Variable continua

CEPAL

14/2/2022

Lectura de la base

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

Definir diseño de la muestra con `srvyr`

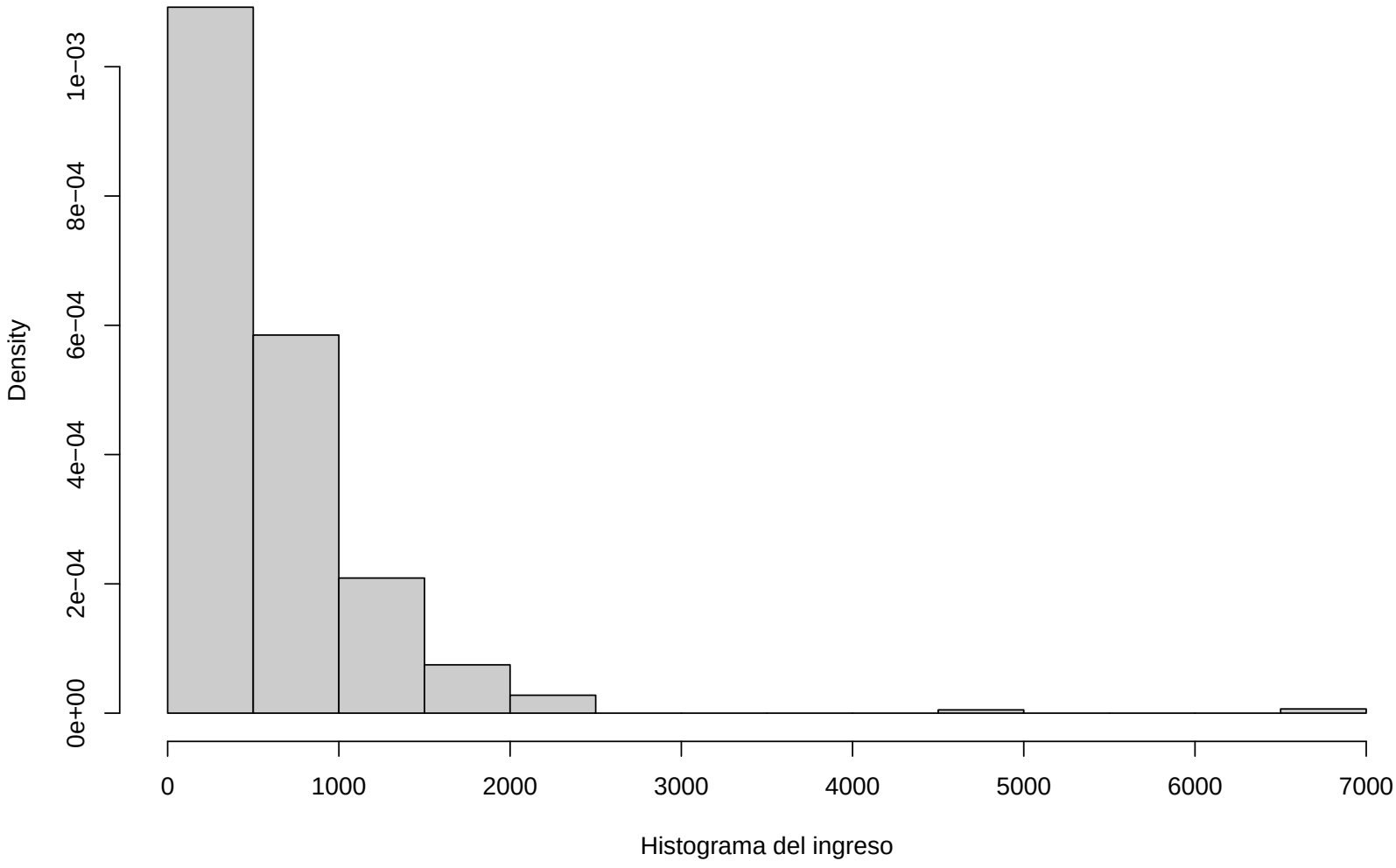
```
library(srvyr)

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"
)
```

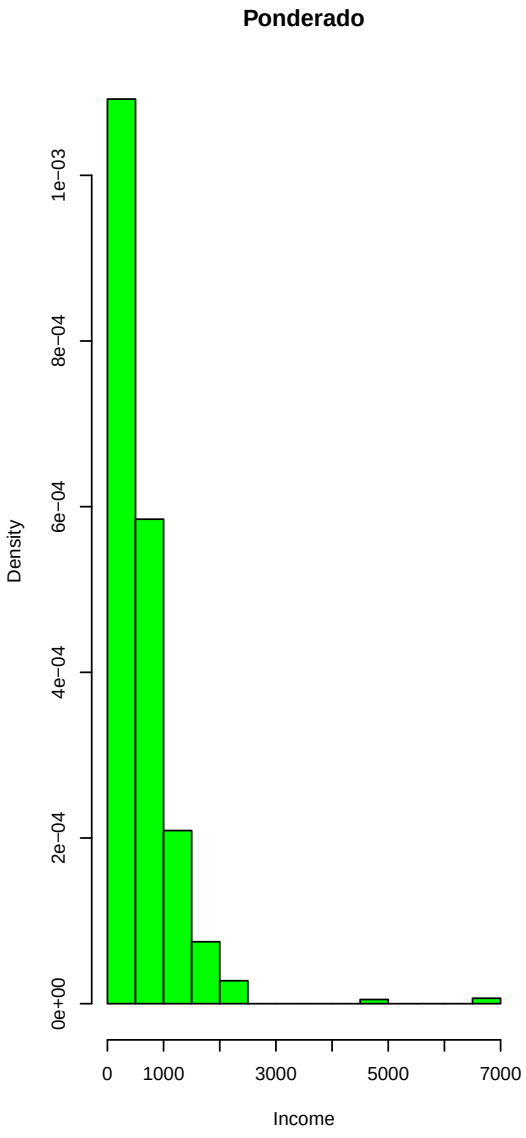
Histograma ponderado para la variable ingreso

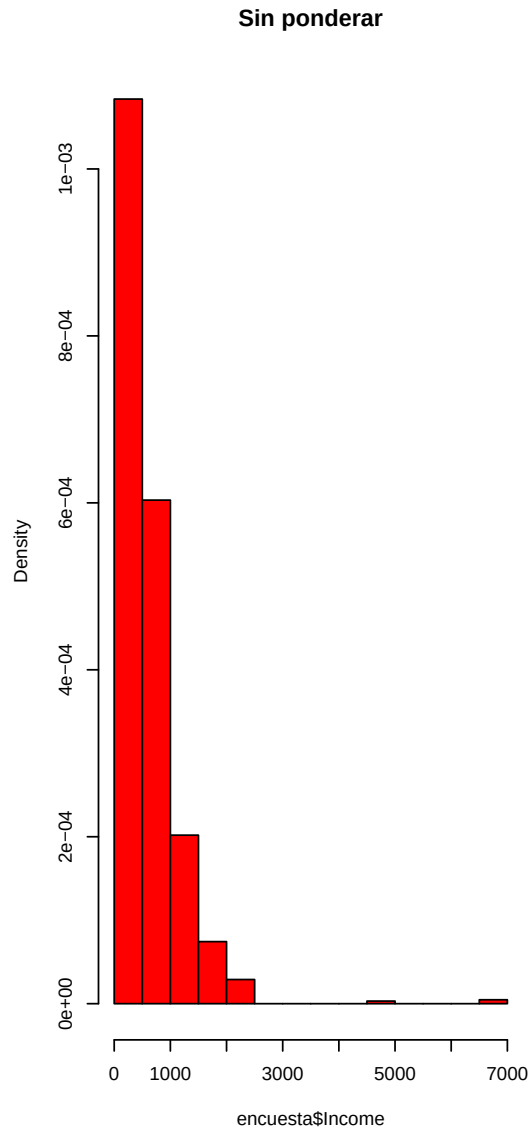


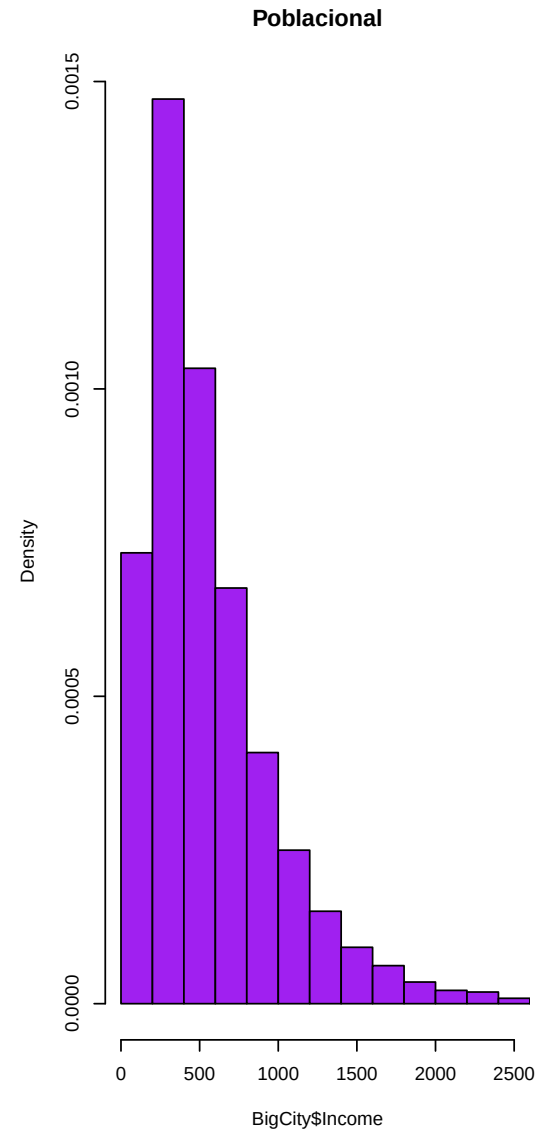
Comparación de histogramas

```
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.

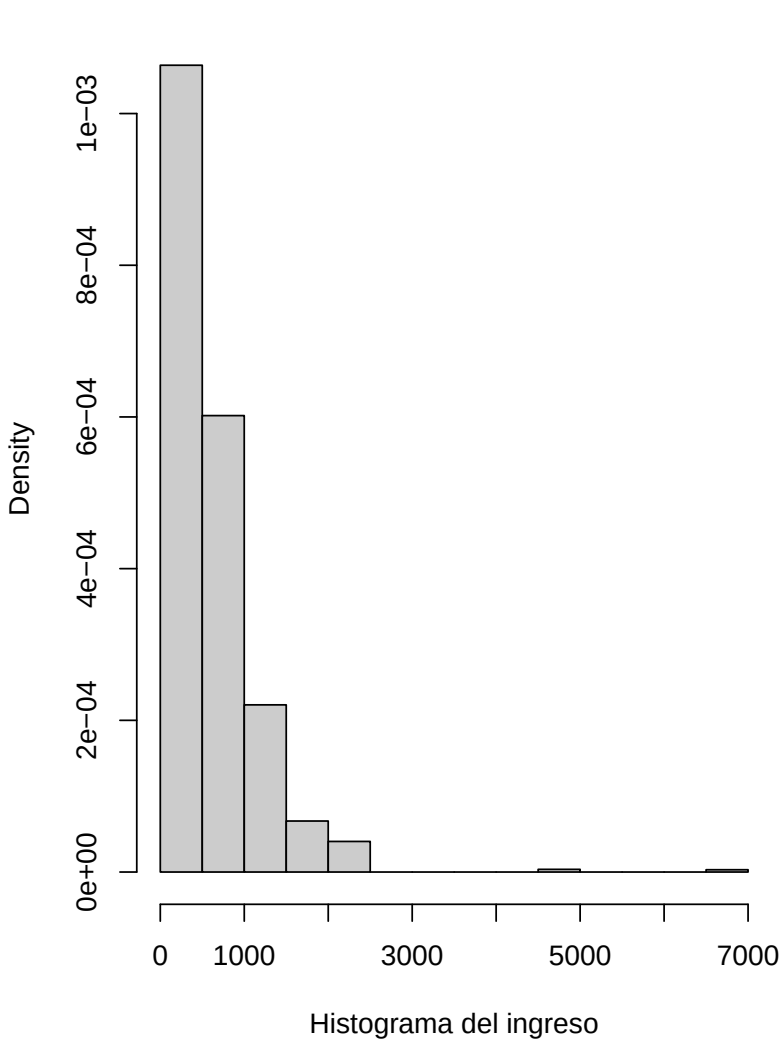
```
sub_Urbano <- diseno %>%  filter(Zone == "Urban")
sub_Rural  <- diseno %>%  filter(Zone == "Rural")
sub_Mujer  <- diseno %>%  filter(Sex == "Female")
sub_Hombre <- diseno %>%  filter(Sex == "Male")
```

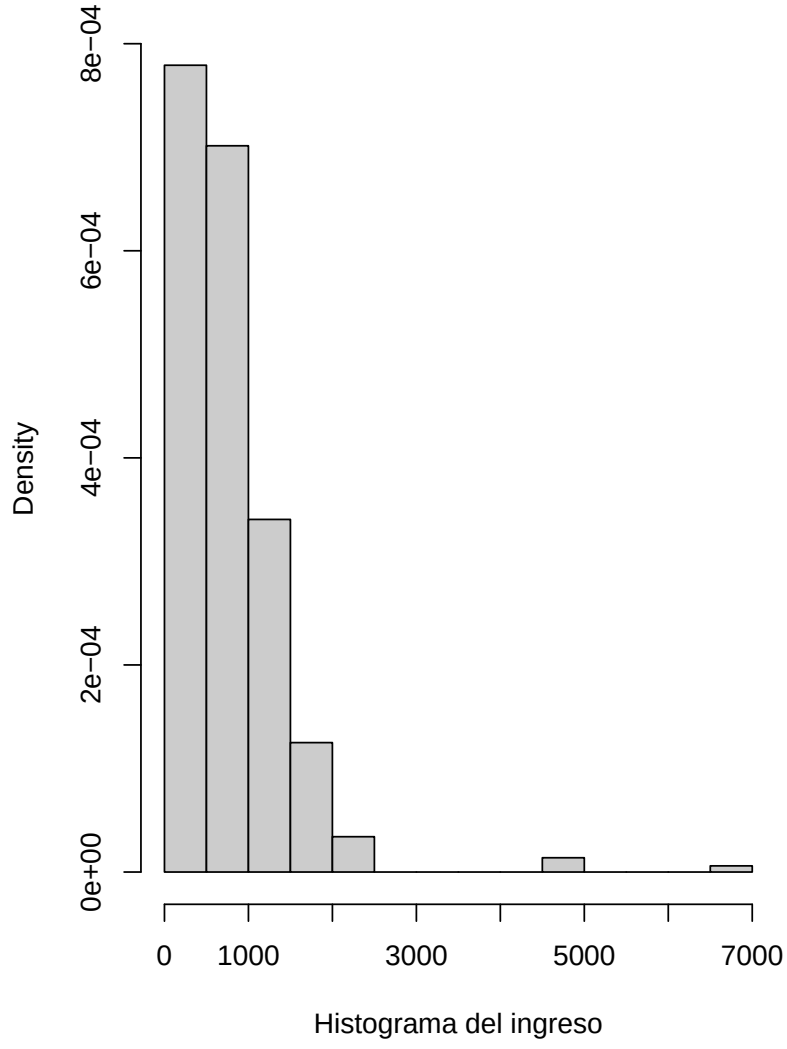
Histograma ponderado en sub-grupos

```
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



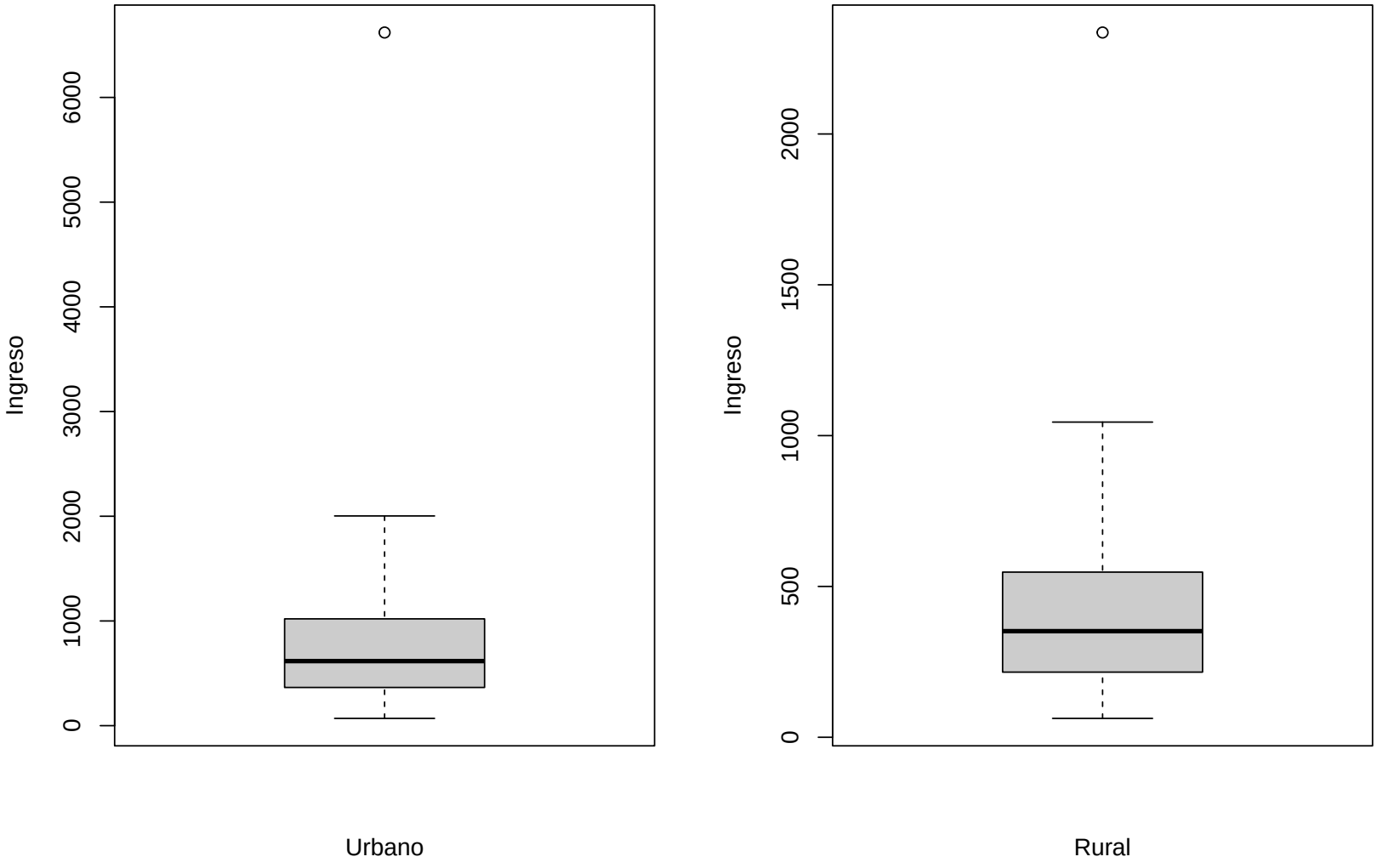


Boxplot ponderado del ingreso por sub-grupos

```
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"
)
```

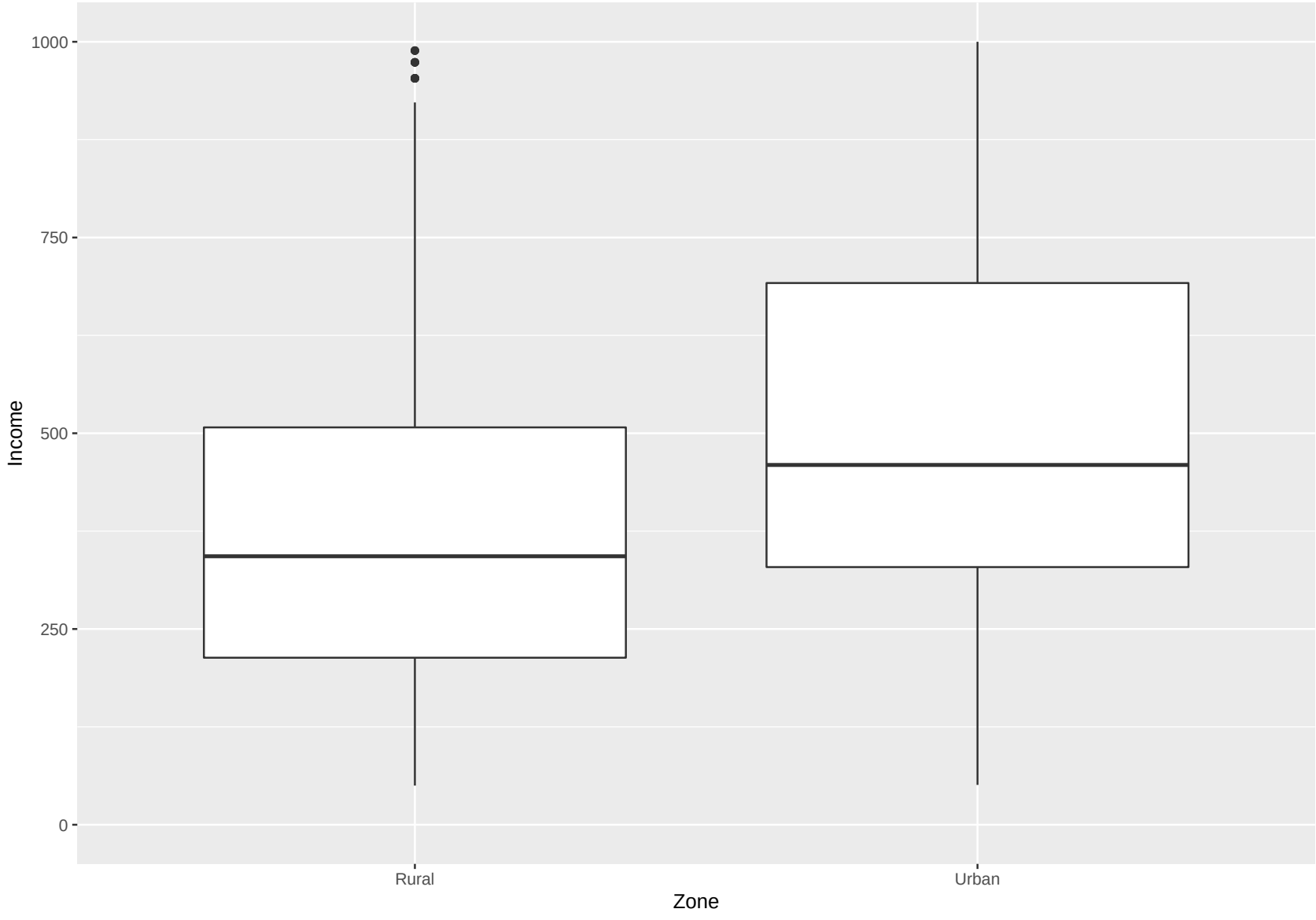
Boxplot ponderado del ingreso por sub-grupos



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	7.408137

```
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	16.24008

```
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	625.5024	23.43734	3.988896

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

	2.5 %	97.5 %
Income	579.5661	671.4387

Estimación de la media e intervalo de confianza del gasto

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

	mean	Expenditure	deff
Expenditure	387.0579	13.5102	7.065132

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

	2.5 %	97.5 %
Expenditure	360.5784	413.5374

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	388.7651	13.64647	361.7438	415.7865
Male	385.1557	14.74954	355.9501	414.3613

```
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	286.7022	12.52363	261.9042	311.5002
Urban	479.6305	20.97185	438.1041	521.1568

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	288.9143	13.66839	261.8495	315.9791
Rural	Male	284.3395	12.33807	259.9089	308.7701
Urban	Female	477.3961	19.77585	438.2379	516.5542
Urban	Male	482.2190	24.31450	434.0738	530.3641

Estimación de la proporción de hombres

```
diseno %>% mutate(Hombres = (Sex == "Male")) %>%
  summarise(
    Prop = survey_mean(Hombres, level = 0.95,
                       vartype =  c("se", "ci"))) %>%
  data.frame()
```

Prop	Prop_se	Prop_low	Prop_upp
0.4730012	0.0111468	0.4509294	0.495073

Estimación de la proporción de hombres en la zona rural

```
sub_Rural %>% mutate(Hombres = (Sex == "Male")) %>%
  summarise(
    Prop = survey_mean(Hombres, level = 0.95,
                       vartype =  c("se", "ci"))) %>%
  data.frame()
```

Prop	Prop_se	Prop_low	Prop_upp
0.4835372	0.018217	0.4470296	0.5200448

Estimación de la proporción de mujeres menor de 18 años

```
sub_Mujer %>% mutate(Menor_18 = Age < 18) %>%
  summarise(
    Prop = survey_mean(Menor_18, level = 0.95,
                       vartype =  c("se", "ci"))) %>%
  data.frame()
```

Prop	Prop_se	Prop_low	Prop_upp
0.3118204	0.0154622	0.2811982	0.3424425

Estimación de la varianza y desviación estándar de los ingresos

```
(tab_var <- diseno %>% summarise(
    Var = survey_var(Income, level = 0.95,
                             vartype =  c("se", "ci"),
                           deff = TRUE)))
```

Var	Var_se	Var_low	Var_upp
358076.3	110493.9	139287.3	576865.2

```
sqrt(tab_var)
```

Var	Var_se	Var_low	Var_upp
598.3947	332.4062	373.2122	759.5164

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	337.4783	183.6105	217.1102	425.0251
Urban	725.3440	457.4290	334.3729	969.7642

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	334.7518	194.0307	193.6806	431.9784
Rural	Male	340.5046	176.4571	232.9996	421.4237
Urban	Female	739.3898	502.3652	216.7439	1022.9451
Urban	Male	709.2652	422.6540	386.4444	925.6214

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
310.58	11.72163	289.48	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	257.79	12.79912	231.53	282.83
Urban	399.42	24.91078	359.30	458.83

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	328.66	12.37347	294.63	343.64
Male	299.46	11.90640	286.67	333.83

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
310.58	12.11808	289.48	337.47

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	217.66	10.11640	197.30	237.37
Male	205.29	10.61881	191.62	233.68

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	173.06	10.798162	149.3	192.58
Urban	272.85	8.917623	259.0	294.63

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
0.6187952	0.0195401	0.5801039	0.6574865

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
1.114159	0.0498225	1.015506	1.212813

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
1.068093	0.0779141	0.9119499	1.224236

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
0.6248777	0.0228829	0.5795594	0.670196

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
0.6120947	0.018535	0.575387	0.6488023

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()
```

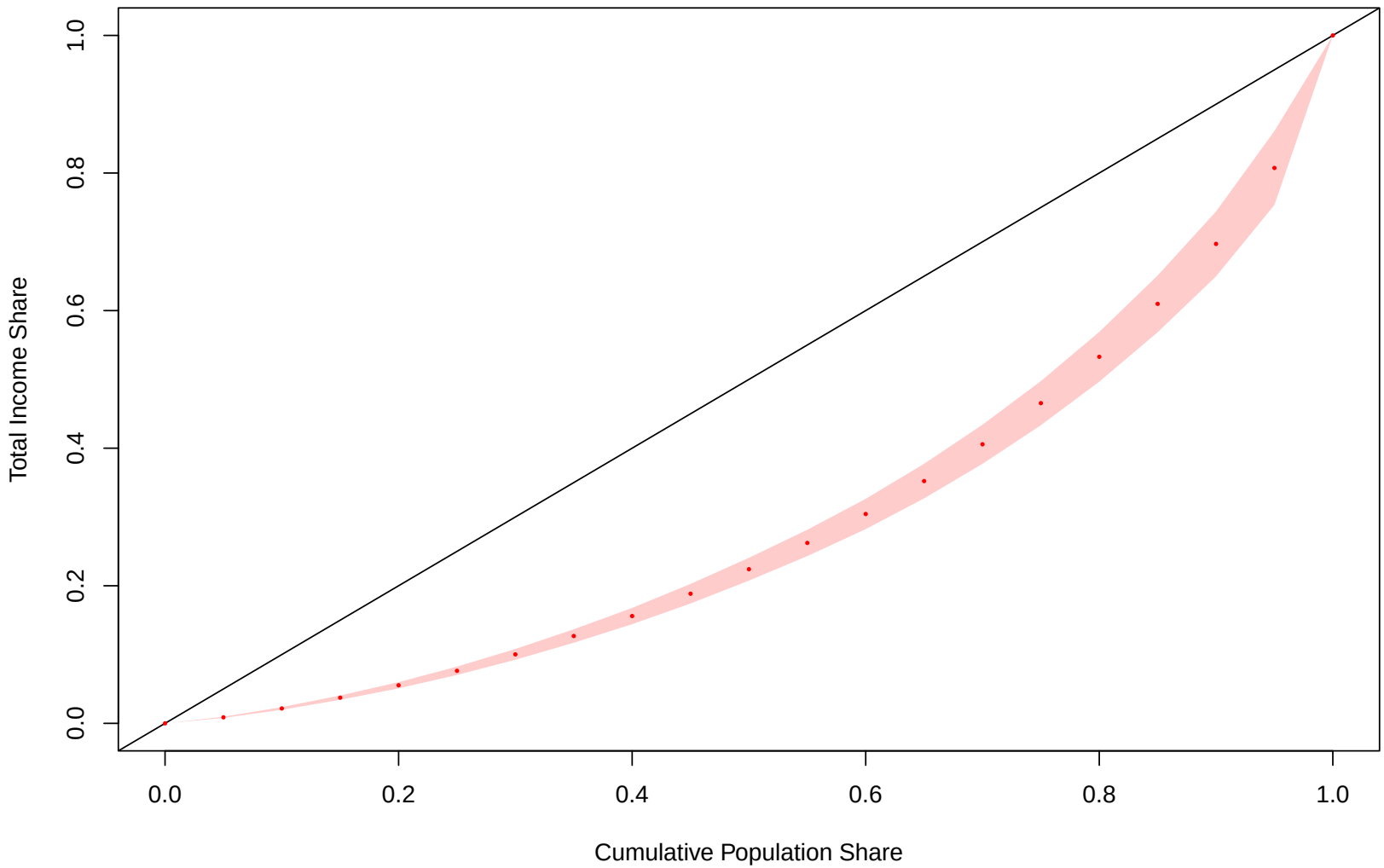
	gini	Income
Income	0.4102334	0.0159258

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

	gini	Expenditure
Expenditure	0.34273	0.0106959

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 = 0.34692, df = 118, p-value = 0.7293
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -33.40897  47.60117
## sample estimates:
## difference in mean
##           7.096097
```

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 = 0.43427, df = 63, p-value = 0.6656
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -53.83886  83.73612
## sample estimates:
## difference in mean
##           14.94863
```

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 = 0.51261, df = 118, p-value = 0.6092
## alternative hypothesis: true difference in mean is not e
## 95 percent confidence interval:
##  -39.21926  66.61545
## sample estimates:
## difference in mean
##            13.6981
```

Contrastes

```
(prom_region <- svyby(~Income, ~Region, diseno, svymean, na
```

	Region	Income	se
Norte	Norte	527.4315	43.69260
Sur	Sur	627.0919	53.43962
Centro	Centro	767.9655	82.86492
Occidente	Occidente	575.4733	53.07773
Oriente	Oriente	650.2452	32.69869

```
# 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	1909.043	0.000	0.000	0.000	0.000
Sur	0.000	2855.793	0.000	0.000	0.000
Centro	0.000	0.000	6866.595	0.000	0.000
Occidente	0.000	0.000	0.000	2817.246	0.000
Oriente	0.000	0.000	0.000	0.000	1069.204

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

```
## [1] 69.02898
```

contrastes

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

	contrast	diff_NS
diff_NS	-99.66039	69.02779

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	1154.5234	69.02779
centro	767.9655	82.86492
Occidente_Oriente	1225.7185	62.34140

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

```
## [1] 69.02898
```

```
sqrt(6867)
```

```
## [1] 82.86736
```

Contrastes no independiente

```
prom_sex <- svyby(~Income, ~Sex, diseno, svymean, na.rm=T,c
svycontrast(prom_sex, list(diff_Sexo = c(1,-1)))
```

```
##             contrast     SE
## diff_Sexo  -7.0961 20.454
```

Contrastes no independiente

```
vcov(prom_sex)
```

	Female	Male
Female	582.2594	448.5481
Male	448.5481	733.2148

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

```
## [1] 20.42058
```

Contrastes no independiente

```
sum_region <- svyby(~Income, ~Region, diseno, svytotal, na

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] 20.42058
```

Contrastes

```
vcov(sum_region)
```

	Norte	Sur	Centro	Occid
Norte	2.805154e+12	0.000000e+00	0.000000e+00	0.000000e
Sur	0.000000e+00	3.839259e+12	0.000000e+00	0.000000e
Centro	0.000000e+00	0.000000e+00	7.559033e+12	0.000000e
Occidente	0.000000e+00	0.000000e+00	0.000000e+00	6.571797e
Oriente	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e

```
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	602.1618	69.91311
16-30	16-30	654.7618	25.41812
31-45	31-45	655.5259	40.13589
46-60	46-60	614.1928	46.87734
6-15	6-15	595.4982	34.35879
Más de 60	Más de 60	580.3491	74.51824

```
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	617.0816	25.66667

Contrastes no independiente

```
vcov(prom_edad)
```

	0-5	16-30	31-45	46-60	
0-5	4887.84261	22.11573	-1289.3652	863.5245	-1377.
16-30	22.11573	646.08088	453.7178	441.4990	347.
31-45	-1289.36522	453.71776	1610.8896	290.4031	819.
46-60	863.52447	441.49902	290.4031	2197.4850	102.
6-15	-1377.75616	347.08754	819.6384	102.6295	1180.
Más de 60	198.02049	856.17028	1336.7419	487.5931	268.

```
(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] 25.64826
```

Contrastes no independiente

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

	Sex	Income/Expenditure	se.Income/Expenditure	c
Female	Female	1.600313	0.0586030	1.485
Male	Male	1.633734	0.0494716	1.536

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

	contrast	diff_sexo
diff_sexo	-0.0334212	0.038261

Contrastes no independiente

```
vcov(razon_sexo)
```

	Female	Male
Female	0.0034343	0.0022089
Male	0.0022089	0.0024474

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

```
## [1] 0.03741657
```

Correlación de variables

```
library(jtools)

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

	Income	Expenditure
Income	1.000000	0.708192
Expenditure	0.708192	1.000000

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

	Income	Expenditure
Income	1.0000000	0.7333025
Expenditure	0.7333025	1.0000000

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

	Income	Expenditure