

# Análisis de encuestas de hogares con R

## Modulo 3: Procesamiento de múltiples encuestas de hogares en R

Andrés Gutiérrez, Ph.D.  
Stalyn Guerrero M.Sc.

CEPAL - Unidad de Estadísticas Sociales

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

1 Variables continuas

2 Variables categóricas

# Lecturas de múltiples archivos

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En algunas ocasiones es necesario procesar más de un archivo para obtener algunos resultados de interés, y contamos con la fortuna que los activos poseen la misma estructura (extensión, nombre de columnas). Por ejemplo, cuando tenemos el censo del país dividido por región o departamento otro caso puede ser las encuestas tomadas en diferentes periodos. En estos casos podemos realizar un procesamiento por medio de loop o alternativas como el proceso por mapeo de listas o procesos en paralelo.

# Identificando las rutas de los archivos

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para realizar la lectura de múltiples bases debemos conocer las rutas donde estas están guardadas, lo que podemos simplificar con la función `file.list` del paquete `base`, que nos permite tener un listado completo de los archivos.

```
(data_path <- list.files("Z:/BC/",  
  full.names = TRUE,  
  pattern = "2020"  
)[1:5])
```

# Identificando las rutas de los archivos

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
## [1] "Z:/BC/ARG_2020N.dta" "Z:/BC/BOL_2020N.dta" "Z:/BC/BRA_2020N1.dta"
## [4] "Z:/BC/CHL_2020N.dta" "Z:/BC/COL_2020N1.dta"
```

Identificando los países.

Para poder identificar que archivo estamos procesando se crea una columna adicional con el nombre del país.

```
data_path <- tibble(path = data_path) %>%  
  mutate(pais = gsub("Z:\\\\BC\\\\(.*)_.*", "\\1",  
    x = path  
  ))
```

# Identificando los países.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

path	pais
Z:/BC/ARG_2020N.dta	ARG
Z:/BC/BOL_2020N.dta	BOL
Z:/BC/BRA_2020N1.dta	BRA
Z:/BC/CHL_2020N.dta	CHL
Z:/BC/COL_2020N1.dta	COL

# Función para operar listas

## Función lapply y sapply

Estas funciones están disponible en el paquete base de R.  
Sintaxis,

```
## Cinco muestras aleatorias con n = 10  
lapply(rep(10, 5), rnorm )  
sapply(rep(10, 5), rnorm , simplify = TRUE)
```

## Familia de funciones map

La familia map amplio las posibilidades para trabajar con los elementos de las listas. map hace parte del paquete purrr, el cual cuenta con funciones adicionales como:

- map permite ingresar un listas que será mapeada elemento a elemento, por ejemplo.

```
map(.x = rep(10,5), rnorm)
```

# Familia de funciones map

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

- `map2` permite ingresar dos listas que interactúen entre sí para obtener un resultado, por ejemplo.

```
map2(.x = rep(10,5),  
     .y = 1:5,  
     ~rnorm(n=.x, mean = .y))
```



# Familia de funciones map

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

- `map2` permite ingresar dos listas que interactúen entre sí para obtener un resultado, por ejemplo.

```
map2(.x = rep(10,5),  
     .y = 1:5,  
     ~rnorm(n=.x, mean = .y))
```

- `pmap` : esta permiten operar con listas de parámetro. Por ejemplo:

```
pmap(list(  
  n = rep(10, 5),  
  mean = 1:5,  
  sd = 1 / 1:5  
),  
rnorm)
```

# Función para operar listas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

## Familia de funciones `future_map`

La familia `future_map` amplió las posibilidades de realizar procesamientos en paralelo, como podemos notar por el nombre, tiene la misma lógica de uso que la familia `map`. Estas funciones las podemos encontrar en el paquete `furrr`.

# Función para operar listas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

## Familia de funciones `future_map`

La familia `future_map` amplio las posibilidades de realizar procesamientos en paralelo, como podemos notar por el nombre, tiene la misma lógica de uso que la familia `map`. Estas funciones las podemos encontrar en el paquete `furrr`.

La familia de funciones `map` y `future_map` tiene una extensión más, la cual permiten organizar los resultados teniendo en cuenta el tipo de resultados que se este reportando, por ejemplo:

- `_dbl` unifica salidas tipo numérica en un vector.

# Función para operar listas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

## Familia de funciones `future_map`

La familia `future_map` amplio las posibilidades de realizar procesamientos en paralelo, como podemos notar por el nombre, tiene la misma lógica de uso que la familia `map`. Estas funciones las podemos encontrar en el paquete `furrr`.

La familia de funciones `map` y `future_map` tiene una extensión más, la cual permiten organizar los resultados teniendo en cuenta el tipo de resultados que se este reportando, por ejemplo:

- `_dbl` unifica salidas tipo numérica en un vector.
- `_df` unifica `data.frame`, equivale hacer `rbind`.

# Lectura de encuestas con map

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para la lectura de los archivos, se procede de la siguiente forma.

```
require(purrr)
require(haven)
data_path %<>%
  mutate(encuesta = path %>% map(~ read_dta(.x) %>%
    transmute(
      upm = `_upm`,
      estrato = `_estrato`,
      sexo, areageo2, lp, li, ingcorte,
      fep = `_fep`
    )))
```

Las variables seleccionadas son sexo, área geográfica (areageo2), Línea de pobreza (lp), Línea de indigencia (li), Ingreso por persona (ingcorte) y factor de expansión por persona (fep)

# Lectura de encuestas (resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

	path <chr>	pais <chr>	encuesta <list>
1	Z:/BC/ARG_2020N.dta	ARG	<tibble [43,767 x 6]>
2	Z:/BC/BOL_2020N.dta	BOL	<tibble [37,092 x 6]>
3	Z:/BC/BRA_2020N1.dta	BRA	<tibble [355,436 x 6]>
4	Z:/BC/CHL_2020N.dta	CHL	<tibble [185,437 x 6]>
5	Z:/BC/COL_2020N1.dta	COL	<tibble [747,822 x 6]>
6	Z:/BC/CRI_2020N1.dta	CRI	<tibble [25,530 x 6]>
7	Z:/BC/DOM_2020N1.dta	DOM	<tibble [71,378 x 6]>
8	Z:/BC/ECU_2020N.dta	ECU	<tibble [30,646 x 6]>
9	Z:/BC/MEX_2020N1.dta	MEX	<tibble [315,743 x 6]>
10	Z:/BC/PER_2020N.dta	PER	<tibble [120,346 x 6]>
11	Z:/BC/PRY_2020N.dta	PRY	<tibble [17,582 x 6]>
12	Z:/BC/SLV_2020N.dta	SLV	<tibble [37,030 x 6]>
13	Z:/BC/URY_2020N.dta	URY	<tibble [145,166 x 6]>

El resultado es objeto tipo tibble el cual permite observar de forma compacta el contenido de una lista e indica el tipo y tamaño de cada objeto en la contenido en la lista.

# Lectura de encuestas future\_map

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
require(furrr)
require(haven)
plan(multisession, workers = 2)
data_path %<>%
  mutate(encuesta = path %>%
    future_map(~ read_dta(.x) %>%
      transmute(
        upm = `_upm`,
        estrato = `_estrato`,
        sexo, areageo2, lp, li, ingcorte,
        fep = `_fep`
      )))
```

# Definir el diseño

Después de tener las bases disponibles en R en momento de definir un diseño muestral para cada encuesta, para nuestro ejemplo se define el siguiente:

```
options(survey.lonely.psu = "adjust")
plan(multisession, workers = 2)
data_path %<>% mutate(
  disenno = encuesta %>%
    future_map(~ as_survey_design(
      .data = .x,
      ids = upm,
      strata = estrato,
      weights = fep,
      nest = T
    ))
)
```



# Definir el diseño (resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
# A tibble: 13 x 4
```

	path <chr>	pais <chr>	encuesta <list>	diseño <list>
1	Z:/BC/ARG_2020N.dta	ARG	<tibble [43,767 x 6]>	<tbl_svy[,6]>
2	Z:/BC/BOL_2020N.dta	BOL	<tibble [37,092 x 6]>	<tbl_svy[,6]>
3	Z:/BC/BRA_2020N1.dta	BRA	<tibble [355,436 x 6]>	<tbl_svy[,6]>
4	Z:/BC/CHL_2020N.dta	CHL	<tibble [185,437 x 6]>	<tbl_svy[,6]>
5	Z:/BC/COL_2020N1.dta	COL	<tibble [747,822 x 6]>	<tbl_svy[,6]>
6	Z:/BC/CRI_2020N1.dta	CRI	<tibble [25,530 x 6]>	<tbl_svy[,6]>
7	Z:/BC/DOM_2020N1.dta	DOM	<tibble [71,378 x 6]>	<tbl_svy[,6]>
8	Z:/BC/ECU_2020N.dta	ECU	<tibble [30,646 x 6]>	<tbl_svy[,6]>
9	Z:/BC/MEX_2020N1.dta	MEX	<tibble [315,743 x 6]>	<tbl_svy[,6]>
10	Z:/BC/PER_2020N.dta	PER	<tibble [120,346 x 6]>	<tbl_svy[,6]>
11	Z:/BC/PRY_2020N.dta	PRY	<tibble [17,582 x 6]>	<tbl_svy[,6]>
12	Z:/BC/SLV_2020N.dta	SLV	<tibble [37,030 x 6]>	<tbl_svy[,6]>
13	Z:/BC/URY_2020N.dta	URY	<tibble [145,166 x 6]>	<tbl_svy[,6]>

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

# Variables continuas

# Estimación de promedios en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
library(tidyr)
data_path %>% mutate(ingreso_medios = diseno %>%
  map(~.x %>%
    summarise(
      ingreso = survey_mean(ingcorte/lp)))) %>%
  select(pais, ingreso_medios) %>%
  unnest(ingreso_medios)
```

# Estimación de promedios en multiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	ingreso	ingreso_se
ARG	1.749	0.0000
BOL	1.943	0.0382
BRA	3.397	0.0465
CHL	3.267	0.0396
COL	2.107	0.0000

# Estimación de promedios por sexo en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
data_path %>% mutate(  
  ingreso_medios = disenno %>%  
    map( ~ .x %>% group_by(sexo) %>%  
      summarise(ingreso = survey_mean(ingcorte / lp)))  
  select(pais, ingreso_medios) %>%  
  unnest(ingreso_medios) %>%  
  head(14)
```

# Estimación de promedios por sexo en multiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	sexo	ingreso	ingreso_se
ARG	1	1.750	0.0000
ARG	2	1.749	0.0000
BOL	1	2.014	0.0403
BOL	2	1.873	0.0397
BRA	1	3.481	0.0510
BRA	2	3.316	0.0454
CHL	1	3.414	0.0456
CHL	2	3.140	0.0401
COL	1	2.132	0.0000
COL	2	2.083	0.0000

# Estimación de promedios por área en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
data_path %>% mutate(ingreso_medios = diseno %>%  
  map(~.x %>% group_by(areageo2) %>%  
    summarise(  
      ingreso = survey_mean(ingcorte/lp)))) %>%  
  select(pais, ingreso_medios) %>%  
  unnest(ingreso_medios) %>%  
  head(14)
```

# Estimación de promedios por área en multiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	areageo2	ingreso	ingreso_se
ARG	1	1.749	0.0000
BOL	1	2.137	0.0425
BOL	2	1.486	0.0828
BRA	1	3.567	0.0539
BRA	2	2.361	0.0315
CHL	1	3.281	0.0427
CHL	2	3.154	0.1026
COL	1	2.267	0.0000
COL	2	1.555	0.0000



# Pruebas de diferencia de medias por grupo en múltiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Diferencia de ingresos promedio por sexo.

```
data_path[-1,] %>% mutate(pvalor = map_dbl(disenos,
  ~svyttest(formula = ingcorte/lp~sexo,
             .x, na = TRUE)$p.value
) ) %>% select(pais, pvalor)
```

Nota: ARG se quita porque las upm y los estratos no están debidamente identificados, tienen valores contrantes.

# Pruebas de diferencia de medias por grupo en múltiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	pvalor
BOL	0
BRA	0
CHL	0
COL	NaN

# Estimación del índice de GINI en múltiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
library(convey)
data_path %>% mutate(
  diseno_gini = map(diseno, convey_prep),
  gini = map(diseno_gini,
    ~svygini( ~ingcorte, design = .x) %>%
      data.frame()
  )) %>%
  select(pais, gini) %>% unnest(gini)
```

# Estimación del índice de GINI en múltiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	gini	ingcorte
ARG	0.3999	0.0000
BOL	0.4510	0.0065
BRA	0.5202	0.0048
CHL	0.4756	0.0040
COL	0.5540	0.0000

# Estimación de la curva de lorenz en múltiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
par(mfrow = c(2,2))
temp <- data_path[1:3,] %>% mutate(
  diseno_gini = map(diseno, convey_prep),
  lorenz = map2(diseno_gini, pais,
    function(diseno_i, pais_i){
      svylorenz( ~ingcorate, diseno_i,
        seq(0,1,.05),
        alpha = .01, plot = TRUE ,
        add = FALSE)
      title(pais_i) })
)
```

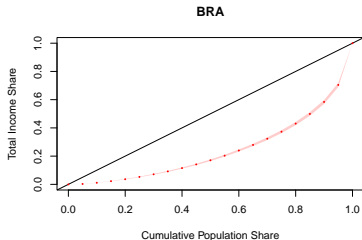
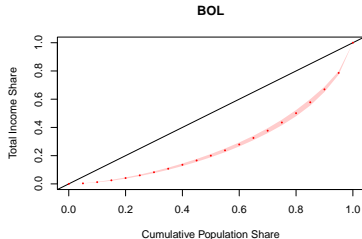
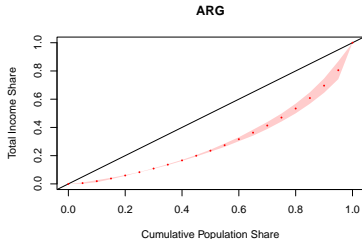
# Estimación de la curva de lorenz en múltiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas



# Procesando múltiples variables en la encuesta.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En estas ocasiones es de interés procesar obtener resultados para múltiples preguntas que tienen características similares, para este ejemplo tomaremos las variables, `sys_pe` , `gan_pe` , `yemp_pe` , `yjub_pe`, `yotr_pe` y `yto_pe`. Para esto procedemos así:

```
name_var <- c("sys_pe" , "gan_pe" , "yemp_pe" ,  
              "yjub_pe", "yotr_pe", "yto_pe")
```

# Leer encuesta y definir diseño muestral

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
# Leer encuesta
encuesta_BOL2020 <- data_path$path[2] %>%
  read_dta(.) %>%
  mutate(
    upm = `_upm`,
    estrato = `_estrato`,
    fep = `_fep`)
# definir disenno
diseno = as_survey_design(
  .data = encuesta_BOL2020,
  ids = upm,
  strata = estrato,
  weights = fep, nest = T)
```



# Estimando los resultados.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
resul_mult <- disenno %>%  
  filter(edad >= 15, sexo == 2) %>%  
  mutate_at(.vars = all_of(name_var),  
            ~ ifelse(. > 0, 1, 0)) %>%  
  summarise_at(.vars = all_of(name_var),  
               ~survey_mean(.))
```

variable	estimado	se
gan_pe	0.2408	0.0056
sys_pe	0.1566	0.0050
yemp_pe	0.3921	0.0059
yjub_pe	0.0359	0.0026
yotr_pe	0.7073	0.0057
yto_pe	0.8820	0.0038

# Estimando los resultados, condicionada.

En esta rutina como obtener estimaciones cuando tenemos que hacer filtro para cada variable

```
resul_mult_cond <- map_dfc(  
  name_var,  
  ~diseno %>%  
  filter(edad >= 15, sexo == 2) %>%  
  filter_at(all_of(.x), ~ . > 0) %>%  
  summarise_at(all_of(.x), ~survey_mean(.))  
)
```

variable	estimado	se
gan_pe	2054.7	85.488
sys_pe	3413.1	70.230
yemp_pe	2626.4	63.870
yjub_pe	2691.3	95.536

# Estimando los resultados por grupo.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
resul_mult_grupo <- diseno %>%  
  filter(edad >= 15, sexo == 2) %>%  
  mutate_at(.vars = all_of(name_var),  
            ~ ifelse(. > 0, 1, 0)) %>%  
  group_by(areageo2) %>%  
  summarise_at(.vars = vars(name_var),  
              ~survey_mean(.))
```

# Estimando los resultados por grupo (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

areageo2	variable	estimado	se
1	gan_pe	0.2261	0.0053
1	sys_pe	0.2051	0.0053
1	yemp_pe	0.4249	0.0060
1	yjub_pe	0.0450	0.0030
1	yotr_pe	0.6835	0.0054
1	yto_pe	0.8935	0.0036
2	gan_pe	0.2788	0.0137
2	sys_pe	0.0314	0.0051
2	yemp_pe	0.3075	0.0140
2	yjub_pe	0.0124	0.0051
2	yotr_pe	0.7686	0.0132
2	yto_pe	0.8524	0.0106

# Estimando de resultados en múltiples variables en múltiples archivos

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
data_path %<>%  
  mutate(  
    encuesta = path %>% map(  
      # Lectura de archivo  
      ~ read_dta(.x) %>%  
        mutate(upm = `_upm`,  
                estrato = `_estrato`, fep = `_fep`)),  
    # Definicion del diseño  
    diseno = encuesta %>% map(  
      ~ as_survey_design(  
        .data = .x,  
        ids = upm,  
        strata = estrato,  
        weights = fep,    nest = T  
      )  
    )  
  )
```

# Procesando los multiples archivos en múltiples variables.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
multi_archivo <- data_path %>% transmute(pais,  
  promedio = diseno %>%  
    map(~.x %>% filter(edad >= 15, sexo == 2) %>%  
      mutate_at(.vars = all_of(name_var),  
                ~ ifelse(. > 0, 1, 0)) %>%  
      summarise_at(.vars = all_of(name_var),  
                    ~survey_mean(.)))) %>%  
  unnest(cols = "promedio")
```

# Procesando los múltiples archivos en múltiples variables (Resultado 1).

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	variable	estimado	se
ARG	gan_pe	0.1171	0.0000
ARG	sys_pe	0.3062	0.0000
ARG	yemp_pe	0.4280	0.0000
ARG	yjub_pe	0.2214	0.0000
ARG	yotr_pe	0.2097	0.0000
ARG	yto_pe	0.7241	0.0000
BOL	gan_pe	0.2408	0.0056
BOL	sys_pe	0.1566	0.0050
BOL	yemp_pe	0.3921	0.0059

# Procesando los múltiples archivos en múltiples variables (Resultado 2).

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

	pais	variable	estimado	se
10	BOL	yjub_pe	0.0359	0.0026
11	BOL	yotr_pe	0.7073	0.0057
12	BOL	yto_pe	0.8820	0.0038
13	BRA	gan_pe	0.1021	0.0012
14	BRA	sys_pe	0.3076	0.0021
15	BRA	yemp_pe	0.4106	0.0022
16	BRA	yjub_pe	0.1722	0.0016
17	BRA	yotr_pe	0.2274	0.0022
18	BRA	yto_pe	0.7296	0.0019
19	CHL	gan_pe	0.1009	0.0017
20	CHL	sys_pe	0.2727	0.0027
21	CHL	yemp_pe	0.4396	0.0032
22	CHL	yjub_pe	0.1454	0.0022
23	CHL	yotr_pe	0.4547	0.0031
24	CHI	yto_pe	0.7530	0.0031



# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.
- **Paso 3** Procesar información.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.
- **Paso 3** Procesar información.
- **Paso 4** Organizar y presentar resultados.

# Creando función para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
options(survey.lonely.psu="adjust")
Promedio_aux <- function(input_file){
  pais = gsub("Z:\\\\BC\\\\(.*)_.*", "\\1",
    x = input_file)
  ## Paso 1
  encuesta <- read_dta(input_file) %>%
    transmute(upm = `_upm`,
              estrato = `_estrato`,
              sexo, areageo2,lp,li,ingcorte,
              fep = `_fep`)
  ## Paso 2
  diseno <- as_survey_design(.data = encuesta,
                             ids = upm,
                             strata = estrato,
                             weights = fep,
                             nest = T
  )
  ## Paso 3
  diseno %>% summarise(
    ingreso = survey_mean(ingcorte/lp)) %>%
    data.frame() %>% mutate(pais = pais)
}
```

# Procesando encuestas múltiples

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para el *Paso 4* realizamos la siguiente sintaxis.

```
list.files("Z:/BC/",full.names = TRUE,  
          pattern = "2020")[1:3] %>%  
  map_df(~Promedio_aux(.x))
```

ingreso	ingreso_se	pais
1.749	0.0000	ARG
1.943	0.0382	BOL
3.397	0.0465	BRA

# Obteniendo algunos resultados simples

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Si el interés es obtener una medición simple de las bases, como lo puedes ser un promedio ponderado, es posible que utilicemos la función `weighted.mean` con la siguiente instrucción

```
plan(multisession, workers = 2)
data_path %>%
  transmute(pais,
             promedio = encuesta %>%
               future_map_dbl(
                 ~ .x %>% summarise(
                   media = weighted.mean(ingcorte/lp, fep)) %>%
                   as.numeric()))
```

Variables  
continuas

Variables  
categóricas

# Obteniendo alguno resultados simples

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	promedio
ARG	1.749
BOL	1.943
BRA	3.397
CHL	3.267
COL	2.107



# Obteniendo algunos resultados simples por grupo

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
plan(multisession, workers = 2)
data_path %>%
  transmute(pais,
             promedio = encuesta %>%
               future_map(
                 ~ .x %>% group_by(sexo) %>%
                   summarise(
                     media = weighted.mean(ingcorte/lp, fep))) %>%
               unnest(promedio) %>%
               spread(key = "sexo", value = "media")
```

# Obteniendo alguno resultados simples por grupo

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	1	2
ARG	1.750	1.749
BOL	2.014	1.873
BRA	3.481	3.316
CHL	3.414	3.140
COL	2.132	2.083

# Obteniendo estimaciones complejas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Si el interés es obtener un resultado un poco más complejo, por ejemplo una razón de dos totales, es necesario que definamos nuestra propia función

```
weighted.ratio <- function(num, den, w){  
  num = sum(num*w)  
  den = sum(den*w)  
  data.frame(num,den,ratio = num/den)  
}
```

```
data_path %>%  
  transmute(pais,  
    ratio = encuesta %>%  
      map(~ .x %>%  
        summarise(  
          ratio = weighted.ratio(sexo == 1, sexo == 2, fep))) %>%  
      unnest(ratio) %>% unnest(ratio)
```

# Obteniendo estimaciones complejas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	num	den	ratio
ARG	14032381	14707283	0.9541
BOL	5786921	5962117	0.9706
BRA	103205385	107891068	0.9566
CHL	9006641	10539158	0.8546
COL	24389491	25037775	0.9741

# Obteniendo estimaciones complejas con restricciones.

Al definir nuestra función es posible que necesitemos hacer algunas restricciones para el calculo. Por ejemplo, ingresos mayores a cero.

```
weighted_mean_cond <- function(x, w){  
  xi <- x[x>0]  
  wi <- w[x>0]  
  sum(xi*wi)/sum(wi)  
}
```

```
data_path %>%  
  transmute(pais,  
    promedio = encuesta %>%  
      map(~ .x %>%group_by(sexo) %>%  
        summarise(  
          media = weighted.mean(ingcorte/lp, fep),  
          media_cond = weighted_mean_cond(ingcorte/lp, fep))) %>%  
    unnest(promedio)
```

# Obteniendo estimaciones complejas con restricciones.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	sexo	media	media_cond
ARG	1	1.750	1.765
ARG	2	1.749	1.762
BOL	1	2.014	2.015
BOL	2	1.873	1.873
BRA	1	3.481	3.543
BRA	2	3.316	3.374
CHL	1	3.414	3.459
CHL	2	3.140	3.178
COL	1	2.132	2.179
COL	2	2.083	2.133

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

# Variables categóricas

# Lectura de múltiples bases

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para realizar la lectura de múltiples bases debemos conocer las rutas donde estas están guardadas para ello empleamos la función `file.list` del paquete `base`, que nos permite tener un listado completo de los archivos.

```
library(tibble)
(data_path <- list.files("Z:/BC/", full.names = TRUE,
                        pattern = "2020") %>% tibble
  mutate(pais = gsub("Z:\\\\BC\\\\/(.*)_.*", "\\1",
                    x = path)))
```

*Note* que utiliza la función `gsub` para separar el nombre del país de la ruta.



# Lectura de múltiples bases

path	pais
Z:/BC/ARG_2020N.dta	ARG
Z:/BC/BOL_2020N.dta	BOL
Z:/BC/BRA_2020N1.dta	BRA
Z:/BC/CHL_2020N.dta	CHL
Z:/BC/COL_2020N1.dta	COL
Z:/BC/CRI_2020N1.dta	CRI
Z:/BC/DOM_2020N1.dta	DOM
Z:/BC/ECU_2020N.dta	ECU
Z:/BC/MEX_2020N1.dta	MEX
Z:/BC/PER_2020N.dta	PER
Z:/BC/PRY_2020N.dta	PRY
Z:/BC/SLV_2020N.dta	SLV
Z:/BC/URY_2020N.dta	URY

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

# Lectura de encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para la lectura de los archivos, se procede de la siguiente forma.

```
require(purrr)
require(haven)
data_path %<>%
  mutate(encuesta = path %>% map(~read_dta(.x) %>%
    transmute(upm = `_upm`,
               estrato = `_estrato`,
               sexo, areageo2, lp, li, ingcorte,
               fep = `_fep`)))
```

# Definir el diseño

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Ahora se debe definir un diseño para cada encuesta, para nuestro ejemplo se define el diseño muestral.

```
options(survey.lonely.psu="adjust")
data_path %<>% mutate(
  disenno = encuesta %>%
    map(~as_survey_design(.data = .x,
      ids = upm,
      strata = estrato,
      weights = fep,
      nest = T
    )))
```

# Estimación de las personas debajo de la linea de pobreza en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
library(tidyr)
data_path %>% mutate(Pobreza = diseno %>%
  map(~.x %>%
    summarise(
      Prop = survey_mean(
        ingcorte<lp,
        vartype = c("se","cv") )))) %>%
  select(pais,Pobreza) %>%
  unnest(Pobreza) %>%
  head(14)
```

# Estimación de las personas con ingresos de bajo de la linea de pobreza en multiples encuestas (Resultado)

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	Prop	Prop_se	Prop_cv
ARG	0.3435	0.0000	0.0000
BOL	0.3229	0.0092	0.0285
BRA	0.1837	0.0023	0.0125
CHL	0.1420	0.0025	0.0176
COL	0.3974	0.0000	0.0000
CRI	0.1942	0.0066	0.0337
DOM	0.2183	0.0063	0.0287
ECU	0.3061	0.0000	0.0000
MEX	0.3737	0.0033	0.0088
PER	0.2834	0.0054	0.0192
PRY	0.2229	0.0107	0.0480
SLV	0.3074	0.0000	0.0000
URY	0.0518	0.0000	0.0000

# Estimación de las personas con ingresos de bajo de la linea de pobreza en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
data_path %>% mutate(Pobreza = disenno %>%  
  map(~.x %>% group_by(sexo) %>%  
    cascade(  
      Prop = survey_mean(  
        ingcorte<lp,  
        vartype = c("se","cv")),  
      .fill = 3 ))) %>%  
  select(pais,Pobreza) %>%  
  unnest(Pobreza) %>%  
  head(14)
```

`.fill = 3` permite poner un nombre al resultado global

# Estimación de las personas debajo de la linea de pobreza en multiples encuestas (Resultado)

pais	sexo	Prop	Prop_se	Prop_cv
ARG	1	0.3485	0.0000	0.0000
ARG	2	0.3387	0.0000	0.0000
ARG	3	0.3435	0.0000	0.0000
BOL	1	0.3123	0.0093	0.0298
BOL	2	0.3331	0.0097	0.0290
BOL	3	0.3229	0.0092	0.0285
BRA	1	0.1783	0.0024	0.0134
BRA	2	0.1888	0.0024	0.0129
BRA	3	0.1837	0.0023	0.0125
CHL	1	0.1391	0.0026	0.0186
CHL	2	0.1445	0.0027	0.0190
CHL	3	0.1420	0.0025	0.0176
COL	1	0.3893	0.0000	0.0000
COL	2	0.4052	0.0000	0.0000

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

# Estimación de las personas con ingresos de bajo de la linea de pobreza en multiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
data_path %>% mutate(Pobreza = diseno %>%  
  map(~.x %>% group_by(areageo2) %>%  
    cascade(  
      Prop = survey_mean(  
        ingcorte<lp,  
        vartype = c("se","cv")),  
      .fill = 3 ))) %>%  
  select(pais,Pobreza) %>%  
  unnest(Pobreza) %>%  
  head(14)
```



# Estimación de las personas debajo de la linea de pobreza en multiples encuestas (Resultado)

pais	areageo2	Prop	Prop_se	Prop_cv
ARG	1	0.3435	0.0000	0.0000
ARG	3	0.3435	0.0000	0.0000
BOL	1	0.2540	0.0076	0.0300
BOL	2	0.4848	0.0259	0.0535
BOL	3	0.3229	0.0092	0.0285
BRA	1	0.1708	0.0025	0.0146
BRA	2	0.2618	0.0055	0.0212
BRA	3	0.1837	0.0023	0.0125
CHL	1	0.1464	0.0028	0.0188
CHL	2	0.1079	0.0049	0.0455
CHL	3	0.1420	0.0025	0.0176
COL	1	0.3785	0.0000	0.0000
COL	2	0.4626	0.0000	0.0000
COL	3	0.3974	0.0000	0.0000

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

# Prueba de independencia en múltiples encuestas

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Se retira Argentina del ejemplo por no tener información en el área rural.

```
data_path[-1,] %>%  
  mutate(test_chisq = map_dbl(disenos,  
    ~svychisq(~areageo2+sexo,  
              design = .x,  
              statistic="F"))$p.value  
)) %>% select(pais, test_chisq)
```

# Estimación de las personas con ingresos de bajo de la línea de pobreza en múltiples encuestas (Resultado)

Análisis de encuestas de hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

pais	test_chisq
BOL	0.0078
BRA	0.0000
CHL	0.0000
COL	NaN
CRI	0.0000
DOM	0.0010
ECU	NaN
MEX	0.0000
PER	0.0007
PRY	0.0000
SLV	NaN
URY	NaN

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.
- **Paso 3** Procesar información.

# Alternativa para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

En ocasiones solo se desea obtener un resultado rápido para realizar un reporte o una comparación rápida de información, en estas ocasiones no es necesario guardar en la memoria de R toda la encuesta, por esta razón se ilustra una alternativa de procesamiento de múltiples archivos.

- **Paso 1** Leer archivo y organizar encuestas.
- **Paso 2** Definir diseño muestral.
- **Paso 3** Procesar información.
- **Paso 4** Organizar y presentar resultados.

# Creando función para el procesamiento de múltiples archivos.

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
options(survey.lonely.psu="adjust")
Promedio_aux <- function(input_file){
  ## Paso 1
  encuesta <- read_dta(input_file) %>%
    transmute(upm = `_upm`,
              estrato = `_estrato`,
              sexo, areageo2,lp,li,ingcorte,
              fep = `_fep`)

  ## Paso 2
  diseno <- as_survey_design(.data = encuesta,
                             ids = upm,
                             strata = estrato,
                             weights = fep,
                             nest = T)

  %>% mutate(pobreza = ifelse(ingcorte<lp,1,0))
  ## Paso 3
  svyby(~as.factor(pobreza), ~sexo,
        diseno, svymean) %>%
    tibble()
}
```



# Procesando encuestas múltiples

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Para el *Paso 4* realizamos la siguiente sintaxis.

```
list.files("Z:/BC/",full.names = TRUE,  
          pattern = "2020")[1:3] %>%  
  map_df(~Promedio_aux(.x))
```

Los resultados se muestran en el orden de lectura de los archivos

sexo	as.factor(pobreza)0	as.factor(pobreza)1	se.as.factor(pobreza)0	se.as.factor(pobreza)1
1	0.6515	0.3485	0.0000	0.0000
2	0.6613	0.3387	0.0000	0.0000
1	0.6877	0.3123	0.0093	0.0093
2	0.6669	0.3331	0.0097	0.0097
1	0.8217	0.1783	0.0024	0.0024
2	0.8112	0.1888	0.0024	0.0024

# Ejemplo Bolivia 2015 a 2020

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

Lectura de las bases de datos.

```
# lista con la ruta de los archivos
```

```
input_file <- list.files("Z:/BC/",full.names = TRUE,  
                        pattern = "(BOL_201[5-9]|BOL_2020)
```

```
## [1] "Z:/BC/BOL_2015N.dta" "Z:/BC/BOL_2016N.dta" "Z:/BC/BOL_2017N.dta"  
## [4] "Z:/BC/BOL_2018N.dta" "Z:/BC/BOL_2019N.dta" "Z:/BC/BOL_2020N.dta"
```

```
# identificando el año
```

```
anio <-gsub(x = input_file, pattern = ".*(\\d{4}).*", "\\1")
```

```
# definir las variables de interes.
```

```
name_var <- c("sys_pe" , "gan_pe" ,  
             "yemp_pe" , "yjub_pe",  
             "yotr_pe", "yto_pe")
```

# Bolivia 2015 a 2020

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

```
library(furrr)
# definir los nucleos para el procesamiento
plan(multisession, workers = 3)
# Lectura, filtro de la base y selección de variables por archivo.

encuesta <- setNames(input_file,anio) %>% ## años de los archivos
  future_map(~read_dta(.x) %>%           ## lectura de archivos
    mutate(fep = `_fep`,                 ## renombrando variable
           upm = `_upm`,
           estrato = `_estrato`,) %>%
  filter(edad >= 15, sexo == 2) %>%      ## filtro de edad y sexo
  select(upm, estrato,fep,all_of(name_var)) ## seleccion de variable
  )
```

# Bolivia 2015 a 2020

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

## Definir el diseño

```
disenio = encuesta %>% future_map(  
  ~ as_survey_design(  
    .data = .x,  
    ids = upm,  
    strata = estrato,  
    weights = fep,    nest = T  
  )  
)
```

# Bolivia 2015 a 2020

# Análisis de encuestas de hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

### Variables continuas

## Variables categóricas

```
diseno %>%
  future_map_dfr(~.x %>%
    mutate_at(.vars = all_of(name_var),
              ~ ifelse(. > 0, 1, 0)) %>%
    summarise_at(.vars = all_of(name_var),
                  ~survey_mean(.)), .id = "anio")
```

	anio	sys	pays	pegas	gran	pevesa	yamp	peulse	peub	peote	peotr	peyte	peyo	pe
2015	0.178	0.0044	0.222	0.0053	0.396	0.0060	0.025	0.0017	0.348	0.0060	0.649	0.0064		
2016	0.208	0.0049	0.232	0.0052	0.433	0.0059	0.026	0.0016	0.386	0.0064	0.693	0.0057		
2017	0.189	0.0047	0.230	0.0054	0.413	0.0060	0.025	0.0017	0.372	0.0059	0.678	0.0059		
2018	0.181	0.0044	0.233	0.0050	0.408	0.0058	0.025	0.0017	0.359	0.0055	0.664	0.0054		
2019	0.194	0.0052	0.250	0.0055	0.439	0.0058	0.032	0.0020	0.356	0.0057	0.684	0.0062		
2020	0.156	0.0050	0.240	0.0056	0.392	0.0059	0.035	0.0026	0.707	0.0057	0.882	0.0038		

# Bolivia 2015 a 2020

# Análisis de encuestas de hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

### Variables continuas

## Variables categóricas

```
future_map_dfr(disenso, function(dis){
  future_map_dfc(name_var, function(var_id){
    dis %>% filter_at(all_of(var_id), ~ . > 0) %>%
    summarise_at(all_of(var_id), ~survey_mean(.))
  })
}, .id = "anio")
```

	anio	sys_pays	pegas	pgan	peveap	year	pepulse	pepub	peote	peotr	peyte	peyto	pe
	2015	2734	54.07	1888	58.79	2296	42.74	2207	122.08	387.4	15.082	1751	34.36
	2016	2850	54.39	1745	46.96	2308	38.77	2221	75.11	377.4	12.730	1785	32.37
	2017	2989	57.58	1836	49.62	2395	44.45	2417	66.06	453.1	16.466	1870	36.65
	2018	3174	48.98	1790	40.43	2434	37.21	2564	78.24	490.7	17.078	1935	32.24
	2019	3180	60.80	1787	42.84	2434	42.74	2496	80.01	513.6	19.852	2014	37.86
	2020	3413	70.23	2055	85.49	2626	63.87	2691	95.54	233.1	6.799	1519	37.09

# ¡Gracias!

Análisis de  
encuestas de  
hogares con R

Andrés  
Gutiérrez,  
Ph.D.  
Stalyn  
Guerrero  
M.Sc.

Variables  
continuas

Variables  
categóricas

*Email:* [andres.gutierrez@cepal.org](mailto:andres.gutierrez@cepal.org)