Análisis de encuestas de hogares con R

Modulo 6: Modelos lineales generalizados

CEPAL - Unidad de Estadísticas Sociales

Tabla de contenidos I

Introducción

Introducción

- Los Modelos Lineales Generalizados (MLGs) son una aproximación unificada a la mayoría de los procedimientos utilizados en estadística aplicada.
- Generalizan los modelos lineales clásicos que se basan en la suposición de una distribución normal para la variable respuesta.
- Los MLGs son ampliamente utilizados en diversas disciplinas y presentan un marco teórico unificado para estimar parámetros.
- La genialidad de Nelder & Wedderburn (1972) radica en demostrar que muchos métodos estadísticos aparentemente no relacionados se pueden abordar con un mismo marco teórico.

Introducción

- Los MLGs son especialmente útiles cuando la suposición de normalidad en la variable respuesta no es razonable, como en el caso de respuestas categóricas, proporciones o conteos.
- Estos modelos son adecuados para datos con no normalidad y varianza no constante, lo que es común en encuestas de hogares.
- Las variables en las encuestas de hogares a menudo son de tipo conteo, binarias, etc., lo que hace que el análisis mediante MLGs sea relevante y útil.

Lectura de las bases de datos y definición del diseño muestral.

```
library(srvyr)
library(survey)
encuesta <- readRDS("../Data/encuesta.rds")
data("BigCity", package = "TeachingSampling")
diseno <- encuesta %>%
  as_survey_design(
    strata = Stratum,
    ids = PSU,
    weights = wk,
    nest = T
  )
```

Creación de nuevas variables.

Las nuevas variables son definidas de la siguiente forma.

```
diseno <- diseno %>% mutate(
  pobreza = ifelse(Poverty != "NotPoor", 1, 0),
  desempleo = ifelse(Employment == "Unemployed", 1, 0))
```

Tablas de doble entrada para el tamaño El cálculo de tablas de doble entrada las obtenemos con así:

```
(tab_pobreza_sexo <- svyby(~factor(pobreza), ~Sex,
      FUN = svytotal, design = as.svrepdesign(diseno),
      se=F, na.rm=T, ci=T, keep.var=TRUE))
```

	Sex	factor(pobreza)0	factor(pobreza)1	se1	se2
Female	Female	48366	30824	2411	2916
Male	Male	43032	28044	2522	3095

Tablas de doble entrada para el tamaño

Sin embargo para la estimación de tamaños más simples podemos emplear la función.

```
tab <- svytable(~pobreza + Sex, design = diseno)
data.frame(tab)
```

pobreza	Sex	Freq
0	Female	48366
1	Female	30824
0	Male	43032
1	Male	28044

Tablas de doble entrada para el proporción

Al hacer uso de la función `svymean` es posible estimar al proporciones.

```
(tab_pobreza_sexo <- svyby(~factor(pobreza), ~Sex,
      FUN = svymean, design = as.svrepdesign(diseno),
      se=F, na.rm=T, ci=T, keep.var=TRUE))
```

	Sex	factor(pobreza)0	factor(pobreza)1	se1	se2
Female	Female	0.6108	0.3892	0.0316	0.0316
Male	Male	0.6054	0.3946	0.0366	0.0366

Tablas de doble entrada para el proporción

En forma alternativa es posible usar la función `prop.table` del paquete base.

```
prop.table(tab, margin = 2) %>% data.frame()
```

pobreza	Sex	Freq
0	Female	0.6108
1	Female	0.3892
0	Male	0.6054
1	Male	0.3946

Estas diferentes formas de proceder son de mucha importancia al momento de hacer uso de pruebas de independencia en tablas cruzadas.

Prueba de independencia F

Prueba de independencia F

La prueba de independencia F de Fisher permite analizar si dos variables dicotómicas están asociadas cuando la muestra a estudiar es demasiado pequeña y no se cumplen las condiciones para aplicar la prueba χ^2. Para utilizar esta técnica, tengamos en cuenta que la probabilidad estimada se escribe como:

$$\hat{\pi}_{rc} = \frac{n_{r+}}{n_{++}} \times \frac{n_{+c}}{n_{++}}$$

Prueba de independencia F

Teniendo en cuenta esta expresión, la estadística $\chi 2$ de Pearson se define de la siguiente manera:

$$\chi^2_{pearsom} = n_{++} \times \sum_r \sum_c \left(\frac{\left(p_{rc} - \hat{\pi}_{rc}\right)^2}{\hat{\pi}_{rc}} \right)$$

y la estadística de razón de verosimilitud se define como:

$$G^2 = 2 \times n_{++} \times \sum_r \sum_c p_{cr} \times \ln\left(\frac{p_{rc}}{\hat{\pi}_{rc}}\right)$$

donde, r es el número de filas y c representa el número de columnas, la prueba tiene $(R-1) \times (C-1)$ grados de libertad.

Correcciones del Estadístico Chi-Cuadrado en Encuestas

- La corrección del estadístico chi-cuadrado de Pearson se utiliza en análisis de datos de encuestas para ajustar el efecto de diseño.
- *Fay (1979, 1985)* y *Fellegi (1980)* fueron pioneros en proponer correcciones basadas en un efecto de diseño generalizado (GDEFF).
- *Rao y Scott (1984)*, junto con *Thomas y Rao (1987)*, ampliaron la teoría de las correcciones del efecto de diseño generalizado.
- El método de Rao-Scott es un estándar para el análisis de datos de encuestas categóricas en software como Stata y SAS.

Estadísticos de Prueba

- Los estadísticos de prueba Rao-Scott Pearson y razón de verosimilitud chi-cuadrado se utilizan para analizar la asociación en datos de encuestas categóricas.
- Estos estadísticos se calculan mediante correcciones basadas en efectos de diseño generalizados.
- Las correcciones de Rao-Scott son analíticamente más complicadas que el enfoque de Fellegi, pero se consideran más precisas.
- Son ampliamente utilizados en el análisis de datos de encuestas, especialmente en software estadístico como Stata, SAS y R.

Estadísticos de Prueba χ^2 y G^2

Los estadísticos de prueba Rao-Scott Pearson ajustados por diseño y razón de verosimilitud chi-cuadrado se calculan de la siguiente manera:

$$\chi^2_{(R-S)} = \chi^2_{(Pearson)} \Big/ GDEFF$$

y, para la estadística basada en la razón de verosimilitud se calcula como:

$$G^2_{(R-S)} = G^2 \Big/ GDEFF$$

donde el efecto generalizado del diseño ($GDEFF$) de Rao–Scott, está dado por

$$GDEFF = \frac{\sum_r \sum_c \left(1 - p_{rc}\right) d^2 \left(p_{rc}\right) - \sum_r \left(1 - p_{r+}\right) d^2 \left(p_{r+}\right) - \sum_c \left(1 - p_{+c}\right) d^2 \left(p_{+c}\right)}{\left(R - 1\right) \left(C - 1\right)}$$

Prueba de independencia F

La estadística F para independencia basada en la chi-cuadrado de Pearson se calcula como sigue:

$$F_{R-S,Pearson} = \chi^2_{R-S} \Big/ \left[(R-1)(C-1)\right] \sim F_{(R-1)(C-1),(R-1)(C-1)df}$$

y, la estadística F para independencia basada en la razón de verosimilitudes se calcula como sigue:

$$F_{R-S,LRT} = G^2_{R-S} \Big/ (C-1) \sim F_{(C-1),df}$$

donde C es el número de columnas de la tabla cruzada

Prueba de independencia ChiSq

En `R`, el cálculo de las estadísticas chi-cuadrado y F se calculan usando la función `summary` como se muestra a continuación:

```
summary(tab, statistic = "Chisq")
```

```
       Sex
pobreza Female  Male
      0  48366 43032
      1  30824 28044

    Pearson's X^2: Rao & Scott adjustment

data:  NextMethod()
X-squared = 0.077, df = 1, p-value = 0.8
```

Se puede concluir que el estado de pobreza y el sexo no están relacionados con una confianza del 95%.

Prueba de independencia F

```r
summary(tab, statistic = "F")
```

```
        Sex
pobreza Female  Male
      0  48366 43032
      1  30824 28044

    Pearson's X^2: Rao & Scott adjustment

data:  NextMethod()
F = 0.056, ndf = 1, ddf = 119, p-value = 0.8
```

Estadístico de Wald

Estadístico de Wald

Este estadístico se aplica cuando ya se ha elegido un modelo estadístico (regresión lineal simple, regresión logística, entre otros).

El estadístico de prueba de Wald χ^2 para la hipótesis nula de independencia de filas y columnas en una tabla de doble entrada se define de la siguiente manera:

$$Q_{wald} = \hat{Y^t} \left(H\hat{V}\left(\hat{N}\right) H^t \right)^{-1} \hat{Y}$$

donde,

$$\hat{Y} = \left(\hat{N} - E\right)$$

es un vector de $R \times C$ de diferencias entre los recuentos de celdas observadas y esperadas, esto es, $\hat{N}_{rc} - E_{rc}$

La matriz $H\hat{V}\left(\hat{N}\right) H^t$, representa la matriz de varianza-covarianza estimada para el vector de diferencias.

Estadístico de Wald

La matriz H es la inversa de la matriz J dada por:

$$J = -\left[\frac{\delta^2 \ln PL\left(B\right)}{\delta^2 B}\right] \mid B = \hat{B}$$

$$\sum_h \sum_a \sum_i x_{hai}^t x_{hai} w_{hai} \hat{\pi}_{hai}\left(B\right)\left(1 - \hat{\pi}_{hai}\left(B\right)\right)$$

Bajo la hipótesis nula de independencia, el estadístico de wald se distribuye chi cuadrado con $(R-1) \times (C-1)$ grados de libertad,

$$Q_{wald} \sim \chi^2_{(R-1)\times(C-1)}$$

Estadístico de Wald

La transformación F del estadístico de Wald es:

$$F_{wald} = Q_{wald} \times \frac{df - (R-1)(C-1) + 1}{(R-1)(C-1)\, df} \sim F_{(R-1)(C-1), df-(R-1)(C-1)+1}$$

Prueba de independencia Wald

En R, para calcular el estadístico de Wald se hace similarmente al cálculo de los estadísticos anteriores usando la función summary como sigue:

```
summary(tab, statistic = "Wald")
```

```
       Sex
pobreza Female  Male
      0  48366 43032
      1  30824 28044

    Design-based Wald test of association

data:  NextMethod()
F = 0.056, ndf = 1, ddf = 119, p-value = 0.8
```

Se puede concluir que, con una confianza del 95% y basado en la muestra no hay relación entre el estado de pobreza y el sexo.

Prueba de independencia adjWald

El estadístco de Wald ajustado en R se se calcula similarmente al anterior y los resultados fueron similares:

```r
summary(tab, statistic = "adjWald")
```

```
        Sex
pobreza Female  Male
      0  48366 43032
      1  30824 28044

    Design-based Wald test of association

data:  NextMethod()
F = 0.056, ndf = 1, ddf = 119, p-value = 0.8
```

Modelo log lineal

Modelo log lineal para tablas de contingencia

El término modelo log-lineal, que básicamente describe el papel de la función de enlace que se utiliza en los modelos lineales generalizados. Iniciaremos esta sección con los modelos log-lineales en tablas de contingencia. El modelo estadístico es el siguiente:

$$\log(p_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY},$$

donde:

- p_{ijk} = la proporción esperada en la celda bajo el modelo.
- $\mu = \log(p_0) = \frac{1}{\# \ de \ celdas}$

Modelo log lineal para tablas de contingencia

El modelo log-lineal en R se ajusta utilizando la función svyloglin como sigue:

```
mod1 <- svyloglin(~pobreza+Sex + pobreza:Sex , diseno)
(s1 <- summary(mod1))
```

```
Loglinear model: svyloglin(~pobreza + Sex + pobreza:Sex, diseno)
                  coef      se        p
pobreza1      0.219673 0.06778 0.001192
Sex1          0.052843 0.01625 0.001145
pobreza1:Sex1 0.005583 0.02350 0.812175
```

Los resultados muestran que, con una confianza del 95% el estado de pobreza es independiente del sexo, como se ha mostrado con las pruebas anteriores.

Modelo log lineal para tablas de contingencia

En la salida anterior se pudo observar que la interacción es no significativa, entonces, ajustemos ahora el modelo sin interacción:

```r
  mod2 <- svyloglin(~pobreza+Sex, diseno)
(s2 <- summary(mod2))
```

```
Loglinear model: svyloglin(~pobreza + Sex, diseno)
             coef      se          p
pobreza1 0.21997 0.06752 0.0011230
Sex1     0.05405 0.01577 0.0006076
```

Modelo log lineal para tablas de contingencia

Mediante un análisis de varianza es posible comparar los dos modelos.

```r
anova(mod1, mod2)
```

```
Analysis of Deviance Table
 Model 1: y ~ pobreza + Sex
Model 2: y ~ pobreza + Sex + pobreza:Sex
Deviance= 0.07719 p= 0.8827
Score= 0.07719 p= 0.8827
```

De la anterior salida se puede concluir que, con una confianza del 95%, la interacción no es significativa en el modelo log-lineal ajustado.

Modelo de regresión logistica

Modelo de regresión logistica

Un modelo de regresión logística es un modelo matemático que puede ser utilizado para describir la relación entre un conjunto de variables independientes y una variable dicotomica Y. El modelo logístico se describe a continuación:

$$g(\pi(x)) = logit(\pi(x))$$

De aquí,

$$z = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = B_0 + B_1 x_1 + \cdots + B_p x_p$$

Modelo de regresión logistica

La probabilidad estimada utilizando el modelo logístico es la siguiente:

$$\hat{\pi}\left(x\right)=\frac{\exp\left(X\hat{B}\right)}{1-\exp\left(X\hat{B}\right)}=\frac{\exp\left(\hat{B}_0+\hat{B}_1x_1+\cdots+\hat{B}x_p\right)}{1-\exp\left(\hat{B}_0+\hat{B}_1x_1+\cdots+\hat{B}x_p\right)}$$

$$\pi\left(x_i\right)=\frac{\exp\left(x_iB\right)}{1-\exp\left(x_iB\right)}$$

La varianza del modelo de regresión logistica

La varianza de los parámetros estimados se calcula como sigue:

$$var\left(\hat{B}\right) = J^{-1} var\left(S\left(\hat{B}\right)\right) J^{-1}$$

con,

$$S\left(B\right) = \sum_{h} \sum_{a} \sum_{i} w_{hai} D_{hai}^{t} \left[\left(\pi_{hai}\left(B\right)\right)\left(1-\pi_{hai}\left(B\right)\right)\right]^{-1}\left(y_{hai}-\pi_{hai}\left(B\right)\right) = 0$$

$$D_{hai} = \frac{\delta\left(\pi_{hai}\left(B\right)\right)}{\delta B_{j}}$$

donde $j = 0, \dots, p$

Prueba de Wald para los parámetros del modelo

El estadístico de Wald para la significancia de los parámetros del modelo se utiliza la razón de verosimilitud. En este caso se contrastan el modelo con todos los parámetros (modelo full) versus el modelo reducido, es decir, el modelo con menos parámetros (modelo reduced),

$$G = -2\ln\left[\frac{L\left(\hat{\beta}_{MLE}\right)_{reduced}}{L\left(\hat{\beta}_{MLE}\right)_{full}}\right]$$

Intervalo de confianza

Para construir los intervalos de confianza se debe aplicar el función exponencial a cada parámetro,

$$\hat{\psi} = \exp\left(\hat{B}_1\right)$$

por ende, el intervalo de confianza es:

$$CI\left(\psi\right) = \exp\left(\hat{B}_j \pm t_{df,1-\frac{\alpha}{2}} se\left(\hat{B}_j\right)\right)$$

Modelo log lineal ajustado

En R se muestra el ajuste de un modelo logístico teniendo e cuenta el diseño muestral

```
mod_loglin <- svyglm(
  pobreza ~ Sex + Zone + Region,
  family=quasibinomial, design=diseno)
tidy(mod_loglin)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.4082	0.2640	-1.5464	0.1248
SexMale	0.0086	0.0915	0.0945	0.9249
ZoneUrban	-0.4378	0.2418	-1.8106	0.0729
RegionSur	0.0063	0.3140	0.0201	0.9840
RegionCentro	0.1915	0.4279	0.4476	0.6553
RegionOccidente	0.2319	0.3144	0.7377	0.4622
RegionOriente	0.3699	0.4259	0.8686	0.3869

La salida muestra que ninguna de las covariables son significativas con una confianza del 95%.

Intervalo de confianza para el modelo

Los intervalos de confianza en los cuales se pueden concluir que en todos los parámetros el cero se encuentra dentro del intrevalo:

```
bind_cols(
  data.frame(exp_estimado = exp(coef(mod_loglin))),
  as.data.frame(exp(confint(mod_loglin)))
)
```

	exp_estimado	2.5 %	97.5 %
(Intercept)	0.6648	0.3941	1.122
SexMale	1.0087	0.8414	1.209
ZoneUrban	0.6454	0.3997	1.042
RegionSur	1.0063	0.5402	1.875
RegionCentro	1.2111	0.5188	2.827
RegionOccidente	1.2611	0.6764	2.351
RegionOriente	1.4476	0.6226	3.366

Plot de la distribución de los betas

```
plot_summs(mod_loglin,
           scale = TRUE, plot.distributions = TRUE)
```

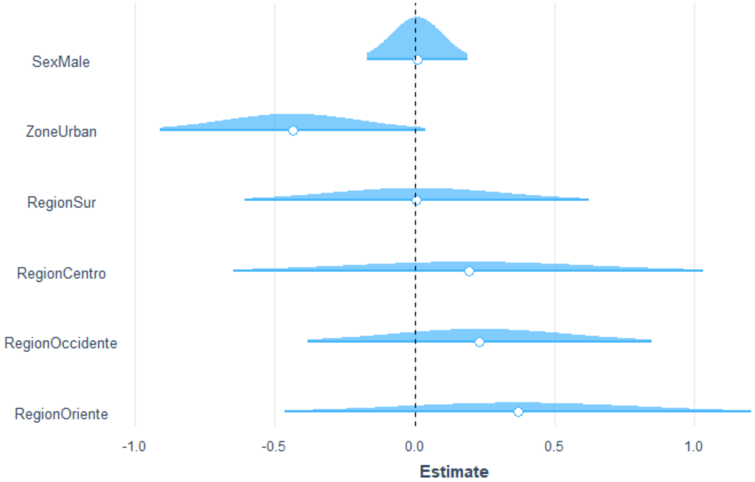

Figura 1: Intervalo de confianza para los coeficiente

Estadístico de Wald sobre los parámetros

El estadístico de Wald para el cada una de las variables del modelo se calcula a continuación con la función `regTermTest` para las variables del modelo:

```r
regTermTest(model = mod_loglin, ~Sex)
```

```
Wald test for Sex
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno,
    family = quasibinomial)
F =  0.00893  on  1  and  113  df: p= 0.92
```

```r
regTermTest(model = mod_loglin, ~Zone)
```

```
Wald test for Zone
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno,
    family = quasibinomial)
F =  3.278  on  1  and  113  df: p= 0.073
```

Estadístico de Wald sobre los parámetros

```r
regTermTest(model = mod_loglin, ~Region)
```

```
Wald test for Region
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno,
    family = quasibinomial)
F =  0.3654  on  4  and  113  df: p= 0.83
```

Efecto del modelo.

Para evaluar los efectos de la variable en el modelo:

```
effe_sex <- effect_plot(mod_loglin, pred = Sex,
                        interval = TRUE)
effe_Zona <-effect_plot(mod_loglin, pred = Zone,
                        interval = TRUE)
effe_Region <- effect_plot(mod_loglin, pred = Region,
                           interval = TRUE)
(effe_sex |effe_Zona)/effe_Region
```

Efecto del modelo.

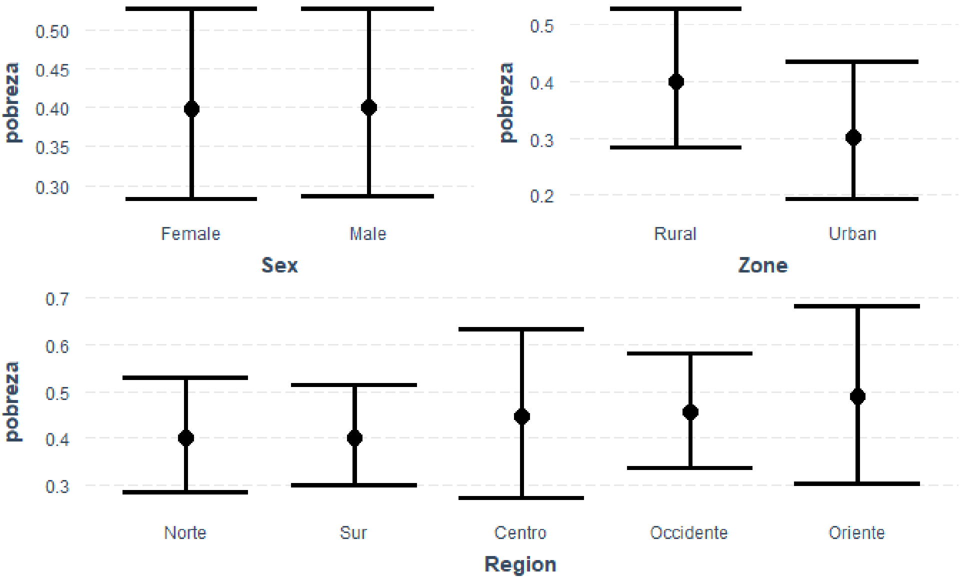

Figura 2: Efectos del modelo

Modelo log lineal ajustado con interacciones

```
mod_loglin_int <- svyglm(
  pobreza ~ Sex + Zone + Region +
    Sex:Zone + Sex:Region,
  family = quasibinomial,
  design = diseno
)
```

Modelo log lineal ajustado con interacciones

term	estimate	std.error	statistic	p.value
(Intercept)	-0.4289	0.2849	-1.5055	0.1351
SexMale	0.0478	0.1994	0.2399	0.8109
ZoneUrban	-0.4248	0.2562	-1.6580	0.1002
RegionSur	-0.1325	0.3464	-0.3825	0.7028
RegionCentro	0.2466	0.4560	0.5408	0.5897
RegionOccidente	0.3342	0.3783	0.8835	0.3790
RegionOriente	0.3843	0.4279	0.8980	0.3712
SexMale:ZoneUrban	-0.0154	0.1872	-0.0824	0.9345
SexMale:RegionSur	0.2871	0.2774	1.0348	0.3031
SexMale:RegionCentro	-0.1162	0.2791	-0.4162	0.6781
SexMale:RegionOccidente	-0.2302	0.2868	-0.8026	0.4240
SexMale:RegionOriente	-0.0304	0.2878	-0.1057	0.9161

Plot de la distribución de los betas

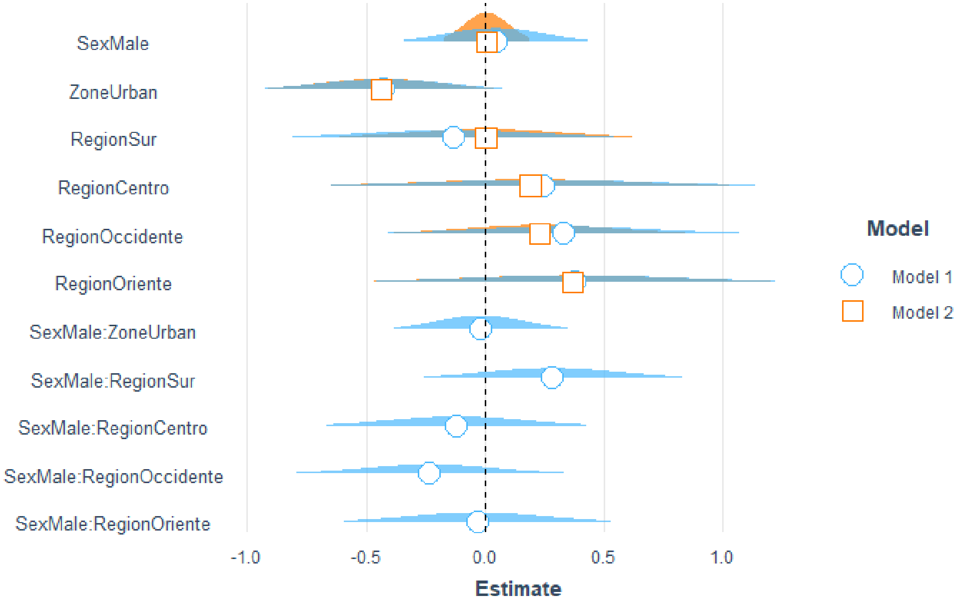

Figura 3: Comparando los modelos

Modelo log lineal ajustado

Observándose que con una confianza del 95% ninguno de los parámetros del modelo es significativo.

	exp_estimado	2.5 %	97.5 %
(Intercept)	0.6512	0.3702	1.145
SexMale	1.0490	0.7065	1.557
ZoneUrban	0.6539	0.3935	1.087
RegionSur	0.8759	0.4408	1.740
RegionCentro	1.2797	0.5183	3.160
RegionOccidente	1.3968	0.6599	2.957
RegionOriente	1.4685	0.6288	3.430
SexMale:ZoneUrban	0.9847	0.6795	1.427
SexMale:RegionSur	1.3325	0.7689	2.309
SexMale:RegionCentro	0.8903	0.5120	1.548
SexMale:RegionOccidente	0.7944	0.4499	1.403
SexMale:RegionOriente	0.9701	0.5484	1.716

Estadístico de Wald sobre los parámetros

Evaluando las variables en el modelo

```
regTermTest(model = mod_loglin_int, ~Sex)
```

```
Wald test for Sex
 in svyglm(formula = pobreza ~ Sex + Zone + Region + Sex:Zone + Sex:Region
    design = diseno, family = quasibinomial)
F =  0.05753  on  1  and  108  df: p= 0.81
```

```
regTermTest(model = mod_loglin_int, ~Zone)
```

```
Wald test for Zone
 in svyglm(formula = pobreza ~ Sex + Zone + Region + Sex:Zone + Sex:Region
    design = diseno, family = quasibinomial)
F =  2.749  on  1  and  108  df: p= 0.1
```

Estadístico de Wald sobre los parámetros

Evaluando las variable región en el modelo

```
regTermTest(model = mod_loglin_int, ~Region)
```

```
Wald test for Region
 in svyglm(formula = pobreza ~ Sex + Zone + Region + Sex:Zone + Sex:Region
    design = diseno, family = quasibinomial)
F =  0.8999  on  4  and  108  df: p= 0.47
```

Estadístico de Wald sobre los parámetros

Evaluando la interacción de los modelos.

```
regTermTest(model = mod_loglin_int, ~Sex:Zone)
```

```
Wald test for Sex:Zone
 in svyglm(formula = pobreza ~ Sex + Zone + Region + Sex:Zone + Sex:Region
    design = diseno, family = quasibinomial)
F =  0.006789  on  1  and  108  df: p= 0.93
```

```
regTermTest(model = mod_loglin_int, ~Sex:Region)
```

```
Wald test for Sex:Region
 in svyglm(formula = pobreza ~ Sex + Zone + Region + Sex:Zone + Sex:Region
    design = diseno, family = quasibinomial)
F =  1.058  on  4  and  108  df: p= 0.38
```

Efecto del modelo.

Evaluando los efectos en el modelo.

```
effe_sex <- effect_plot(mod_loglin_int,
                        pred = Sex,
                        interval = TRUE)
effe_Zona <-effect_plot(mod_loglin_int,
                        pred = Zone,
                        interval = TRUE)
effe_Region <- effect_plot(mod_loglin_int,
                           pred = Region,
                           interval = TRUE)
(effe_sex |effe_Zona)/effe_Region
```

Efecto del modelo.

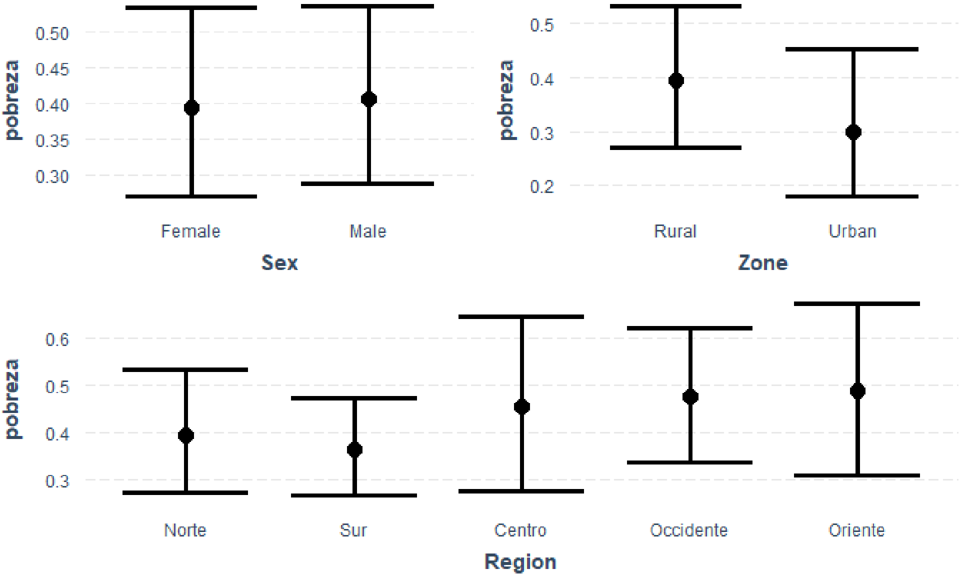

Figura 4: Efectos del modelo

Modelo log lineal ajustado con Q-Weighting

Realizando el modelo con los Q-Weighting

```
fit_wgt <- lm(wk ~  Sex + Zone + Region ,
              data = encuesta)
wgt_hat <- predict(fit_wgt)
encuesta %<>% mutate(wk2 = wk/wgt_hat)

diseno_qwgt <- encuesta %>%
  as_survey_design(
    strata = Stratum,
    ids = PSU,
    weights = wk2,
    nest = T
  )
```

Modelo log lineal ajustado con Q_Weighting

Defiendo la variable pobreza dentro de la base de datos.

```
diseno_qwgt <- diseno_qwgt %>% mutate(
  pobreza = ifelse(Poverty != "NotPoor", 1, 0))
# Estimando el modelo.
mod_loglin_qwgt <- svyglm(pobreza ~ Sex + Zone + Region,
                          family=quasibinomial,
                          design=diseno_qwgt)
(tab_mod <- tidy(mod_loglin_qwgt) )
```

Modelo log lineal ajustado con Q_Weighting

term	estimate	std.error	statistic	p.value
(Intercept)	-0.4644	0.2630	-1.7656	0.0802
SexMale	0.0241	0.0883	0.2726	0.7857
ZoneUrban	-0.3445	0.2311	-1.4903	0.1389
RegionSur	-0.0041	0.3116	-0.0130	0.9896
RegionCentro	0.1613	0.4270	0.3778	0.7063
RegionOccidente	0.2424	0.3147	0.7705	0.4426
RegionOriente	0.3937	0.4319	0.9115	0.3639

Plot de la distribución de los betas

```
plot_summs(mod_loglin, mod_loglin_qwgt,
           scale = TRUE, plot.distributions = TRUE)
```

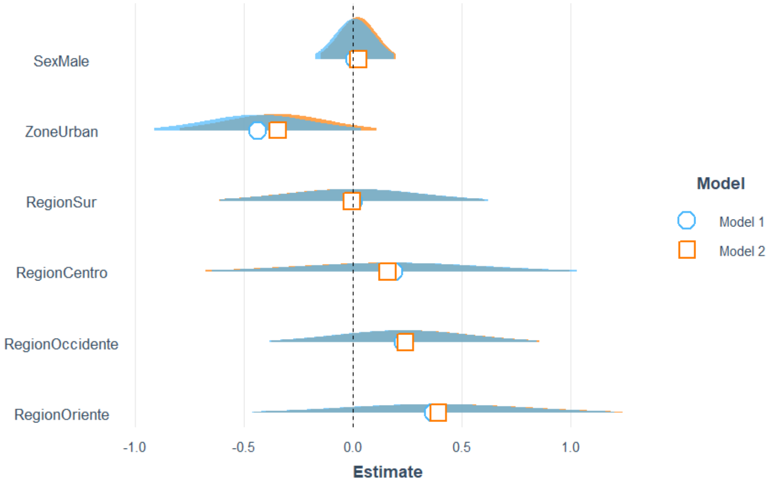

Figura 5: Comparando los modelos

Modelo log lineal ajustado

	exp_estimado	2.5 %	97.5 %
(Intercept)	0.6285	0.3732	1.058
SexMale	1.0244	0.8600	1.220
ZoneUrban	0.7086	0.4482	1.120
RegionSur	0.9960	0.5371	1.847
RegionCentro	1.1750	0.5043	2.738
RegionOccidente	1.2744	0.6832	2.377
RegionOriente	1.4824	0.6301	3.488

Estadístico de Wald sobre los parámetros

```r
regTermTest(model = mod_loglin_qwgt, ~Sex)
```

```
Wald test for Sex
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno_qwgt,
    family = quasibinomial)
F =  0.0743  on  1  and  113  df: p= 0.79
```

```r
regTermTest(model = mod_loglin_qwgt, ~Zone)
```

```
Wald test for Zone
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno_qwgt,
    family = quasibinomial)
F =  2.221  on  1  and  113  df: p= 0.14
```

Estadístico de Wald sobre los parámetros

```r
regTermTest(model = mod_loglin_qwgt, ~Region)
```

```
Wald test for Region
 in svyglm(formula = pobreza ~ Sex + Zone + Region, design = diseno_qwgt,
    family = quasibinomial)
F =  0.4156  on  4  and  113  df: p= 0.8
```

Efecto del modelo.

```
effe_sex <- effect_plot(mod_loglin_qwgt,
                        pred = Sex,
                        interval = TRUE)
effe_Zona <-effect_plot(mod_loglin_qwgt,
                        pred = Zone,
                        interval = TRUE)
effe_Region <- effect_plot(mod_loglin_qwgt,
                           pred = Region,
                           interval = TRUE)
(effe_sex |effe_Zona)/effe_Region
```

Efecto del modelo.

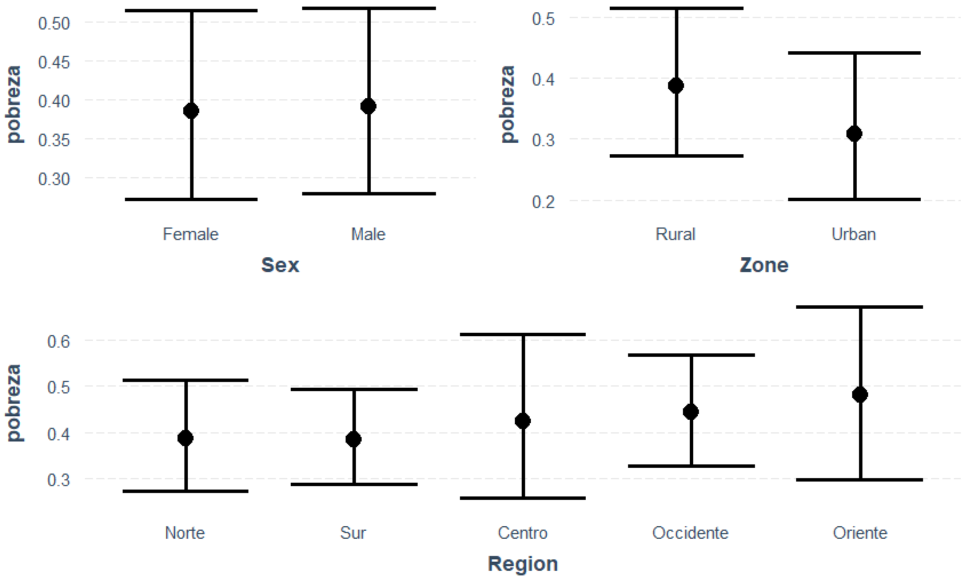

Figura 6: Efecto del modelo

¡Gracias!

Email: andres.gutierrez@cepal.org