



Maestría en Estadística Aplicada  
Universidad Nacional de Córdoba  
Cohorte 2024 -2025

Proyecto de Análisis de Datos I

Docentes a cargo: Dra. María Gabriela Palacio – Mgtr. Sergio Martín Buzzi

---

**Guía Nº 2**

Esta segunda guía consiste en dos ejercicios, uno de análisis sobre experimentos factoriales y otro de análisis multivariante.

Cada grupo deberá subir un archivo .pdf con un desarrollo completo y un script o notebook de R o de Python.

Los archivos deberán ser subidos por un solo integrante de cada grupo y los nombres de estos deberán seguir el formato: GrupoXApellidoApellidoApellidoApellido.

Fecha de entrega: **lunes 12 de mayo de 2025.**

Por cualquier eventualidad, pueden escribirnos a nuestros correos electrónicos: María Gabriela Palacio ([gpalacio@exa.unrc.edu.ar](mailto:gpalacio@exa.unrc.edu.ar)) y Sergio Martín Buzzi ([sergio.buzzi@unc.edu.ar](mailto:sergio.buzzi@unc.edu.ar)).

**Problema 1**

Las lesiones cerebrales en los recién nacidos de pretérmino (RNPT) constituyen un importante problema de salud, tanto por el número de niños afectados como por la magnitud de las secuelas secundarias que producen. La Hemorragia Intracraneana (HIC) del prematuro se presenta frecuentemente y es una importante causa de deterioro neurológico en el recién nacido. Se reconoce su origen multifactorial, pero sobresalen algunos factores involucrados, como las características anatómicas del cerebro neonatal, factores perinatales y algunos fenómenos vasculares muy específicos del prematuro.

En el estudio realizado en el Servicio de Neonatología del Hospital San Antonio de Padua de Río Cuarto entre julio de 2005 y abril de 2007, se consideraron los recién nacidos con peso menor o igual a 1750 grs. internados en dicho servicio.

En la base de datos “HIC” de Excel se incluyeron las variables: Sexo (S), Peso al nacer (P), Edad gestacional en semanas (E), HIC (HIC), Maduración Pulmonar completa (MP), forma de Parto (PAR), Distres respiratorio (D), Sepsis neonatal (SEP), necesidad de Asistencia Respiratoria Mecánica (ARM), administración de Surfactante (SU), Apnea (AP), Hipotensión arterial (HI), Acidosis Metabólica Severa (AC), Hiperglucemia (HIP).



Maestría en Estadística Aplicada  
Universidad Nacional de Córdoba  
Cohorte 2024 -2025

Se solicita analizar los datos usando un **arreglo factorial (con 2 factores) sobre un diseño completamente aleatorizado**, considerando al factor HIC como uno de los factores.

I) Indicar la variable respuesta que considere más adecuada, justificando.

II) Elegir un factor con el cual tenga sentido cruzar a HIC, teniendo precaución con la cantidad de observaciones que quedan en cada tratamiento y considerando que para algunos hay varios datos perdidos. Para este caso:

- a) Escribir brevemente el objetivo específico.
- b) Indicar por qué puede considerarse que los factores son fijos y están cruzados (no anidados), justificando.
- c) ¿Cuántos tratamientos hay en estudio? Mencionarlos.
- d) Escribir el modelo de efectos, indicando qué representa cada uno de sus términos.
- e) Escribir las hipótesis a probar estadísticamente y en palabras.
- f) Se solicita específicamente lo siguiente (aunque pueden agregarse análisis extras):
  1. Realizar el análisis descriptivo (numérico y gráfico) interpretando los resultados obtenidos según el objetivo planteado y el análisis inferencial a realizar.
  2. Mostrar la tabla de medias, indicando la relación de los valores de la misma con los estimadores de los parámetros del modelo.
  3. Mostrar los estimadores de los parámetros del modelo.
  4. Realizar el análisis inferencial y obtener conclusiones.
- g) ¿Son válidas las conclusiones del ANOVA? Si no lo son proponer alguna alternativa posible de análisis.
- h) ¿Los factores actúan de manera independiente? Indique lo que esto significa para el problema. Si fuera de otro modo ¿cómo procedería?



# Maestría en Estadística Aplicada

## Universidad Nacional de Córdoba

### Cohorte 2024 -2025

- i) Según los resultados obtenidos ¿es necesario realizar comparaciones múltiples de medias? Justifique.
- j) En caso de que la interacción fuera significativa, ¿qué análisis completaría al ya realizado?, ¿qué conclusiones se podrían obtener?

## Problema 2

En base a datos del Censo Provincial 2008 de la Provincia de Córdoba, se busca realizar un análisis multivariante sobre la situación socioeconómica de la Ciudad de Córdoba.

Con tal finalidad, a partir de un conjunto de variables registradas por barrio, se calcularon diversas proporciones y se las guardó en el archivo “proporciones.xlsx”.

Estas proporciones fueron calculadas dividiendo la cantidad de personas, hogares, etc. que corresponden a una categoría, sobre el total de personas, hogares, etc. Por ejemplo, a partir de la cantidad de jefes o jefas de hogar que tienen nivel universitario completo en cada barrio, al dividir por la cantidad de jefes o jefas de hogar de cada barrio se obtuvo la proporción de jefes o jefas de hogar con nivel educativo universitario completo de cada barrio.

En forma resumida las proporciones son:

- Proporción de los jefes o jefas de hogar según nivel educativo: inicial (incluye sin instrucción), priminc, primcomp, secinc, seccomp, terinc, tercomp, univinc, univcomp, posinc, poscomp.
- Proporción de las personas de 3 años o más, según si saben leer y escribir: sabeleer, nosabeleer.
- Proporción de personas según cobertura de salud / obra social / prepaga: cobertura, nocobertura.
- Proporción de personas de 14 años o más según condición de ocupación: ocupados, desocupados, inactivos.



## Maestría en Estadística Aplicada

### Universidad Nacional de Córdoba

#### Cohorte 2024 -2025

- Proporción de personas por tipo de necesidades básicas insatisfechas: nbi1, nbi2, nbi3, nbi4, nbi5.
- Proporción de personas según índice de privación material en el hogar: sin pro, privpat, privrec, privconv.
- Proporción de hogares por tipo: casa, rancho, casilla, departamento, piezaolocal.
- Proporción de hogares con determinada cantidad de habitantes: uno, dos, tres, cuatro, cinco, seis, 7, ochoomas.

A partir de estos datos, se pide:

- a) Realizar un análisis descriptivo univariado breve de las variables empleando medidas de resumen y gráficos. ¿ Hay algún problema en la base de datos?
- b) Realizar un análisis descriptivo multivariado.
  - 1) ¿Hay outliers multivariantes?
  - 2) ¿La distribución de los datos es Normal multivariante?
- c) Obtener los componentes principales.
  - 1) ¿Cuántos componentes principales bastan para retener suficiente información sobre los datos? Fundamente empleando los criterios estudiados.
  - 2) Indique qué variables son las más importantes en la determinación del primer componente principal.
- d) Efectuar un análisis factorial exploratorio.
  - 1) Probar diversas rotaciones.
  - 2) Intentar encontrar nombres intuitivos para los factores.
  - 3) Graficar.