

Estudios Transversales en R

Msc Juan José Vindell

27/1/2023

INTRODUCCIÓN

Los estudios epidemiológicos clásicamente se dividen en Experimentales y No experimentales. En los estudios experimentales hay asignación de una exposición determinada o sea que el investigador “impone” el factor al sujeto de investigación. Cuando el experimento no es posible se diseñan estudios no experimentales que simulan de alguna forma el experimento que no se ha podido realizar.



Figure 1: Esquema de los estudios de investigación epidemiológica

ESTUDIOS TRANSVERSALES

Se consideran transversales los estudios en los que los datos de cada sujeto representan esencialmente un momento del tiempo. Estos datos pueden corresponder a la presencia, ausencia o diferentes grados de una característica o enfermedad (como ocurre, por ejemplo, en los estudios de prevalencia de un problema de salud en una comunidad determinada), o bien examinar la relación entre diferentes variables en una población

definida en un momento de tiempo determinado.

Dado que las variables se han medido de forma simultánea, no puede establecerse la existencia de una secuencia temporal entre ellas y, por tanto, estos diseños no permiten abordar el estudio de una presunta relación causa-efecto. Así pues, los estudios transversales son por definición descriptivos.

APLICACIÓN EN R

Instalando paquetes

```
#install.packages("Nombre del paquete que se desea utilizar")
```

Abriendo paquetes

```
#library("Nombre el paquete que se desea utilizar")
```

EJEMPLO

```
#install.packages("tidyverse")
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.0      v dplyr  1.0.8
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5
## Warning: package 'readr' was built under R version 4.0.5
## Warning: package 'dplyr' was built under R version 4.0.5
## Warning: package 'stringr' was built under R version 4.0.5

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

IMPORTANDO LA BASE DE DATOS

Consideraciones:

- Tener la base de datos en una carpeta.
- Conocer el formato de la base de datos.
- .xls

Instalando paquete readxl

Ahora, necesitamos instalar el paquete readxl para importar datos de Excel. Éste es, en mi opinión, el paquete más sencillo de usar para importar este tipo de datos.

Instalamos el paquete con el siguiente código:

```
#install.packages("readxl")
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.5
```

Por otra parte, necesitamos conocer la ruta del archivo de Excel. La ruta, o path en inglés, es la dirección/ubicación donde está el archivo en tu computadora u ordenador.

Ruta del archivo para importar de R a excel

Primero, veremos cómo sacar esta ruta con una función muy chévere de R (Es de esos secretos que revelamos aquí para ser mejores con R).

Así, la función se llama `file.choose()`. Luego, esta función te permite abrir una ventana de Windows para buscar un archivo y regresa una variable tipo carácter con la ruta del archivo. Por ejemplo, en R.

```
datos=file.choose()
```

Antes de cargar los datos a R, veamos cuales son las hojas que tiene este Excel. Luego, podemos utilizar el código

```
excel_sheets(datos)
```

```
## [1] "HOJA1"
```

Para leer los datos de excel hacia R

```
data1=read_excel(datos)
```

Para leer los encabezados de la data

```
head(data1)
```

```
## # A tibble: 6 x 14
##   EDAD GENERO  ESCUELA  HTA  PESO TALLA calcimc IMC_REAL  GLUCOSA GLICEMIA
##   <dbl> <chr>   <chr>   <chr> <dbl> <dbl>   <dbl> <chr>    <dbl> <chr>
## 1    60 Masculi~ Administr~ Si    68.8  1.7    23.8 Normal    79.2 No
## 2    56 Masculi~ Administr~ No    68.8  1.6    26.9 Sobrepeso  72.3 No
## 3    63 Masculi~ Administr~ No    69.9  1.58   28.0 Sobrepeso  80.8 No
## 4    66 Masculi~ Administr~ No    57.9  1.61   22.3 Normal    81.4 No
## 5    64 Masculi~ Administr~ No    60    1.64   22.3 Normal   127. Si
## 6    45 Masculi~ Administr~ No    87.5  1.66   31.8 Obesidad~  89  No
## # ... with 4 more variables: COLESTEROL <dbl>, COLESTEROL_TOTAL <chr>,
## # TRIGLICÉRIDOS <dbl>, TRIGLICERIDOS_COD <chr>
```

Para ver el nombre de las variables que tiene mi data

```
names(data1)
```

```
## [1] "EDAD"           "GENERO"          "ESCUELA"
## [4] "HTA"            "PESO"            "TALLA"
## [7] "calcimc"        "IMC_REAL"        "GLUCOSA"
## [10] "GLICEMIA"       "COLESTEROL"      "COLESTEROL_TOTAL"
## [13] "TRIGLICÉRIDOS"  "TRIGLICERIDOS_COD"
```

Para ver que tipo de variable tiene mi data

```
str(data1)
```

```
## tibble[,14] [210 x 14] (S3: tbl_df/tbl/data.frame)
## $ EDAD          : num [1:210] 60 56 63 66 64 45 72 68 38 48 ...
## $ GENERO        : chr [1:210] "Masculino" "Masculino" "Masculino" "Masculino" ...
## $ ESCUELA       : chr [1:210] "Administrativo" "Administrativo" "Administrativo" "Administrativo" ...
## $ HTA           : chr [1:210] "Si" "No" "No" "No" ...
## $ PESO          : num [1:210] 68.8 68.8 69.9 57.9 60 87.5 80.6 66.5 65.3 74.7 ...
## $ TALLA         : num [1:210] 1.7 1.6 1.58 1.61 1.64 1.66 1.71 1.56 1.61 1.64 ...
## $ calcimc       : num [1:210] 23.8 26.9 28 22.3 22.3 ...
```

```
## $ IMC_REAL      : chr [1:210] "Normal" "Sobrepeso" "Sobrepeso" "Normal" ...
## $ GLUCOSA       : num [1:210] 79.2 72.3 80.8 81.4 127.4 ...
## $ GLICEMIA      : chr [1:210] "No" "No" "No" "No" ...
## $ COLESTEROL    : num [1:210] 221 181 247 236 201 ...
## $ COLESTEROL_TOTAL : chr [1:210] "Si" "No" "Si" "Si" ...
## $ TRIGLICÉRIDOS : num [1:210] 192.5 80.5 241 108.2 124.2 ...
## $ TRIGLICERIDOS_COD: chr [1:210] "Si" "No" "Si" "No" ...
```

ANÁLISIS DESCRIPTIVO

Variables cualitativas

- Frecuencia `table()`
- Proporciones `prop.table()`

EJEMPLO

```
table(data1$GENERO)
```

```
##
## Femenino Masculino
##      110      100
```

```
prop.table(table(data1$GENERO))*100
```

```
##
## Femenino Masculino
## 52.38095 47.61905
```

Si por ejemplo quisieramos hacer una tabla de contingencia ocupamos el siguiente comando

```
table(data1$GENERO,data1$HTA)
```

```
##
##           No  Si
## Femenino  100  9
## Masculino  77  23
```

Variable cuantitativa

- Un cuadro resumen `summary()`
- Medidas descriptivas como `mean()` y `sd()`

##EJEMPLO

```
summary(data1)
```

```
##      EDAD      GENERO      ESCUELA      HTA
## Min.   :22.00  Length:210    Length:210    Length:210
## 1st Qu.:38.00  Class :character  Class :character  Class :character
## Median :48.00  Mode  :character  Mode  :character  Mode  :character
## Mean   :47.66
## 3rd Qu.:56.00
## Max.   :77.00
##
##      PESO      TALLA      calcimc      IMC_REAL
## Min.   : 47.50  Min.   :1.440  Min.   :19.54  Length:210
## 1st Qu.: 65.78  1st Qu.:1.530  1st Qu.:26.42  Class :character
```

```
## Median : 74.10      Median :1.600      Median :29.02      Mode  :character
## Mean   : 75.84      Mean   :1.601      Mean   :29.49
## 3rd Qu.: 86.08      3rd Qu.:1.670      3rd Qu.:32.07
## Max.   :122.60      Max.   :1.800      Max.   :50.91
##
##      GLUCOSA          GLICEMIA          COLESTEROL      COLESTEROL_TOTAL
## Min.   : 54.10      Length:210      Min.   : 67.4      Length:210
## 1st Qu.: 70.80      Class :character 1st Qu.:110.8      Class :character
## Median : 77.90      Mode  :character Median :133.8      Mode  :character
## Mean   : 82.57
## 3rd Qu.: 86.40
## Max.   :253.90
## NA's   :1
## TRIGLICÉRIDOS      TRIGLICERIDOS_COD
## Min.   : 29.1      Length:210
## 1st Qu.:113.0      Class :character
## Median :152.1      Mode  :character
## Mean   :166.8
## 3rd Qu.:198.9
## Max.   :478.7
## NA's   :1
```

```
mean(data1$EDAD)
```

```
## [1] 47.65714
```

```
sd(data1$EDAD)
```

```
## [1] 11.91704
```

Para facilitarnos todo este trabajo ocuparemos el comando **tableone()**

¿Qué es tableone?

El paquete tableone es un paquete R que facilita la construcción de la “Tabla 1”, es decir, la tabla de características iniciales del paciente que se encuentra comúnmente en los artículos de investigación biomédica.

Los paquetes pueden resumir variables continuas y categóricas mezcladas en una tabla. Las variables categóricas se pueden resumir como recuentos y / o porcentajes. Las variables continuas se pueden resumir de forma “normal” (medias y desviaciones estándar) o “no normal” (medianas y rangos intercuartílicos).

```
#install.packages("tableone")
library(tableone)
```

```
## Warning: package 'tableone' was built under R version 4.0.5
```

Resumen de un solo grupo

Caso de uso más simple

El caso de uso más simple es resumir todo el conjunto de datos. Puede simplemente alimentar en el marco de datos a la función principal **CreateTableOne()**. Puede ver que hay 210 pacientes en el conjunto de datos.

```
tab=CreateTableOne(data = data1)
tab
```

```
##
##                               Overall
##      n                               210
```

```
## EDAD (mean (SD))          47.66 (11.92)
## GENERO = Masculino (%)    100 (47.6)
## ESCUELA (%)
##   Administrativo          112 (53.3)
##   Escuela de Medicina     45 (21.4)
##   ETM                      53 (25.2)
## HTA = Si (%)              32 (15.3)
## PESO (mean (SD))          75.84 (14.87)
## TALLA (mean (SD))          1.60 (0.08)
## calcimc (mean (SD))        29.49 (4.79)
## IMC_REAL (%)
##   Normal                   38 (18.1)
##   Obesidad I                66 (31.4)
##   Obesidad II               16 ( 7.6)
##   Obesidad M3rbida          7 ( 3.3)
##   Sobrepeso                 83 (39.5)
## GLUCOSA (mean (SD))        82.57 (24.05)
## GLICEMIA = Si (%)          14 ( 6.7)
## COLESTEROL (mean (SD))     137.37 (38.41)
## COLESTEROL_TOTAL = Si (%)  18 ( 8.6)
## TRIGLIC3RIDOS (mean (SD))  166.82 (78.89)
## TRIGLICERIDOS_COD = Si (%) 107 (51.2)
```

Como observamos en la tabla **tab** no muestra todos los niveles de las variables categ3ricas por lo tanto debemos de ocupar un comando que nos muestre todos esos niveles, el comando a utilizar es **print()**

```
print(tab,showAllLevels=T)
```

```
##
##              level              Overall
## n              210
## EDAD (mean (SD)) 47.66 (11.92)
## GENERO (%)       Femenino    110 (52.4)
##                  Masculino    100 (47.6)
## ESCUELA (%)      Administrativo 112 (53.3)
##                  Escuela de Medicina 45 (21.4)
##                  ETM          53 (25.2)
## HTA (%)          No          177 (84.7)
##                  Si          32 (15.3)
## PESO (mean (SD)) 75.84 (14.87)
## TALLA (mean (SD)) 1.60 (0.08)
## calcimc (mean (SD)) 29.49 (4.79)
## IMC_REAL (%)     Normal      38 (18.1)
##                  Obesidad I    66 (31.4)
##                  Obesidad II   16 ( 7.6)
##                  Obesidad M3rbida 7 ( 3.3)
##                  Sobrepeso     83 (39.5)
## GLUCOSA (mean (SD)) 82.57 (24.05)
## GLICEMIA (%)       No          195 (93.3)
##                  Si          14 ( 6.7)
## COLESTEROL (mean (SD)) 137.37 (38.41)
## COLESTEROL_TOTAL (%) No          191 (91.4)
##                  Si          18 ( 8.6)
## TRIGLIC3RIDOS (mean (SD)) 166.82 (78.89)
## TRIGLICERIDOS_COD (%) No          102 (48.8)
```

Si 107 (51.2)

Si necesita información más detallada, incluido el número / proporción que falta. Utilice el método **summary()** en el objeto de resultado. Las variables continuas se muestran primero y las variables categóricas se muestran en segundo lugar.

summary(tab)

```
##
##      ### Summary of continuous variables ###
##
## strata: Overall
##      n miss p.miss mean      sd median p25 p75 min max skew kurt
## EDAD      210    0    0.0  48 11.92    48  38  56  22  77 0.09 -0.6
## PESO      210    0    0.0  76 14.87    74  66  86  48 123 0.60  0.2
## TALLA      210    0    0.0   2  0.08     2   2   2   1   2 0.09 -0.9
## calcimc    210    0    0.0  29  4.79    29  26  32  20  51 0.86  1.8
## GLUCOSA    210    1    0.5  83 24.05    78  71  86  54 254 4.06 21.3
## COLESTEROL 210    1    0.5 137 38.41   134 111 153  67 275 1.05  1.1
## TRIGLICÉRIDOS 210    1    0.5 167 78.89   152 113 199  29 479 1.28  2.0
##
## =====
##
##      ### Summary of categorical variables ###
##
## strata: Overall
##      var      n miss p.miss      level freq percent cum.percent
##      GENERO 210    0    0.0      Femenino 110    52.4    52.4
##      Masculino 100    47.6    100.0
##
##      ESCUELA 210    0    0.0      Administrativo 112    53.3    53.3
##      Escuela de Medicina 45    21.4    74.8
##      ETM 53    25.2    100.0
##
##      HTA 210    1    0.5      No 177    84.7    84.7
##      Si 32    15.3    100.0
##
##      IMC_REAL 210    0    0.0      Normal 38    18.1    18.1
##      Obesidad I 66    31.4    49.5
##      Obesidad II 16    7.6    57.1
##      Obesidad Mórbida 7    3.3    60.5
##      Sobrepeso 83    39.5    100.0
##
##      GLICEMIA 210    1    0.5      No 195    93.3    93.3
##      Si 14    6.7    100.0
##
##      COLESTEROL_TOTAL 210    1    0.5      No 191    91.4    91.4
##      Si 18    8.6    100.0
##
##      TRIGLICERIDOS_COD 210    1    0.5      No 102    48.8    48.8
##      Si 107    51.2    100.0
##
```

Ahora bien, si lo que deseamos es comparar todo este conjunto de datos con una variable dependiente, utilizamos el comando **strata**

```
tab1=CreateTableOne(data = data1,strata = "HTA")
tab1
```

	Stratified by HTA			
	No	Si	p	test
n	177	32		
EDAD (mean (SD))	47.00 (11.56)	51.09 (13.53)	0.074	
GENERO = Masculino (%)	77 (43.5)	23 (71.9)	0.006	
ESCUELA (%)			0.305	
Administrativo	90 (50.8)	21 (65.6)		
Escuela de Medicina	40 (22.6)	5 (15.6)		
ETM	47 (26.6)	6 (18.8)		
HTA = Si (%)	0 (0.0)	32 (100.0)	<0.001	
PESO (mean (SD))	74.34 (13.69)	84.19 (18.52)	0.001	
TALLA (mean (SD))	1.60 (0.08)	1.63 (0.09)	0.044	
calcimc (mean (SD))	29.09 (4.49)	31.58 (5.90)	0.007	
IMC_REAL (%)			0.072	
Normal	34 (19.2)	4 (12.5)		
Obesidad I	55 (31.1)	10 (31.2)		
Obesidad II	11 (6.2)	5 (15.6)		
Obesidad Mórbida	4 (2.3)	3 (9.4)		
Sobrepeso	73 (41.2)	10 (31.2)		
GLUCOSA (mean (SD))	82.34 (23.90)	83.42 (25.51)	0.816	
GLICEMIA = Si (%)	13 (7.4)	1 (3.1)	0.616	
COLESTEROL (mean (SD))	136.09 (38.52)	143.46 (38.03)	0.319	
COLESTEROL_TOTAL = Si (%)	15 (8.5)	3 (9.4)	1.000	
TRIGLICÉRIDOS (mean (SD))	166.71 (79.67)	169.09 (76.31)	0.876	
TRIGLICERIDOS_COD = Si (%)	92 (52.3)	15 (46.9)	0.712	

```
print(tab1,showAllLevels=T,quote = TRUE,formatOptions = list(big.mark = ","))
```

	"Stratified by HTA"	
	"level"	"No"
"n"	" "	" 177"
"EDAD (mean (SD))"	" "	" 47.00 (11.56)"
"GENERO (%)"	"Femenino"	" 100 (56.5) "
" "	"Masculino"	" 77 (43.5) "
"ESCUELA (%)"	"Administrativo"	" 90 (50.8) "
" "	"Escuela de Medicina"	" 40 (22.6) "
" "	"ETM"	" 47 (26.6) "
"HTA (%)"	"No"	" 177 (100.0) "
" "	"Si"	" 0 (0.0) "
"PESO (mean (SD))"	" "	" 74.34 (13.69)"
"TALLA (mean (SD))"	" "	" 1.60 (0.08)"
"calcimc (mean (SD))"	" "	" 29.09 (4.49)"
"IMC_REAL (%)"	"Normal"	" 34 (19.2) "
" "	"Obesidad I"	" 55 (31.1) "
" "	"Obesidad II"	" 11 (6.2) "
" "	"Obesidad Mórbida"	" 4 (2.3) "
" "	"Sobrepeso"	" 73 (41.2) "
"GLUCOSA (mean (SD))"	" "	" 82.34 (23.90)"
"GLICEMIA (%)"	"No"	" 163 (92.6) "
" "	"Si"	" 13 (7.4) "
"COLESTEROL (mean (SD))"	" "	"136.09 (38.52)"
"COLESTEROL_TOTAL (%)"	"No"	" 161 (91.5) "


```
##      ""                "Si"                " 15 ( 8.5) "
##      "TRIGLICÉRIDOS (mean (SD))" ""                "166.71 (79.67)"
##      "TRIGLICERIDOS_COD (%)" "No"                " 84 ( 47.7) "
##      ""                "Si"                " 92 ( 52.3) "
##      "Stratified by HTA"
##      ""                "Si"                "p"                "test"
##      "n"                " 32"                ""                ""
##      "EDAD (mean (SD))" " 51.09 (13.53)" " 0.074" ""
##      "GENERO (%)" " 9 ( 28.1) " " 0.006" ""
##      "" " 23 ( 71.9) " "" ""
##      "ESCUELA (%)" " 21 ( 65.6) " " 0.305" ""
##      "" " 5 ( 15.6) " "" ""
##      "" " 6 ( 18.8) " "" ""
##      "HTA (%)" " 0 ( 0.0) " "<0.001" ""
##      "" " 32 (100.0) " "" ""
##      "PESO (mean (SD))" " 84.19 (18.52)" " 0.001" ""
##      "TALLA (mean (SD))" " 1.63 (0.09)" " 0.044" ""
##      "calcimc (mean (SD))" " 31.58 (5.90)" " 0.007" ""
##      "IMC_REAL (%)" " 4 ( 12.5) " " 0.072" ""
##      "" " 10 ( 31.2) " "" ""
##      "" " 5 ( 15.6) " "" ""
##      "" " 3 ( 9.4) " "" ""
##      "" " 10 ( 31.2) " "" ""
##      "GLUCOSA (mean (SD))" " 83.42 (25.51)" " 0.816" ""
##      "GLICEMIA (%)" " 31 ( 96.9) " " 0.616" ""
##      "" " 1 ( 3.1) " "" ""
##      "COLESTEROL (mean (SD))" "143.46 (38.03)" " 0.319" ""
##      "COLESTEROL_TOTAL (%)" " 29 ( 90.6) " " 1.000" ""
##      "" " 3 ( 9.4) " "" ""
##      "TRIGLICÉRIDOS (mean (SD))" "169.09 (76.31)" " 0.876" ""
##      "TRIGLICERIDOS_COD (%)" " 17 ( 53.1) " " 0.712" ""
##      "" " 15 ( 46.9) " "" ""
```

Es posible que desee ver solo las variables categóricas o continuas. Puede hacer esto accediendo a la parte **CatTable** y la parte **ContTable** del objeto TableOne de la siguiente manera. Los métodos summary () se definen para ambos, así como el método print () con varios argumento

```
#Solamente la parte categórica
```

```
print(tab1$CatTable)
```

```
##      Stratified by HTA
##      No      Si      p      test
##      n      177     32
##      GENERO = Masculino (%) 77 (43.5) 23 ( 71.9) 0.006
##      ESCUELA (%)          0.305
##      Administrativo 90 (50.8) 21 ( 65.6)
##      Escuela de Medicina 40 (22.6) 5 ( 15.6)
##      ETM 47 (26.6) 6 ( 18.8)
##      HTA = Si (%) 0 ( 0.0) 32 (100.0) <0.001
##      IMC_REAL (%)          0.072
##      Normal 34 (19.2) 4 ( 12.5)
##      Obesidad I 55 (31.1) 10 ( 31.2)
##      Obesidad II 11 ( 6.2) 5 ( 15.6)
##      Obesidad Mórbida 4 ( 2.3) 3 ( 9.4)
##      Sobrepeso 73 (41.2) 10 ( 31.2)
```

```
## GLICEMIA = Si (%)      13 ( 7.4)   1 ( 3.1)   0.616
## COLESTEROL_TOTAL = Si (%) 15 ( 8.5)   3 ( 9.4)   1.000
## TRIGLICERIDOS_COD = Si (%) 92 (52.3)  15 ( 46.9)  0.712
```

#Solamente la parte continua

```
print(tab1$ContTable)
```

```
##                               Stratified by HTA
##                               No           Si           p       test
##  n                               177          32
##  EDAD (mean (SD))             47.00 (11.56)  51.09 (13.53)  0.074
##  PESO (mean (SD))             74.34 (13.69)  84.19 (18.52)  0.001
##  TALLA (mean (SD))             1.60 (0.08)   1.63 (0.09)   0.044
##  calcimc (mean (SD))          29.09 (4.49)   31.58 (5.90)  0.007
##  GLUCOSA (mean (SD))          82.34 (23.90)  83.42 (25.51)  0.816
##  COLESTEROL (mean (SD))       136.09 (38.52) 143.46 (38.03)  0.319
##  TRIGLICÉRIDOS (mean (SD))    166.71 (79.67) 169.09 (76.31)  0.876
```