



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE CIENCIAS
MODELOS LINEALES Y DISEÑO DE EXPERIMENTOS
TRABAJO EN R 02



Fecha entrega: 2015/08/02 23:59 horas

INSTRUCCIONES

1. Realice un Fork del repositorio **Regresion-Lineal-Multiple**.
2. Genere un archivo `rlm.R`, de tal forma que le permita responder las siguientes preguntas:

- 2.1 Leer los archivos `poblacion1.xlsx` y `poblacion2.xlsx`, y analizar sus dimensiones.
- 2.2 Una los archivos leídos en un mismo objeto llamado `poblacion`.
- 2.2 Cree un código que identifique la clase de cada variable y genere diagramas de cajas para variables continuas y diagramas de barras para variables discretas.
- 2.3 Cree un código que calcule automáticamente el mínimo, media, máximo, desviación estándar, primer cuartil de cada variable numérica y la frecuencia en el caso de variables categóricas.

```
quantile(x, probs = seq(0, 1, 0.25), na.rm = FALSE)
```

- 2.4 Calcule la correlación entre la variable dependiente `poblacion` y cada una de las variables explicativas (numéricas).
- 2.5 Considere la variable categórica `serv.bas.compl` con una confiabilidad del 90 %, ¿Puede asumirse que la media de la variable `poblacion` en el grupo `serv.bas.compl: SI` es distinta a la media del grupo `serv.bas.compl: NO`? Utilice la función:

```
t.test(x, y, alternative = c("two.sided", "less", "greater"),  
       mu = 0, paired = FALSE, var.equal = FALSE,  
       conf.level = 0.95)
```

- 2.6 Considerando los cálculos anteriores genere el modelo de regresión lineal múltiple que mejor se ajuste a los datos. Interprete los coeficientes obtenidos.
 - 2.7 Interprete el R^2 .
 - 2.8 Analice la significancia de la regresión y de cada uno de los parámetros individuales.
 - 2.9 Realice un análisis detallado de los residuos.
3. Genere un informe dinámico (archivo `rlm.Rmd`) en el cual se detalle cada uno de los pasos ejecutados en la generación del modelo de regresión lineal múltiple, las conclusiones y resultados obtenidos.

DESCRIPCIÓN DE LA DATA

Para el presente trabajo se dispone de dos archivos en Excel que contienen información relacionada con el número de habitantes en un determinado sector.

1. poblacion1.xlsx

Este archivo contiene las siguientes variables:

- a. identificador:** Variable que identifica al sector, única para cada sector.
- b. poblacion:** Número de habitantes en cada sector.
- c. var.pobl.mayor:** Variación en población adulta (en cuanto se incremento el número de personas mayores a 65 años)
- d. menores.18:** Porcentaje de sujetos menores a 18 años en cada sector.

1. poblacion2.xlsx

Este archivo contiene las siguientes variables:

- a. identificador:** Variable que identifica al sector, única para cada sector.
- b. part.almz.escl:** Porcentaje de participación en almuerzo escolar.
- c. var.ingresos:** Variación del ingreso promedio en cada sector.
- d. tasa.crimen:** Tasa de crimen en cada sector.
- e. var.tasa.crimen:** Variación de la tasa de crimen en cada sector.
- f. region:** Región en la cual se ubica el sector.
- g. serv.bas.compl:** Variable que especifica si el sector dispone de todos los servicios básicos.

Recordar que el trabajo es sobre dos puntos de la nota final, si no se entrega hasta la fecha límite no recibirá calificación alguna (No existe opción de entrega retrasada).