

# **Analítica de datos aplicada a estudios sobre desarrollo**

## **Segundo examen**

Docente: Giovany Babativa-Márquez

Monitor: David Felipe Tautiva.

Octubre 2025

## Tabla de Contenidos

1	Instrucciones.....	2
2	Entregables.....	3
3	Desarrollo .....	3
4	Ejercicios de apoyo para la construcción del paper.....	3
4.1	Descripción de los datos .....	4
4.2	Estrategia de analítica .....	6

## 1 Instrucciones

El trabajo debe ser desarrollado por los grupos conformados al inicio del curso. Se deberá entregar el *script* o archivo de **R** utilizado para generar los resultados, asegurando su reproducibilidad.

Además, se deberá presentar un documento con una extensión máxima de *10 páginas*, estructurado en formato de artículo científico y debe tener como mínimo, las siguientes secciones:

- Resumen
- Introducción
- Materiales y métodos
- Resultados
- Conclusiones
- Bibliografía

Estas secciones servirán como rúbrica de evaluación, considerando, entre otros aspectos, la calidad del contenido en términos de su relevancia, capacidad de análisis, uso de fuentes bibliográficas y calidad retórica.

El documento debe incluir los resultados que se consideren más relevantes, desarrollados a partir de los ejercicios propuestos más adelante. Dichos ejercicios han sido diseñados con el propósito de orientar la construcción de una estructura retórica coherente. Es decir, *no se debe entregar por separado el desarrollo de cada ejercicio*, sino integrar todos los elementos en un solo documento articulado.

La fecha límite para la entrega será el *07 de noviembre del 2025* y se debe cargar por la plataforma de Bloque Neón. *Los trabajos enviados por correo electrónico no serán considerados para su evaluación.*

Tenga en cuenta que:

1. Se hará una comparación entre los trabajos y de encontrar una alta similitud entonces la nota será dividida entre el número de personas involucradas.
2. Si el documento tienen más de las páginas permitidas (sin contar la bibliografía) tendrá una penalidad de 5 décimas por hoja adicional.

## 2 Entregables

1. Documento en formato Word o PDF tipo artículo académico siguiendo las instrucciones mencionadas previamente.
2. Script de R o archivo .Rmd que permita reproducir los resultados presentados.

## 3 Desarrollo

El Proyecto de Opinión Pública Latinoamericana (LAPOP por sus siglas en inglés), es una encuesta realizada desde el año 2004 que busca medir los valores, actitudes y comportamientos democráticos. Para el año 2015 ya se realizaba en 28 países con un tamaño de muestra de más de 50.000 encuestas y considera un diseño muestral probabilístico en cada país. La ronda de 2021, corresponde al último estudio realizado y se llevó a cabo en 22 países con más de 64.000 encuestas.

Para este examen se han descargado los datos de 2.993 encuestas realizadas en Colombia en la ronda 2021 y el cuestionario que fue utilizado, los cuales fueron descargados de la página de LAPOP en este [enlace](#). Asimismo, los estudiantes pueden acceder a los materiales de forma directa desde el repositorio de GitHub de este curso:

- *Ficha técnica*: haga clic [aquí](#)
- *Cuestionario*: haga clic [aquí](#)
- *Conjunto de datos en formato stata*: haga clic [aquí](#)

## 4 Ejercicios de apoyo para la construcción del paper

### Ejercicio 1

Este ejercicio tiene como objetivo proporcionar el contexto necesario para la redacción de la introducción del documento.

Se le ha solicitado construir tres (3) índices:

- Índice de confianza en instituciones
- Índice de democracia

- Índice de valores antidemocráticos

Para cada uno de los índices, defina claramente el concepto que busca medir, sustentando su enfoque a partir de referencias bibliográficas pertinentes. Asegúrese de incluir las citas y referencias correspondientes en el documento, siguiendo un formato de citación adecuado.

## Ejercicio 2

Este ejercicio tiene como propósito apoyar la elaboración de la sección de materiales y métodos del documento.

Investigue en qué consiste el estudio, su periodicidad y otros aspectos relevantes. Descargue el cuestionario desde el enlace indicado anteriormente e identifique las variables que emplearía para la construcción de cada uno de los tres (3) índices definidos en el ejercicio anterior.

En esta sección, deberá describir el método de análisis que utilizará, sin entrar en detalles técnicos, pero incluyendo las referencias bibliográficas pertinentes que sustenten su enfoque.

### 4.1 Descripción de los datos

El siguiente paso consiste en construir la base de datos. Para ello se llevaron a cabo algunas tareas que permiten importar, ordenar y transformar las variables que serán relevantes en nuestro análisis, el script utilizado puede ser descargado de [acá](#).

El conjunto de datos obtenido, luego del preprocesamiento, contiene 16 variables que fueron construidas desde las preguntas de la encuesta LAPOP, y en el que finalmente se tuvieron en cuenta 2.971 encuestas que pueden ser descargadas [acá](#). A continuación se describen las variables utilizadas.

Nombre	Descripción	Preguntas
<b>just_golpe</b>	Circunstancias en que se justificaría que los militares de este país tomen el poder por un golpe de Estado.	<i>jc13, jc13covid</i>
<b>just_cierre_cong</b>	Justificación para que el presidente del país cierre el Congreso y gobierne sin Congreso	<i>jc15a</i>
<b>conf_gobierno_nal</b>	Mide en una escala del 1 al 4 la confianza en que el gobierno nacional hace lo correcto.	<i>anestg</i> en escala invertida
<b>conf_instit</b>	Mide en una escala del 1 al 7 el nivel de respeto a las instituciones políticas en el país del encuestado.	<i>b2</i>

<b>conf_alcaldia</b>	Mide en una escala de 1 a 7 el nivel de confianza en la Alcaldía.	<i>b32</i>
<b>conf_elecciones</b>	Mide en una escala de 1 a 7 el nivel de confianza en las elecciones	<i>b47a.</i>
<b>conf_policia</b>	Mide en una escala de 1 a 7 el nivel de confianza en la policía	<i>b18</i>
<b>conf_medios</b>	Mide en una escala del 1 al 7 el nivel de confianza en los medios de comunicación	<i>b37</i>
<b>conf_fuerzas_mil</b>	Mide en una escala del 1 al 7 el nivel de confianza en las fuerzas militares.	<i>b12</i>
<b>sat_democracia</b>	Variable dicotómica que mide el nivel de satisfacción con la democracia de los encuestados.	<i>pn4</i> en escala invertida
<b>prot_derechos</b>	Mide en una escala del 1 al 7 la protección de los derechos básicos del ciudadano por parte del	
sistema político colombiano	<i>b3</i>	
<b>orgullo_sistema</b>	Mide en una escala del 1 al 7 qué tanto se siente usted orgulloso	
de vivir bajo el sistema político colombiano	<b>d4</b>	

Para realizar los análisis cargue los paquetes necesarios y el conjunto de datos, recuerde que puede hacerlo así:

```
rm(list = ls())

options(scipen = 999)
library(pacman)

p_load(tidyverse, janitor, corrplot, haven,
       devtools, FactoMineR, factoextra,
       ggcorrplot, GGally)

url <- "https://github.com/jgbabatvam/AnaDatos/raw/main/datos/df_colombia.dta"

datos <- read_dta(url)
```

Considerando la naturaleza de las variables que fueron elegidas, describa por qué un análisis de componentes principales es apropiado y describa brevemente en qué consiste el método (esto le debe permitir complementar su sección de materiales y métodos).

## 4.2 Estrategia de analítica

Los ejercicios propuestos deben orientarle en la elaboración de la sección de resultados del documento.

Siga las normas para llevar adecuadamente la numeración de las gráficas y tablas, y recuerde que todas las que decida incluir deben contar con una descripción o interpretación dentro del texto.

Procure utilizar una retórica clara y coherente que facilite la lectura de los resultados, de modo que el lector pueda seguir un hilo conductor atractivo y comprensible. En lo posible, intente contar una historia a través de los hallazgos obtenidos.

A partir de la base de datos construida, desarrolle los siguientes ejercicios con el fin de alimentar su paper:

### Análisis de correlación

3. Analice la relación lineal entre las variables. Para ello, calcule la matriz de correlación entre las variables cuantitativas y represente el resultado mediante un diagrama utilizando los paquetes `corrplot` o `ggcorrplot`. Concluya brevemente sobre los patrones observados en estos resultados.

### Análisis de componentes principales

Realice un Análisis de Componentes Principales (PCA) sobre las variables cuantitativas.

4. Discuta sobre la cantidad de información del conjunto de los datos es explicado por las dos primeras dimensiones. Presente un gráfico que lo ilustre.

Ayuda: use la función `fviz_screplot()` del paquete `factoextra`.

5. Haga una gráfica del plano factorial generado por las dimensiones 1 y 2 para las variables. Concluya sobre la asociación de las variables indicando cuáles presentan fuertes correlaciones y apuntan en la misma dirección representando el mismo concepto, indique también si considera que hay variables que no se representen bien en el plano de las primeras dos dimensiones.

Ayuda: use la función `fviz_pca_var()` del paquete `factoextra`.

6. Use el siguiente comando para visualizar de forma simultánea a los encuestados y las variables

```
fviz_pca_biplot(res, repel = F, col.var = "black", col.ind = "gray")
```

Tenga en cuenta que `res` corresponde al objeto que contiene los resultados del PCA. De manera visual, y considerando la densidad de puntos grises (encuestados), reflexione: ¿podría conjeturar que existe una buena confianza en las instituciones? Explique su razonamiento.

7. Explique qué representa un puntaje alto en la primera componente principal. Asigne un nombre apropiado al índice que se obtendría a partir de esta dimensión. Apoye su conclusión con el gráfico generado por el siguiente comando:

```
fviz_contrib(res, choice = "var", axes = 1, top = 10)
```

8. Explique qué representa un puntaje alto en la segunda componente principal. Asigne un nombre al índice correspondiente y fundamente su elección apoyándose en el gráfico generado con:

```
fviz_contrib(res, choice = "var", axes = 2, top = 10)
```

9. A partir de los resultados, explique que representa un puntaje alto en la tercera componente principal. Asigne un nombre al índice que se obtendría desde esta dimensión. Apoye su conclusión en el gráfico que resulta del siguiente comando:

```
fviz_contrib(res, choice = "var", axes = 3, top = 10)
```

10. Construya los índices formados por las tres (3) primeras componentes principales. Debe generar un índice por cada componente y agregarlos al conjunto de datos.

Ayuda: puede usar el siguiente código como guía:

#### ##### Indices

```
index1 <- as.data.frame(res$ind$coord[,1]) |>
  rename(score = `res$ind$coord[, 1]`) |>
  mutate(Indice1 = round(GGally::rescale01(score)*100, 1)) |>
  select(-score)

index2 <- as.data.frame(res$ind$coord[,2]) |>
  rename(score = `res$ind$coord[, 2]`) |>
  mutate(Indice2 = round(GGally::rescale01(score)*100, 1)) |>
  select(-score)

index3 <- as.data.frame(res$ind$coord[,3]) |>
  rename(score = `res$ind$coord[, 3]`) |>
  mutate(Indice3 = round(GGally::rescale01(score)*100, 1)) |>
  select(-score)
```

```
df_index <- bind_cols(datos, index1, index2, index3)
```

11. Realice análisis descriptivos de los valores de los índices por región, sexo, edad y nivel educativo.

Ayuda: puede calcular promedios o utilizar representaciones gráficas que le permitan identificar si existe una mayor confianza en determinadas regiones, rangos de edad o niveles educativos.

12. Elabore las conclusiones generales del ejercicio, integrando los hallazgos obtenidos y reflexionando sobre las implicaciones de los resultados en relación con los conceptos de confianza, democracia y valores antidemocráticos.

Recuerden que nos pueden escribir para cualquier duda o programar una sesión de monitoria:

[j.babativamarquez@uniandes.edu.co](mailto:j.babativamarquez@uniandes.edu.co)

[d.tautivan@uniandes.edu.co](mailto:d.tautivan@uniandes.edu.co)