

**EXAMEN FINAL - TERCER CORTE**  
**ESTADÍSTICA EXPLORATORIA (25%)**

Fecha de entrega: Viernes 11 de Noviembre 2022. Enviar al correo [ednamoreno@usantotomas.edu.co](mailto:ednamoreno@usantotomas.edu.co) en formato pdf o html.

**En el asunto del correo por favor colocar:** Examen Final, Tercer Corte-Exploratoria2A

**Nombre del archivo:** pepitoPerez.pdf (Colocar primer nombre y primer apellido) El trabajo es individual

**Base de datos** Buscar una base de datos **con mínimo** dos variables cuantitativas y dos variables cualitativas. Base de datos estructurada o formarla, es decir en las filas los individuos o unidades poblacionales y en las columnas las variables, deben tener por lo menos 80 filas. Esta base de datos debe ser real, y deben indicar la fuente de los datos.

Pueden recurrir a los siguientes enlaces:

<https://toolbox.google.com/datasetsearch> (La nueva funcionalidad de google, hagan el ejercicio de colocar Colombia en la búsqueda y les saldrán las bases de datos libres para nuestro país.)

<http://www.dane.gov.co/>

<http://www.banrep.gov.co/>

<https://goo.gl/Cx5djr>

<https://datos.bancomundial.org/>

- Reportar las cifras utilizando la cantidad adecuada de decimales, dependiendo de lo que se quiera mostrar y las necesidades del problema.
- El objetivo principal de este trabajo es la claridad lógica y **la interpretación de los resultados**. El informe no necesita ser extenso. Recuerde ser minimalista escribiendo el reporte. Se deben incluir solo aquellos gráficos y tablas (y valores en la tabla) que son relevantes para la discusión. **Por favor interprete las salidas y mencione lo más relevante.**
- Trate de plantear preguntas interesantes sobre su base de datos y responderlas con las técnicas vistas en el curso. Use filtros en su base de datos con consultas en sql o con la función subset.
- El trabajo debe ser elaborado en RMarkdown o LaTeX, se debe adjuntar el archivo pdf o html pues estos son los que se califican y como prueba el .Rmd o .Tex. La habilidad de usar herramientas como LaTeX y html para la elaboración de informes es muy importante y por esto si el trabajo no es realizado en LaTeX o Rmarkdown baja un punto.

**Realizar un análisis exploratorio y descriptivo de la base de datos.**

Debe tener suficientes gráficos, tablas y medidas descriptivas, que resuelvan preguntas de interés en el contexto en el que estén trabajando.

**TABLAS: (1.0)** Elabore las tablas cruzadas que considere importantes, **donde en por lo menos una saque porcentaje del total fila o columna. Interprete y resalte los resultados más importantes.**

**MEDIDAS DESCRIPTIVAS: (1.0)** Para una de las variables cuantitativas calcule TODAS las estadísticas descriptivas aprendidas (medidas de tendencia central, medidas de variabilidad, medidas de forma) en términos de las categorías de una variable cualitativa.

Por ejemplo: las descriptivas del salario en términos del género.

**Interprete muy bien cada uno de estos resultados.**

**GRAFICOS: (2.0)** (En esta parte se califica que todos los gráficos tengan un título apropiado, el nombre de los ejes de forma adecuada, el uso de colores y formatos de edición del gráfico y además se califica que todo gráfico sea acompañado por medidas descriptivas apropiadas que apoyen su interpretación, haga uso de la librería ggplot2)

- **Elabore e interprete un Boxplot comparativo de una cuantitativa en términos de las categorías de una variable cualitativa.** Acompañe este boxplot de un summary para facilitar la interpretación y redacte una muy buena interpretación.
- **Elabore un histograma y acompañelo de medidas descriptivas pertinentes para su interpretación.**
- Elabore un diagrama de barras apilado apropiado
- **Elabore un diagrama de dispersión** interesante y acompañelo de una medida descriptiva que ayude en su interpretación.

Tenga en cuenta que las siguientes recomendaciones:

La presentación del trabajo es muy importante, cuide mucho su redacción, deben tener títulos apropiados, párrafos introductorios y de enlace, párrafos de comparación o contraste.

\*Todo gráfico y tabla debe llevar un título apropiado y se debe hacer referencia con claridad a ellos en el análisis.

\*Es muy importante que enmarquen sus análisis en el contexto de la base de datos que escogieron, eviten hacer interpretaciones como “La media de los datos es....”

\*Es importante hacer buen uso de las referencias bibliográficas y escribir cuál es la fuente de los datos.

\*Las conclusiones deben ser un resumen claro de los resultados más importantes.

## ANEXO 2

Recomendación para la interpretación de los resultados: Por lo general, los temas que se desarrollan en los análisis de datos se centran alrededor de las siguientes preguntas, que debe tratar de responder:

- Localización: ¿Dónde están ubicados los valores de las variables? ¿Cuáles valores informan sobre los extremos posibles de la distribución? ¿Cuáles son los valores o los intervalos de valores que se repiten más? ¿Se puede identificar un valor que los resuma de manera conveniente? ¿Cuáles valores permiten repartir en fracciones especiales el conjunto de datos para identificar, por ejemplo, la mitad o la décima parte superior?
- Dispersión: ¿Qué tan diferentes son los valores de una variable en una agrupación específica? ¿Qué tan diferentes son con respecto a un valor que se toma como punto de referencia? ¿Presentan más diferencias en un grupo de datos que en otro?
- Forma: ¿Qué tanta concentración se puede apreciar alrededor de unos pocos valores? (Apuntamiento)? ¿Los valores se distribuyen de manera simétrica alrededor de un valor central.

**RETO: (1.0)** **Elabora un gráfico diferente y poco común** con tu base de datos utilizando alguna librería diferente a las vistas en clase. Consulta distintas fuentes en internet, acá dejo un página super interesante de gráficos avanzados. <https://r-graph-gallery.com/>