

Laboratorio 3

Periodismo de Datos, MPPP.
Febrero, 2020

M.C. JORGE JUVENAL CAMPOS FERREIRA.

Asistente de investigación.

Laboratorio Nacional de Políticas Públicas

CIDE

Preguntas:

1. Qué **han hecho estos días para esta materia?**

Preguntas:

2. Qué planean hacer esta semana para esta materia?

Preguntas:

3. Qué impedimentos se les han presentado durante la semana?



Laboratorio 3.

Contenido:

En este laboratorio abordaremos los siguientes temas:

1. Las librerías `sjmisc` y `sjlabelled`.
2. ¿Qué significa usar labels?

Referencias:

- <https://strengjacke.github.io/sjmisc/articles/exploringdatasets.html>



1. Las librerías `sjmisc` y `sjlabelled`.

Las librerías `sjmisc` y `sjlabelled` son librerías que nos permiten realizar diversas funciones con bases de datos. Se usan, especialmente, para trabajar con datos de encuestas.

El paquete `sjmisc` presenta una ventaja comparativa al poder trabajar bien con datos etiquetados, debido a que ofrece opciones para transformación de datos con estas características.

Para el ejemplo de hoy, utilizaremos los datos de la encuesta `efc`, la cual contiene información sobre la situación de los cuidadores de adultos mayores en Europa (proyecto EUROFAMCARE).

Datos labelados

```
# Librerias  
library(sjmisc)  
library(dplyr)  
data(efc)  
  
# Exploramos la base de datos  
head(efc)
```

- **¿Como se ve la base de datos?**
- **¿Qué tiene de diferente con las que hemos visto hasta ahorita?**

Motivación del etiquetado



<p>1.</p> <p>NOMBRE DE LA CIUDAD), es...</p> <p>CIRCULA UN SOLO CÓDIGO</p> <p>seguro? 1</p> <p>inseguro? 2</p> <p>No sabe o no responde 9</p>	<p>2. Pensando en las condiciones de delincuencia en (NOMBRE DE LA CIUDAD), ¿considera que en los próximos 12 meses...</p> <p>CIRCULA UN SOLO CÓDIGO</p> <p>mejorará? 1</p> <p>seguirá igual de bien? 2</p> <p>seguirá igual de mal? 3</p> <p>empeorará? 4</p> <p>No sabe o no responde 9</p>
<p>3. En los últimos tres meses, ¿ha escuchado o ha visto en los alrededores de su vivienda situaciones como...</p> <p>REGISTRA EN CADA CASO EL CÓDIGO CORRESPONDIENTE</p> <p>Sí 1</p> <p>No 2</p> <p>No sabe o no responde 9</p> <p>1 vandalismo (grafitis, pintas, vidrios quebrados o daños) en las viviendas o negocios? <input type="checkbox"/></p> <p>2 consumo de alcohol en las calles? <input type="checkbox"/></p> <p>3 robos o asaltos? <input type="checkbox"/></p> <p>4 bandas violentas o pandillerismo? <input type="checkbox"/></p> <p>5 venta o consumo de droga? <input type="checkbox"/></p>	<p>4. En este mismo periodo de tres meses, por temor a sufrir algún delito (robo, asalto, secuestro, entre otros), ¿usted cambió sus hábitos respecto a...</p> <p>REGISTRA EN CADA CASO EL CÓDIGO CORRESPONDIENTE</p> <p>Sí 1</p> <p>No 2</p> <p>No aplica 3</p> <p>No sabe o no responde 9</p> <p>1 llevar cosas de valor como joyas, dinero o tarjetas de crédito? <input type="checkbox"/></p> <p>2 caminar por los alrededores de su vivienda, pasadas las ocho de la noche? <input type="checkbox"/></p> <p>3 visitar a parientes o amigos? <input type="checkbox"/></p> <p>4 permitir que salgan de su vivienda sus hijos menores? <input type="checkbox"/></p>

Fuente: Cuestionario de la ENSU (Encuesta Nacional de Seguridad Pública Urbana) INEGI, 2015

Ver y tocar las etiquetas



```
# Obteniendo las etiquetas:  
library(sjlabelled)
```

```
# Funcion para obtener las etiquetas de una variable  
sjlabelled::get_labels(efc$e42dep)
```

```
## [1] "independent"          "slightly dependent"    "moderately dependent"  
## [4] "severely dependent"
```

```
# Funcion para obtener las etiquetas de todas las variables del data.frame  
sjlabelled::get_label(efc)
```

```
##                                c12hour  
##                                "average number of hours of care per week"  
##                                e15relat  
##                                "relationship to elder"  
##                                e16sex  
##                                "elder's gender"
```

Funciones

Imprimir frecuencias con etiquetas.

Qué contiene esta base de datos?

Lo primero que nos podría interesar es probablemente la distribución del sexo. Podemos checar las frecuencias de los datos etiquetados con la función `frq`. Esta función requiere un vector o data.frame como entrada e imprime la etiqueta de la variable en la primera línea, seguida de una tabla de frecuencia con valores, etiquetas, conteos y porcentajes de los datos del vector.

```
# Imprimimos la variable de sexo?  
frq(efc$c16lsex)
```

Funciones



```
# Imprimimos la variable de sexo?  
frq(efc$cl6lsex)
```

```
##  
## carer's gender (x) <numeric>  
## # total N=908  valid N=901  mean=1.76  sd=0.43  
##  
##   val  label frq raw.prc valid.prc cum.prc  
##   1   Male 215  23.68    23.86    23.86  
##   2 Female 686  75.55    76.14   100.00  
##  NA  <NA>   7   0.77      NA      NA
```

Las mujeres son el 75.55 % de la muestra.

Encontrando variables en un data.frame

Para encontrar variables en un data.frame etiquetado, utilizamos la función `find_var()`:

```
# Esta función nos encuentra la columna con la palabra especificada  
find_var(efc, "dependency", out = "table")
```

```
##   col.nr var.name          var.label  
## 1      5   e42dep elder's dependency
```

```
# Esta función nos da la columna con la palabra especificada  
find_var(efc, "dependency", out = "df") %>% head()
```

Tablas cruzadas.

El paquete nos permite cruzar dos variables para ver la relación que hay entre ellas. Por ejemplo, a partir de cruzar la variable de sexo con la de dependencia del cuidado respecto al cuidador.

```
# Tabla que muestra la distribución de personas con cuidados con respecto al sexo del  
flat_table(efc, e42dep, c161sex)
```

```
##           c161sex Male Female  
## e42dep  
## independent           18      48  
## slightly dependent     54     170  
## moderately dependent   80     226  
## severely dependent     63     241
```

Tablas cruzadas.

El paquete nos permite cruzar dos variables para ver la relación que hay entre ellas. Por ejemplo, a partir de cruzar la variable de sexo con la de dependencia del cuidado respecto al cuidador.

```
# Proporciones, en función del sexo,  
flat_table(efc, e42dep, c161sex, margin = "col")
```

```
##           c161sex  Male Female  
## e42dep  
## independent      8.37    7.01  
## slightly dependent 25.12   24.82  
## moderately dependent 37.21   32.99  
## severely dependent 29.30   35.18
```

Recodificando variables.

Vamos a generar una nueva variable llamada “carga” a partir de la variable `neg_c_7` (Impacto negativo con 7 items).

Para esto, vamos a utilizar la función de recodificación, `rec`, para asignar etiquetas a los valores numéricos de una variable.

```
# generamos la nueva variable, con rec
efc$carga = rec(efc$neg_c_7,
  rec = c("min:9=1 [baja]; 10:12=2[moderada]; 13:max=3[alta]; else=NA"),
  var.label = "Carga subjetiva",
  as.num = FALSE # Para que nos de un factor
)
```

Formato: intervalo_1=valor_numerico [categoria];

Comprobación



```
# Obtenemos etiquetas  
get_labels(efc$carga)
```

```
## [1] "baja"      "moderada" "alta"
```

```
get_values(efc$carga)
```

```
## [1] 1 2 3
```

```
# Obtenemos frecuencias a partir de las etiquetas  
frq(efc$carga)
```

```
##  
## Carga subjetiva (x) <category>  
## # total N=908  valid N=892  mean=2.03  sd=0.81  
##  
##   val   label frq raw.prc valid.prc cum.prc  
##   1     baja 280  30.84    31.39   31.39  
##   2  moderada 301  33.15    33.74   65.13  
##   3     alta 311  34.25    34.87  100.00  
##  NA      <NA>  16   1.76         NA    NA
```