

Facultad de Medicina

Epidemiología- MEDI-1105



Profesora: Olga Lucía Sarmiento M.D M.P.H Ph.D

2025-1

GUIA DE TALLERES

Objetivos: El taller de epidemiología tiene como objetivo complementar las clases magistrales. Los objetivos son:

1. Desarrollar un entendimiento básico de procedimientos para analizar bases de datos comúnmente usados en investigación epidemiológica.
2. Aprender los procedimientos básicos de epidemiología usando el lenguaje de programación R, en el ambiente de programación RStudio.

Estructura: El laboratorio consiste en un ejercicio analítico de un estudio epidemiológico. Durante este laboratorio

los estudiantes trabajarán en grupos de cuatro a cinco personas para desarrollar el análisis de esta base de datos.

Trabajos para entregar: Tres talleres serán entregados como resultado de este análisis. Cada trabajo incluye una serie de ejercicios y preguntas que los guiarán a través del análisis. Todos los resultados deben tener las anotaciones pertinentes de acuerdo a la pregunta.

Cálculos Estadísticos: Algunos de los trabajos requieren que usted calcule a mano los resultados. El propósito es que entiendan los cálculos y procedimientos de los análisis de datos que el programa de computador realiza. Necesitarán una calculadora.

TALLER 1

MODELO CONCEPTUAL, DESCRIPCIÓN DE LAS VARIABLES Y CARACTERIZACIÓN DE LA POBLACIÓN A ESTUDIO

En este taller se evaluará el modelo conceptual en el cual se basa el análisis que se realizará de la ENSIN 2010. Adicionalmente se examinarán las variables a estudiar considerando el tipo de variables en la base de datos, su distribución y valores faltantes.

Objetivos específicos:

- Entender por qué y para qué estamos realizando este análisis (usando el modelo conceptual).
 - Familiarizarse con la base de datos y comandos básicos de R.
 - Aprender la importancia de evaluar los valores faltantes de las variables.
 - Calcular las medidas de tendencia central y dispersión de las variables continuas bajo estudio.
 - Aprender a realizar una tabla y a describir la población a estudio como si fuese a ser publicada en el American Journal of Epidemiology (AJE).

Pasos:

1. Lea:

- Meisel, J. D., Esguerra, V., Giraldo, J. K., Montes, F., Stankov, I., Meisel, C. A., Sarmiento, O. L., & Valdivia, J. A. (2023). Understanding the dynamics of the obesity transition associated with physical activity, sedentary lifestyle, and consumption of ultra-processed foods in Colombia. *Preventive medicine*, 177, 107720. <https://doi.org/10.1016/j.ypmed.2023.107720>
 - González, S., Sarmiento, O. L., Lozano, O., Ramírez, A., & Grijalba, C. (2014). Niveles de actividad física de la población colombiana: desigualdades por sexo y condición socioeconómica. *Biomédica*, 34(3). <https://doi.org/10.7705/biomedica.v34i3.2258>
2. Lea el reporte de la ENSIN de sobrepeso y obesidad que se encuentra en esta página web en el apartado 2.1 de antropometría: <https://www.icbf.gov.co/bienestar/nutricion/encuesta-nacional-situacion-nutricional#ensin2>
 3. Elabore el modelo conceptual del este estudio que estamos analizando. Recuerde que el objetivo es evaluar los factores sociodemográficos asociados a sobrepeso y obesidad en la población colombiana. (utilice las variables del codebook para el modelo conceptual)
 4. Describa el tipo de estudio

5. Tabla #1: Describa el tipo de variables que se utilizarán en el análisis. Clasifique cada variable de la base de datos en nominal categórica, ordinal categórica o continua.

VARIABLE LABEL	DESCRIPCIÓN PREGUNTA FORMULADA	NOMBRE DE LA VARIABLE ORIGINAL PARA USAR EN R	CÓDIGO DE LA VARIABLE ORIGINAL	NOMBRE DE LA NUEVA VARIABLE	NUEVO CÓDIGO
EDAD	Edad del adulto	EDAD	RANGO DE 18 - 64	EDADNEW	1 = 18 - 25 años 2 = > 25 años
SEXO	Sexo del adulto	SEXO	1 = Hombre 2 = Mujer	SEXONEW	0 = Hombre 1 = Mujer

ETNIA	Etnia del adulto	ETNIA	1= Indígena 2= Gitanos 3= Raizal del archipiélago o 4= Palenquero 5= Negro/ Mulato/ Adrofes / Afrocol 6= Otros NA= missing	ETNIA NEW	0= Indígena 0= Gitanos 1= Raizal del archipiélago o 1= Palenquero 1= Negro/ Mulato/ Adrofes / Afrocol 0= Otros NA= missing
ESCOLARIDAD	Máximo nivel educativo alcanzado del adulto	ESCOLARIDAD	0= Ninguna 1= Preescolar 2= Primaria 3= Secundaria 4= Técnico o tecnológico 5= Universidad 6= Posgrado 8= No sabe NA= missing	ESCOLARIDAD NEW	0= Ninguna 0= Preescolar 0= Primaria 0= Secundaria 1= Técnico o tecnológico 1= Universidad 1= Posgrado NA= No sabe NA= missing
OCUPACION	Ocupación en la semana anterior	OCUPACION	1= Trabajó en la semana anterior 2= No trabajó, pero tenía trabajo 3= Buscando trabajo 4= Estudió 5= Oficios del hogar 6= Pensionado 7= Discapacidad 8= Otro 96= Otro 98= No sabe NA= missing	OCUPACION NEW	0= Trabajó en la semana anterior 0= No trabajó, pero tenía trabajo 1= Buscando trabajo 1= Estudió 1= Oficios del hogar 1= Pensionado 1= Discapacidad NA= Otro NA= Otro NA= No sabe NA=

					missing
DISCAPACIDAD	Cualquier tipo de discapacidad física y/o sensorial	QH42	1= Si 2= No NA= missing	DISCAPACIDAD NEW	0= No 1= Sí NA= missing

NIVEL DE SISBEN	Nivel de Sisben del hogar del adulto	SISBEN	1= sisben 1 2= sisben 2 3= sisben 3 4= sisben 4-6 NA= missing	SISBENNEW	0= sisben 1 0= sisben 2 0= sisben 3 1= sisben 4-6 NA= missing
ACTIVIDAD FISICA EN TIEMPO LIBRE	Cumplir con la recomendación de 150 minutos de AF moderada o 75 minutos de AF vigorosa en la semana	Tlmeett	0= No cumple con las recomendaciones <150 mins 1= Cumple con recomendaciones caminando en tiempo libre ≥150 mins NA= missing	TlmeettNEW	0= No cumple con las recomendaciones <150 mins 1= Cumple con recomendaciones caminando en tiempo libre ≥150 mins NA= missing
IMC	Indice de masa corporal <i>Peso/Talla²</i>	imcNEW	0= delgadez 1= Normal 2= Sobrepeso 3= Obesidad NA= missing	¿Cómo lo recategorizaría?	Estudiantes creen sus propias categorías y justifiquen

Inicio de R

6. Baje la base de datos de BLOQUE NEÓN y guárdela en una carpeta nueva. NO LA GUARDE EN EL ESCRITORIO.
7. Abra un nuevo archivo .R y guárdelo en la misma carpeta en la que está la base de datos.
8. Asigne la dirección de la base de datos en el archivo .R mediante el comando `setwd()`. En esta dirección se guardarán los resultados. Recuerde que la dirección debe ir con '/' o con doble '\' (sólo para windows)

```
setwd("/Users/mariaalejandra/Taller1")
```

o

```
setwd("\\Users\\mariaalejandra\\Taller1")
```

9. Instale las siguientes librerías para poder usar las funciones y comandos necesarios para el taller:

```
epiDisplay,  
car, plyr, psych, modeest
```

Para instalar estas librerías puede ir al botón *Tools* en la barra de herramientas y seleccionar el comando *Install Packages...* , o usar la siguiente línea de código. Recuerde que esto se debe realizar una sola vez por computador.

```
install.packages(c("epiDisplay", "car", "plyr", "psych", "modeest"))
```

Los nombres de las librerías necesarias están contenidos en un vector denotado como `c()`

10. Cargue las librerías para poder usarlas

```
library(epiDisplay) #Esta librería contiene funciones necesarias para análisis epidemiológicos
```

```
library(car) #Esta librería contiene funciones necesarias para análisis estadísticos library(plyr) #Esta librería contiene funciones para transformación de datos library(psych) #Esta librería contiene funciones análisis de variables continuas library(modeest) #Esta librería contiene funciones para estimar la moda
```

11. Cargar la base de datos y asignarle un nombre

```
base<-read.csv("af_estudiantes.csv") #La base guardada tiene el nombre: taller1.csv
```

12. Revise que las variables del diccionario de variables se encuentran en el data set

```
summary(base)
```

13. Convierta las variables necesarias a factores. Este es el código que necesita:

```
base$variable<-as.factor(base$variable)
```

14. Re-codifique las variables a estudiar utilizando SOLO los códigos nuevos de la tabla #1. Cree variables duplicadas para no alterar la base de datos original.

Para las variables numéricas necesita el siguiente código:

```
base$variableNueva<-cut(base$VariableVieja,breaks = c(min,corte,max), labels = c("1","2"),include.lowest = TRUE)
```

La función cut() vuelve la variable numérica en una variable categórica. Los breaks = c(min,corte,max) generan los cortes necesarios a las categorías y labels=c("1","0") le da nombres a las categorías creadas.

Para las variables categóricas necesita el siguiente código:

```
base$variableNueva<-recode(base$variableVieja,"c('1','2','3')=0;c('4','5','6') = '1';else=NA")
```

La función recode() recodifica las variable categórica. c('1','2','3') son las categorías antiguas y '0' es la categoría nueva. Cada recodificación se separa con un ; y debe ir dentro de comillas.

15. Verifique que los códigos asignados son correctos mediante la comparación de variables originales y las variables nuevas. Para variables categóricas use este código:

```
table(base$variableVieja,base$variableNueva,useNA = "always") summary(base$variableVieja) summary(base$variableNueva)
```

16. Tabule las frecuencias, **incluyendo** los valores **missing** de todas las **variables nuevas (solo variables nuevas)**.

La función que necesitará para este procedimiento es:

```
table(base$variableNueva,useNA = "always")
```

17. Tabule las frecuencias, **excluyendo** los valores **missing** de todas las **variables nuevas**.

```
table(base$variableNueva)
```

ANALISIS UNIVARIADO

18. Grafique y examine la distribución de las variables continuas. La función que necesitará es:

```
hist(base$variable)
```

19. Determine la media, la desviación estándar, el rango, la mediana, u los percentiles 5th y 95th para las variables continuas.

```
describe(base$variable)
```

O

```
summary(base$variable)
```

```
#media mean(base$variable) #desviación
```

```
estandar sd(base$variable) #mediana
```

```
median(base$variable)
```

```
#moda mlv(base$variable, method = "mfv")
```

```
#percentiles 5 y 95
```

```
quantile(base$variable,c(0.05, .95))
```


TRABAJO ESCRITO PARA ENTREGAR (OJO, SOLO ENTREGAR LAS RESPUESTAS A ESTAS PREGUNTAS, NO ENTREGAR CODIGO) NO SE RECIBEN TRABAJOS POR EMAIL

1. ¿El tipo de estudio que representa la ENSIN puede demostrar causalidad? (Describa el tipo de estudio y justifique su respuesta)
2. ¿Qué características debe tener un estudio para poder demostrar causalidad?
3. ¿Cuál es la **muestra** de la población a estudio de la ENSIN para el análisis de sobrepeso y obesidad?
4. Entregar el gráfico del modelo conceptual del análisis que estamos realizando, teniendo en cuenta solo las variables de la base de datos. (Tenga en cuenta el ejemplo de la doctora en clase)
5. Justificar **TODAS** las relaciones representadas en el modelo conceptual (la justificación incluye la **dirección, el sentido y la magnitud** de las asociaciones descritas en el modelo) (esta justificación no debe ser más larga de 200 palabras)
6. Escribir la pregunta de estudio del análisis que estamos realizando (Pregunta de investigación).
7. Escribir la hipótesis nula y alterna del estudio y análisis que estamos realizando.
8. Describa el tipo de las variables que se están analizando (Tipo de numérica, tipo de categórica) (describa **TODAS** las **variables originales y nuevas** del diccionario de variables)
9. Haga una tabla con los porcentajes y frecuencias de las variables como si fuera para publicación en el AJE. En la tabla recuerde la importancia del título y el nombre de las variables. **Consulte el ejemplo de tabla de un artículo publicado en el AJE que está en BLOQUE NEÓN para el formato de tabla correcto.**
10. **SÓLO** en un párrafo, describa los resultados de la tabla de la pregunta anterior (p.9)
11. Hacer una tabla con sólo los porcentajes de missings de cada variable nueva
12. ¿Cree usted que el porcentaje de missings de alguna variable es significativo y puede alterar la validez de los resultados?
13. Describa la media, desviación estándar, mediana, moda, rango y rango Inter cuantil de las variables continuas. Escriba la utilidad de cada una de estas medidas en estadística.
14. Colocar las gráficas de la distribución de las variables numéricas y determinar qué tipo de distribución tienen. Analice y describa la kurtosis de las gráficas.
15. Comente acerca de la categorización de la variable escolaridad y ocupación. ¿Está usted de acuerdo o en desacuerdo? ¿Por qué?
16. ¿Qué categorización utilizó para IMC teniendo en cuenta que el análisis será enfocado a sobrepeso y obesidad? Justifique su respuesta.