

Análisis de encuestas de hogares con R

Módulo 2: Análisis de variables categóricas

CEPAL - Unidad de Estadísticas Sociales

- 1 Motivación
- 2 Lectura y procesamientos de encuestas con R
- 3 Estimación del tamaño.
- 4 Estimación de la proporción.
- 5 Tablas cruzadas.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Motivación

Motivación

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Esta sección esta enfocada en los procedimientos para análisis simples univariados, bivariados y multivariados seleccionados para respuestas de encuestas categóricas, centrándose en la adaptación de técnicas de inferencia estadística a datos de encuestas de muestras complejas

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Lectura y procesamientos de encuestas con R

Lectura de la base

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Iniciemos con la lectura de la encuesta.

```
encuesta <- readRDS("../Data/encuesta.rds")
```

El paso siguiente es realizar declaración del objeto tipo diseño.

Definir diseño de la muestra con srvyr

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

Para el desarrollo de la presentación se define el diseño muestral con la función `as_survey_design`.

```
options(survey.lonely.psu = "adjust")
library(srvyr)
diseno <- encuesta %>% # Base de datos.
  as_survey_design(
    strata = Stratum, # Id de los estratos.
    ids = PSU, # Id para las observaciones.
    weights = wk, # Factores de expansión.
    nest = TRUE # Valida el anidado dentro
                # del estrato
  )
```

Creación de nuevas variables

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

Durante los análisis de encuesta surge la necesidad de crear nuevas variables a partir de las existentes, aquí mostramos la definición de algunas de ellas.

```
disenio <- disenio %>% mutate(  
  pobreza = ifelse(Poverty != "NotPoor", 1, 0),  
  desempleo = ifelse(Employment == "Unemployed", 1, 0),  
  edad_18 = case_when(Age < 18 ~ "< 18 años",  
                       TRUE ~ ">= 18 años")  
)
```

Se ha introducido la función `case_when` la cual es una extensión del a función `ifelse` que permite crear múltiples categorías a partir de una o varias condiciones.

Dividiendo la muestra en Sub-grupos

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

En ocasiones se desea realizar estimaciones por sub-grupos de la población, en este caso se extraer 4 sub-grupos de la encuesta.

```
sub_Urbano <- diseno %>% filter(Zone == "Urban")
sub_Rural  <- diseno %>% filter(Zone == "Rural")
sub_Mujer  <- diseno %>% filter(Sex == "Female")
sub_Hombre <- diseno %>% filter(Sex == "Male")
```

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

**Estimación del
tamaño.**

Estimación de
la proporción.

Tablas
cruzadas.

Estimación del tamaño.

Estimación del tamaño.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

El primer parámetro estimado serán los tamaños de la población y subpoblaciones.

Estimación de tamaño

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(tamano_zona <- diseno %>% group_by(Zone) %>%  
  summarise(  
    n = unweighted(n()), # Observaciones en la muestra  
    Nd = survey_total(vartype = c("se", "ci"))))
```

Zone	n	Nd	Nd_se	Nd_low	Nd_upp
Rural	1297	72102	3062	66039	78165
Urban	1308	78164	2847	72526	83802

En la tabla n denota el número de observaciones en la muestra por Zona y Nd denota la estimación del total de observaciones en la población.

Estimación de tamaño

Empleando una sintaxis similar es posible estimar el número de personas en condición de pobreza extrema, pobreza y no pobres.

```
(tamano_pobreza <- diseno %>% group_by(Poverty) %>%  
  summarise(  
    Nd = survey_total(vartype = c("se","ci"))))
```

Poverty	Nd	Nd_se	Nd_low	Nd_upp
NotPoor	91398	4395	82696	100101
Extreme	21519	4949	11719	31319
Relative	37349	3695	30032	44666

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Estimación de tamaño

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

En forma similar es posible estimar el número de personas debajo de la línea de pobreza.

```
(tamano_pobreza <- disenio %>%  
  group_by(pobreza) %>%  
  summarise(  
    Nd = survey_total(vartype = c("se","ci"))))
```

pobreza	Nd	Nd_se	Nd_low	Nd_upp
0	91398	4395	82696	100101
1	58868	5731	47519	70216

Estimación de tamaño

Otra variable de interés es conocer el estado de ocupación de la personas.

```
(tamano_ocupacion <- disenio %>%  
  group_by(Employment) %>%  
  summarise(  
    Nd = survey_total(vartype = c("se","ci"))))
```

Employment	Nd	Nd_se	Nd_low	Nd_upp
Unemployed	4635	760.6	3129	6141
Inactive	41465	2162.8	37183	45748
Employed	61877	2540.1	56847	66907
NA	42289	2780.0	36784	47794

Estimación de tamaño

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Utilizando la función `group_by` es posible obtener resultados por más de un nivel de agregación.

```
(tamano_ocupacion_pobreza <- diseno %>%  
  group_by(Employment, Poverty) %>%  
  cascade(  
    Nd = survey_total(vartype = c("se","ci")),  
    .fill = "Total") %>%  
  data.frame()  
)
```


Estimación de tamaño

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Employment	Poverty	Nd	Nd_se	Nd_low	Nd_upp
Unemployed	NotPoor	1768	405.4	965.7	2571
Unemployed	Extreme	1169	348.1	479.9	1859
Unemployed	Relative	1697	457.8	790.7	2604
Unemployed	Total	4635	760.6	3128.7	6141
Inactive	NotPoor	24346	1736.3	20908.0	27784
Inactive	Extreme	6422	1320.7	3806.6	9037
Inactive	Relative	10697	1460.3	7805.9	13589
Inactive	Total	41465	2162.8	37182.7	45748
Employed	NotPoor	44600	2596.2	39459.6	49741
Employed	Extreme	5128	1121.6	2906.6	7349
Employed	Relative	12149	1346.6	9482.7	14816
Employed	Total	61877	2540.1	56847.4	66907
Total	Total	150266	4181.4	141986.5	158546
NA	NotPoor	20684	1256.6	18195.4	23172
NA	Extreme	8800	2979.9	2899.7	14701
NA	Relative	12805	1551.0	9733.9	15876
NA	Total	42289	2780.0	36784.3	47794

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

**Estimación de
la proporción.**

Tablas
cruzadas.

Estimación de la proporción.

Estimación de la proporción.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Otro resultado de interés es la estimación de las proporciones, dado que estas entregan una mayor información sobre las tendencias en la población, siendo de mucha importancia en la toma de decisiones.

Estimación de proporción de urbano y rural

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

El procedimiento estándar para el cálculo de proporciones es crear una *variable dummy* y sobre ella realizar las operaciones. Sin embargo, la librería *srvy* nos simplifica el cálculo, mediante la sintaxis.

```
(prop_zona <- diseno %>% group_by(Zone) %>%  
  summarise(  
    prop = survey_mean(vartype = c("se", "ci"),  
      proportion = TRUE )))
```

Zone	prop	prop_se	prop_low	prop_upp
Rural	0.4798	0.014	0.4523	0.5075
Urban	0.5202	0.014	0.4925	0.5477

Note que, se utilizó la función *survey_mean* para la estimación.

Estimación de proporción de urbano y rural

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

La función idónea para realizar la estimación de las proporciones es `survey_prop` y la sintaxis es como sigue:

```
(prop_zona2 <- diseno %>% group_by(Zone) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci") )))
```

Zone	prop	prop_se	prop_low	prop_upp
Rural	0.4798	0.014	0.4522	0.5075
Urban	0.5202	0.014	0.4925	0.5478

Proporción de hombres y mujeres en la zona urbana

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

Si el interés es obtener la estimación para una subpoblación, procedemos así:

```
(prop_sexou <- sub_Urbano %>% group_by(Sex) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se","ci"))))
```

Sex	prop	prop_se	prop_low	prop_upp
Female	0.5367	0.013	0.5108	0.5627
Male	0.4633	0.013	0.4373	0.4892

Proporción de hombres y mujeres en la zona urbana

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

Si el interés es obtener la estimación para una subpoblación, procedemos así:

```
(prop_sexoU <- sub_Urbano %>% group_by(Sex) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se","ci"))))
```

Sex	prop	prop_se	prop_low	prop_upp
Female	0.5367	0.013	0.5108	0.5627
Male	0.4633	0.013	0.4373	0.4892

Ejercicio

¿Cómo estimar el Proporción de hombres dado que están en zona rural?

Proporción de hombres y mujeres en la zona rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_sexor <- sub_Rural %>% group_by(Sex) %>%  
  summarise(  
    n = unweighted(n()),  
    prop = survey_prop(vartype = c("se", "ci"))))
```

Sex	n	prop	prop_se	prop_low	prop_upp
Female	679	0.5165	0.0082	0.500	0.533
Male	618	0.4835	0.0082	0.467	0.500

Proporción de hombres y mujeres en la zona rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_sexor <- sub_Rural %>% group_by(Sex) %>%  
  summarise(  
    n = unweighted(n()),  
    prop = survey_prop(vartype = c("se", "ci"))))
```

Sex	n	prop	prop_se	prop_low	prop_upp
Female	679	0.5165	0.0082	0.500	0.533
Male	618	0.4835	0.0082	0.467	0.500

Ejercicio

¿Cómo estimar el Proporción de hombres en la zona rural dado que es hombre?

Proporción de hombres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaH <- sub_Hombre %>% group_by(Zone) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci"))))
```

Zone	prop	prop_se	prop_low	prop_upp
Rural	0.4905	0.0178	0.4552	0.5258
Urban	0.5095	0.0178	0.4742	0.5448

Proporción de hombres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaH <- sub_Hombre %>% group_by(Zone) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci"))))
```

Zone	prop	prop_se	prop_low	prop_upp
Rural	0.4905	0.0178	0.4552	0.5258
Urban	0.5095	0.0178	0.4742	0.5448

Ejercicio

¿Cómo estimar el Proporción de mujeres en la zona rural dado que es mujer?

Proporción de mujeres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaM <- sub_Mujer %>% group_by(Zone) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se","ci"))))
```

Zone	prop	prop_se	prop_low	prop_upp
Rural	0.4702	0.014	0.4425	0.4980
Urban	0.5298	0.014	0.5020	0.5575

Proporción de hombres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Con el uso de la función `group_by` es posible estimar un mayor número de niveles de agregación al combinar dos o más variables.

```
(prop_ZonaH_Pobreza <- sub_Hombre %>%  
  group_by(Zone, Poverty) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci")))%>%  
  data.frame())
```

Proporción de hombres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Zone	Poverty	prop	prop_se	prop_low	prop_upp
Rural	NotPoor	0.5488	0.0626	0.4248	0.6729
Rural	Extreme	0.1975	0.0675	0.0640	0.3311
Rural	Relative	0.2536	0.0372	0.1799	0.3274
Urban	NotPoor	0.6599	0.0366	0.5874	0.7324
Urban	Extreme	0.1129	0.0245	0.0643	0.1614
Urban	Relative	0.2272	0.0260	0.1757	0.2788

Proporción de mujeres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaM_Pobreza <- sub_Mujer %>%  
  group_by(Zone, Poverty) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci"))) %>%  
  data.frame())
```

Zone	Poverty	prop	prop_se	prop_low	prop_upp
Rural	NotPoor	0.5539	0.0557	0.4436	0.6642
Rural	Extreme	0.1600	0.0557	0.0496	0.2704
Rural	Relative	0.2861	0.0436	0.1998	0.3724
Urban	NotPoor	0.6612	0.0322	0.5974	0.7251
Urban	Extreme	0.1094	0.0221	0.0656	0.1531
Urban	Relative	0.2294	0.0266	0.1768	0.2820

Proporción de hombres en la zona y empleado

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaH_Ocupacion <- sub_Hombre %>%  
  group_by(Zone, Employment) %>%  
  summarise(  
    prop = survey_prop(vartype = c("se", "ci")))%>%  
  data.frame())
```

Zone	Employment	prop	prop_se	prop_low	prop_upp
Rural	Unemployed	0.0513	0.0157	0.0201	0.0824
Rural	Inactive	0.1035	0.0203	0.0634	0.1436
Rural	Employed	0.5225	0.0265	0.4700	0.5750
Rural	NA	0.3227	0.0350	0.2534	0.3920
Urban	Unemployed	0.0437	0.0085	0.0269	0.0606
Urban	Inactive	0.1633	0.0181	0.1275	0.1991
Urban	Employed	0.5134	0.0236	0.4666	0.5602
Urban	NA	0.2796	0.0221	0.2358	0.3233

Proporción de mujeres en la zona urbana y rural

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(prop_ZonaM_Ocupacion <- sub_Mujer %>%  
  group_by(Zone, Employment) %>%  
  summarise(  
    prop = survey_prop(  
      vartype = c("se", "ci")) %>%  
    data.frame())
```

Zone	Employment	prop	prop_se	prop_low	prop_upp
Rural	Unemployed	0.0102	0.0055	-0.0008	0.0211
Rural	Inactive	0.4472	0.0352	0.3774	0.5170
Rural	Employed	0.2400	0.0392	0.1625	0.3175
Rural	NA	0.3026	0.0308	0.2417	0.3636
Urban	Unemployed	0.0211	0.0060	0.0093	0.0329
Urban	Inactive	0.3645	0.0214	0.3220	0.4069
Urban	Employed	0.3846	0.0195	0.3460	0.4231
Urban	NA	0.2299	0.0139	0.2025	0.2573

Estimación de la proporción de personas menor a 18 años

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
diseno %>%  
group_by(edad_18, pobreza) %>%  
  summarise(  
    Prop = survey_prop(  
      vartype = c("se", "ci")) %>%  
    data.frame()
```

Estimación de la proporción de personas menor a 18 años

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

edad_18	pobreza	Prop	Prop_se	Prop_low	Prop_upp
< 18 anios	0	0.4985	0.0373	0.4246	0.5723
< 18 anios	1	0.5015	0.0373	0.4277	0.5754
>= 18 anios	0	0.6646	0.0298	0.6056	0.7236
>= 18 anios	1	0.3354	0.0298	0.2764	0.3944

Estimación de la proporción de personas menor a 18 años

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
diseno %>%  
  group_by(edad_18, desempleo) %>%  
  summarise(  
    Prop = survey_prop(  
      vartype = c("se", "ci")) %>%  
    data.frame()
```

Estimación de la proporción de personas menor a 18 años

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

edad_18	desempleo	Prop	Prop_se	Prop_low	Prop_upp
< 18 años	0	0.1667	0.0149	0.1373	0.1961
< 18 años	1	0.0037	0.0020	-0.0002	0.0076
< 18 años	NA	0.8296	0.0150	0.7998	0.8593
>= 18 años	0	0.9552	0.0076	0.9403	0.9702
>= 18 años	1	0.0448	0.0076	0.0298	0.0597

Estimación de la proporción de personas menor a 18 años en zona rural

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

```
sub_Rural %>%  
  group_by(edad_18) %>%  
  summarise(  
    Prop = survey_prop(  
      vartype = c("se", "ci")) %>%  
    data.frame()
```

edad_18	Prop	Prop_se	Prop_low	Prop_upp
< 18 años	0.3712	0.0302	0.3106	0.4317
>= 18 años	0.6288	0.0302	0.5683	0.6894

Estimación de la proporción de mujeres rango de edad

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

```
sub_Mujer %>% mutate(edad_rango = case_when(
  Age >= 18 & Age <= 35 ~ "18 - 35",
  TRUE ~ "Otro")) %>%
group_by(edad_rango, Employment) %>%
summarise(
  Prop = survey_prop(
    vartype = c("se", "ci"))) %>%
data.frame()
```

edad_rango	Employment	Prop	Prop_se	Prop_low	Prop_upp
18 - 35	Unemployed	0.0289	0.0091	0.0108	0.0470
18 - 35	Inactive	0.5165	0.0379	0.4415	0.5916
18 - 35	Employed	0.4545	0.0357	0.3839	0.5252
Otro	Unemployed	0.0102	0.0040	0.0022	0.0181
Otro	Inactive	0.3527	0.0207	0.3117	0.3937
Otro	Employed	0.2548	0.0217	0.2119	0.2978
Otro	NA	0.3823	0.0223	0.3381	0.4265

Estimación de la proporción de hombres rango de edad

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

```
sub_Hombre %>% mutate(edad_rango = case_when(
  Age >= 18 & Age <= 35 ~ "18 - 35",
  TRUE ~ "Otro")) %>%
group_by(edad_rango, Employment) %>%
summarise(
  Prop = survey_prop(
    vartype = c("se", "ci"))) %>%
data.frame()
```

edad_rango	Employment	Prop	Prop_se	Prop_low	Prop_upp
18 - 35	Unemployed	0.0964	0.0182	0.0603	0.1324
18 - 35	Inactive	0.0894	0.0164	0.0568	0.1219
18 - 35	Employed	0.8142	0.0230	0.7687	0.8598
Otro	Unemployed	0.0261	0.0072	0.0119	0.0403
Otro	Inactive	0.1534	0.0199	0.1141	0.1928
Otro	Employed	0.3885	0.0203	0.3484	0.4286
Otro	NA	0.4320	0.0211	0.3902	0.4738

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

**Tablas
cruzadas.**

Tablas cruzadas.

Tabla Zona Vs Sexo

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

Haciendo uso de la función `group_by` organizada en forma de `data.frame`.

```
(  
  prop_sexo_zona <- diseno %>%  
    group_by(pobreza, Sex) %>%  
    summarise(  
      prop = survey_prop(vartype = c("se", "ci"))) %>%  
    data.frame()  
)
```

Esta forma de organizar la información es recomendable cuando el realizar el análisis sobre las estimaciones puntuales.

Tabla Zona Vs Sexo

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

pobreza	Sex	prop	prop_se	prop_low	prop_upp
0	Female	0.5292	0.0124	0.5046	0.5538
0	Male	0.4708	0.0124	0.4462	0.4954
1	Female	0.5236	0.0159	0.4922	0.5550
1	Male	0.4764	0.0159	0.4450	0.5078

Tablas de doble entrada.

Una alternativa es utilizar la función `svyby` con la siguiente sintaxis.

```
tab_Sex_Pobr <- svyby(~Sex, ~pobreza, diseno, svymean)
tab_Sex_Pobr %>% select(-se.SexFemale, -se.SexMale)
```

	pobreza	SexFemale	SexMale
0	0	0.5292	0.4708
1	1	0.5236	0.4764

```
tab_Sex_Pobr %>% select(-SexFemale, -SexMale)
```

	pobreza	se.SexFemale	se.SexMale
0	0	0.0124	0.0124
1	1	0.0159	0.0159

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Tablas de doble entrada.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Para la estimación de los intervalos de confianza utilizar la función `confint`.

```
confint(tab_Sex_Pobr) %>% as.data.frame()
```

	2.5 %	97.5 %
0:SexFemale	0.5048	0.5535
1:SexFemale	0.4925	0.5547
0:SexMale	0.4465	0.4952
1:SexMale	0.4453	0.5075

Prueba de independencia.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Para realizar la prueba de independencia χ^2 puede ejecuta la siguiente linea de código.

```
svychisq(~Sex + pobreza, diseno, statistic="F")
```

```
##
```

```
## Pearson's X^2: Rao & Scott adjustment
```

```
##
```

```
## data: NextMethod()
```

```
## F = 0.056, ndf = 1, ddf = 119, p-value = 0.8
```

Más adelante se profundiza en la metodología de esta prueba.

Tablas de doble entrada.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(tab_Sex_Ocupa <- svyby(~Sex, ~Employment,  
                          diseno, svymean))
```

	Employment	SexFemale	SexMale	se.SexFemale	se.SexMale
Unemployed	Unemployed	0.2727	0.7273	0.0535	0.0535
Inactive	Inactive	0.7703	0.2297	0.0234	0.0234
Employed	Employed	0.4052	0.5948	0.0185	0.0185

Tablas de doble entrada

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
confint(tab_Sex_Ocupa) %>% as.data.frame()
```

	2.5 %	97.5 %
Unemployed:SexFemale	0.1678	0.3776
Inactive:SexFemale	0.7245	0.8162
Employed:SexFemale	0.3689	0.4415
Unemployed:SexMale	0.6224	0.8322
Inactive:SexMale	0.1838	0.2755
Employed:SexMale	0.5585	0.6311

Prueba de independencia.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

La prueba de independencia χ^2 se obtiene con la siguiente línea de código.

```
svychisq(~Sex + Employment,  
          design = diseneno, statistic="F")
```

```
##  
## Pearson's X^2: Rao & Scott adjustment  
##  
## data: NextMethod()  
## F = 62, ndf = 1.7, ddf = 200.7, p-value <2e-16
```

Tablas de doble entrada.

Dado que la variable *pobreza* es de tipo numérica, es necesario convertirla en factor.

```
tab_region_pobreza <-  
  svyby(~as.factor(pobreza), ~Region, diseno, svymean)  
tab_region_pobreza %>%  
  select("-se.as.factor(pobreza)0",  
         "-se.as.factor(pobreza)1")
```

	Region	as.factor(pobreza)0	as.factor(pobreza)1
Norte	Norte	0.6410	0.3590
Sur	Sur	0.6562	0.3438
Centro	Centro	0.6346	0.3654
Occidente	Occidente	0.5992	0.4008
Oriente	Oriente	0.5482	0.4518

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Tablas de doble entrada.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
tab_region_pobreza %>%  
  select("se.as.factor(pobreza)0",  
         "se.as.factor(pobreza)1")
```

	se.as.factor(pobreza)0	se.as.factor(pobreza)1
Norte	0.0555	0.0555
Sur	0.0435	0.0435
Centro	0.0786	0.0786
Occidente	0.0467	0.0467
Oriente	0.0885	0.0885

Prueba de independencia.

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Una vez más la prueba de independencia es:

```
svychisq(~Region + pobreza,  
          design = diseno,  statistic="F")
```

```
##
```

```
## Pearson's  $X^2$ : Rao & Scott adjustment
```

```
##
```

```
## data:  NextMethod()
```

```
## F = 0.49, ndf = 3, ddf = 358, p-value = 0.7
```

Razón de obbs

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
(tab_Sex <- svyby(~pobreza, ~Sex, diseno,  
  svymean, vartype = c("se", "ci")))
```

	Sex	pobreza	se	ci_l	ci_u
Female	Female	0.3892	0.0316	0.3273	0.4512
Male	Male	0.3946	0.0366	0.3228	0.4664

```
svycontrast(tab_Sex, quote(`Female`/`Male`))
```

```
##          nlcon    SE  
## contrast 0.987 0.12
```

Razón de obbs

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
tab_Sex_Pobr <-  
  svymean(~interaction (Sex, pobreza), diseno,  
           se=T, na.rm=T, ci=T, keep.vars=T)  
tab_Sex_Pobr %>% as.data.frame()
```

	mean	SE
interaction(Sex, pobreza)Female.0	0.3219	0.0178
interaction(Sex, pobreza)Male.0	0.2864	0.0177
interaction(Sex, pobreza)Female.1	0.2051	0.0166
interaction(Sex, pobreza)Male.1	0.1866	0.0178

Razón de obbs

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Suponga que se desea calcular la siguiente razón de obbs.

$$\frac{\frac{P(\text{Sex}=\text{Female}|\text{pobreza}=0)}{P(\text{Sex}=\text{Female}|\text{pobreza}=1)}}{\frac{P(\text{Sex}=\text{Male}|\text{pobreza}=1)}{P(\text{Sex}=\text{Male}|\text{pobreza}=0)}}$$

La forma de cálculo en sería:

```
svycontrast(tab_Sex_Pobr,  
  quote(`interaction(Sex, pobreza)Female.0`/  
        `interaction(Sex, pobreza)Female.1`)/  
  (`interaction(Sex, pobreza)Male.0`/  
    `interaction(Sex, pobreza)Male.1`) ))
```

```
##          nlcon  SE  
## contrast  1.02 0.1
```

Razón de obbs

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Ahora, se desea realizar la siguiente razón:

$$\frac{\frac{P(\text{Sex}=\text{Male}|\text{pobreza}=1)}{P(\text{Sex}=\text{Female}|\text{pobreza}=1)}}{\frac{P(\text{Sex}=\text{Male}|\text{pobreza}=0)}{P(\text{Sex}=\text{Female}|\text{pobreza}=0)}}$$

```
svycontrast(tab_Sex_Pobr,  
             quote(`interaction(Sex, pobreza)Male.1`/  
                   `interaction(Sex, pobreza)Female.1`)/  
             (`interaction(Sex, pobreza)Male.0`/  
               `interaction(Sex, pobreza)Female.0`)))
```

```
##          nlcon  SE  
## contrast  1.02 0.1
```


Contrastes

El interés ahora es realizar en contraste de proporciones, por ejemplo: $\hat{p}_F - \hat{p}_M$

```
(tab_sex_pobreza <- svyby(~pobreza, ~Sex,
                          diseno ,
                          svymean, na.rm=T,
                          covmat = TRUE,
                          vartype = c("se", "ci")))
```

	Sex	pobreza	se	ci_l	ci_u
Female	Female	0.3892	0.0316	0.3273	0.4512
Male	Male	0.3946	0.0366	0.3228	0.4664

- *Paso 1:* Calcular la diferencia de estimaciones

0.3892 - 0.3946

Contrastes de la diferencia de proporciones

Con la función `vcov` obtiene la matriz de covarianzas

```
library(kableExtra)
vcov(tab_sex_pobreza)%>% data.frame() %>%
  kable(digits = 10,
        format.args = list(scientific = FALSE))
```

	Female	Male
Female	0.0009983	0.0009183
Male	0.0009183	0.0013416

■ Paso 2: El cálculo del error estándar es:

```
sqrt(0.0009983 + 0.0013416 - 2*0.0009183)
```

```
## [1] 0.02243
```

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Contrastes de la diferencia de proporciones en R

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Para realizar la diferencia de proporciones se hace uso de la función `svycontrast`.

```
svycontrast(tab_sex_pobreza,  
             list(diff_Sex = c(1, -1))) %>%  
             data.frame()
```

	contrast	diff_Sex
diff_Sex	-0.0053	0.0224

Contrastes de la diferencia de proporciones

Diferencia en desempleo por sexo.

```
(tab_sex_desempleo <- svyby(  
  ~desempleo, ~Sex,  
  diseno %>% filter(!is.na(desempleo)) ,  
  svymean, na.rm=T, covmat = TRUE,  
  vartype = c("se", "ci")))
```

	Sex	desempleo	se	ci_l	ci_u
Female	Female	0.0217	0.0056	0.0107	0.0326
Male	Male	0.0678	0.0122	0.0440	0.0917

■ Paso 1: Diferencia de las estimaciones

0.02169 - 0.06783

[1] -0.04614

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Contrastes de la diferencia de proporciones

Estimación de la matriz de covarianza:

```
vcov(tab_sex_desempleo) %>% data.frame() %>%  
  kable(digits = 10,  
        format.args = list(scientific = FALSE))
```

	Female	Male
Female	0.00003114	0.00002081
Male	0.00002081	0.00014789

■ *Paso 2:* Estimación del error estándar.

```
sqrt(0.00003114 + 0.00014789 - 2*0.00002081)
```

```
## [1] 0.01172
```

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Contrastes de la diferencia de proporciones en R

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Siguiendo el ejemplo anterior se tiene que:

```
svycontrast(tab_sex_desempleo,  
             list(diff_Sex = c(-1, 1))) %>%  
             data.frame()
```

	contrast	diff_Sex
diff_Sex	0.0461	0.0117

Contrastes

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

La tabla de desempleo por región se obtiene como:

```
(tab_region_desempleo <- svyby(
  ~desempleo, ~Region,
  diseno %>% filter(!is.na(desempleo)) ,
  svymean, na.rm=T, covmat = TRUE,
  vartype = c("se", "ci")))
```

	Region	desempleo	se	ci_l	ci_u
Norte	Norte	0.0488	0.0200	0.0095	0.0880
Sur	Sur	0.0656	0.0238	0.0191	0.1122
Centro	Centro	0.0387	0.0124	0.0144	0.0630
Occidente	Occidente	0.0400	0.0123	0.0159	0.0641
Oriente	Oriente	0.0295	0.0126	0.0049	0.0541

Creado una matriz de contrastes

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Ahora, el interés es realizar los contrastes siguientes para desempleo:

$$\hat{p}_{Norte} - \hat{p}_{Centro} = 0.01004, \hat{p}_{Sur} - \hat{p}_{Centro} = 0.02691$$

$$\hat{p}_{Occidente} - \hat{p}_{Oriente} = 0.01046$$

Escrita de forma matricial es:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Contrastes múltiples

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
vcov(tab_region_desempleo)%>%  
  data.frame() %>%  
  kable(digits = 10,  
        format.args = list(scientific = FALSE))
```

Contrastes múltiples

Análisis de encuestas de hogares con R

Motivación

Lectura y procesamiento de encuestas con R

Estimación del tamaño.

Estimación de la proporción.

Tablas cruzadas.

	Norte	Sur	Centro	Occidente	Oriente
Norte	0.0004009	0.0000000	0.0000000	0.0000000	0.000000
Sur	0.0000000	0.0005641	0.0000000	0.0000000	0.000000
Centro	0.0000000	0.0000000	0.0001538	0.0000000	0.000000
Occidente	0.0000000	0.0000000	0.0000000	0.0001512	0.000000
Oriente	0.0000000	0.0000000	0.0000000	0.0000000	0.000158

```
sqrt(0.0002981 + 0.0002884 - 2*0) ;
```

```
## [1] 0.02422
```

```
sqrt(0.0001968 + 0.0002884 - 2*0);
```

```
## [1] 0.02203
```

```
sqrt(0.0001267 + 0.0004093 - 2*0)
```

```
## [1] 0.02315
```

Creado una matriz de contrastes en R

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
svycontrast(tab_region_desempleo, list(  
  Norte_sur = c(1, 0, -1, 0, 0),  
  Sur_centro = c(0, 1, -1, 0, 0),  
  Occidente_Oriente = c(0, 0, 0, 1, -1)  
)) %>% data.frame()
```

	contrast	SE
Norte_sur	0.0100	0.0236
Sur_centro	0.0269	0.0268
Occidente_Oriente	0.0105	0.0176

Ejercicio.

Repetir el contraste anterior para pobreza.

```
(tab_region_pobreza <- svyby(  
  ~pobreza, ~Region,  
  diseno %>% filter(!is.na(desempleo)) ,  
  svymean, na.rm=T, covmat = TRUE,  
  vartype = c("se", "ci"))
```

	Region	pobreza	se	ci_l	ci_u
Norte	Norte	0.3263	0.0480	0.2322	0.4204
Sur	Sur	0.2947	0.0479	0.2007	0.3886
Centro	Centro	0.3234	0.0721	0.1820	0.4647
Occidente	Occidente	0.3673	0.0440	0.2811	0.4536
Oriente	Oriente	0.3871	0.0916	0.2075	0.5666

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Creado una matriz de contrastes

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Ahora, el interés es realizar los contrastes siguientes para pobreza:

$$\hat{p}_{Norte} - \hat{p}_{Centro}, \hat{p}_{Sur} - \hat{p}_{Centro}$$
$$\hat{p}_{Occidente} - \hat{p}_{Oriente}$$

Escrita de forma matricial es:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Creado una matriz de contrastes en R

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

```
svycontrast(tab_region_pobreza, list(  
  Norte_sur = c(1, 0, -1, 0, 0),  
  Sur_centro = c(0, 1, -1, 0, 0),  
  Occidente_Oriente = c(0, 0, 0, 1, -1)  
)) %>% data.frame()
```

	contrast	SE
Norte_sur	0.0029	0.0866
Sur_centro	-0.0287	0.0866
Occidente_Oriente	-0.0197	0.1016

¡Gracias!

Análisis de
encuestas de
hogares con R

Motivación

Lectura y
procesamien-
tos de
encuestas con
R

Estimación del
tamaño.

Estimación de
la proporción.

Tablas
cruzadas.

Email: andres.gutierrez@cepal.org