Variable continua

CEPAL

14/2/2022

Lectura de la base

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

Definir diseño de la muestra con `srvyr`

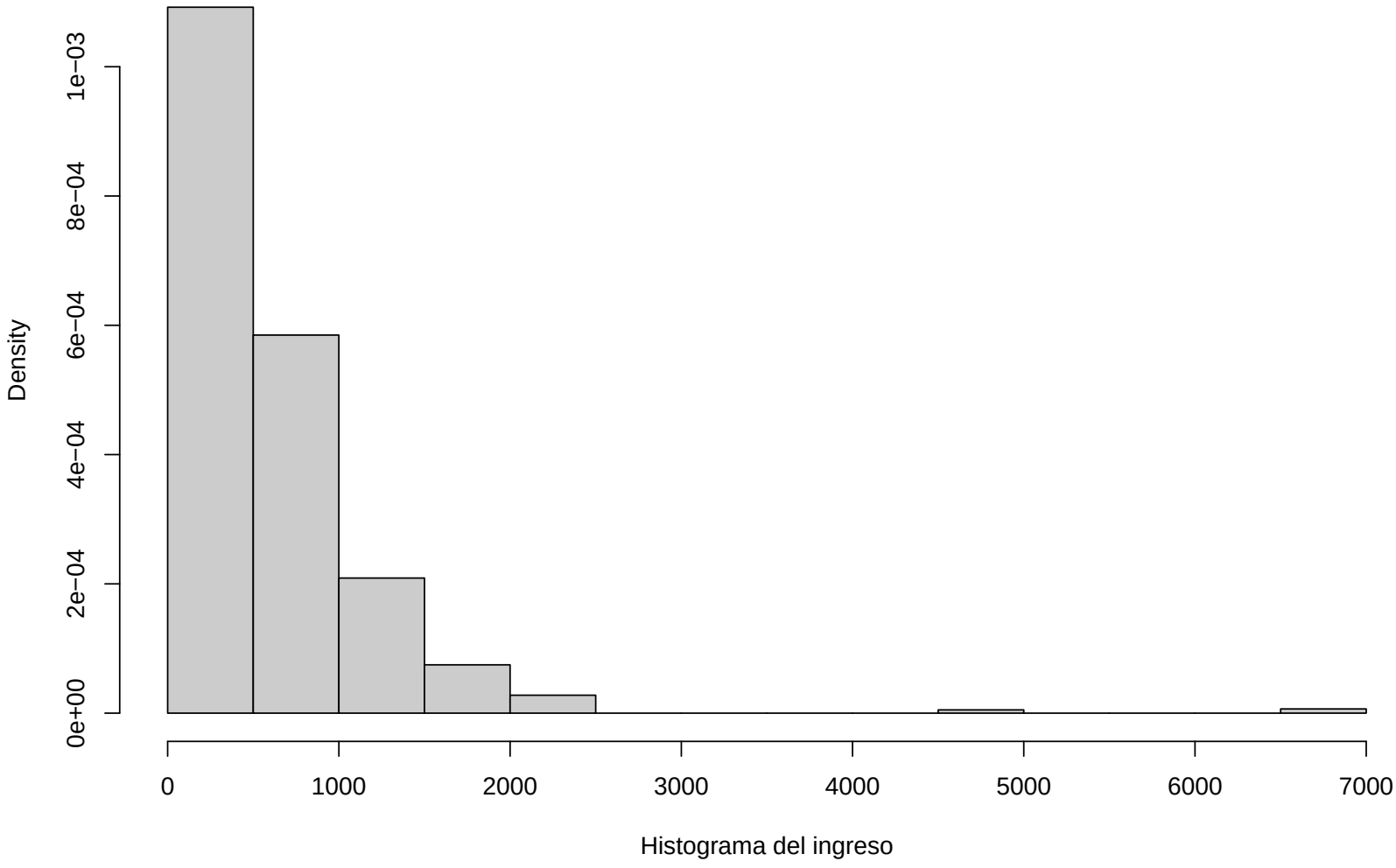
```
library(srvyr)

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

Histograma ponderado para la variable ingreso

```
svyhist(
  ~ Income ,
  diseno,
  main = "",
  col = "grey80",
  xlab = "Histograma del ingreso"
)
```

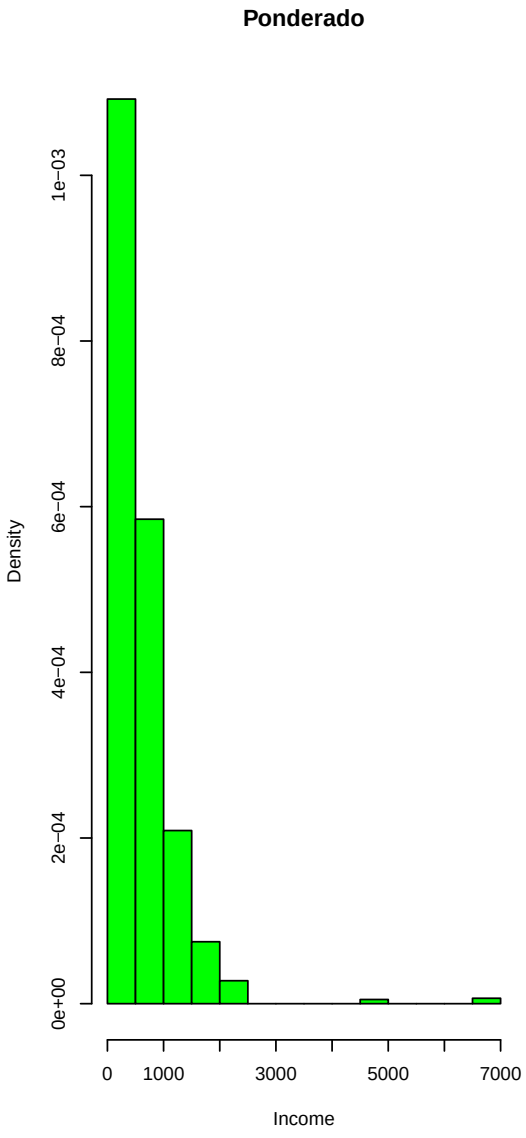
Histograma ponderado para la variable ingreso

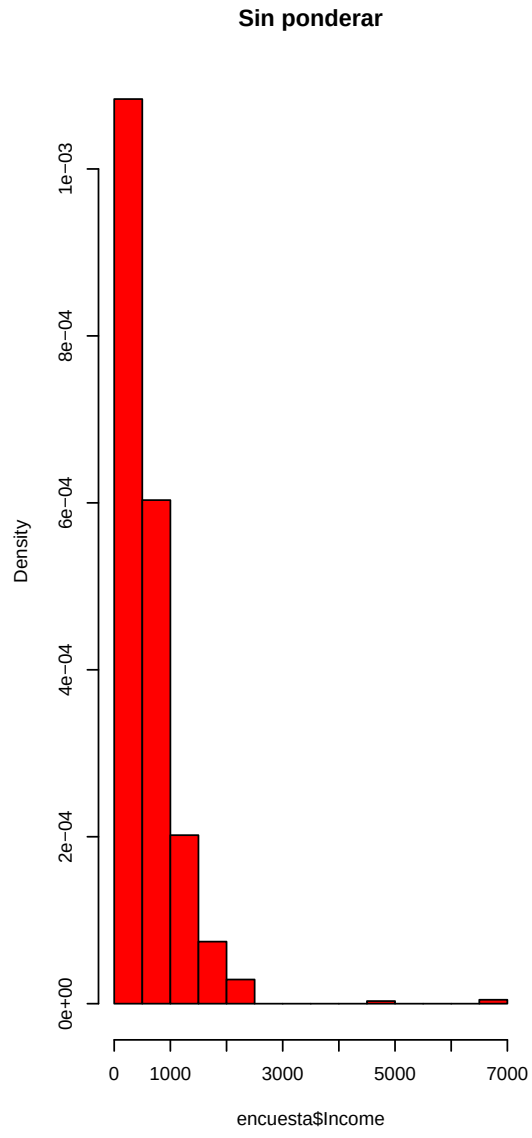


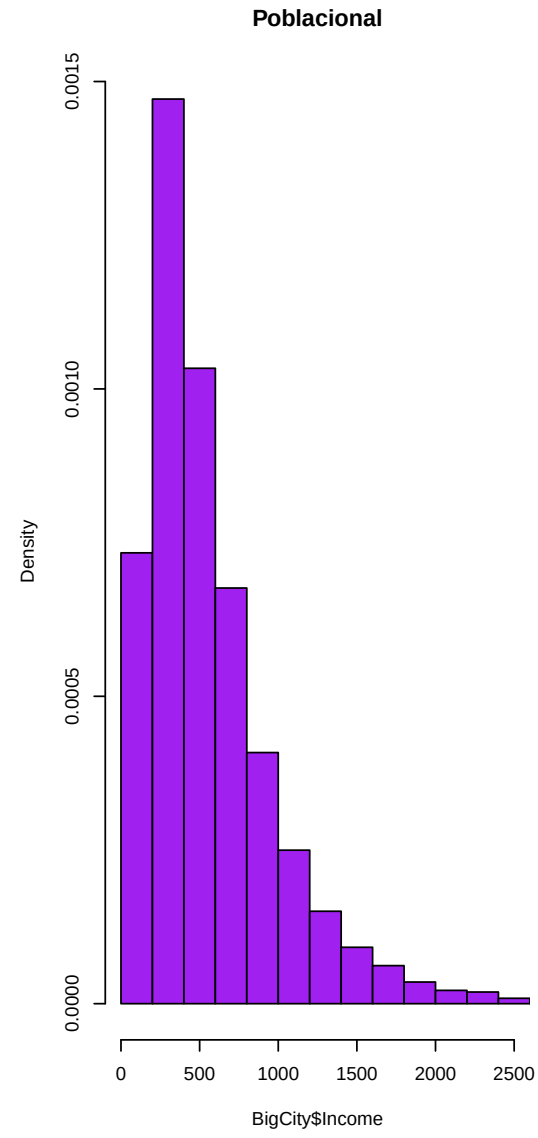
Comparación de histogramas

```
data("BigCity", package = "TeachingSampling")
par(mfrow = c(1,3))
svyhist(   ~ Income,
  diseno, main = "Ponderado",
  col = "green"
)
hist( encuesta$Income,
  main = "Sin ponderar",
  col = "red", prob = TRUE
)
hist(  BigCity$Income,
  main = "Poblacional",
  col = "purple", prob = TRUE,
  xlim = c(0, 2500), breaks = 200
)
```

Comparación de histogramas







Sub-grupos

Extraer sub-grupos de la encuesta.

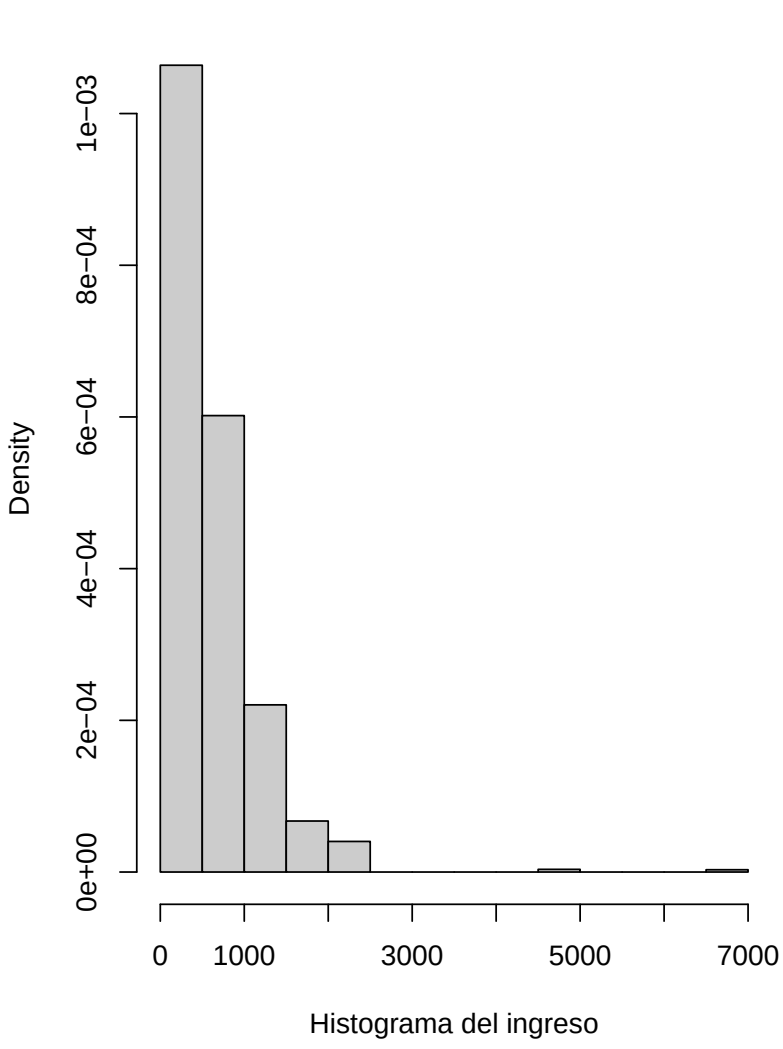
```
sub_Urbano <- diseno %>%  filter(Zone == "Urban")
sub_Rural  <- diseno %>%  filter(Zone == "Rural")
sub_Mujer  <- diseno %>%  filter(Sex == "Female")
sub_Hombre <- diseno %>%  filter(Sex == "Male")
```

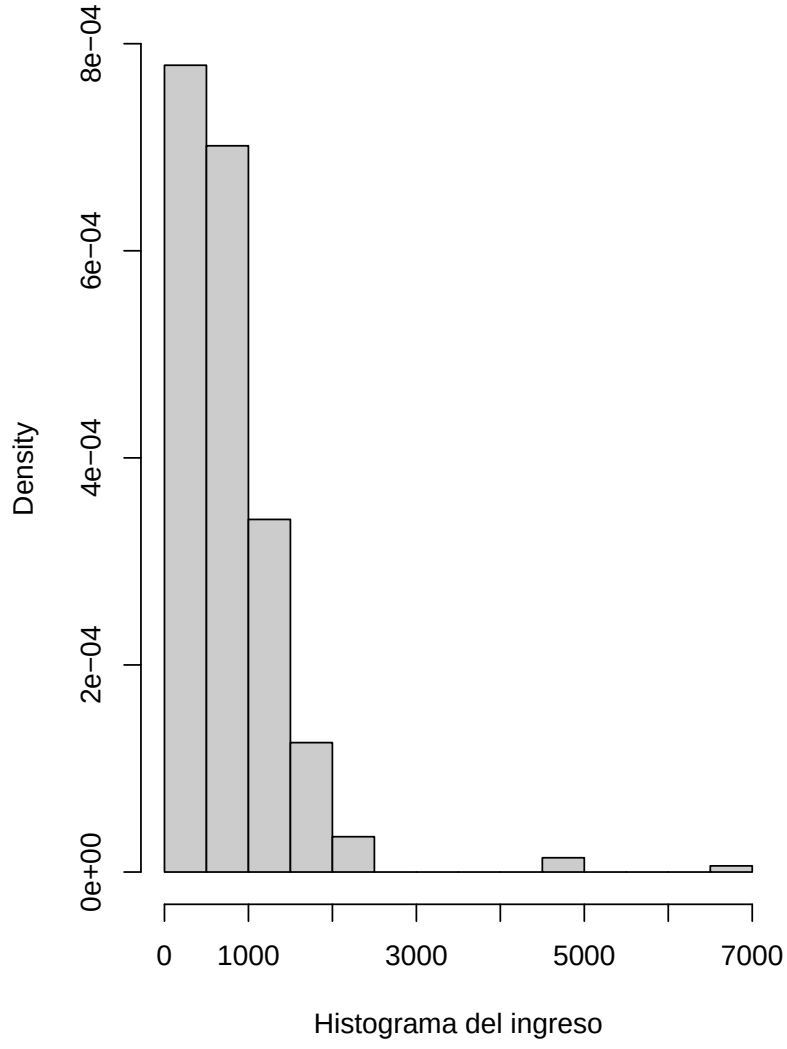
Histograma ponderado en sub-grupos

```
par(mfrow = c(1,2))
svyhist(
  ~ Income ,
  subset (sub_Mujer, Age >= 18),
  main = "",
  col = "grey80",
  xlab = "Histograma del ingreso"
)

svyhist(
  ~ Income ,
  subset (sub_Urbano, Age >= 18),
  main = "",
  col = "grey80",
  xlab = "Histograma del ingreso"
)
```

Histograma ponderado en sub-grupos



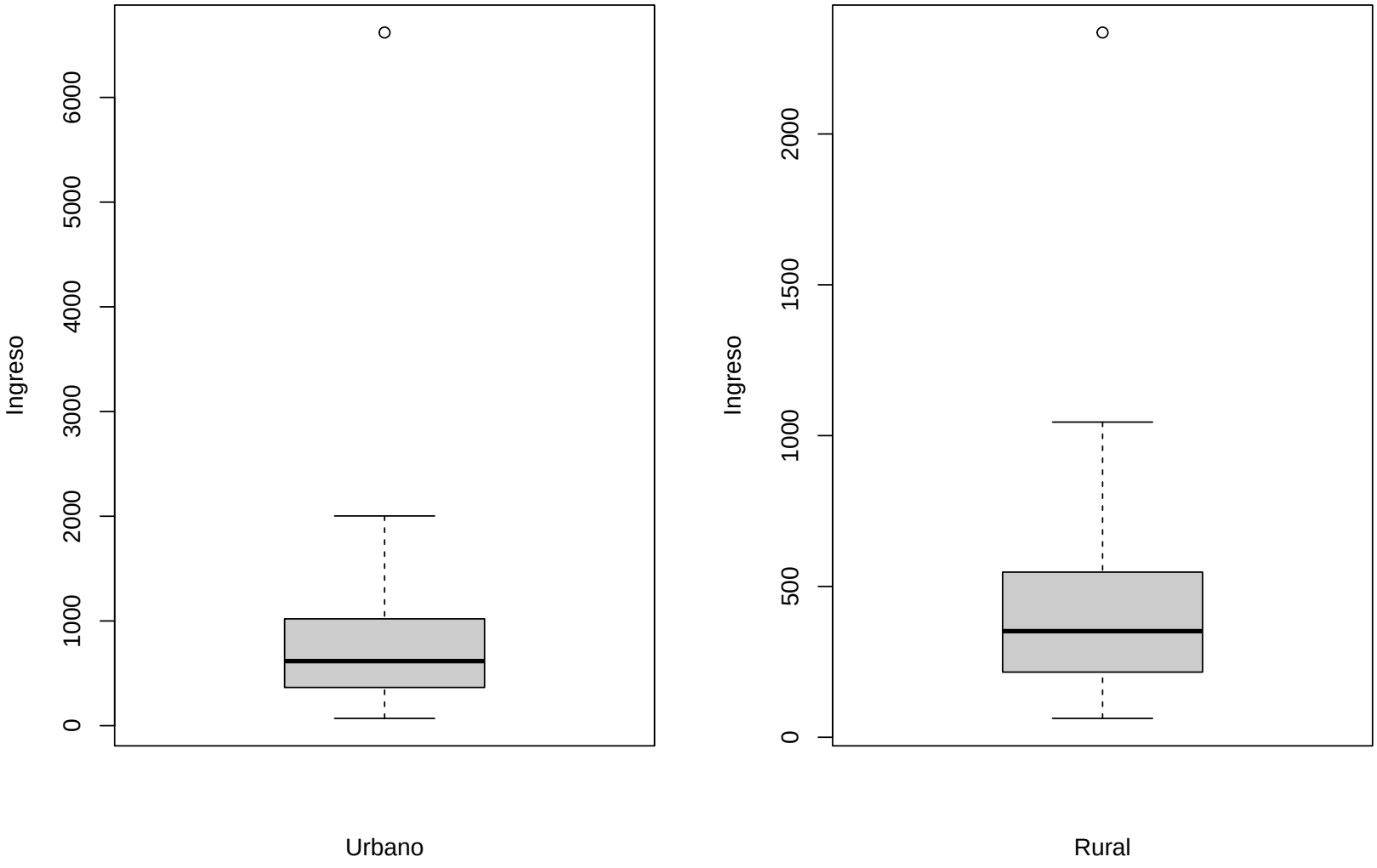


Boxplot ponderado del ingreso por sub-grupos

```
par(mfrow = c(1,2))
svyboxplot(
  Income ~1 ,
  sub_Urbano,
  col = "grey80",
  ylab = "Ingreso",
  xlab = "Urbano")

svyboxplot(
  Income ~ 1 ,
  sub_Rural,
  col = "grey80",
  ylab = "Ingreso",
  xlab = "Rural"
)
```

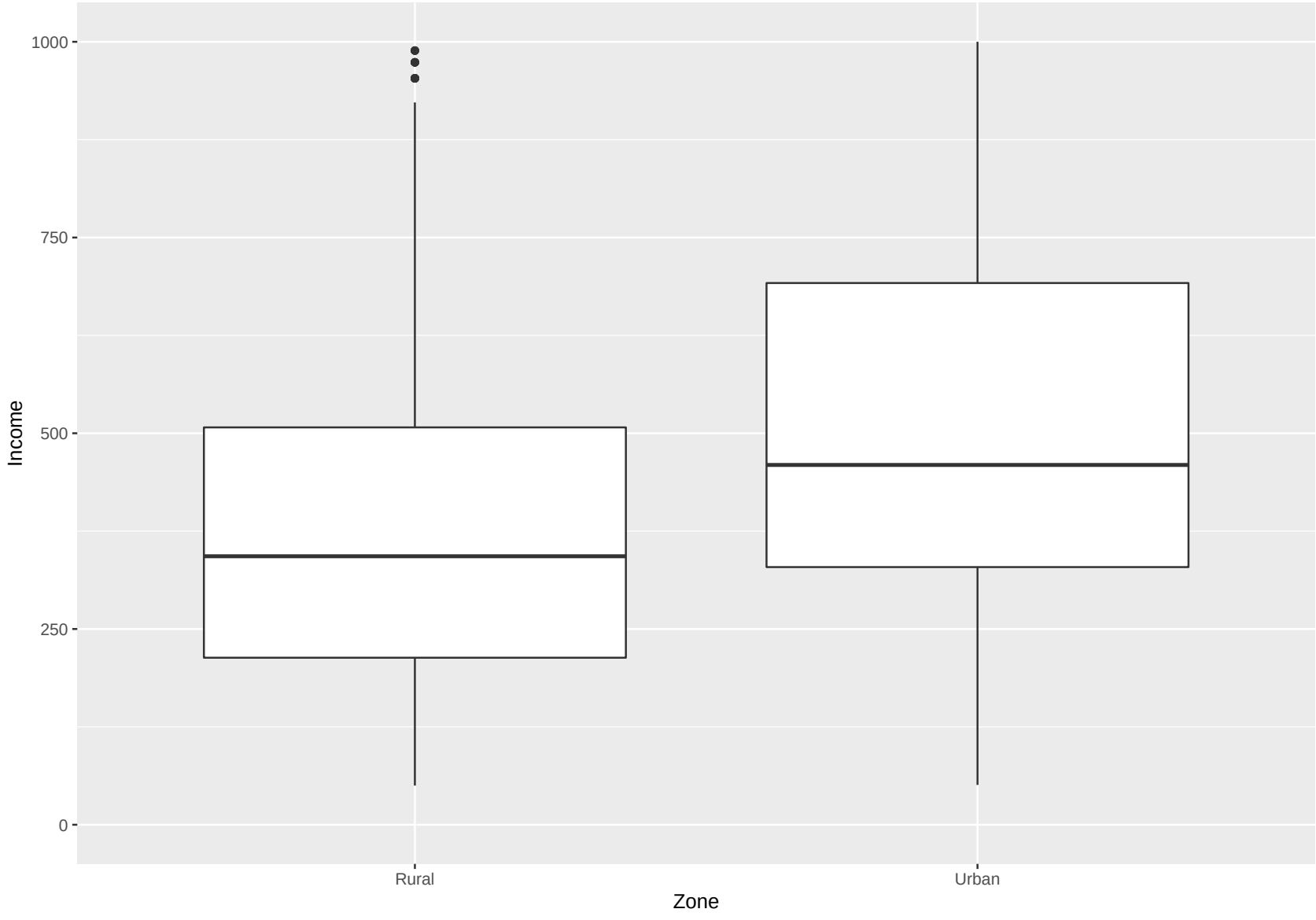
Boxplot ponderado del ingreso por sub-grupos



Boxplot ponderado del ingreso por sub-grupos alternativa

```
library(ggplot2)
ggplot(data = encuesta,
       aes(y = Income, x = Zone, weight = wk)) +
  geom_boxplot() + ylim(c(0, 1000))
```

Boxplot ponderado del ingreso por sub-grupos alternativa



Estimación de totales e intervalos de confianza del ingreso

```
svytotal (~Income, diseno, deff=T) %>%
  data.frame()
```

	total	Income	deff
Income	93991744	4799513	7.408

```
confint(svytotal (~Income, diseno, deff=T))
```

	2.5 %	97.5 %
Income	84584871	103398616

Estimación de totales e intervalos de confianza del gasto

```
svytotal (~Expenditure, diseno, deff=T) %>%
  data.frame()
```

	total	Expenditure	deff
Expenditure	58161637	3077914	16.24

```
confint(svytotal (~Expenditure, diseno, deff=T))
```

	2.5 %	97.5 %
Expenditure	52129036	64194238

Estimación de totales por sub-grupos

```
diseno %>% group_by(Sex) %>%
  summarise(
    Total = survey_total(Income, level = 0.95,
                         vartype =  c("se", "ci") ))
```

Sex	Total	Total_se	Total_low	Total_upp
Female	49267737	2684653	43951855	54583618
Male	44724007	2693747	39390119	50057895

Estimación de la media e intervalo de confianza del ingreso

```
svymean(~Income, diseno, deff=T) %>%
  data.frame()
```

	mean	Income	deff
Income	625.5	23.44	3.989

```
confint(svymean (~Income, diseno, deff=T))
```

	2.5 %	97.5 %
Income	579.6	671.4

Estimación de la media e intervalo de confianza del gasto

```
svymean (~Expenditure, diseno, deff=T) %>%
  data.frame()
```

	mean	Expenditure	deff
Expenditure	387.1	13.51	7.065

```
confint(svymean (~Expenditure, diseno, deff=T))
```

	2.5 %	97.5 %
Expenditure	360.6	413.5

Estimación de la media por sub-grupos

```
diseno %>% group_by(Sex) %>%
  summarise(Media = survey_mean(Expenditure, level = 0.95,
                                vartype =  c("se", "ci")))
```

Sex	Media	Media_se	Media_low	Media_upp
Female	388.8	13.65	361.7	415.8
Male	385.2	14.75	356.0	414.4

```
diseno %>% group_by(Zone) %>%
  summarise(Media = survey_mean(Expenditure, level = 0.95,
                                vartype =  c("se", "ci")))
```

Zone	Media	Media_se	Media_low	Media_upp
Rural	286.7	12.52	261.9	311.5
Urban	479.6	20.97	438.1	521.2

Estimación de medias por sub-grupos

```
diseno %>% group_by(Zone, Sex) %>%
  summarise(
    Media = survey_mean(Expenditure, level = 0.95,
                        vartype =  c("se", "ci"))) %>%
  data.frame()
```

Zone	Sex	Media	Media_se	Media_low	Media_upp
Rural	Female	288.9	13.67	261.8	316.0
Rural	Male	284.3	12.34	259.9	308.8
Urban	Female	477.4	19.78	438.2	516.6
Urban	Male	482.2	24.31	434.1	530.4

Estimación de la desviación estándar de los ingresos por sub-grupo

```
(tab_sd <- diseno %>% group_by(Zone) %>%
   summarise(Sd = sqrt(
  survey_var(
    Income,
    level = 0.95,
    vartype =  c("se", "ci"),
  )
)))
```

Zone	Sd	Sd_se	Sd_low	Sd_upp
Rural	337.5	183.6	217.1	425.0
Urban	725.3	457.4	334.4	969.8

Estimación de la desviación estándar de los ingresos por sub-grupo

```
(tab_sd <- diseno %>% group_by(Zone, Sex) %>%
   summarise(Sd = sqrt(
  survey_var(
    Income,
    level = 0.95,
    vartype =  c("se", "ci"),
   )
))) %>% data.frame()
```

Zone	Sex	Sd	Sd_se	Sd_low	Sd_upp
Rural	Female	334.8	194.0	193.7	432.0
Rural	Male	340.5	176.5	233.0	421.4
Urban	Female	739.4	502.4	216.7	1022.9
Urban	Male	709.3	422.7	386.4	925.6

Estimación de la mediana para gastos

```
diseno %>% summarise(Mediana =
  survey_median(
    Expenditure,
    level = 0.95,
    vartype =  c("se", "ci"),
   ))
```

Mediana	Mediana_se	Mediana_low	Mediana_upp
310.6	11.72	289.5	335.9

Estimación de la mediana por sub-grupo

```
diseno %>% group_by(Zone) %>%
  summarise(Mediana =
  survey_median(
    Expenditure,
    level = 0.95,
    vartype =  c("se", "ci"),
   ))
```

Zone	Mediana	Mediana_se	Mediana_low	Mediana_upp
Rural	257.8	12.80	231.5	282.8
Urban	399.4	24.91	359.3	458.8

Estimación de la mediana por sub-grupo

```
diseno %>% group_by(Sex) %>%
  summarise(Mediana =
  survey_median(
    Expenditure,
    level = 0.95,
    vartype =  c("se", "ci"),
   ))
```

Sex	Mediana	Mediana_se	Mediana_low	Mediana_upp
Female	328.7	12.37	294.6	343.6
Male	299.5	11.91	286.7	333.8

Estimación de quantile para el gasto

```
diseno %>%
  summarise(
    Q =  survey_quantile(
    Expenditure,
    quantiles = 0.5,
    level = 0.95,
    vartype =  c("se", "ci"),
    interval_type = "score"
   ))
```

Q_q50	Q_q50_se	Q_q50_low	Q_q50_upp
310.6	12.12	289.5	337.5

Estimación de quantile para el gasto por sub-grupo

```
diseno %>% group_by(Sex) %>%
  summarise(
    Q =  survey_quantile(
    Expenditure,
    quantiles = 0.25,
    level = 0.95,
    vartype =  c("se", "ci"),
    interval_type = "score"
   ))
```

Sex	Q_q25	Q_q25_se	Q_q25_low	Q_q25_upp
Female	217.7	10.12	197.3	237.4
Male	205.3	10.62	191.6	233.7

Estimación de quantile para el gasto por sub-grupo

```
diseno %>% group_by(Zone) %>%
  summarise(
    Q =  survey_quantile(
    Expenditure,
    quantiles = 0.25,
    level = 0.95,
    vartype =  c("se", "ci"),
    interval_type = "score"
   ))
```

Zone	Q_q25	Q_q25_se	Q_q25_low	Q_q25_upp
Rural	173.1	10.798	149.3	192.6
Urban	272.9	8.918	259.0	294.6

Estimación de la razón entre el gasto y el ingreso

```
diseno %>% summarise(
    Razon =  survey_ratio(
      numerator = Expenditure,
      denominator = Income,
      level = 0.95,
    vartype =  c("se", "ci")
    ))
```

Razon	Razon_se	Razon_low	Razon_upp
0.6188	0.0195	0.5801	0.6575

Estimación de la razón entre hombres y mujeres

```
diseno %>% summarise(
    Razon =  survey_ratio(
      numerator = (Sex == "Female"),
      denominator = (Sex == "Male"),
      level = 0.95,
    vartype =  c("se", "ci")
    ))
```

Razon	Razon_se	Razon_low	Razon_upp
1.114	0.0498	1.016	1.213

Estimación de la razón entre hombres y mujeres en la zona rural

```
sub_Rural %>% summarise(
    Razon =  survey_ratio(
      numerator = (Sex == "Female"),
      denominator = (Sex == "Male"),
      level = 0.95,
    vartype =  c("se", "ci")
    ))
```

Razon	Razon_se	Razon_low	Razon_upp
1.068	0.0779	0.9119	1.224

Estimación de la razón del gastos y los ingreso entre las mujeres

```
sub_Mujer %>% summarise(
    Razon =  survey_ratio(
      numerator = Expenditure,
      denominator = Income,
      level = 0.95,
    vartype =  c("se", "ci")
    ))
```

Razon	Razon_se	Razon_low	Razon_upp
0.6249	0.0229	0.5796	0.6702

Estimación de la razón del gasto y los ingresos entre los hombres

```
sub_Hombre %>% summarise(
    Razon =  survey_ratio(
      numerator = Expenditure,
      denominator = Income,
      level = 0.95,
    vartype =  c("se", "ci")
    ))
```

Razon	Razon_se	Razon_low	Razon_upp
0.6121	0.0185	0.5754	0.6488

Estimación del índice de GINI

La estimación del índice de GINI se realiza haciendo uso de la librería convey

```
library(convey)
diseno_gini <- convey_prep(diseno)
svygini( ~Income, design = diseno_gini) %>%
  data.frame()
```

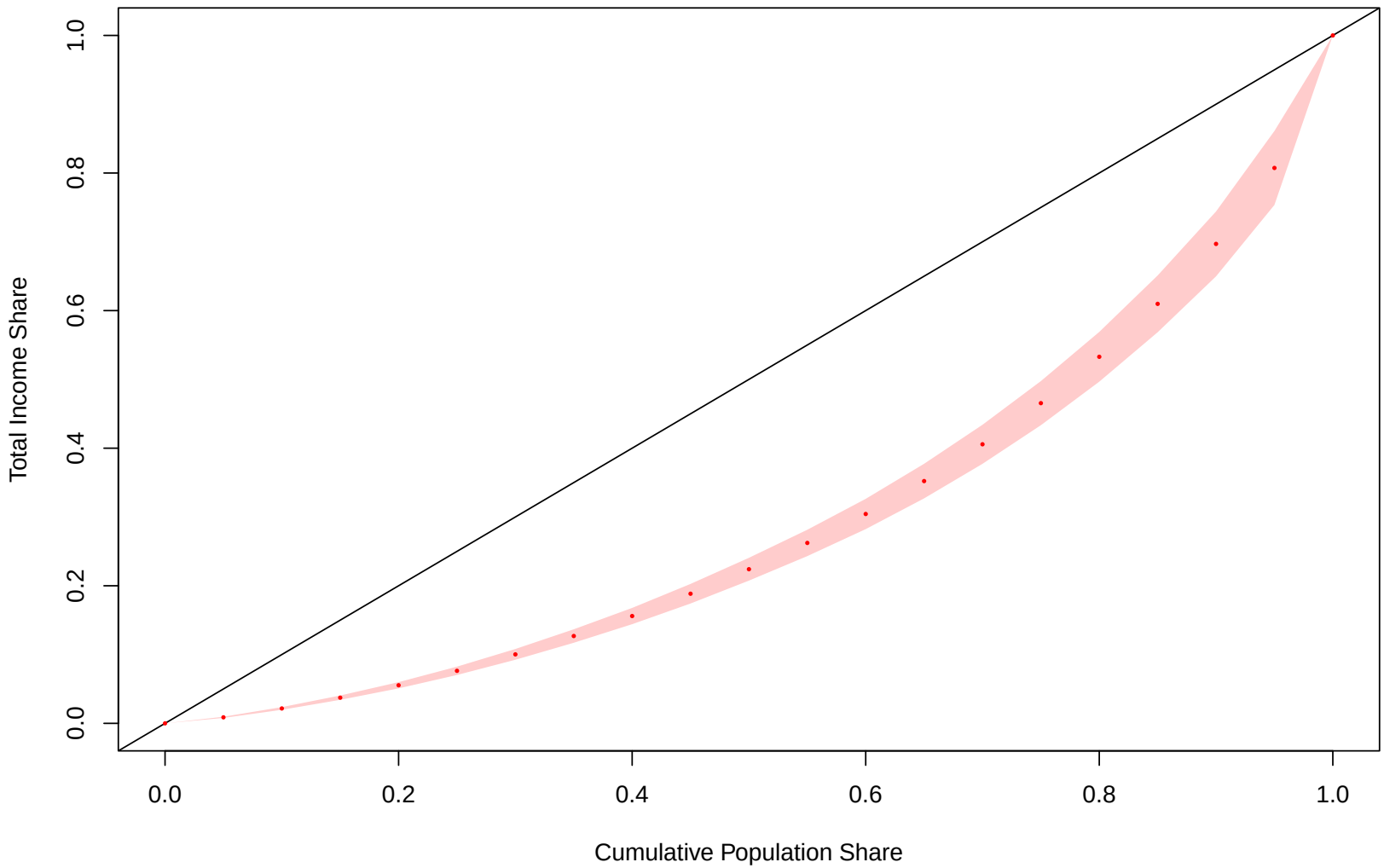
	gini	Income
Income	0.4102	0.0159

```
svygini( ~Expenditure, design = diseno_gini) %>%
  data.frame()
```

	gini	Expenditure
Expenditure	0.3427	0.0107

Estimación del índice de GINI por sub-grupo

```
svylorenz( ~Income, diseno_gini,
           seq(0,1,.05), alpha = .01 )
```



Pruebas de diferencia medias de los ingresos entre hombres y mujeres

```
svyttest(Income ~ Sex, diseno)
```

```
##
##  Design-based t-test
##
## data:  Income ~ Sex
## t = 0.35, df = 118, p-value = 0.7
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -33.41  47.60
## sample estimates:
## difference in mean
##              7.096
```

Pruebas de diferencia medias de los ingresos entre hombres y mujeres en la zona urbana

```
svyttest(Income ~ Sex, sub_Urbano)
```

```
##
##  Design-based t-test
##
## data:  Income ~ Sex
## t = 0.43, df = 63, p-value = 0.7
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -53.84  83.74
## sample estimates:
## difference in mean
##              14.95
```

Pruebas de diferencia medias de los ingresos entre hombres y mujeres mayores a 18 años

```
svyttest(Income ~ Sex, diseno %>% filter(Age > 18))
```

```
##
##  Design-based t-test
##
## data:  Income ~ Sex
## t = 0.51, df = 118, p-value = 0.6
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -39.22  66.62
## sample estimates:
## difference in mean
##               13.7
```

Contrastes

```
(prom_region <- svyby(~Income, ~Region, diseno,
                      svymean, na.rm=T, covmat = TRUE,
                      vartype = c("se", "ci")))
```

	Region	Income	se	ci_l	ci_u
Norte	Norte	527.4	43.69	441.8	613.1
Sur	Sur	627.1	53.44	522.4	731.8
Centro	Centro	768.0	82.86	605.6	930.4
Occidente	Occidente	575.5	53.08	471.4	679.5
Oriente	Oriente	650.2	32.70	586.2	714.3

```
# Paso 1: diferencia de estimaciones (Norte - Sur)
527 - 627
```

```
## [1] -100
```

contrastes

```
# Paso 2: error estándar de la diferencia
vcov(prom_region)
```

	Norte	Sur	Centro	Occidente	Oriente
Norte	1909	0	0	0	0
Sur	0	2856	0	0	0
Centro	0	0	6867	0	0
Occidente	0	0	0	2817	0
Oriente	0	0	0	0	1069

```
sqrt(1909 + 2856 - 2*0)
```

```
## [1] 69.03
```

contrastes

```
svycontrast(prom_region,
            list(diff_NS = c(1, -1, 0, 0, 0))) %>%
  data.frame()
```

	contrast	diff_NS
diff_NS	-99.66	69.03

Contrastes

```
svycontrast(prom_region, list(
                Norte_sur = c(1, 1, 0, 0, 0),
                centro = c(0, 0, 1, 0, 0),
                Occidente_Oriente = c(0, 0, 0, 1, 1)
              )) %>% data.frame()
```

	contrast	SE
Norte_sur	1155	69.03
centro	768	82.86
Occidente_Oriente	1226	62.34

```
sqrt(1909 + 2856 - 2*0) ; sqrt(6867);
```

```
## [1] 69.03
```

```
## [1] 82.87
```

```
sqrt(2817 + 1069 - 2*0)
```

Contrastes no independiente

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

	Sex	Income	se	ci_l	ci_u
Female	Female	622.1	24.13	574.9	669.4
Male	Male	629.2	27.08	576.2	682.3

Contrastes no independiente

```
svycontrast(prom_sexo, list(diff_Sexo = c(1, -1))) %>%
  data.frame()
```

	contrast	diff_Sexo
diff_Sexo	-7.096	20.45

Contrastes no independiente

```
vcov(prom_sexo)
```

	Female	Male
Female	582.3	448.5
Male	448.5	733.2

```
# Note que el error estandar  de la diferencia es igual a
sqrt(582 + 733 - 2*449)
```

```
## [1] 20.42
```

Contrastes no independiente

```
(sum_region <- svyby( ~ Income,  ~ Region,
                      diseno, svytotal, na.rm = T,
                      covmat = TRUE,
                      vartype = c("se", "ci")))
```

	Region	Income	se	ci_l	ci_u
Norte	Norte	14023740	1674859	10741075	17306404
Sur	Sur	16623420	1959403	12783061	20463779
Centro	Centro	18143894	2749370	12755229	23532560
Occidente	Occidente	20164710	2563552	15140241	25189179
Oriente	Oriente	25035980	1503357	22089454	27982505

Contrastes no independiente

```
svycontrast(sum_region,
            list(
              Agregado_NCS = c(1, 1, 1, 0, 0)
              )) %>% data.frame()
```

	contrast	Agregado_NCS
Agregado_NCS	48791054	3768746

```
# Note que el error estandar  de la diferencia es igual a
sqrt(582 + 733 - 2*449)
```

```
## [1] 20.42
```

Contrastes

```
vcov(sum_region)
```

	Norte	Sur	Centro	Occidente	Orien
Norte	2.805e+12	0.000e+00	0.000e+00	0.000e+00	0.00e+
Sur	0.000e+00	3.839e+12	0.000e+00	0.000e+00	0.00e+
Centro	0.000e+00	0.000e+00	7.559e+12	0.000e+00	0.00e+
Occidente	0.000e+00	0.000e+00	0.000e+00	6.572e+12	0.00e+
Oriente	0.000e+00	0.000e+00	0.000e+00	0.000e+00	2.26e+

```
sqrt(2805154074898 + 3839259031856 + 7559032807016 )
```

```
## [1] 3768746
```

Contrastes no independiente

```
(prom_edad <- svyby(~Income, ~CatAge, diseno, svymean, na.r
```

	CatAge	Income	se
0-5	0-5	602.2	69.91
16-30	16-30	654.8	25.42
31-45	31-45	655.5	40.14
46-60	46-60	614.2	46.88
6-15	6-15	595.5	34.36
Más de 60	Más de 60	580.3	74.52

```
svycontrast(prom_edad,
            list(
              agregado_edad = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/
              )) %>% data.frame()
```

	contrast	agregado_edad
agregado_edad	617.1	25.67

Contrastes no independiente

```
vcov(prom_edad)
```

	0-5	16-30	31-45	46-60	6-15	Más de 60
0-5	4887.84	22.12	-1289.4	863.5	-1377.8	198.0
16-30	22.12	646.08	453.7	441.5	347.1	856.2
31-45	-1289.37	453.72	1610.9	290.4	819.6	1336.7
46-60	863.52	441.50	290.4	2197.5	102.6	487.6
6-15	-1377.76	347.09	819.6	102.6	1180.5	268.1
Más de 60	198.02	856.17	1336.7	487.6	268.1	5553.0

```
(1/6)*sqrt(4888 + 646 + 1611 + 2197 + 1181 + 5553
       + 2*22 + 2*(-1289) + 2*864 + 2*(-1387) + 2*189 +
       2*454 + 2*441 + 2*347 + 2*856 +
       2*290 + 2*820 + 2*1337+
       2*103 + 2*488 +
       2*268)
```

```
## [1] 25.65
```

Contrastes no independiente

```
(razon_sexo <- svyby(~Income, ~Sex, denominator = ~Expendit
                     diseno, svyratio, na.rm=T, covmat = TR
```

	Sex	Income/Expenditure	se.Income/Expenditure	ci_l
Female	Female	1.600	0.0586	1.486
Male	Male	1.634	0.0495	1.537

```
svycontrast(razon_sexo,
            list(
              diff_sexo = c(1, -1)
              )) %>% data.frame()
```

	contrast	diff_sexo
diff_sexo	-0.0334	0.0383

Contrastes no independiente

```
vcov(razon_sexo)
```

	Female	Male
Female	0.0034	0.0022
Male	0.0022	0.0024

```
sqrt(0.0034 + 0.0024 - 2*0.0022)
```

```
## [1] 0.03742
```

Correlación de variables

```
library(jtools)

svycor( ~ Income + Expenditure, design = diseno)$cors %>% c
```

	Income	Expenditure
Income	1.0000	0.7082
Expenditure	0.7082	1.0000

```
svycor( ~ Income + Expenditure, design = sub_Hombre)$cors %
```

	Income	Expenditure
Income	1.0000	0.7333
Expenditure	0.7333	1.0000

```
svycor( ~ Income + Expenditure, design = sub_Mujer)$cors %>
```

	Income	Expenditure