza para la proporción al 95 % e indica cuál es el error máximo cometido.