

Análisis e Interpretación de Datos

MÁSTER UNIVERSITARIO EN ANÁLISIS Y VISUALIZACIÓN DE DATOS
MASIVOS / VISUAL ANALYTICS AND BIG DATA

Miller Janny Ariza Garzón

Actividad 1. Estadística descriptiva: información a partir del análisis crítico de datos

Objetivo

Ser capaz de:

1. Aplicar conceptos básicos de la **estadística descriptiva** en un contexto de un problema real.
2. **Comunicar resultados** teniendo como base **elementos técnicos** bien definidos en el Área de Análisis e Interpretación de Datos
3. Presentar un informe que muestre la funcionalidad de los modelos (**estrategias**) propuestos para ofrecer un **análisis crítico y con fundamento técnico** sobre el **problema** planteado.

Fecha de entrega

viernes, 16 de diciembre de 2022, 23:59

Selección de datos y temática

Desde comienzos del año 2020 el mundo se ha visto “paralizado” por la incidencia de la pandemia conocida como **covid-19**. En esta actividad debes trabajar para echar luz sobre alguna de las tantas cuestiones que surgen de la mera “visualización” de los efectos de esta pandemia.

Pueden ampliar la temática a cualquier campo de interés. Inclusive, pueden buscar datos asociados con su TFM.

Selección de datos (algunas fuentes si el tema es covid-19)

- <https://www.ecdc.europa.eu/en/covid-19/data>
- <https://data.world/datasets/covid-19>
- <https://github.com/owid/covid-19-data/tree/master/public/data/>
- <https://cloud.google.com/blog/products/data-analytics/free-public-datasets-for-covid19>
- https://www.reddit.com/r/bigquery/comments/fxac2k/bigquery_public_datasets_covid19_related_datasets/
- <https://datascience.nih.gov/covid-19-open-access-resources>
- <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datalist?filter=datasets>
- <https://datascience.wisc.edu/institute/covid19/#localviz>
- <https://covid19.who.int/data>
- <https://ourworldindata.org/coronavirus-source-data>
- <https://datos.gob.es/es>

Selección de datos y temática

17 objetivos para transformar nuestro mundo. La Agenda 2030 para el Desarrollo Sostenible



1. Problema a tratar

Define un problema estadístico que consideres puede ser relevante para entender a nivel global el impacto del covid-19. **Puede ser de otra temática.**

(Revisa las orientaciones dadas en clase respecto a la “definición de problema estadístico”. En resumen, se trata de “una línea” tipo: estudio de ..., análisis de ..., impacto de ...)

Revisa y asegura la disponibilidad de datos

1. Problema a tratar

Temáticas:

Pinzón, J. E. D. (2020). **Estudio de los resultados del contagio por COVID-19 a nivel mundial.** *Revista Repertorio de Medicina y Cirugía*, 65-71.

Erades, N., & Sabuco, A. M. (2020). **Impacto psicológico del confinamiento por la COVID-19 en niños españoles: un estudio transversal.** *Revista de Psicología Clínica con Niños y Adolescentes*, 7(3), 27-34.

Sandín, B., Valiente, R. M., García-Escalera, J., & Chorot, P. (2020). **Impacto psicológico de la pandemia de COVID-19: Efectos negativos y positivos en población española asociados al periodo de confinamiento nacional.** *Revista de Psicopatología y Psicología Clínica*, 25(1).

Vega Falcón, V., Castro Sánchez, F., & Romero Fernández, A. (2020). **Impacto de la COVID-19 en el turismo mundial** (Impact of COVID-19 on World Tourism). *Impacto de la Covid-19 en el turismo mundial. Revista Universidad y Sociedad*, 12, 207-216.

Gago-García, C., González-Relaño, R., Cambronero, M. S., & Babinger, F. (2021). **Impacto de la crisis de la COVID-19 en el empleo del sector turístico en España: perspectivas territorial y de género.** *Boletín de la Asociación de Geógrafos Españoles*, (91).

Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H. K., Morgan, R., & Klein, S. L. (2020). **Impact of sex and gender on COVID-19 outcomes in Europe.** *Biology of sex differences*, 11(1), 1-13.

1. Problema a tratar

Piensa en una temática y pregunta general que pueda responderse con análisis de datos. Para esto, verifica los siguientes puntos (Sea lo más específico posible)

- Debe incluir: ¿Qué? (temática-problemática), ¿Dónde? (lugar-contexto), ¿Cuándo? (tiempo) ¿Cuánto(s)?(desea medir-evaluar)

Construya un párrafo que describa toda la problemática

Se espera un **problema** con respaldo a partir de **citas**, componente temporal, periodicidad, componente espacial, mencione las variables o constructos más relevantes a analizar. Debe ser alcanzable y pertinente.

1. Problema a tratar. Aspectos a tener en cuenta

- ¿Hay datos?. Fuentes.
- ¿Cómo va a analizar los datos? (estrategias o modelos estadísticos). Con la respuesta anterior, ¿ayuda a entender el problema o a dar soluciones a la problemática? ¿Cuál método ...?
- Describa las variables a usar (no necesariamente son todas las que conforman un dataset)
- Indicadores y medidas (preguntas cuantitativas)
 - ¿Qué tan grande es..?, ¿Cuánto? ¿Con que frecuencia..? ¿Cuál es el impacto...? ¿Cuáles son los determinantes de..?, ¿Cuál es la diferencia...?,...
- Visualización de resultados: tablas, cuadros de salida, gráficos

Tenga en cuenta estos aspectos al escoger la temática y definir el problema

2. Cita una referencia (o más)

Cita una referencia (o más) que consideres relevante para entender la importancia o justificar tu selección. (Recuerda las citas deben tener un formato adecuado, recomendamos APA).

https://bibliografiaycitas.unir.net/documentos/APA_7ed_UNIR.pdf

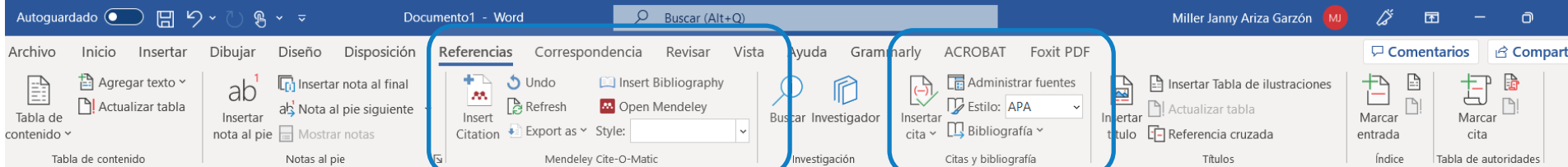
Tenga en cuenta citas parentéticas y citas narrativas:

- En las citas parentéticas, el nombre del autor y la fecha de publicación aparecen entre paréntesis. El machine learning carece de...(Ariza, 2021)
- En las citas narrativas, el nombre del autor se incorpora al texto como parte de la frase y el año sigue entre paréntesis. Se menciona el autor directamente...En su documento Ariza (2021) propone...

APA aplica también para la inclusión de tablas y gráficos

2. Cita una referencia (o más)

<https://biblioguias.ucm.es/mendeley/mendeley-cite>



Utilizar herramientas de word...

3 y 4. Conjunto de datos. Modelos (estrategias) estadísticos de trabajo

- Define el conjunto de datos y describe las variables a utilizar.
- Define ~~al menos dos~~ (un) modelos estadísticos de trabajo. Estrategia a utilizar.

4. -> 3. Detalla el conjuntos de datos. Fuentes (detallar, citar). Variables elegidas (tipos), número de registros, temporalidad, variable(s) objetivo (si la hay). Datos útiles para estudiar el problema planteado. Justifica la elección de variables en función del problema definido. Si hubo elección de una muestra, menciónalo.

Opcional: acompaña con el código de R que carga el conjunto de datos correspondiente en un anexo (si es de fuentes directas).

- Identificación tratamiento de NaNs,
- Tratamiento de registros duplicados
- Identificación y tratamiento de outliers.

3. -> 4. Detalla la estrategia estadística-en relación con el problema:

- Estrategia (metodología):
 - Comparar (impacto, evaluación,): Variables objetivo, variable grupo que define la comparación, medidas que resumen la comparación (media, mediana, desviación estándar, ..., nuevos indicadores), tablas y gráficos a realizar (histogramas, densidades, box-plots, barras, ...).
 - Identificar determinantes de..., predecir... : Estadísticas (media, ...) de la variable objetivo por variable categórica, tablas, gráficos y medidas de asociación como....
- Justifica cada elección (robustez, ...) y la estructura de la estrategia. Si usas tablas o gráficos, explícalos antes de interpretarlos.

5. “fotografía” de tus datos.

- Realiza el EDA incluyendo el análisis estadístico descriptivo. Tablas.
- Comenta tus resultados. Debes ser capaz de hacer un análisis crítico de los resultados y le debes dar alguna respuesta al problema inicialmente planteado.
- Debes dejar claramente expresadas las ventajas y limitaciones de tu estrategia estadística. Qué cosas se podrían mejorar

ISSUE TUTOR

6. Informe gráfico.

- Incluye en el informe gráficos que ilustren la secuencia de análisis de tu problema.
- Comenta tus resultados. Debes ser capaz de hacer un análisis crítico de los gráficos y le debes dar alguna respuesta al problema inicialmente planteado.

7. Conclusiones

- Incluye en el informe unas conclusiones compactas, claras y que en base a los resultados obtenidos respondan a la cuestión/problema inicial que hayas planteado.
- Relaciona estas conclusiones con las potencialidades y limitaciones de la estrategia elegida.
- Relaciona estas conclusiones con los referentes bibliográficos utilizados.

Formato

- Word o PDF. Si se utiliza RMD, anexar el archivo.
- Extensión: cinco páginas (una página de portada, una página de índice y tres páginas para desarrollar la actividad, como máximo).
- Una página adicional de bibliografía.
- Fuente Georgia 11 e interlineado 1,5.
- La base de datos generada se debe adjuntar en un archivo separado.
- Código R o python completo en un archivo separado.
- $2 \leq \text{archivos obligatorios} \leq 3$ (word o pdf, datos o **enlace directo a los datos** y código).

Rúbrica

Estadística descriptiva: información a partir del procesamiento numérico de datos.	Descripción	Puntuación máxima (puntos)	Peso %
Criterio 1	Definición del problema y referencias	2	20
Criterio 2	Elección de base de datos	2	20
Criterio 3	Análisis descriptivo numérico	2	20
Criterio 4	Análisis descriptivo gráfico	2	20
Criterio 5	Discusión y conclusiones de los resultados	2	20
		10	100 %

Próxima sesión

□ Tema 4: Regresión y correlación

- Correlación.
- Regresión lineal.
- Gráfico de residuos.

Learn by **DOING**.





www.unir.net