

## Variable categórica

CEPAL

17/2/2022

## Lectura de la base

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

## Definir diseño de la muestra con srvyr

```
library(srvyr)

diseno <- encuesta %>%
  as_survey_design(
    strata = Stratum,
    ids = PSU,
    weights = wk,
    nest = T
  )
```

## definir nuevas variables

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

# Sub-grupos

Extraer 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")
```

## Estimación de tamaño

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

Zone	n	Nd	Nd_se	Nd_low	Nd_upp
Rural	1238	72102	2973	66215	77989
Urban	1184	78164	2683	72852	83476

## Estimación de tamaño

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

Poverty	Nd	Nd_se	Nd_low	Nd_upp
NotPoor	86922	4141	78723	95122
Extreme	9935	1605	6756	13113
Relative	53409	5591	42338	64481

## Estimación de tamaño

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

pobreza	Nd	Nd_se	Nd_low	Nd_upp
0	86922	4141	78723	95122
1	63344	5660	52137	74550



## Estimación de tamaño

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

Employment	Nd	Nd_se	Nd_low	Nd_upp
Unemployed	5412	832.7	3764	7061
Inactive	43771	2021.0	39769	47773
Employed	61021	2454.7	56160	65881
NA	40062	1974.9	36151	43972

## Estimación de tamaño

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

## Estimación de tamaño

Employment	Poverty	Nd	Nd_se	Nd_low	Nd_upp
Unemployed	NotPoor	2898.7	613.0	1684.9	4113
Unemployed	Extreme	973.4	309.4	360.8	1586
Unemployed	Relative	1540.3	356.1	835.2	2245
Unemployed	Total	5412.4	832.7	3763.7	7061
Inactive	NotPoor	24884.8	1772.5	21375.1	28395
Inactive	Extreme	3137.9	650.6	1849.8	4426
Inactive	Relative	15748.4	1673.0	12435.7	19061
Inactive	Total	43771.2	2021.0	39769.3	47773
Employed	NotPoor	40952.3	1784.3	37419.1	44485
Employed	Extreme	1773.0	362.1	1055.9	2490
Employed	Relative	18295.4	2545.0	13256.0	23335
Employed	Total	61020.6	2454.7	56160.1	65881
Total	Total	150266.0	4004.6	142336.5	158195
NA	NotPoor	18186.4	1680.3	14859.3	21514
NA	Extreme	4050.2	784.6	2496.5	5604
NA	Relative	17825.1	1852.6	14156.7	21493
NA	Total	40061.7	1974.9	36151.3	43972

## Estimación de proporción de urbano y rural

```
(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	0.0134	0.4534	0.5064
Urban	0.5202	0.0134	0.4936	0.5466

## Estimación de proporción de urbano y rural

```
(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	0.0134	0.4533	0.5063
Urban	0.5202	0.0134	0.4937	0.5467

## Propoción de hombres y mujeres en la zona urbana y rural

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

Sex	n	prop	prop_se	prop_low	prop_upp
Female	631	0.5367	0.0147	0.5073	0.5661
Male	553	0.4633	0.0147	0.4339	0.4927

## Propoción de hombres y mujeres en la zona urbana y rural

```
(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	647	0.5165	0.0198	0.4768	0.5561
Male	591	0.4835	0.0198	0.4439	0.5232

## Propoción de hombres en la zona urbana y rural

```
(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	0.0221	0.4468	0.5342
Urban	0.5095	0.0221	0.4658	0.5532



## Proporción de mujeres en la zona urbana y rural

```
(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	0.0138	0.4430	0.4975
Urban	0.5298	0.0138	0.5025	0.5570

## Propoción de hombres en la zona urbana y rural

```
(prop_ZonaH_Pobreza <- sub_Hombre %>%  
  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	0.5227	0.0655	0.3929	0.6524
Rural	Extreme	0.0765	0.0176	0.0416	0.1114
Rural	Relative	0.4008	0.0709	0.2604	0.5412
Urban	NotPoor	0.5718	0.0370	0.4985	0.6452
Urban	Extreme	0.0748	0.0189	0.0373	0.1123
Urban	Relative	0.3534	0.0326	0.2888	0.4179

## Propoción de mujeres en la zona urbana y rural

```
(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	0.5760	0.0472	0.4826	0.6694
Rural	Extreme	0.0629	0.0149	0.0333	0.0925
Rural	Relative	0.3612	0.0496	0.2630	0.4594
Urban	NotPoor	0.6328	0.0299	0.5735	0.6920
Urban	Extreme	0.0529	0.0144	0.0244	0.0813
Urban	Relative	0.3144	0.0279	0.2591	0.3697

## Propoción de hombres en la zona y empleado

```
(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	0.0163	0.0191	0.0835
Rural	Inactive	0.1180	0.0218	0.0749	0.1612
Rural	Employed	0.5094	0.0265	0.4569	0.5619
Rural	NA	0.3213	0.0268	0.2681	0.3744
Urban	Unemployed	0.0555	0.0107	0.0343	0.0768
Urban	Inactive	0.1806	0.0202	0.1406	0.2207
Urban	Employed	0.4727	0.0213	0.4306	0.5149
Urban	NA	0.2911	0.0220	0.2475	0.3347

## Propoción de mujeres en la zona urbana y rural

```
(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	0.0209	0.0079	0.0053	0.0365
Rural	Inactive	0.4790	0.0201	0.4393	0.5188
Rural	Employed	0.2156	0.0158	0.1844	0.2468
Rural	NA	0.2845	0.0170	0.2508	0.3181
Urban	Unemployed	0.0199	0.0077	0.0047	0.0351
Urban	Inactive	0.3641	0.0220	0.3205	0.4078
Urban	Employed	0.4318	0.0244	0.3834	0.4801
Urban	NA	0.1842	0.0182	0.1482	0.2201

## Estimación de la proporción de personas menor a 18 años

```
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

edad_18	pobreza	Prop	Prop_se	Prop_low	Prop_upp
< 18 años	0	0.4665	0.0372	0.3929	0.5402
< 18 años	1	0.5335	0.0372	0.4598	0.6071
>= 18 años	0	0.6326	0.0300	0.5733	0.6919
>= 18 años	1	0.3674	0.0300	0.3081	0.4267

## Estimación de la proporción de personas menor a 18 años

```
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

edad_18	desempleo	Prop	Prop_se	Prop_low	Prop_upp
< 18 años	0	0.1785	0.0168	0.1453	0.2117
< 18 años	1	0.0040	0.0021	-0.0002	0.0081
< 18 años	NA	0.8175	0.0170	0.7838	0.8512
>= 18 años	0	0.9485	0.0080	0.9327	0.9643
>= 18 años	1	0.0515	0.0080	0.0357	0.0673

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

```
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	0.3618	0.0149	0.3320	0.3917
>= 18 años	0.6382	0.0149	0.6083	0.6680

## Estimación de la proporción de mujeres rango de edad

```
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	0.0512	0.0159	0.0198	0.0827
18 - 35	Inactive	0.4762	0.0358	0.4052	0.5471
18 - 35	Employed	0.4726	0.0291	0.4150	0.5302
Otro	Unemployed	0.0065	0.0029	0.0009	0.0122
Otro	Inactive	0.3922	0.0190	0.3545	0.4299
Otro	Employed	0.2663	0.0178	0.2310	0.3017
Otro	NA	0.3349	0.0180	0.2993	0.3705

## Estimación de la proporción de hombres rango de edad

```
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	0.1106	0.0221	0.0668	0.1544
18 - 35	Inactive	0.1537	0.0296	0.0949	0.2124
18 - 35	Employed	0.7358	0.0324	0.6716	0.7999
Otro	Unemployed	0.0312	0.0082	0.0149	0.0474
Otro	Inactive	0.1485	0.0159	0.1170	0.1799
Otro	Employed	0.3951	0.0202	0.3551	0.4352
Otro	NA	0.4253	0.0216	0.3825	0.4680

## Tabla Zona Vs Sexo

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

## Tabla Zona Vs Sexo

Sex	Zone	pobreza	prop	prop_se	prop_low	prop_upp
Female	Rural	0	0.5760	0.0472	0.4826	0.6693
Female	Rural	1	0.4240	0.0472	0.3307	0.5174
Female	Urban	0	0.6328	0.0299	0.5735	0.6920
Female	Urban	1	0.3672	0.0299	0.3080	0.4265
Male	Rural	0	0.5227	0.0655	0.3930	0.6524
Male	Rural	1	0.4773	0.0655	0.3476	0.6070
Male	Urban	0	0.5718	0.0370	0.4985	0.6451
Male	Urban	1	0.4282	0.0370	0.3549	0.5015

## Tablas de doble entrada.

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

	pobreza	SexFemale	SexMale
0	0	0.5521	0.4479
1	1	0.4925	0.5075

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

	pobreza	se.SexFemale	se.SexMale
0	0	0.0147	0.0147
1	1	0.0169	0.0169

## Tablas de doble entrada.

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

	2.5 %	97.5 %
0:SexFemale	0.5233	0.5810
1:SexFemale	0.4595	0.5256
0:SexMale	0.4190	0.4767
1:SexMale	0.4744	0.5405



## Prueba de independencia.

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

```
##
```

```
## Pearson's X2: Rao & Scott adjustment
```

```
##
```

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

```
## F = 7.7, ndf = 1, ddf = 119, p-value = 0.007
```

## Tablas de doble entrada.

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

	Employment	SexFemale	SexMale	se.SexFemale	se.SexMale
Unemployed	Unemployed	0.2981	0.7019	0.0678	0.0678
Inactive	Inactive	0.7565	0.2435	0.0210	0.0210
Employed	Employed	0.4284	0.5716	0.0186	0.0186

## Tablas de doble entrada

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

	2.5 %	97.5 %
Unemployed:SexFemale	0.1652	0.4310
Inactive:SexFemale	0.7155	0.7976
Employed:SexFemale	0.3919	0.4649
Unemployed:SexMale	0.5690	0.8348
Inactive:SexMale	0.2024	0.2845
Employed:SexMale	0.5351	0.6081

## Prueba de independencia.

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

```
##
```

```
##  Pearson's X2: Rao & Scott adjustment
```

```
##
```

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

```
## F = 59, ndf = 1.8, ddf = 219.1, p-value <2e-16
```

## Tablas de doble entrada.

```
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	0.4118	0.5882
Sur	Sur	0.6083	0.3917
Centro	Centro	0.7630	0.2370
Occidente	Occidente	0.6398	0.3602
Oriente	Oriente	0.5022	0.4978

## Tablas de doble entrada.

```
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	0.0571	0.0571
Sur	0.0594	0.0594
Centro	0.0437	0.0437
Occidente	0.0435	0.0435
Oriente	0.0928	0.0928

## Prueba de independencia.

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

```
##
```

```
##  Pearson's X2: Rao & Scott adjustment
```

```
##
```

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

```
## F = 4.2, ndf = 2.9, ddf = 343.3, p-value = 0.007
```

## Razón de obbs

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

	Sex	pobreza	se	ci_l	ci_u
Female	Female	0.3940	0.0273	0.3405	0.4474
Male	Male	0.4523	0.0377	0.3784	0.5261

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

```
##              nlcon    SE  
## contrast 0.871 0.09
```



## Razón de obbs

```
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	0.3194	0.0191
interaction(Sex, pobreza)Male.0	0.2591	0.0164
interaction(Sex, pobreza)Female.1	0.2076	0.0128
interaction(Sex, pobreza)Male.1	0.2139	0.0206

## Razón de obbs

```
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    1.27 0.11
```

## Razón de obbs

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

```
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  1.27 0.11
```

## Contrastes

```
(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	0.3940	0.0273	0.3405	0.4474
Male	Male	0.4523	0.0377	0.3784	0.5261

```
# Paso 1: diferencia de estimaciones  
0.3937      - 0.4047
```

```
## [1] -0.011
```

## contrastes

```
library(kableExtra)
# Paso 2: error estándar de la diferencia
vcov(tab_sex_pobreza)%>% data.frame() %>%
  kable(digits = 10,
        format.args = list(scientific = FALSE))
```

	Female	Male
Female	0.0007435	0.0008486
Male	0.0008486	0.0014202

```
sqrt(0.0007435 + 0.0014202 - 2*0.0008486)
```

```
## [1] 0.0216
```

## contrastes

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

	contrast	diff_Sex
diff_Sex	-0.0583	0.0216

## Contrastes

```
(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	0.0265	0.0071	0.0127	0.0404
Male	Male	0.0770	0.0143	0.0490	0.1050

*# diferencia de estimaciones*

0.0770 - 0.0265

## [1] 0.0505

# Contrastes

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

	Female	Male
Female	0.000049961	0.000009386
Male	0.000009386	0.000203901

```
sqrt(0.000049961 + 0.000203901 - 2*0.000009386)
```

```
## [1] 0.01533
```



# Contrastes

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

	contrast	diff_Sex
diff_Sex	0.0505	0.0153

## Contrastes

```
(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	0.0545	0.0173	0.0206	0.0883
Sur	Sur	0.0513	0.0140	0.0238	0.0788
Centro	Centro	0.0607	0.0170	0.0274	0.0940
Occidente	Occidente	0.0383	0.0113	0.0162	0.0603
Oriente	Oriente	0.0455	0.0202	0.0059	0.0852

# Contrastes

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

	Norte	Sur	Centro	Occidente	Oriente
Norte	0.0002981	0.0000000	0.0000000	0.0000000	0.0000000
Sur	0.0000000	0.0001968	0.0000000	0.0000000	0.0000000
Centro	0.0000000	0.0000000	0.0002884	0.0000000	0.0000000
Occidente	0.0000000	0.0000000	0.0000000	0.0001267	0.0000000
Oriente	0.0000000	0.0000000	0.0000000	0.0000000	0.0004093

## Contrastes

```
(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	0.5464	0.0634	0.4222	0.6706
Sur	Sur	0.3515	0.0557	0.2424	0.4606
Centro	Centro	0.2020	0.0389	0.1257	0.2783
Occidente	Occidente	0.3104	0.0441	0.2240	0.3968
Oriente	Oriente	0.4679	0.0870	0.2974	0.6384