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
  ) %>% mutate(Uno = 1)
```

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 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	0.0125	0.4551	0.5047
Urban	0.5202	0.0125	0.4953	0.5449

```
(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	0.0125	0.4550	0.5047
Urban	0.5202	0.0125	0.4953	0.5450

Poporsión 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	718	0.5367	0.0119	0.5128	0.5606
Male	614	0.4633	0.0119	0.4394	0.4872

```
(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	633	0.5165	0.0153	0.4858	0.5471
Male	632	0.4835	0.0153	0.4529	0.5142

Poporsión de hombres en la zona urbana y rural

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

Zone	n	prop	prop_se	prop_low	prop_upp
Rural	632	0.4905	0.0149	0.461	0.520
Urban	614	0.5095	0.0149	0.480	0.539

Poporsión de hombres en la zona urbana y rural

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

Zone	n	prop	prop_se	prop_low	prop_upp
Rural	633	0.4702	0.0167	0.4372	0.5033
Urban	718	0.5298	0.0167	0.4967	0.5628

Estimación de la proporción de personas menor a 18 años

```
diseno %>% mutate(edad_18 = case_when(
  Age<18~"menor a 18 años",
  TRUE ~ "mayor o igual a 18 años")) %>%
  group_by(edad_18) %>%
  summarise(
    Prop = survey_prop(vartype =  c("se", "ci"))) %>%
  data.frame()
```

edad_18	Prop	Prop_se	Prop_low	Prop_upp
mayor o igual a 18 años	0.6764	0.0118	0.6531	0.6997
menor a 18 años	0.3236	0.0118	0.3003	0.3469

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

```
sub_Rural %>% mutate(edad_18 = case_when(
  Age<18~"menor a 18 años",
  TRUE ~ "mayor o igual a 18 años")) %>%
  group_by(edad_18) %>%
  summarise(
    Prop = survey_prop(vartype =  c("se", "ci"))) %>%
  data.frame()
```

edad_18	Prop	Prop_se	Prop_low	Prop_upp
mayor o igual a 18 años	0.6442	0.0201	0.6040	0.6844
menor a 18 años	0.3558	0.0201	0.3156	0.3960

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

edad_rango	Prop	Prop_se	Prop_low	Prop_upp
18 - 35	0.3059	0.0122	0.2818	0.3299
Otro	0.6941	0.0122	0.6701	0.7182

Tabla Zona Vs Sexo

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

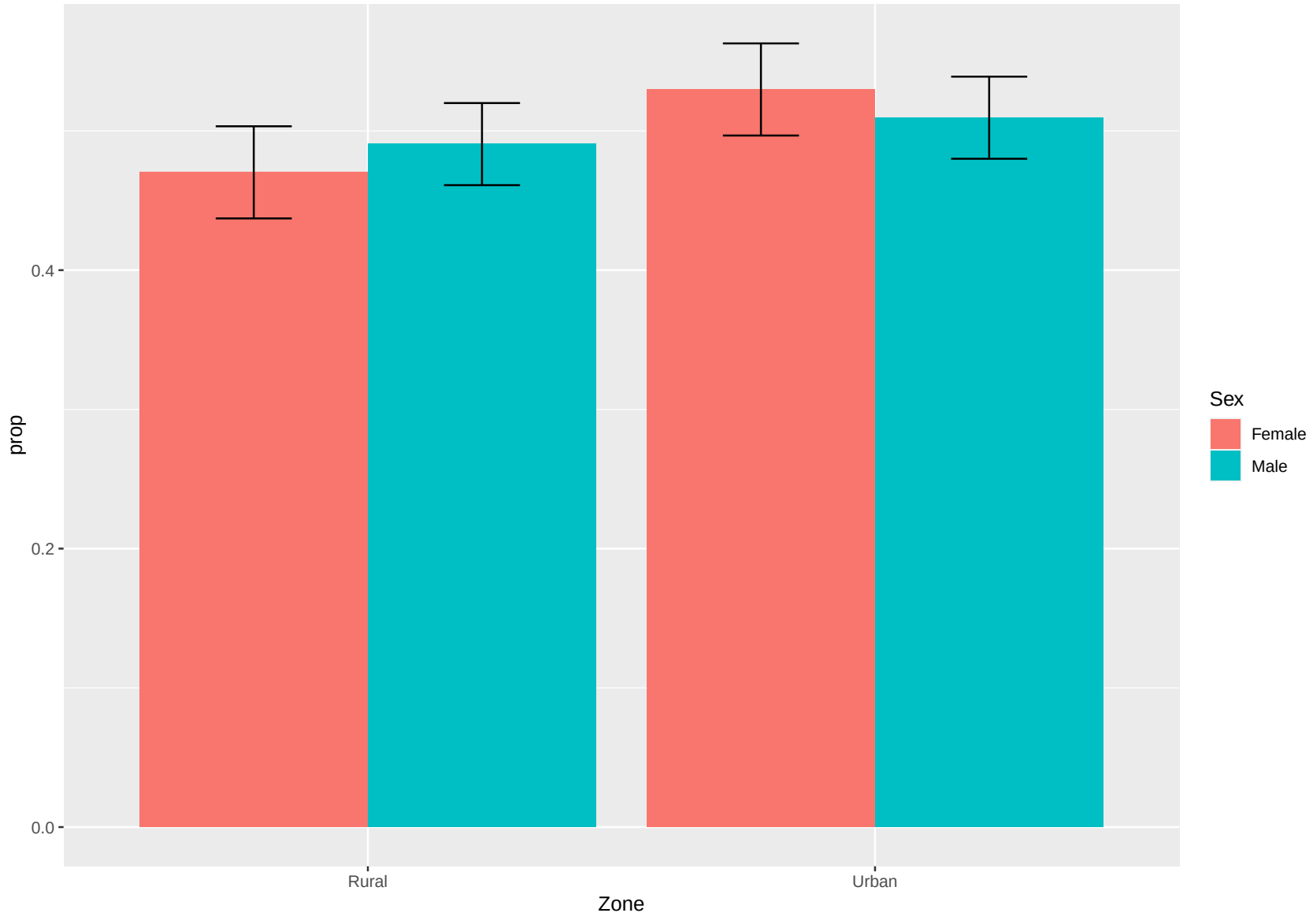
Tabla Zona Vs Sexo

Sex	Zone	prop	prop_se	prop_low	prop_upp	n
Female	Rural	0.4702	0.0167	0.4372	0.5033	633
Female	Urban	0.5298	0.0167	0.4967	0.5628	718
Male	Rural	0.4905	0.0149	0.4610	0.5200	632
Male	Urban	0.5095	0.0149	0.4800	0.5390	614

Diagrama de barras Zona Vs Sexo

```
require(ggplot2)
ggplot(data = prop_sexo_zona,
       aes(
         x = Zone,
         y = prop,
         ymax = prop_upp,
         ymin = prop_low,
         fill = Sex
       )) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(position = position_dodge(width = 0.9),
                width = 0.3)
```

Diagrama de barras Zona Vs Sexo



Prueba de independencia.

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

```
##
##  Pearson's X^2: Rao & Scott adjustment
##
## data:  NextMethod()
## F = 1.1, ndf = 1, ddf = 119, p-value = 0.3
```

Prueba de independencia.

```
svychisq(~Sex + Zone,
         subset(diseno, Region == "Centro"),
         statistic="F")
```

```
##
##  Pearson's X^2: Rao & Scott adjustment
##
## data:  NextMethod()
## F = 2.6, ndf = 1, ddf = 23, p-value = 0.1
```

Prueba de independencia.

```
svychisq(~Sex + Zone,
         subset(diseno, Region == "Sur"),
         statistic="F")
```

```
##
##  Pearson's X^2: Rao & Scott adjustment
##
## data:  NextMethod()
## F = 0.8, ndf = 1, ddf = 24, p-value = 0.4
```

Prueba de independencia.

```
diseno %>% filter(Employment == "Inactive") %>%
  group_by(Sex) %>%
  summarise(
    prop = survey_prop(vartype =  c("se", "ci")))
```

Sex	prop	prop_se	prop_low	prop_upp
Female	0.7668	0.0135	0.7401	0.7935
Male	0.2332	0.0135	0.2065	0.2599

Prueba de independencia.

```
svychisq(~Sex + Zone,
         subset(diseno, Employment == "Inactive"),
         statistic="F")
```

```
##
##  Pearson's X^2: Rao & Scott adjustment
##
## data:  NextMethod()
## F = 15, ndf = 1, ddf = 105, p-value = 2e-04
```

Tabla con más de dos categorías Sexo Vs empleo.

```
(
  prop_sexo_empleo <- diseno %>%
    filter(!is.na(Employment)) %>%
    group_by(Sex, Employment) %>%
    summarise(
      prop = survey_prop(vartype = c("se", "ci")),
                n = unweighted(n())) %>%
    data.frame()
)
```

Tabla del gasto Sexo Vs Rango de edad.

Sex	Employment	prop	prop_se	prop_low	prop_upp	n
Female	Unemployed	0.0285	0.0061	0.0165	0.0406	28
Female	Inactive	0.5505	0.0188	0.5132	0.5878	555
Female	Employed	0.4210	0.0188	0.3838	0.4581	458
Male	Unemployed	0.0724	0.0125	0.0477	0.0971	56
Male	Inactive	0.2037	0.0146	0.1748	0.2326	178
Male	Employed	0.7239	0.0207	0.6830	0.7649	645

Diagrama de barras del gasto Sexo Vs empleo.

```
require(ggplot2)
ggplot(data = prop_sexo_empleo,
       aes(
         x = Sex,
         y = prop,
         ymax = prop_upp,
         ymin = prop_low,
         fill = Employment
       )) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(position = position_dodge(width = 0.9),
                width = 0.3)
```

Diagrama de barras del gasto Sexo Vs Rango de edad.

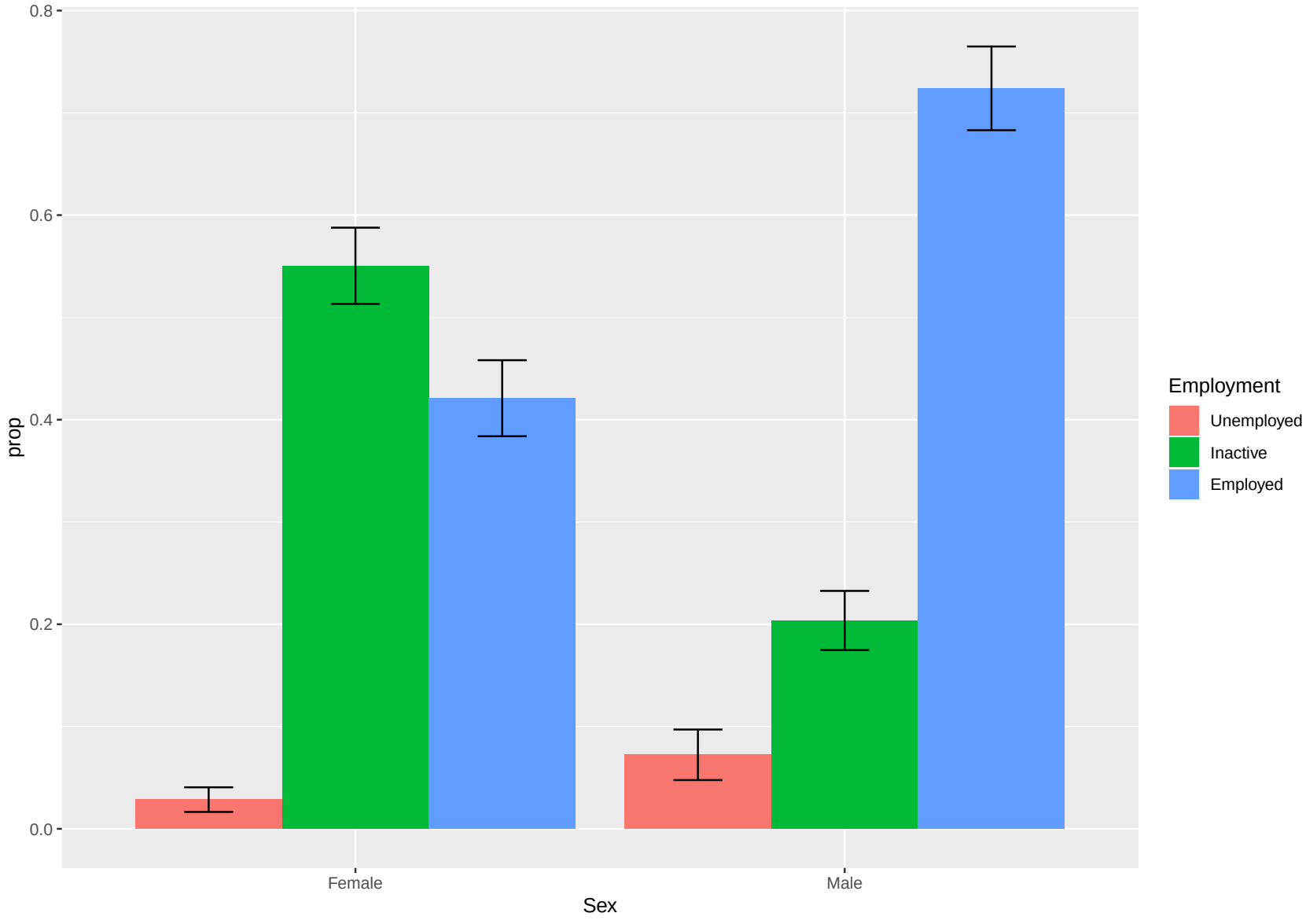


Tabla con más de dos categorías Sexo Vs Region

```
(
  prop_region_sexo <- diseno %>%
    group_by(Region, Sex) %>%
    summarise(
      prop = survey_prop(vartype = c("se", "ci")),
               n = unweighted(n())) %>%
    data.frame()
)
```

Tabla con más de dos categorías Sexo Vs Region

Region	Sex	prop	prop_se	prop_low	prop_upp	n
Norte	Female	0.4985	0.0161	0.4667	0.5303	285
Norte	Male	0.5015	0.0161	0.4697	0.5333	263
Sur	Female	0.5390	0.0241	0.4914	0.5867	290
Sur	Male	0.4610	0.0241	0.4133	0.5086	261
Centro	Female	0.5397	0.0187	0.5027	0.5767	240
Centro	Male	0.4603	0.0187	0.4233	0.4973	210
Occidente	Female	0.5161	0.0156	0.4851	0.5470	275
Occidente	Male	0.4839	0.0156	0.4530	0.5149	260
Oriente	Female	0.5381	0.0247	0.4891	0.5870	261
Oriente	Male	0.4619	0.0247	0.4130	0.5109	252

Diagrama de barras con más de dos categorías Sexo Vs Region

```
ggplot(data = prop_region_sexo,
       aes(
         x = Region,
         y = prop,
         ymax = prop_upp,
         ymin = prop_low,
         fill = Sex
       )) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(position = position_dodge(width = 0.9),
                width = 0.3)
```

Diagrama de barras del gasto Sexo Vs Región

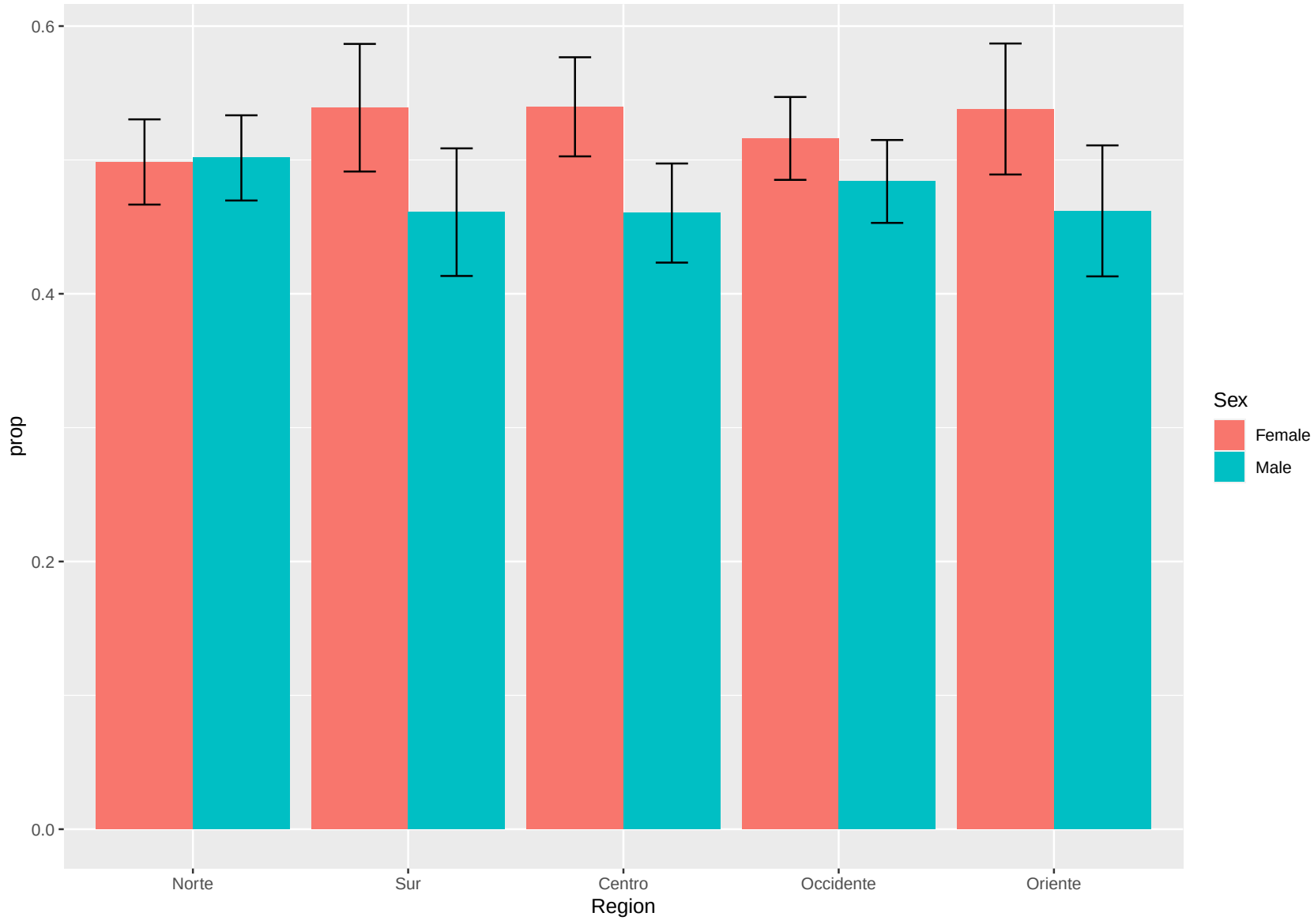


Tabla con más de dos categorías Sexo Vs Region zona urbana

```
(
  prop_region_sexo_zonaU <- sub_Urbano %>%
    group_by(Region, Sex) %>%
    summarise(
      prop = survey_prop(vartype = c("se", "ci")),
               n = unweighted(n())) %>%
    data.frame()
)
```

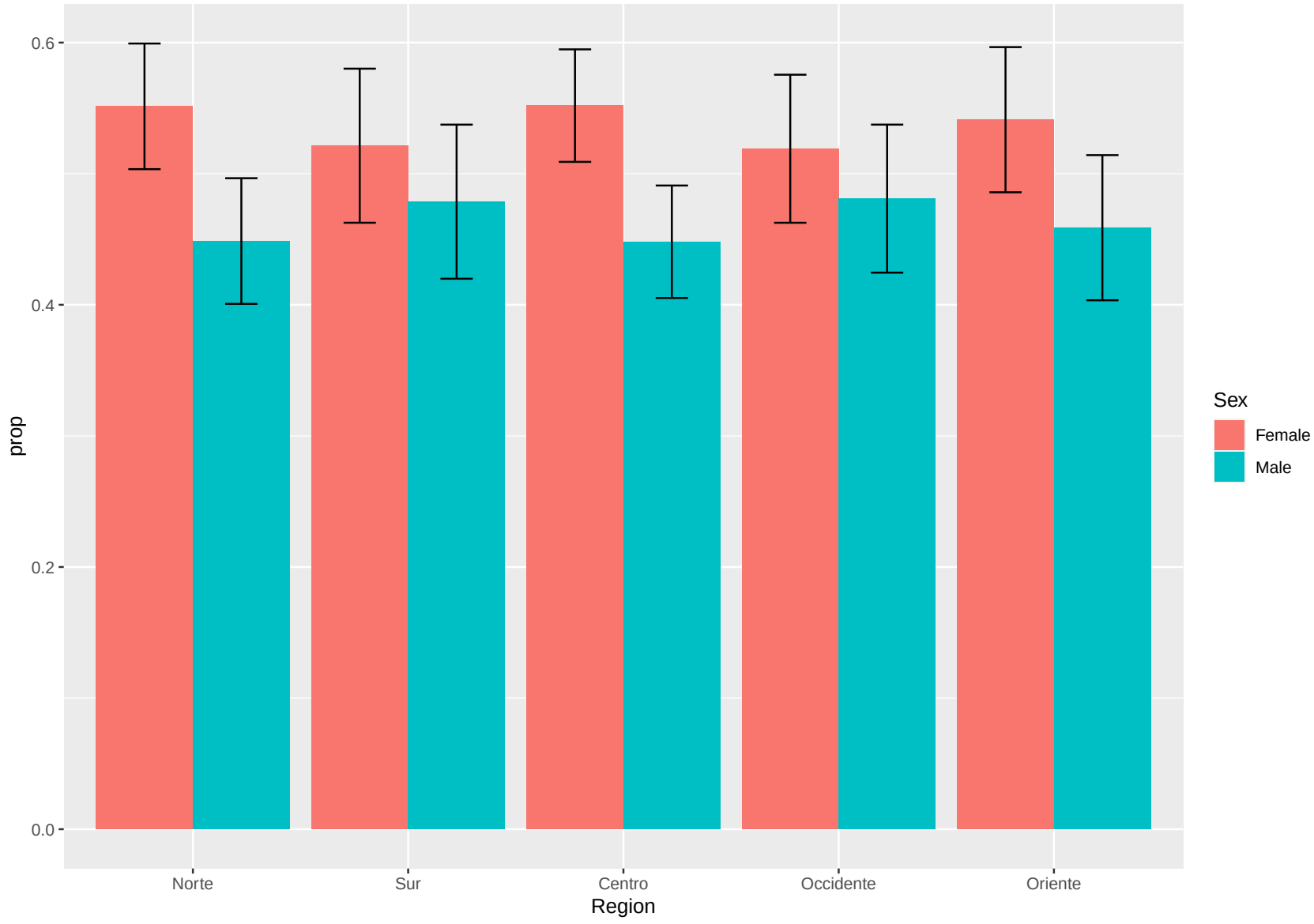
Tabla del gasto Sexo Vs Rango de edad.

Region	Sex	prop	prop_se	prop_low	prop_upp	n
Norte	Female	0.5514	0.0240	0.5034	0.5993	143
Norte	Male	0.4486	0.0240	0.4007	0.4966	111
Sur	Female	0.5213	0.0294	0.4626	0.5801	144
Sur	Male	0.4787	0.0294	0.4199	0.5374	131
Centro	Female	0.5519	0.0215	0.5090	0.5948	167
Centro	Male	0.4481	0.0215	0.4052	0.4910	135
Occidente	Female	0.5190	0.0283	0.4626	0.5755	131
Occidente	Male	0.4810	0.0283	0.4245	0.5374	116
Oriente	Female	0.5412	0.0277	0.4858	0.5966	133
Oriente	Male	0.4588	0.0277	0.4034	0.5142	121

Diagrama de barras del gasto Sexo Vs Región

```
ggplot(data = prop_region_sexo_zonaU,
       aes(
         x = Region,
         y = prop,
         ymax = prop_upp,
         ymin = prop_low,
         fill = Sex
       )) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(position = position_dodge(width = 0.9),
                width = 0.3)
```

Diagrama de barras del gasto Sexo Vs Región



glm del Zona en función del ingreso

```
mod <- svyglm(as.factor(Zone) ~ Income,
              design=diseno, family=quasibinomial)
summary(mod)$coefficients
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.8300	0.1555	-5.336	0
Income	0.0018	0.0003	5.564	0

glm del Zona en función del ingreso

```
mod2 <- svyglm(as.factor(Sex) ~ Income,
               design=diseno, family=quasibinomial)
summary(mod2)$coefficients
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.1476	0.0736	-2.0040	0.0474
Income	0.0001	0.0001	0.7627	0.4472

Tablas con más de dos categorías.

```
(prop_edad <- diseno %>% group_by(CatAge) %>%
    summarise(
      prop = survey_mean(
        vartype = c("se", "ci"))))
```

CatAge	prop	prop_se	prop_low	prop_upp
0-5	0.0977	0.0059	0.0860	0.1094
6-15	0.1902	0.0099	0.1705	0.2099
16-30	0.2569	0.0105	0.2361	0.2777
31-45	0.2070	0.0095	0.1881	0.2258
46-60	0.1333	0.0089	0.1156	0.1510
Más de 60	0.1150	0.0089	0.0974	0.1325