

CONSIGNA PARA EL TRABAJO OBLIGATORIO

El trabajo consiste en la realización de un análisis integrador de los conocimientos adquiridos durante el curso en el área de la Estadística Descriptiva en base a datos observables.

Se realiza en grupos de máximo 3 estudiantes.

Se espera que cada equipo de trabajo entregue, por un lado, **un documento PDF** conteniendo el análisis realizado que incluya los cuadros, gráficos e indicadores con sus correspondientes interpretaciones y, por otro, el **script de R** con el que se generan los resultados como un Anexo.

El informe en PDF debe ser autocontenido y tener un máximo de 15 páginas.

El script de R debe estar bien organizado y permitir localizar fácilmente las sentencias con que se generan los cuadros, gráficos e indicadores que se comentan en el PDF. Todos los resultados que se presenten en el PDF deben poder reproducirse al correr el script entregado.

Fecha de entrega del informe de avance (opcional): a acordar con el docente a cargo de su grupo.

Fecha de entrega: el plazo de entrega del informe final es el 3 de diciembre de 2019 a las 20:30 horas. La entrega se realizará a través de Aulas para lo que se creará una tarea que permitirá subir los archivos solicitados.

La **defensa** del trabajo se realizará en forma escrita y estará **integrada al segundo parcial**.

Cada equipo de trabajo conformado dentro de cada grupo deberá seleccionar una base de datos sobre la que trabajará. Si el equipo de trabajo dispone de una base de datos que le resulta de interés analizar deberá consultar con el docente a cargo del grupo si la misma se adecúa a los objetivos del análisis que se solicita realizar. También se puede recurrir a una de las bases de datos dentro de las opciones¹ que siguen. El docente coordinará la selección de bases de datos que realicen los equipos de trabajo.

Posibles bases contenidas en el paquete UsingR (Para obtener información sobre estas bases consulte el link: <https://cran.r-project.org/web/packages/UsingR/UsingR.pdf>):

- i. babies
- ii. ceo2013
- iii. cfb
- iv. homeprice
- v. Medicare
- vi. SAT

Posibles bases contenidas en el paquete ISLR (Para obtener información sobre estas bases consulte el link: <https://cran.r-project.org/web/packages/ISLR/ISLR.pdf>):

- vii. Wage
- viii. Carseats
- ix. Credit

¹ Una vez que instaló el paquete de preferencia la base de datos puede cargarse de diversas maneras, por ejemplo a través del comando: `fix(nombredelabase)`. Si escribe el nombre de la base en la ayuda obtendrá información sobre ella.

El informe mínimamente deberá contener:

1. Una descripción de la base de datos: de qué trata, cuántos elementos y variables tiene.
2. La clasificación de las variables por tipo: cualitativas (nominales u ordinales) o cuantitativas (de intervalo o de razón, continua o discreta).
3. La determinación del tipo de datos contenidos en la base (de corte transversal o de series de tiempo).
4. Un análisis descriptivo detallado de una de las variables cualitativas y de dos de las variables cuantitativas contenidas en la base de datos. El equipo de trabajo deberá seleccionar estas 3 variables de forma tal que las mismas, a su vez, permitan realizar al menos un análisis de asociación entre la variable cualitativa y una de las variables cuantitativas seleccionadas y un análisis de correlación entre las 2 variables cuantitativas seleccionadas. Siendo así, la selección de variables deberá tener en consideración qué análisis de relación entre variables pueden ser relevantes en el contexto de la descripción de la base de datos y deberá fundamentarse en el informe.
 - a. Se espera que el análisis de la variable cualitativa seleccionada contenga como mínimo: una tabla de frecuencias y el o los gráficos que se consideren más adecuados.
 - b. Se espera que el análisis descriptivo de cada una de las 2 variables cuantitativas seleccionadas contenga como mínimo:
 - i. Una tabla de frecuencias agrupadas por las clases. La elección de la cantidad y el ancho de clase deberá fundamentarse.
 - ii. El cálculo de las medidas de resumen (tendencia central, separación y dispersión).
 - iii. Los gráficos que se consideren relevantes.
 - iv. La determinación de la presencia o ausencia de observaciones atípicas.
5. Una comparación de las distribuciones de las dos variables cuantitativas seleccionadas y un análisis de la distribución por grupos de cada variable cuantitativa por clases de la variable cualitativa.
6. Un análisis de asociación entre la variable cualitativa seleccionada y una de las variables cuantitativas seleccionadas (ya agrupada por clases).
7. Un análisis de correlación entre las 2 variables cuantitativas seleccionadas.

BONUS: Presente en un único gráfico de varios cuadrantes el diagrama de dispersión de las dos variables cuantitativas para cada una de las categorías que toma la variable cualitativa. Pista: Se recomienda leer las páginas 158 y 191 del libro "Using R for Introductory Statistics" de J. Versani.

8. Una reflexión final a modo de conclusión del análisis realizado.