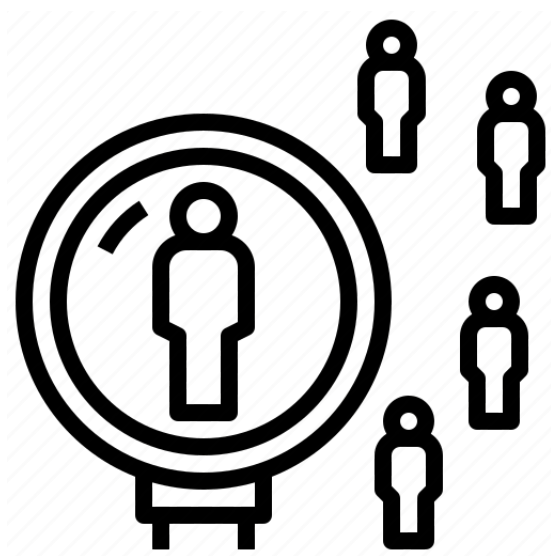


UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONOMICAS Y DE ADMINISTRACIÓN
LICENCIATURA EN ESTADÍSTICA

MUESTREO II



PROYECTO FINAL

Ignacio Acosta - Valentina Caldiroli - Mauro Loprete

Parte 1 : Estimaciones con ponderadores originales

Se calculan las estimaciones con los ponderadores originales, estimaciones de la tasa de desempleo, la proporción de personas pobres e ingreso promedio.

Dada la existencia de no respuesta en la muestra y el tratamiento realizado, estamos frente a **una postura determinística de la no respuesta**.

A continuación se muestra el código utilizado para realizar las diferentes estimaciones :

```
muestra %>%
  as_survey_design(
    ids = 1,
    weight = w0,
    strata = estrato
  ) %T>%
  assign(
    "diseño",
    .,
    envir = .GlobalEnv
  ) %>%
  filter(
    R > 0
  ) %>%
  summarize(
    td = survey_ratio(
      desocupado,
      activo,
      deff = TRUE,
      vartype = c("se", "cv")
    ) * 100,
    pobre = survey_mean(
      pobreza,
      deff = TRUE,
      vartype = c("se", "cv")
    ),
    yprom = survey_mean(
      ingreso,
      deff = TRUE,
      vartype = c("se", "cv")
    )
  ) %>%
  assign(
    "est_originales",
    .,
    envir = .GlobalEnv
  )
```

Los resultados se encuentran en el siguiente cuadro:

Cuadro 1: Estimaciones poblacionales usando ponderadores originales

Variable	Estimación puntual	Error estandar	CV	deff
pobre	0.085	0.002	0.028	1.034
td	8.155	0.331	4.060	103.806
yprom	22037.709	255.394	0.012	0.925

En base al cuadro, podemos ver que los errores estandar son relativamente chicos, en cambio podemos notar una gran variabilidad en las estimaciones para la tasa de desempleo entre muestra y muestra, esto es, si consideramos el coeficiente de variación.

A lo que refiere al efecto diseño, que recordemos, refiere a la reducción (o aumento de varianza) respecto a un diseño simple, se puede ver que son valores coherentes, ya que recordemos que el diseño de esta encuesta cuenta con varias etapas de selección.

Este estadístico tiene el mismo problema que el coeficiente de variación a lo que refiere a las estimaciones de la tasa de desocupación, llamando nuestra atención.

Tasa de no respuesta

Desde un enfoque determinístico de la no respuesta, podemos dividir a la muestra en aquellos individuos que respondieron y en aquellos que no lo hicieron.

Es decir, podemos particionar la muestra en aquellos respondentes r_u y no respondentes $s - r_u$.

Una medida de interés, es ver la proporción de respuestas en nuestra muestra, definida como :

$$p_{r_u} = \frac{n_{r_u}}{n_s}$$

Para nuestra muestra particular, este dato viene dado por :

```
muestra %>%
  summarize.(
    tr = mean(R)
  ) %>%
  mutate.(
    tnr = 1 - tr
  ) %>%
  assign(
    "tasaRespuesta",
    .,
    envir = .GlobalEnv
  )
```

Cuadro 2: Tasa de Respuesta

Tasa de Respuesta	Tasa de No Respuesta
0.54	0.46

En base a este indicador, podemos ver que poco más de la mitad de las personas seleccionadas en la muestra se pudo recabar información.

Por último, podemos ver la tasa de no respuesta poblacional, definida como :

$$\hat{p}_{r_u} = \frac{\sum_{r_u} w_0}{\sum_{r_s} w_0} = \frac{\hat{N}_{r_u}}{\hat{N}_s}$$

```
muestra %>%
  summarize.(
    tr = sum(R*w0) / sum(w0)
  ) %>%
  assign(
    "tasaRespuestapob",
    .,
    envir = .GlobalEnv
  )
```

Cuadro 3: Tasa de Respuesta poblacional

Tasa de respuesta poblacional	Tasa de no respuesta poblacional
0.54	0.46

De manera análoga, la tasa de respuesta es idéntica a la anterior, si consideramos dos cifras significativas. Esta estimación puede tener la siguiente interpretación : *el porcentaje de la población que estoy cubriendo una vez que expanda la muestra*, que para este caso particular, **es sumamente bajo**.

Parte 2

Parte 3

Parte 4

Referencias

Paquetes de R

- [1] Stefan Milton Bache y Hadley Wickham. magrittr: A Forward-Pipe Operator for R. R package version 2.0.1. 2020. URL: <https://CRAN.R-project.org/package=magrittr>.
- [2] Mark Fairbanks. tidytable: Tidy Interface to data.table. R package version 0.6.5. 2021. URL: <https://github.com/markfairbanks/tidytable>.
- [3] Greg Freedman Ellis y Ben Schneider. srvyr: dplyr-Like Syntax for Summary Statistics of Survey Data. R package version 1.1.0. 2021. URL: <https://CRAN.R-project.org/package=srvyr>.
- [4] Kirill Müller. here: A Simpler Way to Find Your Files. R package version 1.0.1. 2020. URL: <https://CRAN.R-project.org/package=here>.
- [5] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: <https://www.R-project.org/>.
- [6] Tyler Rinker y Dason Kurkiewicz. pacman: Package Management Tool. R package version 0.5.1. 2019. URL: <https://github.com/trinker/pacman>.
- [7] Tyler W. Rinker y Dason Kurkiewicz. pacman: Package Management for R. version 0.5.0. Buffalo, New York, 2018. URL: <http://github.com/trinker/pacman>.
- [8] Hadley Wickham. forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. 2021. URL: <https://CRAN.R-project.org/package=forcats>.
- [9] Hadley Wickham y Jennifer Bryan. readxl: Read Excel Files. R package version 1.3.1. 2019. URL: <https://CRAN.R-project.org/package=readxl>.
- [10] Hadley Wickham y col. dplyr: A Grammar of Data Manipulation. R package version 1.0.7. 2021. URL: <https://CRAN.R-project.org/package=dplyr>.
- [11] Hao Zhu. kableExtra: Construct Complex Table with kable and Pipe Syntax. R package version 1.3.4. 2021. URL: <https://CRAN.R-project.org/package=kableExtra>.