Modelo SAE para estimar estadísticas del trabajo

Resultados del procesamiento

CEPAL

3/1/23

Introducción

- La Encuesta de Caracterización Socioeconómica Nacional (CASEN) de Chile 2017.
- Los niveles de estimación o dominios de estudio para los cuales la muestra fue representativa son: nacional, nacional urbano, nacional rural y regional.
- El diseño corresponde a una muestra probabilística, estratificada y bietápica, siendo los estratos conformados por la dupla Comuna-Área.
- La base de datos de la encuesta cuenta con 216439 registros, distribuidos en 1637 upm

Selección de dominios para modelo de área

Después de realizar la estimación directa en los dominios presentes en la muestra se aplican los siguientes criterios para la selección los dominios incluidos en el modelo de área.

- El municipio tiene dos o más upm
- Tener estimación del Deff para las tres tasas estimadas.

El resultado de aplicar estos criterios fue conservar 309 municipios.

Definición del modelo multinomial

- Sea K el número de categorías de la variable de interés $Y \sim multinomial\,(p)$, con $p = (p_1, p_2, \dots, p_k)$ y $\sum_{k=1}^{K} p_k = 1$.
- Sea N_i el número de elementos en el i-ésimo dominio y N_{ik} el número de elementos que tienen la k-ésima categoría, note que $\sum_{k=1}^{K} N_{ik} = N_i$ y $p_{ik} = \frac{N_{ik}}{N_i}$.
- Sea $\hat{p}_{ik}$ la estimación puntual de p_{ik} y $v_{ik} = Var\,(\hat{p}_{ik})$ y denote el estimador de la varianza por $\hat{v}_{ik}$.

Definición del tamaño de muestra efectivo

Note que el efecto diseño cambia entre categoría, por tanto, lo primero será definir el tamaño de muestra efectivo por categoría. Esto es:

La estimación de $\tilde{n}$ esta dado por $\tilde{n}_{ik} = \frac{(\tilde{p}_{ik} \times (1 - \tilde{p}_{ik}))}{\hat{v}_{ik}}$,

$\tilde{y}_{ik} = \tilde{n}_{ik} \times \hat{p}_{ik}$

luego, $\hat{n}_i = \sum_{k=1}^{K} \tilde{y}_{ik}$

de donde se sigue que $\hat{y}_{ik} = \hat{n}_i \times \hat{p}_{ik}$

Definición del modelo multinomial

$p = (p_1, p_2, \dots p_k)^T = \left(\frac{N_{i1}}{N_i}, \frac{N_{i2}}{N_i} \frac{N_{i3}}{N_i}\right)$, entonces el modelo multinomial estaría dado por:

$$(\tilde{y}_{i1}, \tilde{y}_{i2}, \tilde{y}_{i3}) \mid \hat{n}_i, p_i \sim multinomial\,(\hat{n}_i, p_i)$$

donde $\ln\left(\frac{p_{i2}}{p_{i1}}\right) = X_i^T \beta_2 + u_{i2}$ y $\ln\left(\frac{p_{i3}}{p_{i1}}\right) = X_i^T \beta_3 + u_{i3}$

Reescribiendo p

Dada la restricción $1 = p_{i1} + p_{i2} + p_{i3}$ entonces $p_{i1} + p_{i1}e^{X_i^T\beta_2} + p_{i1}e^{X_i^T\beta_3}$ de donde se sigue que

$$p_{i1} = \frac{1}{1 + e^{X_i^T\beta_2} + e^{{X_i}^T\beta_2}}$$

$$p_{i2} = \frac{e^{X_i^T\beta_2}}{1 + e^{X_i^T\beta_2} + e^{{X_i}^T\beta_2}}$$

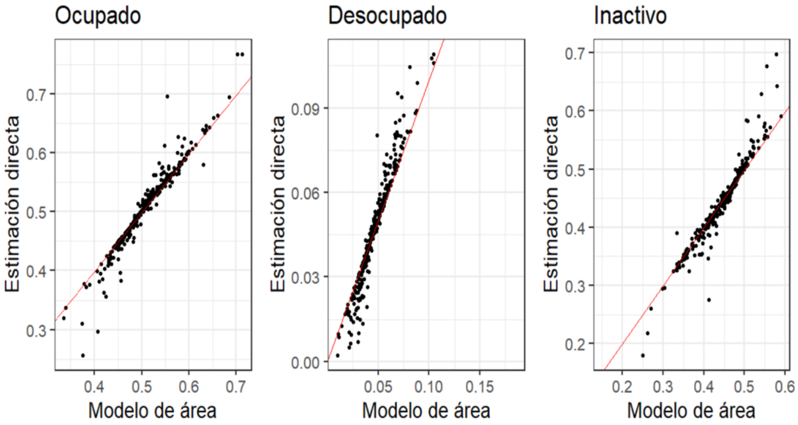
$$p_{i3} = \frac{e^{X_i^T\beta_3}}{1 + e^{X_i^T\beta_2} + e^{{X_i}^T\beta_2}}$$

Covariables del modelo de área

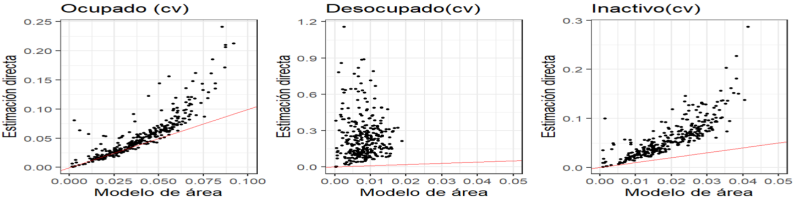
- Luces nocturnas
- Cubrimiento de suelo urbano
- Cubrimiento de suelo cultivos
- Distancia a hospitales
- Distancia a hospitales vehículo no motorizado
- Modificación humana

Resultados de la estimación

Comparando el modelo de área y la estimación directa



Comparando el modelo de área y la estimación directa (cv)



Matriz de correlación

variable	mean	median	sd	mad	q5	q95
Omega[1,1]	1.0000	1.0000	0.0000	0.0000	1.0000	1.0000
Omega[2,1]	0.3437	0.3474	0.0648	0.0636	0.2327	0.4449
Omega[1,2]	0.3437	0.3474	0.0648	0.0636	0.2327	0.4449
Omega[2,2]	1.0000	1.0000	0.0000	0.0000	1.0000	1.0000