

Universidad de Buenos Aires  
Maestría en Estadística Matemática  
Muestreo

**Trabajo Práctico 1**

Fecha de Entrega: 13/11/2018

La presentación de TP deberá incluir las bases de datos que se generen y los códigos utilizados.

1. Se desea realizar una encuesta económica dirigida a las empresas cuyas ramas de actividad principal corresponden al sector industrial. El objetivo es contar con estimaciones de parámetros relacionados a variables de la actividad del sector tales como consumos energéticos, sueldos y cargas sociales abonadas, montos facturados, producción, gastos diversos, valor de materia prima, entre otras.

Como marco muestral, se cuenta con un listado de las 3984 empresas que forman parte de la población objetivo con la información del personal ocupado (PO) promedio de los últimos 12 meses (*Marco.PO.RData*). Se asume que esta variable está correlacionada con las variables que son consideradas de interés para la encuesta.

Si bien no se cuenta con requerimientos específicos, se considera que el coeficiente de variación (CV) del estimador de la media de las variables principales no debería ser superior al 5 %, debiéndose también estudiar la posibilidad de obtener tamaños de muestra que aseguren el 3 % y el 1 % para el mismo estimador. Por otra parte y por experiencias anteriores, se considera que existe una tasa de respuesta del 85 % para encuestas similares, siendo ese porcentaje del 100 % en estratos autorepresentados en encuestas donde se utilizó un diseño estratificado con estratos de ese tipo.

A continuación se pide calcular los tamaños de muestras necesarios para diversos diseños utilizando distintos paquetes de R. Se considera importante brindar conclusiones en cada uno de los puntos.

Se incluye como parte del material para realizar el TP la siguiente bibliografía, además de considerar útil la lectura de las ayudas de los

paquetes *stratification* y *PracTools*.

Baillargeon, S., Rivest, L.P.(2011). The construction of stratified designs in R with the package stratification. *Survey Methodology*, 37, 53-65.

Valliant, R., Dever, J., Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.

- a) Estudiar la asimetría de la variable PO a través de gráficos y medidas.
- b) Calcular el tamaño de muestra necesario para obtener los CV fijados y la respuesta considerada considerando un muestreo simple al azar.
- c) Es usual en este tipo de encuestas que se requieran estimaciones para distintos dominios tales como ramas específicas de actividad o regiones. Especificar qué fórmula de tamaño de muestra se debe utilizar y calcular el tamaño de muestra necesario para los requerimientos planteados, considerando 3 valores posibles de CV de la variable ( $CV_y$ ) para un dominio particular y 3 proporciones distintas de unidades en el dominio.
- d) A partir de este punto se desea considerar diseños estratificados. En primer lugar se utiliza una estratificación que surge de la definición de tamaño de empresa que considera pequeña empresa a todas aquellas que posean menos de 10 empleados, medianas a las que poseen entre 10 y 34 empleados, y grandes a las que tienen entre 35 y más empleados. Calcular el tamaño de muestra para la estimación de la media de la variable PO asumiendo adjudicación proporcional y de Neyman para los CV y la tasa de respuesta consideradas.
- e) En este punto se desea utilizar algunos de los métodos de estratificación apropiados para variables asimétricas. Para ello, asumiendo distintos escenarios que surgen de la combinación de los 3 CV's deseados, la tasa de respuesta considerada, la posibilidad de contar con 3, 4 y 5 estratos, la utilización de asignación proporcional y de Neyman, y los métodos de estratificación geométrico y de Kozak (considerar un estrato de unidades autorepresentadas), comparar

los distintos tamaños de muestra que se obtienen en cada escenario.

- f) Se logró luego de mucho esfuerzo, realizar una prueba piloto (*Prueba.Piloto.RData*) sobre 50 empresas a las cuales se les consultó sobre el personal ocupado promedio de los últimos 12 meses (PO) y el consumo energético promedio en el mismo período (CE). Esta última variable es considerada una de las más importantes entre las de interés para la encuesta. Se sospecha que existe una regresión lineal heteroscedástica entre las variables del tipo  $CE_k = \beta PO_k + \epsilon_k$  donde  $\epsilon_k \sim N(0, \sigma^2 PO_k^\gamma)$ . Verificar que la misma sea cierta a través de gráficos y de algún test que corrobore la sospecha (Breusch y Pagan, por ejemplo).
- g) Considerar esta discrepancia entre la variable de interés y la variable de estratificación y calcular nuevamente los tamaños calculados en el punto e, comprobando de qué forma influye la misma en los tamaños de muestra finales. Para estimar el parámetro *gamma* se puede considerar el punto 3.2 del libro *Practical Tools...*
- h) Por razones presupuestarias, el tamaño de muestra no debe superar las 200 empresas. Escoger la opción más conveniente en términos de precisión en base a los puntos anteriores, incorporar los estratos a la base *Marco.PO.RData*, y seleccionar la muestra de acuerdo al diseño elegido. Presentar un objeto con el marco actualizado y otro con la muestra seleccionada.
- i) Considerando el punto d), se refiere que uno de los indicadores importantes es la diferencia que existe entre la media del estrato de las empresas pequeñas y la media general. Más en general, se está interesado en obtener un estimador de la diferencia  $d_i = \bar{y}_{U_i} - \bar{y}_U$ 
  - i. Construir el estimador HT,  $\hat{d}_{i\pi}$  de  $d_i$  para cualquier adjudicación
  - ii. Desarrollar la variancia de  $\hat{d}_{i\pi}$
  - iii. Cuál es la adjudicación de Neyman para minimizar la variancia de  $\hat{d}_{i\pi}$  para un tamaño de muestra fijo?
  - iv. Cuando existirán diferencias entre esta adjudicación y la de Neyman considerando la variable de estratificación?

- 2. Práctica 2 Ejercicio 7
- 3. Práctica 2 Ejercicio 13
- 4. Práctica 3 Ejercicio 3

## 5. Práctica 3 Ejercicio 4