Variable continua

CEPAL

14/2/2022

Lectura de la base

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

Definir diseño de la muestra con `srvyr`

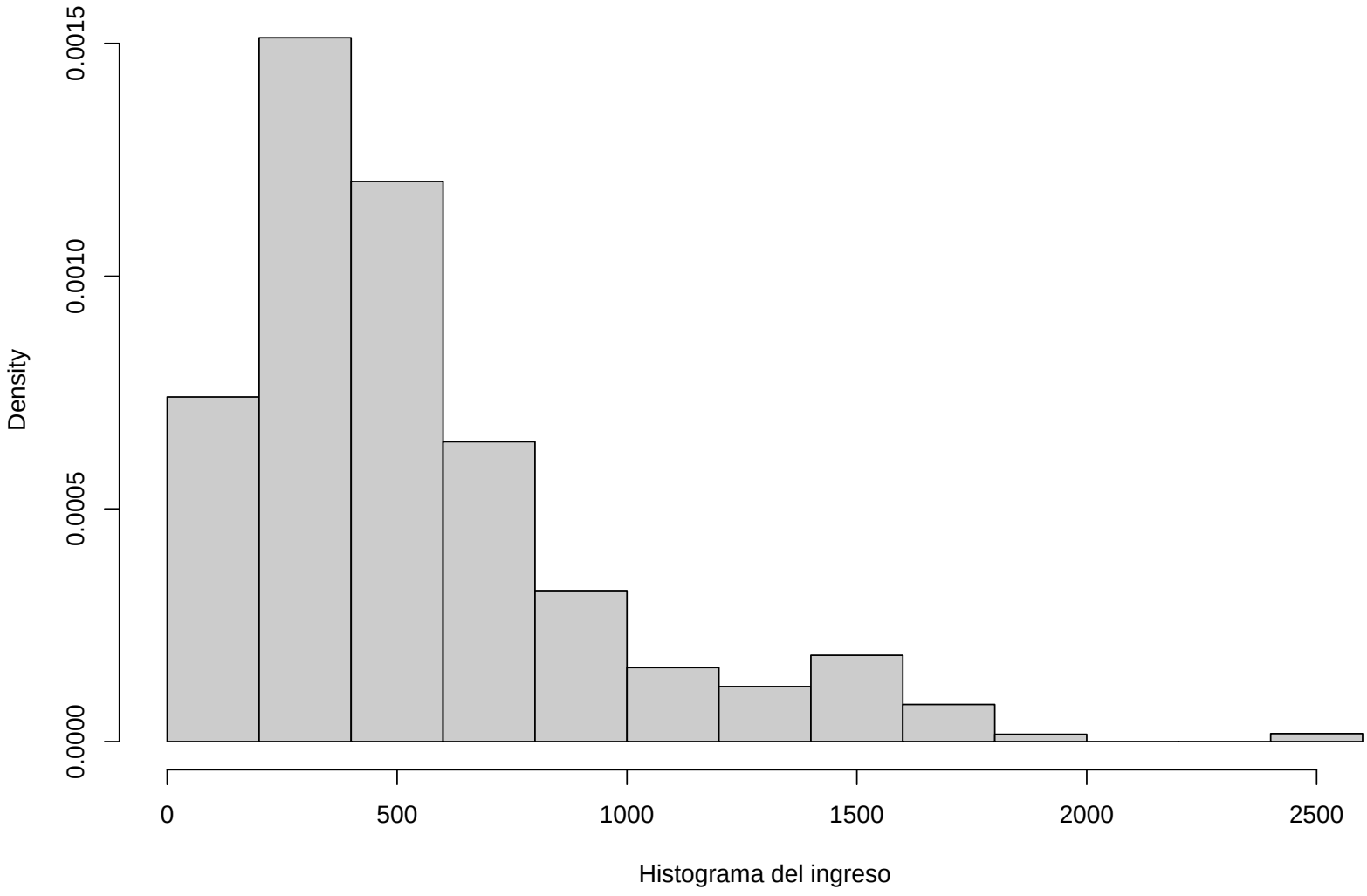
```
library(srvyr)

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"
)
```

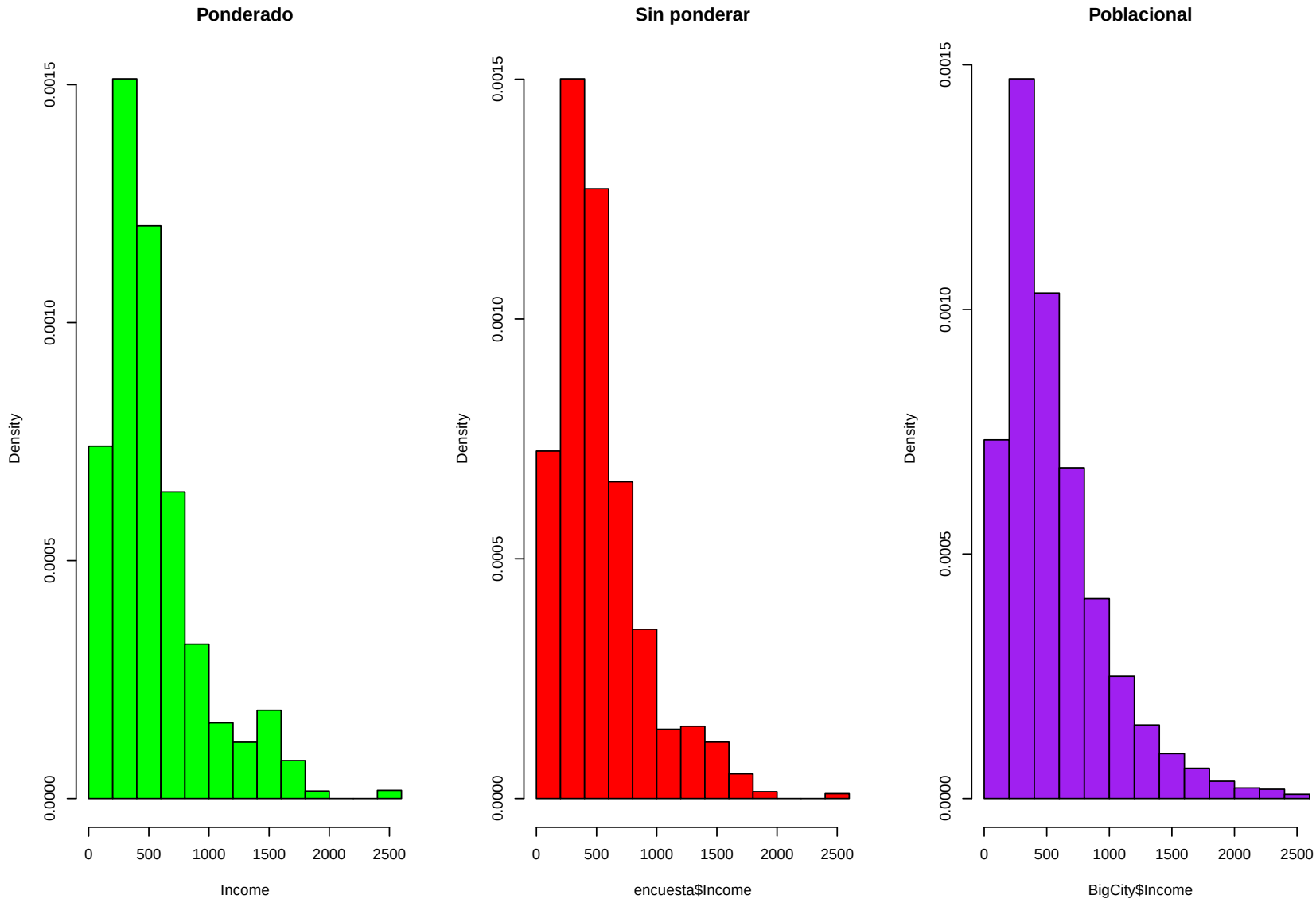
Histograma ponderado para la variable ingreso



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.

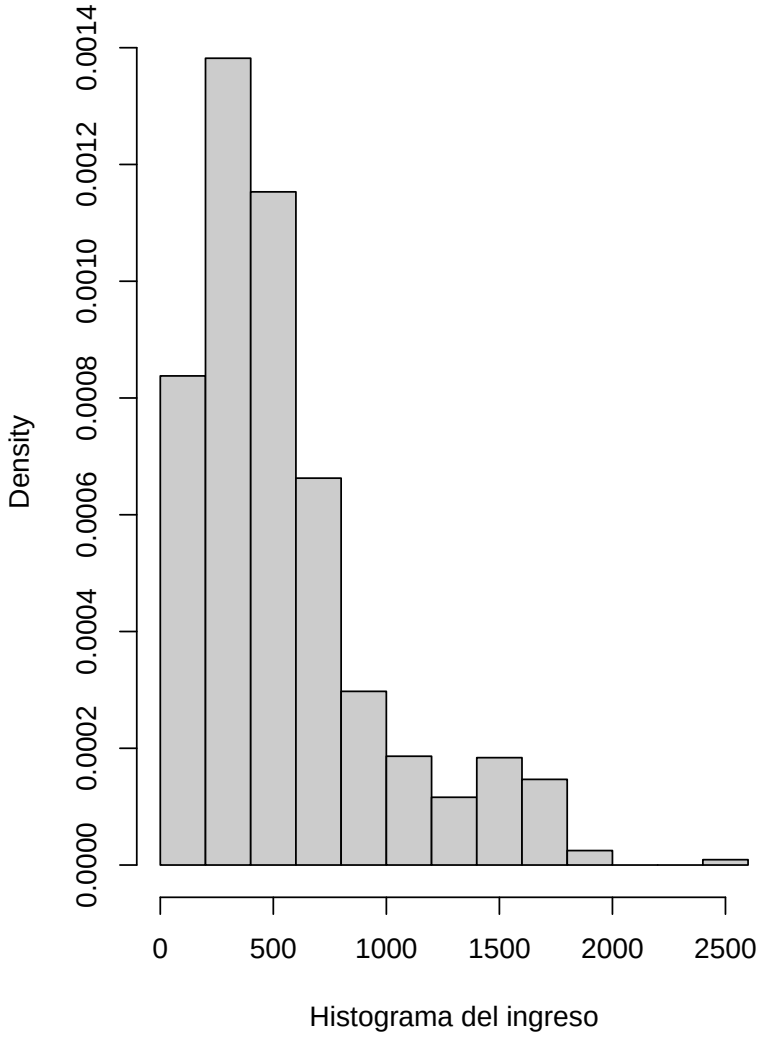
```
sub_Urbano <- diseno %>%  filter(Zone == "Urban")
sub_Rural  <- diseno %>%  filter(Zone == "Rural")
sub_Mujer  <- diseno %>%  filter(Sex == "Female")
sub_Hombre <- diseno %>%  filter(Sex == "Male")
```

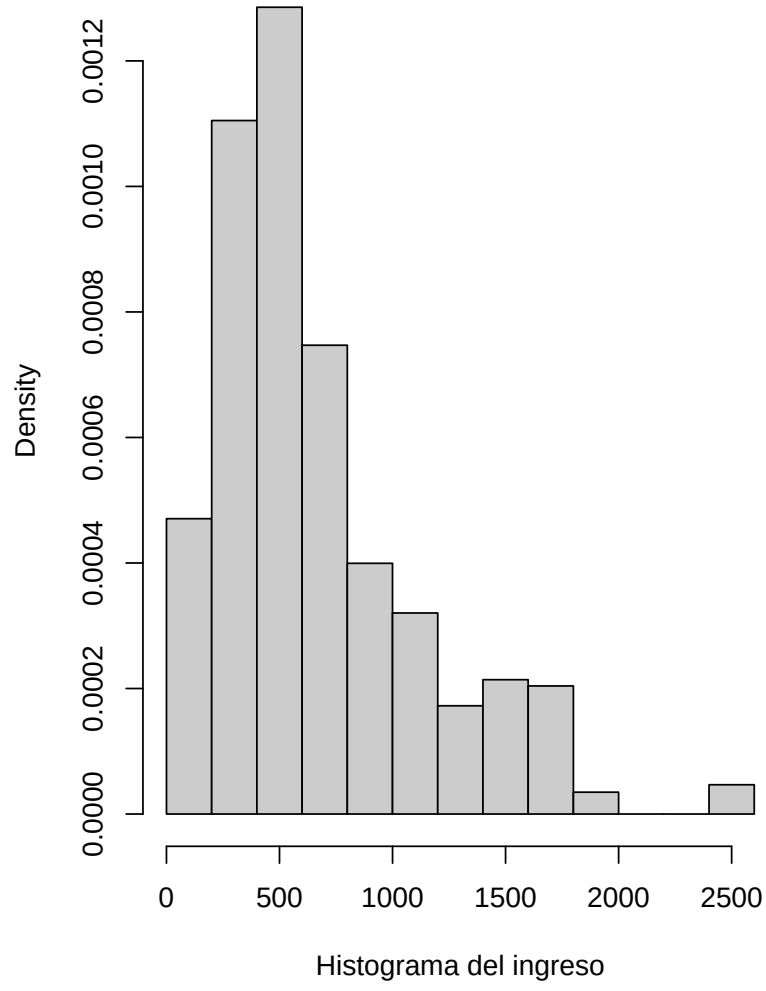
Histograma ponderado en sub-grupos

```
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



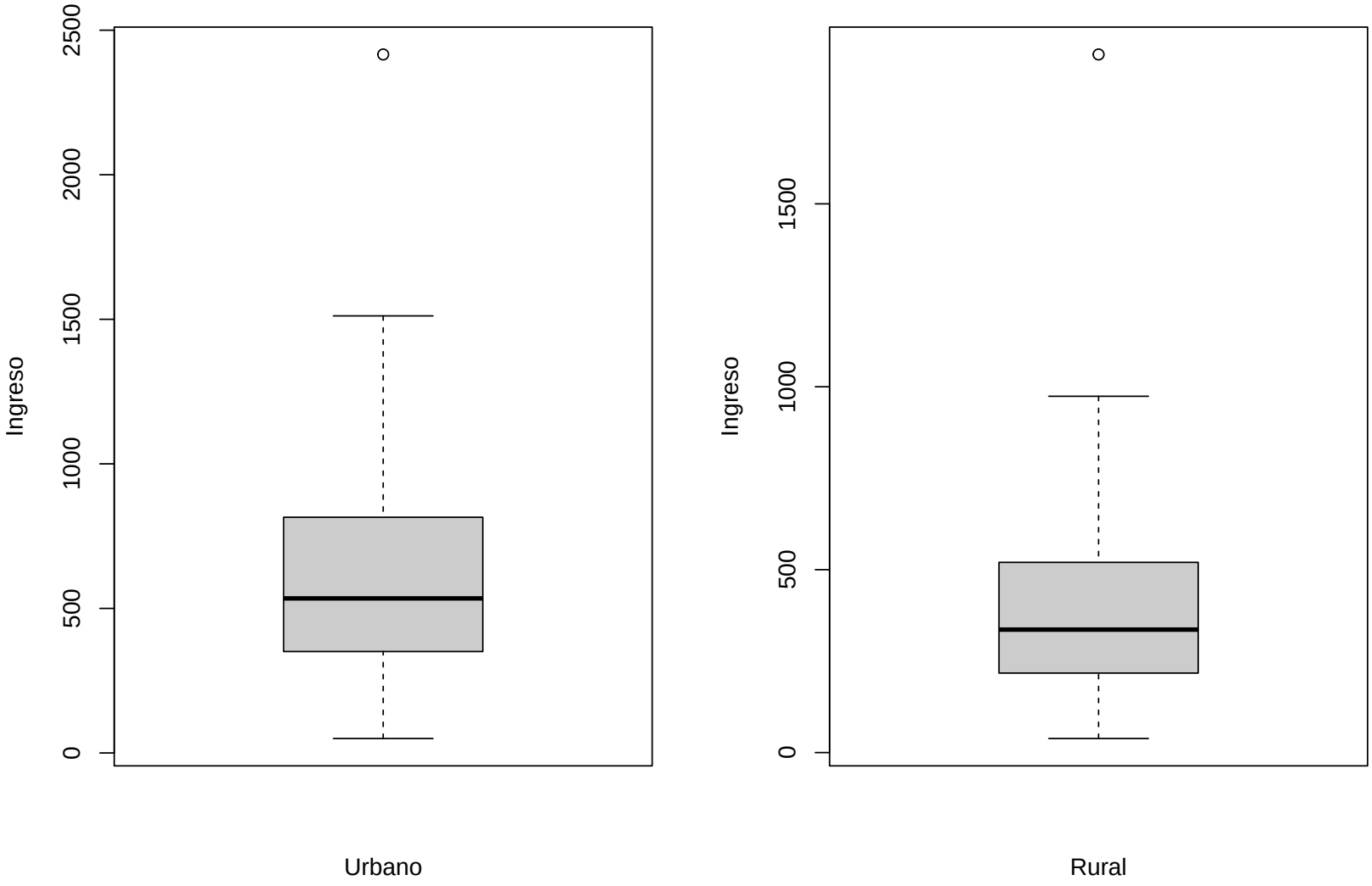


Boxplot ponderado del ingreso por sub-grupos

```
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"
)
```

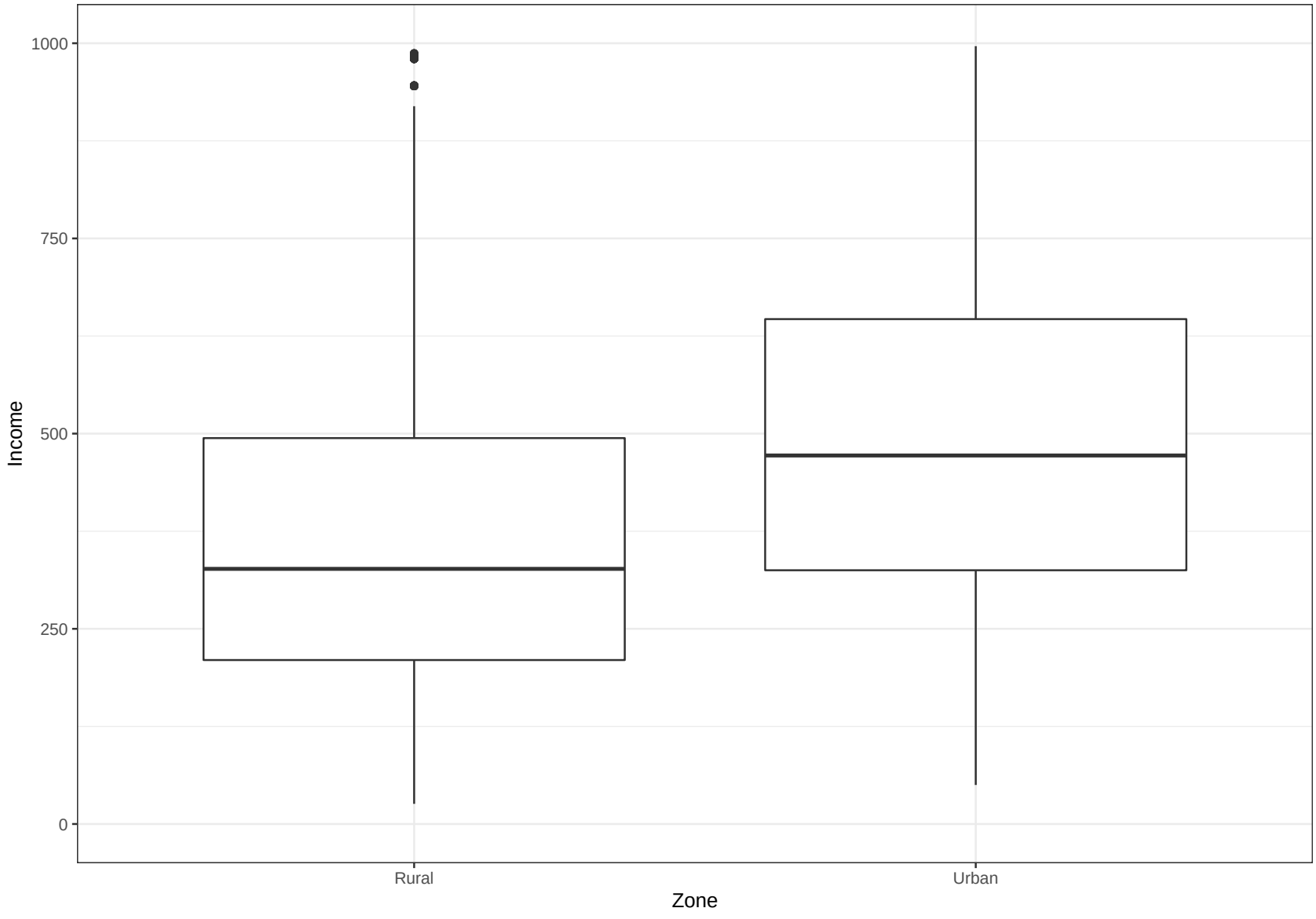
Boxplot ponderado del ingreso por sub-grupos



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)) +
  theme_bw()
```

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	81365001	3966459	11.19

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

	2.5 %	97.5 %
Income	73590884	89139118

Estimación de totales e intervalos de confianza del gasto

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

	total	Expenditure	deff
Expenditure	54611556	2355756	10.13

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

	2.5 %	97.5 %
Expenditure	49994358	59228754

Estimación de totales por sub-grupos

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

Sex	Total	Total_se	Total_low	Total_upp
Female	43374388	2366708	38688069	48060707
Male	37990613	2376145	33285609	42695617
Total ingreso	81365001	3966459	73511016	89218986

Estimación de la media e intervalo de confianza del ingreso

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

	mean	Income	deff
Income	541.5	20.72	6.895

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

	2.5 %	97.5 %
Income	500.9	582.1

Estimación de la media e intervalo de confianza del gasto

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

	mean	Expenditure	deff
Expenditure	363.4	12.48	6.416

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

	2.5 %	97.5 %
Expenditure	339	387.9

Estimación de la media por sub-grupos

```
diseno %>% group_by(Sex) %>%
  cascade(
    Media = survey_mean(
      Expenditure, level = 0.95,
       vartype =  c("se", "ci")),
        .fill = "El gasto medio"  ) %>%
  arrange(desc(Sex))
```

Sex	Media	Media_se	Media_low	Media_upp
Male	356.0	14.00	328.3	383.8
Female	370.1	13.81	342.7	397.4
El gasto medio	363.4	12.48	338.7	388.1

Estimación de la media por sub-grupos

```
diseno %>% group_by(Zone) %>%
  cascade(
    Media = survey_mean(
      Expenditure, level = 0.95,
       vartype =  c("se", "ci")),
        .fill = "El gasto medio")%>%
  arrange(desc(Zone))
```

Zone	Media	Media_se	Media_low	Media_upp
Urban	445.3	20.45	404.8	485.8
Rural	274.7	11.41	252.1	297.3
El gasto medio	363.4	12.48	338.7	388.1

Estimación de medias por sub-grupos

```
diseno %>% group_by(Zone, Sex) %>%
  cascade(
    Media = survey_mean(
      Expenditure, level = 0.95,
       vartype =  c("se", "ci")),
        .fill = "El gasto medio") %>%
  arrange(desc(Zone), desc(Sex)) %>%
  data.frame()
```

Zone	Sex	Media	Media_se	Media_low	Media_upp
Urban	Male	434.4	22.40	390.1	478.8
Urban	Female	454.7	22.96	409.2	500.1
Urban	El gasto medio	445.3	20.45	404.8	485.8
Rural	Male	274.6	12.41	250.0	299.2
Rural	Female	274.8	11.63	251.7	297.8
Rural	El gasto medio	274.7	11.41	252.1	297.3
El gasto medio	El gasto medio	363.4	12.48	338.7	388.1

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	317.7	129.8	259.9	366.4
Urban	423.8	171.3	348.6	487.5

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	318.8	142.1	248.3	376.3
Rural	Male	316.7	123.8	264.5	361.5
Urban	Female	418.2	169.2	343.8	481.3
Urban	Male	430.4	198.0	328.1	512.7

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
303.3	11.52	274.7	320.3

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	241.5	8.558	230.5	264.8
Urban	373.1	18.296	335.8	408.9

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	311.4	11.95	280.7	328.0
Male	290.5	13.12	264.8	316.8

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
303.3	9.838	281.4	320.4

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	205.4	8.213	182.2	214.7
Male	194.4	8.015	182.9	214.6

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	162.0	11.47	132.9	178.9
Urban	252.2	17.83	240.2	311.4

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
0.6712	0.0166	0.6383	0.7041

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	0.0547	1.006	1.222

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	0.0846	0.8985	1.238

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
0.6757	0.0166	0.6427	0.7086

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
0.6661	0.0197	0.627	0.7052

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	0.372	0.0123

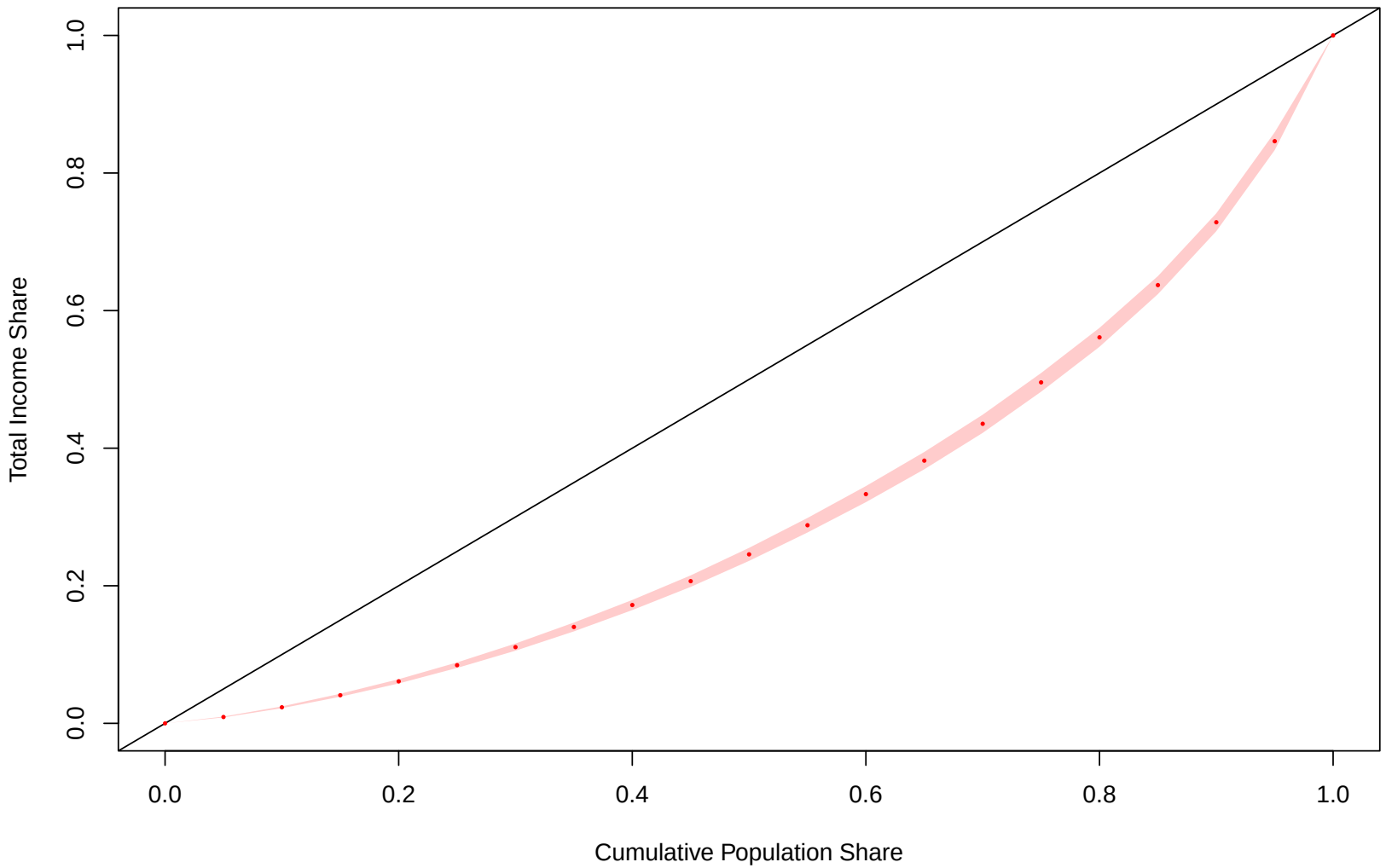
Estimación del índice de GINI

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

	gini	Expenditure
Expenditure	0.3439	0.0121

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 = -0.6, df = 118, p-value = 0.6
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -56.88  30.45
## sample estimates:
## difference in mean
##              -13.22
```

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:
##  -80.17  51.55
## sample estimates:
## difference in mean
##              -14.31
```

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 = 0.26, df = 118, p-value = 0.8
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -51.63  67.49
## sample estimates:
## difference in mean
##              7.927
```

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	420.9	27.56	366.8	474.9
Sur	Sur	563.3	61.28	443.2	683.4
Centro	Centro	738.6	50.93	638.8	838.4
Occidente	Occidente	601.6	50.90	501.8	701.3
Oriente	Oriente	427.6	31.97	364.9	490.3

```
# Paso 1: diferencia de estimaciones (Norte - Sur)
461.9124 - 592.2575
```

```
## [1] -130.3
```

contrastes

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

	Norte	Sur	Centro	Occidente	Oriente
Norte	759.8	0	0	0	0
Sur	0.0	3755	0	0	0
Centro	0.0	0	2594	0	0
Occidente	0.0	0	0	2591	0
Oriente	0.0	0	0	0	1022

```
sqrt(759.8 + 3755 - 2*0)
```

```
## [1] 67.19
```

contrastes

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

	contrast	diff_NS
diff_NS	-142.4	67.19

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	984.1	67.19
centro	738.6	50.93
Occidente_Oriente	1029.2	60.10

```
sqrt(759.8 + 3755 - 2*0) ; sqrt(2594);
```

```
## [1] 67.19
```

```
## [1] 50.93
```

```
sqrt(2591 + 1022 - 2*0)
```

```
## [1] 60.11
```

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	547.7	22.63	503.4	592.1
Male	Male	534.5	24.32	486.8	582.2

Contrastes no independiente

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

	contrast	diff_Sexo
diff_Sexo	13.22	22.05

Contrastes no independiente

```
vcov(prom_sexo)
```

	Female	Male
Female	512.3	308.8
Male	308.8	591.6

```
# Note que el error estandar  de la diferencia es igual a
sqrt(512.3 + 591.6 - 2*308.8)
```

```
## [1] 22.05
```

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	12040478	1165493	9756153	14324804
Sur	Sur	12906341	2042505	8903105	16909576
Centro	Centro	19294196	1808361	15749874	22838518
Occidente	Occidente	20994599	2337484	16413214	25575984
Oriente	Oriente	16129387	1211856	13754193	18504580

Contrastes no independiente

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

	contrast	Agregado_NCS
Agregado_NCS	44241015	2966542

Contrastes

```
require(kableExtra)
# Note que el error estandar  de la diferencia es igual a
vcov(sum_region) %>% data.frame() %>%
  kable(digits = 10,
        format.args = list(scientific = FALSE))
```

	Norte	Sur	Centro	Occidente	Oriente
Norte	1358374844388	0	0	0	0
Sur	0	4171825158601	0	0	0
Centro	0	0	3270169132601	0	0
Occidente	0	0	0	5463833230799	0
Oriente	0	0	0	0	1468594303378

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

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

Contrastes no independiente

```
(prom_edad <- svyby(~Income, ~CatAge, diseno,
                    svymean, na.rm=T,covmat = TRUE))
```

	CatAge	Income	se
0-5	0-5	482.3	25.94
6-15	6-15	503.0	25.67
16-30	16-30	603.1	31.45
31-45	31-45	533.7	24.01
46-60	46-60	591.8	40.69
Más de 60	Más de 60	461.0	37.83

Contrastes no independiente

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

	contrast	agregado_edad
agregado_edad	529.2	20

Contrastes no independiente

```
vcov(prom_edad)
```

	0-5	6-15	16-30	31-45	46-60	Más de 60
0-5	672.84	32.12	223.7	98.01	205.4	86.96
6-15	32.12	658.90	161.1	331.72	415.6	197.11
16-30	223.67	161.12	989.1	173.55	879.0	256.93
31-45	98.01	331.72	173.6	576.57	328.8	369.00
46-60	205.40	415.63	879.0	328.82	1655.7	448.95
Más de 60	86.96	197.11	256.9	369.00	448.9	1430.78

```
(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 = ~Expenditure,
                     diseno, svyratio,
                     na.rm=T, covmat = TRUE,
                     vartype = c("se", "ci")))
```

	Sex	Income/Expenditure	se.Income/Expenditure	ci_l	ci_u
Female	Female	1.480	0.0364	1.409	1.551
Male	Male	1.501	0.0445	1.414	1.589

Contrastes no independiente

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

	contrast	diff_sexo
diff_sexo	-0.0212	0.0332

Contrastes no independiente

```
vcov(razon_sexo)
```

	Female	Male
Female	0.0013	0.0011
Male	0.0011	0.0020

```
sqrt(0.0013 + 0.0020 - 2*0.0011)
```

```
## [1] 0.03317
```