



Universidad de Jaén

Facultad de Ciencias Sociales
y Jurídicas

Trabajo Fin de Grado

APLICACIÓN DE TÉCNICAS DE MUESTREO EN INVESTIGACIÓN DE MERCADOS

Alumno: María del Pilar Navarrete Reche

Julio, 2020

RESUMEN

Se ha creado una guía de muestreo para facilitar el uso de métodos básicos de muestreo. Se ha realizado una serie de pasos importantes para llevar a cabo un correcto muestreo, dónde toma relativa importancia el tamaño de la muestra y la selección de ésta.

El siguiente paso ha sido seleccionar entre los diferentes tipos de muestreo, un muestreo aleatorio simple y un muestreo estratificado. Se han definido y explicado los objetivos y características de dichos muestreos.

Tras explicar ambos métodos se ha llevado a cabo un análisis de datos, ponderando y estimando los errores que podemos encontrar a la hora de realizar un correcto muestreo.

Finalmente se han propuesto una serie de ejemplos que nos permitirá entender mejor cada tipo de muestreo anteriormente citado. La implementación de todas las funciones realizadas se han llevado a cabo con el programa estadístico R.

ABSTRACT

A sampling guide has been created to facilitate the use of basic sampling methods. A number of important steps have been taken to perform proper sampling, where the sample size and sample selection are relatively important.

The next step has been to select between the different types of sampling, simple random sampling, and tiered sampling. The objectives and characteristics of these samplings have been defined and explained.

After explaining both methods, a data analysis has been carried out, weighing and estimating the errors that we can find when performing a correct sampling.

Finally, a number of examples have been proposed that will allow us to better understand each type of sampling mentioned above. The implementation of all the functions performed have been carried out with the statistical program R.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN
2. ¿QUÉ ES EL MUESTREO?
 - 2.1. VENTAJAS E INCONVENIENTES DEL MUESTREO
3. PLAN DE MUESTREO
4. TAMAÑO DE LA MUESTRA
5. SELECCIÓN DE LA MUESTRA
 - 5.1. MÉTODOS DE MUESTREO
 - 5.1.1. MUESTREO ALEATORIO SIMPLE
 - 5.1.2. MUESTREO ESTRATIFICADO
6. ANÁLISIS DE DATOS
 - 6.1. PONDERACIÓN DE DATOS
 - 6.2. ESTIMACIÓN DE ERRORES
7. EJEMPLOS
8. CONCLUSIONES
9. BIBLIOGRAFÍA

1. INTRODUCCIÓN

El propósito del muestreo es reducir los costes que conlleva la recolección de los datos de una población. Hoy en día, las encuestas son la manera más viable para recopilar los datos necesarios o para manifestar la opinión sobre algún tema determinado.

El objetivo de esta guía será conocer que es el muestreo estadístico, como seleccionar y analizar una muestra de una población para que, cuando se apliquen diferentes técnicas de muestreo, proporcione conclusiones eficaces.

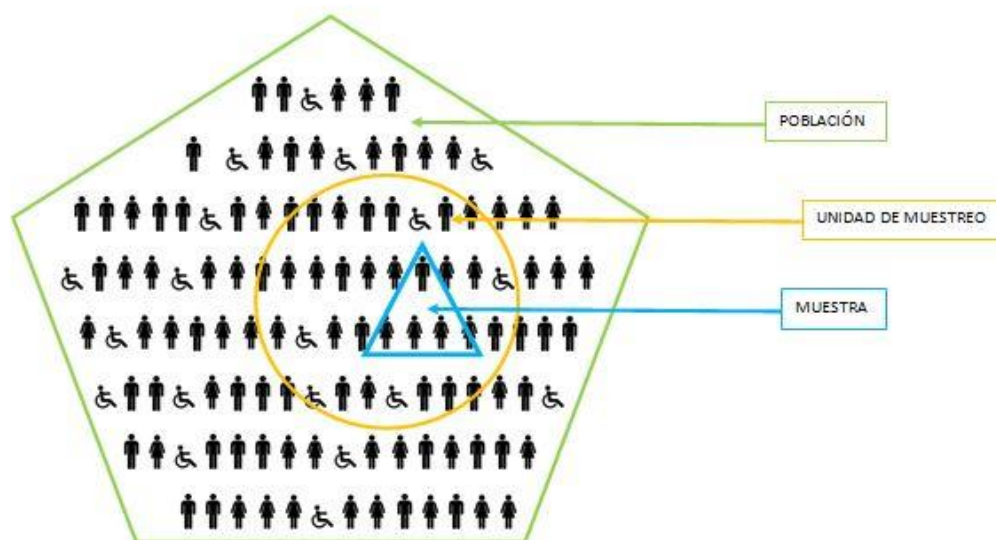
Esta guía está diseñada para aquellas personas con limitados conocimientos de muestreo, aunque, el entendimiento de las técnicas de muestreo más básicas será bastante útil para desarrollar concisión sobre las estimaciones. A lo largo de esta guía se muestran las herramientas necesarias para un buen desarrollo del muestreo estadístico.

“Cuando la estadística no se basa en cálculos precisos, confunde en vez de guiarnos. La mente se deja llevar fácilmente por la falsa apariencia de exactitud que la estadística mantiene en sus errores, y confiadamente adopta errores escondidos bajo la forma de una verdad matemática” — (Alexis de Tocqueville, Democracy in América)

2. ¿QUÉ ES EL MUESTREO?

El muestreo es una técnica de la investigación científica cuyo propósito es seleccionar una población y determinar que parte de esta analizar. Se debe obtener una apropiada representación de la población, en la que se represente de una manera eficiente los atributos esenciales de dicha población que son más importante para la investigación.

MUESTRA



Fuente: Elaboración propia

El objetivo principal del muestreo aleatorio es que cada uno de los individuos que son seleccionados como parte de la población, tiene la misma probabilidad de aparecer en la muestra. Sin embargo, si utilizamos un muestreo intencional, los individuos no tienen que tener la misma probabilidad de aparecer puesto que es el propio investigador el que decide, que individuos formarán la muestra.

Para comprender todos los conceptos que se explicarán en esta guía hay que destacar el siguiente. Una muestra es unos conjuntos de personas, datos, elegidos al azar, representativos y seleccionados para ser estudiados y analizados.

El muestreo aleatorio por lo general trabaja con muestras extensas, sujetas en función del error de muestreo e intervalo de confianza que desee aceptar el investigador. Por el contrario, el muestreo intencional aplica muestras reducidas.

Estos procedimientos de muestreo tienen distintos objetivos. El muestreo aleatorio realizado a través de una investigación cuantitativa (generalización de hechos sociales) y el muestreo intencional a través de una investigación cualitativa (aspectos específicos de realidad social).

2.1. VENTAJAS E INCONVENIENTES DEL MUESTREO

Para realizar un correcto muestreo aleatorio debemos conocer previamente las ventajas y los inconvenientes que se presentan.

- El uso de una muestra pequeña para el muestreo aleatorio presenta varias ventajas.
 - Al escoger un menor número de sujetos al de toda la población, se reduce el tiempo y el dinero.
 - El muestreo puede ser más preciso puesto que no se estudia a toda la población.
 - La utilización de la estadística es mucho más fácil con conjuntos más pequeños de datos para así evitar un error al introducir y analizar estos.

Por otro lado, los inconvenientes son los siguientes:

- El investigador establece reducidos recursos económicos y además trabaja con limitado tiempo, lo que hace que no se pueda estudiar a todos los individuos que forman parte del segmento poblacional de sus investigaciones.
- El muestreo requiere el conocimiento de la estadística y el diseño del estudio depende del método de muestreo a utilizar.
- Hay que tener cuidado a la hora de elegir los sujetos. Se puede dar el caso en que el investigador seleccione a una población en el que todos elijan los datos deseados o que dicha población tienda a seleccionarse ellos mismos. Esto son factores determinantes para nuestro estudio y puede influir de una manera notable.
- Existe el riesgo de rechazar buenos lotes y aceptar malos lotes.

El rigor científico de una investigación no está definido por el tipo de diseño que se utilice, sino, por el empleo de un diseño adecuado, que dé respuesta a los objetivos que se persiguen, y la práctica de un diseño muestral severo. La falta de rigor en una muestra es un defecto que se da con frecuencia, tanto en muestreo aleatorio como en muestreo intencional.

3. PLAN DE MUESTREO

El plan de muestreo es un procedimiento diseñado par seleccionar, determinar, organizar las muestras. El desarrollo del muestreo tiene que estar detallado por escrito y debe incluir:

- Dónde se ha realizado la toma de muestras.
- Cuándo se ha realizado (diariamente, semanalmente, mensualmente, ...)
- Cómo se ha realizado la muestra.
- Quién ha realizado la toma de muestras (analista, inspector de sanidad, etc.)

Un plan de muestreo deberá ser administrativa y económicamente factible.

Para estructurar un plan de muestreo, cobra relativa importancia conocer:

- La naturaleza de la muestra, el equipo de utilización, el material, el tipo de muestra, el estado físico, etc.
- El grado de homogeneidad y heterogeneidad de la muestra.
- El número de submuestras necesarias, determinadas estadísticamente por su tamaño, para que la muestra sea representativa y sin un coste desmesurado económico y personal.
- Las precauciones necesarias para la preparación de la muestra, durante las operaciones de muestreo. Necesario para prever a los analíticos de supuestas alteraciones.

Existen distintas etapas para estudiar en el plan de muestreo. Estas son:

- Determinar los objetivos de nuestro muestreo, clasificando por tipo y por clase. El tipo de muestreo, lo clasificamos:

- **Aleatorio:** Tomar una muestra al azar. Se usa cuando no se tiene mucha información para muestrear, se cogen muestras de franjas seleccionadas, utilizando programas generadores de número aleatorios. Este tipo de muestreo es perfecto para un muestreo homogéneo.

- **Sistemático:** Consiste en coger la primera muestra al azar y después, a partir de un intervalo definido por el investigador, seleccionar al resto de la muestra. Cuando se escogen muestras más grandes, los resultados son más exactos que el muestreo aleatorio. - Por

ejemplo: Suponemos que tenemos 5.000 individuos y deseamos obtener una muestra de 100. En primer lugar, dividimos el marco muestra de 100 en fragmentos de 50. Seguidamente, seleccionamos un número aleatorio entre 1 y 50, para extraer la primera muestra al azar de primer fragmento: por ejemplo el 14. A partir de aquí, quedaría definida la muestra, extrayendo el resto de la muestra con intervalo de 50, como sería: 14, 64, 114, 164, ..., 4.964

- **Estratificado:** Consiste en dividir a la población en varios subgrupos a los que se le aplica un muestreo aleatorio. Este tipo de muestreo se suele utilizar para un muestreo heterogéneo. Como en este muestreo se escogen diferentes subgrupos, es necesario coger un número mayor de muestras.

- **Sistemático-estratificado:** Combina el muestreo sistemático con el estratificado representando el muestreo más adecuado.

- Una vez seleccionado el tipo de muestreo, se escoge la clase de muestra, clasificadas como:
 - *Aleatoria:* Seleccionamos a los individuos de tal forma que todos tienen la misma probabilidad de ser seleccionados.
 - *Estratificada:* Seleccionamos un subgrupo de individuos, los cuales se cogen aleatoriamente.
 - *Sistemática:* Si obtenemos la muestra de un procedimiento sistemático, elegimos la primera muestra al azar y las siguientes de forma ordenada hasta concluir la muestra.
 - *Selectiva:* Se tiene en determinadas partes de material, utilizando un procedimiento dirigido de muestreo.
 - *Compuesta:* Formada por dos o más submuestras. Se refiere a una combinación de muestras sencillas cogidas en un mismo lugar pero a distinta hora.
 - *Representativa:* Se trata de una pequeña cantidad de individuos, que manifiesta, con la mayor precisión, a un grupo más grande.
- Detallar el procedimiento operativo, determinando:
 - El tipo de muestra a utilizar.

- Realización de una lista con los parámetros que se desean medir.
- Identificar el lugar para realizar el muestreo.
- Selección del número de muestras.
- Selección del material para realizar el muestreo.
- Identificar la frecuencia y la cantidad de la muestra.
- Seleccionar un método para la toma de la muestra.
- Envío de las muestras.

4. TAMAÑO DE LA MUESTRA

El primer paso para comenzar una encuesta es determinar la finalidad de la medición. El hecho de que elijamos una muestra no aleatoria no inhabilita los resultados de la investigación, sin embargo, no significa que el investigador pueda elegir libremente, es decir, el investigador no puede elegir a los individuos sin ningún tipo de criterio.

“La lógica del muestreo intencional estriba en seleccionar unidades que te permitirán realizar comparaciones significativas con relación a tus preguntas de investigación, tu teoría y el tipo de explicación que desees desarrollar.” — (Jennifer Mason (1997:96))

El tamaño de la muestra es una relación directa entre el nivel de confianza y error de muestreo aceptado. Es proporcional al nivel de muestreo y al nivel de confianza, sin embargo, cuanto mayor es el error de muestreo, menor es el tamaño de la muestra, por el contrario, cuanto mayor es el nivel de confianza, mayor es el tamaño de la muestra.

El tamaño de la muestra en estudios cuantitativos, a través del muestreo probabilístico, puede componerse de entre uno o pocos entrevistados hasta la inserción de un mayor número de hablantes. En estudios cualitativos, a través del muestreo no probabilístico, la determinación del tamaño de la muestra no puede ser determinado a priori, es decir, en este tipo de diseño el que decide tanto los individuos que formarán la muestra como cuántos la formarán, es el investigador.

En la investigación cuantitativa, cada individuo es un elemento semejante y sustituible mientras que en la investigación cualitativa no se corresponde esta equivalencia.

Es importante señalar que en la investigación cualitativa la repetición de los elementos no es lo único que se debe tomar en cuenta en el análisis de los datos, también, en la población de todo el muestreo en general. Por tanto, cuando el investigador aplique una determinada técnica tanto cualitativa como cuantitativa, debe seguir con el trabajo hasta obtener una certeza de haber agotado las diferencias discursivas presentadas por el tema que analiza.

Como conclusión, podemos decir que la muestra estará formada por el menor número de casos posible, que a su vez, proporcione información suficiente para desarrollar todas las dimensiones posibles de nuestra investigación, de acuerdo con nuestros objetivos. Si seleccionamos una muestra reducida el precio de nuestra investigación será más económico, tardaremos menos tiempo pero tendremos un mayor error de muestreo, sin embargo, a medida que aumentamos la muestra, el costo y el tiempo aumenta, pero, el error de muestreo se reduce. Del mismo modo, una muestra de mayor tamaño nos permite conocer más en profundidad nuestro fenómeno social analizado y proporciona una explicación más compleja.

A medida que aumenta la muestra, el costo y el tiempo requerido para nuestro análisis aumenta y viceversa.

5. SELECCIÓN DE LA MUESTRA

“Toda investigación debe ser transparente, así como estar sujeta a crítica y réplica, y este ejercicio solamente es posible si el investigador delimita con claridad la población estudiada y hace explícito el proceso de selección de su muestra.” — (Roberto Hernández-Sampieri)

Como hemos comentado anteriormente, el objetivo del muestreo se basa en estimar parámetros de una población (el total, la media,...) en función de la información recogida de una muestra.

Para seleccionar una muestra adecuada para la investigación, tenemos que tener en cuenta los siguientes pasos:

1º. Definir el número de casos (objetos, procesos, sucesos,...) para seleccionar los datos necesarios.

2º. Identificar la población.

3º. Seleccionar un método para analizar la muestra: probabilístico o no probabilístico.

4º. Determinar el tamaño de la muestra.

5º. Adaptar el procedimiento de selección a la muestra.

6º. Alcanzar o conseguir la muestra.

Siguiendo los pasos anteriores, aún nos podemos preguntar: ¿Cómo es posible determinar el tipo de procedimiento que debo de utilizar y por cuántas observaciones tiene que estar formada mi muestra? La respuesta a esta pregunta dependerá de la cantidad de información deseemos conseguir. Si por ejemplo llamamos θ a un parámetro de interés y $\hat{\theta}$ es el estimador de θ , es necesario conocer un límite para calcular el error de estimación; es decir, debemos determinar que θ y $\hat{\theta}$ en valor absoluto sea menor que una cantidad B.

$$\text{Error de estimación} = |\theta - \hat{\theta}| < B$$

Del mismo modo, es necesario definir una probabilidad, (1- alfa), que nos indique la proporción de veces que se exige que el error de estimación (calculado anteriormente) sea menor que B. Estableciéndose esta condición como:

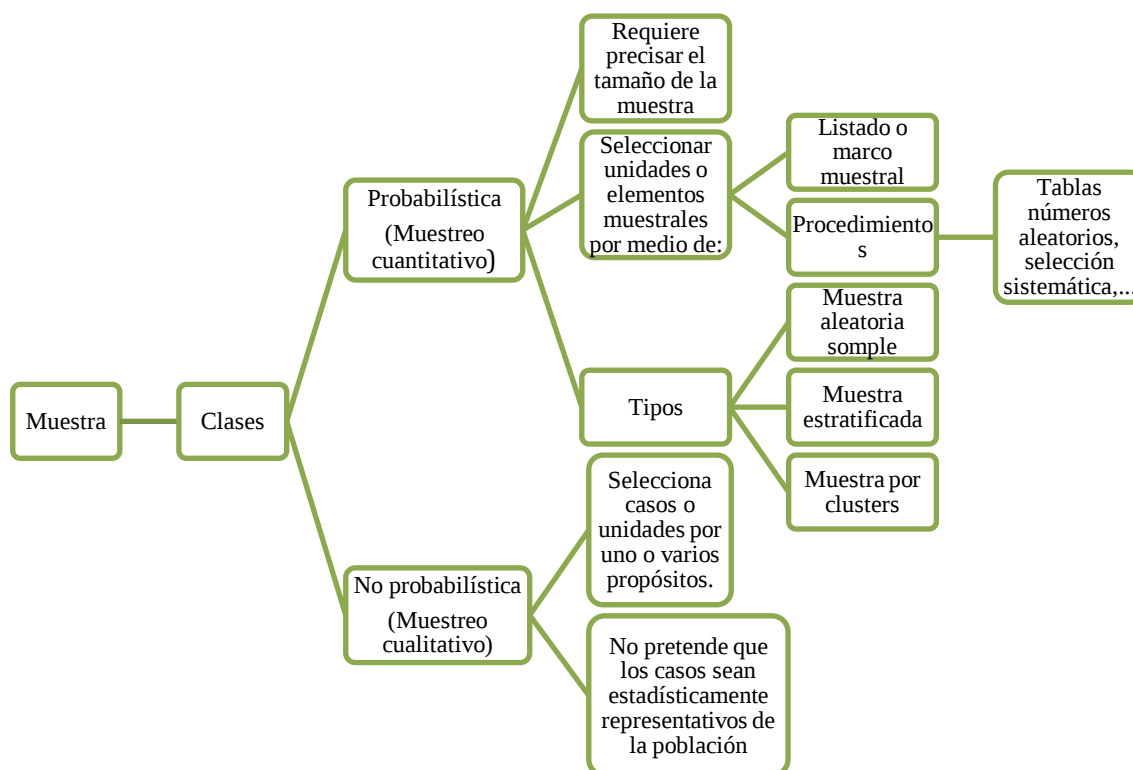
$$P[\text{Error de estimación} < B] = 1 - \alpha$$

Una vez obtenido el límite específico junto a la probabilidad asociada, podemos comparar distintos diseños para determinar el procedimiento que nos proporciona precisión necesaria con coste mínimo.

5.1. MÉTODOS DE MUESTREO

Existen diferentes tipos de muestreo que se utilizan para analizar diferentes tipos de muestras. En esta guía tan sólo nos centraremos en el muestreo aleatorio simple y en el muestreo estratificado.

ESQUEMA MUESTRA



Fuente: elaboración propia

5.1.1. MUESTREO ALEATORIO SIMPLE

Este tipo de muestreo probabilístico consiste en seleccionar a n unidades de muestreo de tal forma que cada muestra extraída de tamaño n , tenga la misma probabilidad de ser seleccionada.

Por ejemplo: Es posible obtener una muestra aleatoria con n votantes en una encuesta, mediante la extracción de nombres de una lista de votantes registrados, de tal forma, que cada muestra de tamaño n , tenga la misma oportunidad de ser seleccionada. Por lo tanto, si seleccionamos una muestra de tamaño n de una población de tamaño N , de forma que cada

muestra de tamaño n tenga la misma probabilidad de ser elegida, el procedimiento de muestreo es denominado muestreo aleatorio simple.

El muestreo aleatorio simple puede realizarse con o sin reposición. Hipotéticamente, la diferencia de un muestreo simple con reposición a otro sin reposición, dice la relación de independencia o dependencia en la probabilidad seleccionada.

Utilizaremos este tipo de muestreo para conseguir estimadores de medias poblacionales, proporcionales y totales. Por ejemplo: Un guardabosques puede estimar la proporción de árboles enfermos y el volumen de madera, recopilando distintos puntos geográficos en el área de un bosque y seguidamente asociando a cada uno de los puntos escogidos, una parcela de tamaño y forma fijos. De esta forma, es posible estudiar todos los distintos árboles dentro de cada parcela seleccionada, por lo que el diseño a utilizar es un diseño aleatorio simple.

Seguidamente se proporcionan una serie de fórmulas necesarias para obtener el mínimo número de sujetos que deben formar nuestra m.a.s. para estimar los parámetros esenciales como son: la media, la proporción y el total de una población.

Denotando los siguientes términos:

n = tamaño de la muestra.

N = tamaño de la población.

$1 - \alpha$ = nivel de confianza.

$z_{1-\alpha}$ = El valor de z (siguiendo una Normal (0,1))

p = proporción.

$q = 1 - p$

e = error de estimación.

Por lo tanto, el tamaño de la muestra para una población infinita es:

$$n = z_{1-\alpha}^2 \frac{pq}{e^2}$$

Y el tamaño de la muestra para una población finita es:

$$n = \frac{z_{1-\alpha}^2 pqN}{e^2(N-1) + z_{1-\alpha}^2 pq}$$

A continuación se muestra un ejemplo realizado con RStudio.

El R base contiene la función denominada [sample](#) dónde podemos conseguir muestras aleatoria simples con reemplazamiento y sin reemplazamiento.

Cargaremos un dataset para cuantificar sus dimensiones.

R source
<pre>crime<-data.frame(crimtab) dim(crime)</pre>
R Output
<pre>[1] 924 3</pre>

Como se puede observar, este dataset está formado por 924 filas y 3 columnas. Seguidamente vamos a seleccionar una muestra de 50 casos sin reemplazamiento.

R source
<pre>n<50 muestraejemplo<- sample(1:nrow(crime),size=n,replace=FALSE) muestraejemplo</pre>
R Output

```
[1] 571 777 353 707 786 823 833 749 321 508 442 477 526 350 331 711 888 130 738 854
[21] 154 173 221 93 524 677 263 259 850 110 692 48 598 733 242 495 249 222 413 297
[41] 530 474 90 478 917 271 20 135 31 561
```

Asignamos los elementos obtenidos de la muestra al data frame de los datos.

R Source

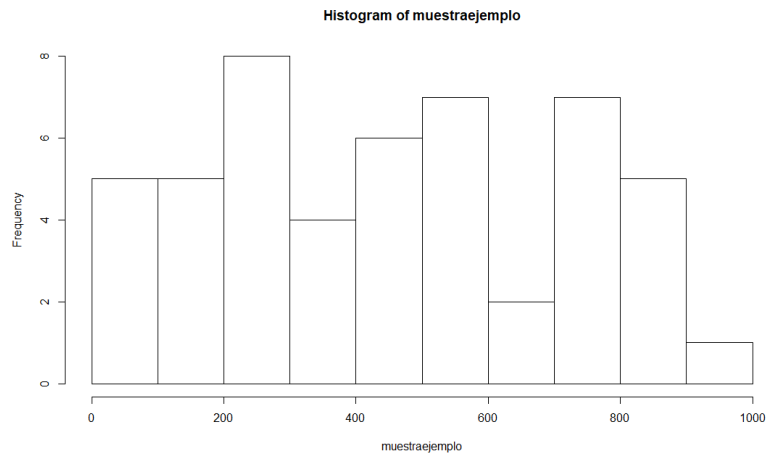
```
crimemuestraejemplo<- crime[muestraejemplo, ]
head(crimemuestraejemplo)
```

R Output

```
  Var1  Var2 Freq
571 11.8 175.26  10
777 11.4 187.96   0
353  11 162.56  37
707 12.8 182.88   3
786 12.3 187.96   0
823 11.8 190.5   0
```

-A continuación se muestra un histograma que nos representa nuestra variable en forma de barras, dónde la superficie (eje vertical) de la barra es representada por la frecuencia de los valores, mientras que el eje horizontal representa los valores de la variable por intervalos.

HISTOGRAMA EJEMPLO



Fuente: Elaboración propia

5.1.2. MUESTREO ESTRATIFICADO

El muestreo estratificado consiste en utilizar información permitiendo agrupar a los elementos de la muestra en diferentes estratos. Un estrato está formado por elementos parecidos y cada estratos es diferente entre sí.

Este tipo de muestreo trata de agrupar los elementos de la muestra en estratos homogéneos para mejorar la precisión y minimizar el costo mediante respecto al muestreo aleatorio simple. El efecto que tiene un diseño estratificado homogéneo es menor que 1, lo que nos muestra que nuestro diseño es más preciso.

Si seleccionamos un número reducido de estratos es probable que no podamos aprovechar de la estratificación y por el contrario, si seleccionamos un número grande de estratos, se puede generar una agrupación pequeña y complicada en los cálculos.

MUESTREO ESTRATIFICADO



Fuente: Google imágenes

El objetivo de que tiene este tipo de muestreo es maximizar la información a un coste dado.

Llamamos afijación de la muestra a la distribución del tamaño muestral entre el número de estratos. Es decir, a la determinación de n_h , $h = 1, 2, \dots, L$ que verifique lo siguiente:

$$n_1 + n_2 + \dots + n_L = n$$

Existen varios tipos de afijación para extraer la muestra de los estratos, siendo estas: la afijación uniforme, la afijación proporcional, la afijación de varianza mínima y la afijación óptima.

En este caso veremos un ejemplo en el apartado 7 de esta guía, utilizando la afijación de mínima varianza o afijación de Neyman para determinar el valor de n (número de valores que son extraídos de cada estrato). Hemos de comentar que las fórmulas son distintas si utilizamos la media o la proporción.

Las fórmulas que utilizaremos para calcular la afijación de Neyman son:

Para la media $\longrightarrow n = \frac{(\sum_{h=1}^L W_h S_h)^2}{e^2 + \frac{1}{N} \sum_{h=1}^L W_h S_h}$ con $n_h = n \frac{N_h \sigma_h}{\sum_h N_h \sigma_h}$

Para la proporción $\longrightarrow n = \frac{(\sum W_h \sqrt{p_h(1-p_h)})^2}{e^2 + \frac{1}{N} \sum W_h p_h(1-p_h)}$ con $n_h = n \frac{N_h \sqrt{p_h q_h}}{\sum N_h \sqrt{p_h q_h}}$

6. ANÁLISIS DE DATOS

El análisis de datos consiste en recolectar datos para la realización de operaciones con el objetivo de obtener conclusiones precisas para la toma de decisiones. Un número elevado de técnicas de análisis de datos utilizadas en una encuesta, aconseja realizar algunas de estas técnicas con la finalidad de explicar alguna de ellas con profundidad. Uno de los criterios más utilizados para diferenciar técnicas de análisis de datos, hace referencia al tipo de métrica de las variables a utilizar diferenciándose en variables cualitativas y cuantitativas.

6.1. PONDERACIÓN DE DATOS

La ponderación de datos es una técnica estadística utilizada para perfeccionar los desequilibrios que se presentan en una muestra una vez recogidos los datos, ya que muchos de estos datos no son representativos para la población.

Para explicar el uso de ponderaciones realizamos un ejemplo a continuación con información ficticia. La tabla que se muestra a continuación nos muestra información muestral y poblacional referida a dos poblaciones de distinto tamaño. En esta tabla podemos ver que el tamaño de la muestra en las distintas poblaciones no es proporcional al tamaño de estas en la población. En consecuencia, se aprecia que ambas muestras son desproporcionadas respecto al tamaño de cada población.

Poblaciones	N	n	% población	% muestra
Utrera	50.000	500	10 %	50%
Huelva	450.000	500	90%	50%

La inferencia de Utrera respecto a Huelva será más precisa en consecuencia de un mayor tamaño relativo de la muestra. Posteriormente, se desarrollan distintos procedimientos para ponderar. El coeficiente de ponderación se obtiene como el cociente del porcentaje de la población respecto al porcentaje de muestra. Los porcentajes para los datos anteriores son:

Poblaciones	Coeficiente de ponderación
Utrera	10/50 = 0.2
Huelva	90/50 = 1.8

Para la práctica, la ponderación de una muestra es recomendable realizarla computacionalmente. Si utilizamos el programa de SPSS la ponderación se llevaría a cabo con la instrucción Weight. La síntesis a llevar a cabo es la siguiente:

```
COMPUTE PONDERA = 1
IF (POBLACION EQ1) PONDERA = 0.2
IF (POBLACION EQ2) PONDERA = 1.8
WEIGHT BY PONDERA
```

Con esta síntesis conseguimos otorgar a cada una de las poblaciones de la muestra el peso correspondiente según el tamaño relativo considerado anteriormente. Como consecuencia, el cálculo de la estimación del parámetro se realiza a partir del estadístico obtenido a partir de los pesos diferenciales de cada una de las poblaciones.

Podemos realizar gráficos ponderados utilizando las funciones hist y boxplot.

6.2. ESTIMACIÓN DE ERRORES

Podemos entender un error de estimación en términos relativos. El error relativo es adimensional (su dimensión es siempre igual a 1) y se puede expresar en forma de porcentaje. Este error es fácil de interpretar y nos permite comparar distintos errores en muestras cuyas variables no tengan el mismo tipo de medida. De la misma forma, leer el error en términos de porcentaje nos permite definir estos errores de muestreo superiores al 10%. La fórmula utilizada para el cálculo del error relativo, anula la unidad de medida que representa al error máximo que es admitido ponderado por la media y por la proporción. A continuación representamos el error relativo para la media y para la proporción.

$$\text{Error relativo media} = \frac{\text{Error absoluto}}{\text{Media}}$$

$$\text{Error relativo proporción} = \frac{\text{Error absoluto}}{\text{Proporción}}$$

Donde el error absoluto es el máximo aceptable en la estimación de la media y de la proporción.

7. EJEMPLOS

Ejemplo 1. Muestreo aleatorio simple

De 5765 ahorristas registrados en una entidad financiera, se escogen aleatoriamente a 200 y se les consulta si vinieron a la entidad por consejo de alguien. De los 200 encuestados, 47 contestaron que si vinieron por consejo de alguien. Se pide:

a) Hacer una estimación de la proporción de las 3765 personas que vinieron por consejo de alguien.

- En primer lugar introducimos los datos proporcionados.

R Source
N=5765 n=200 a=47 NS=0.05;NC=1-NS k=qnorm(1-NS/2);round(k,3)
R Output
[1] 1.96

-Seguidamente calculamos la proporción de personas que vinieron por consejo de alguien.

R Source
$p=a/n;p$
R Output
[1] 0.235

b) Determinar el error de estimación asumiendo un nivel de confianza adecuado.

R Source
$vp=((p*q)/n)*((N-n)/N);vp$
R Output
[1] 0.0008676911

-Calculamos ahora el error de estimación

R Source
$B=k*\text{sqrt}(vp);B$
R Output
[1] 0.05773387

c) ¿Qué opinión tiene sobre el valor obtenido para dicho error?

- Calculamos el error relativo

R Source
$ER = (B/p); ER$
R Output
[1] 0.245676

-Por último calcularemos la opinión que tiene el valor obtenido para dicho error.

R Source
<code>if ((B/p)>0.10) ("La estimacion no es confiable, se sugiere aumentar n") else ("La estimacion es confiable")</code>
R Output
[1] "La estimacion no es confiable, se sugiere aumentar n"

Ejemplo 2. Muestreo estratificado

Dado los siguientes estratos:

Estrato 1	
X	ni
2	20
3	30
4	60
7	100

Estrato 2	
X	ni
10	150
12	200
16	120
20	80
25	50

Estrato 3	
X	ni
30	20
35	18
50	10
60	8
100	4

a) Calcular el tamaño de las muestras usando la Afijación de Neyman.

- En primer lugar introducimos los datos en una tabla Excel y vamos metiendo los datos por estratos (primero el estrato 1, lo guardamos como texto (MS-DOS), después el 2 y por último el 3)

- En segundo lugar abrimos el programa R e introducimos los datos recogidos en excel.

R Source

```
#Lectura de la base de datos  
  
#read.table("nombre.tipo",T)  
  
#definimos la carpeta con la que vamos a trabajar  
setwd("C:/Users/Usuario/Desktop/TFG/ejemplo2")
```

```
est1=read.table("est1.txt",T)
```

```
est1
```

```
est2=read.table("est2.txt",T)
```

```
est2
```

```
est3=read.table("est3.txt",T)
```

```
est3
```

R Output

```
> setwd("C:/Users/Usuario/Desktop/TFG/ejemplo2")
```

```
> est1=read.table("est1.txt",T)
```

```
> est1
```

```
  x  ni
```

```
1 2 20
```

```
2 3 30
```

```
3 4 60
```

```
4 7 100
```

```
> est2=read.table("est2.txt",T)
```

```
> est2
```

```
  x  ni
```

```
1 10 150
```

```
2 12 200
```

```
3 16 120
```

```
4 20  80
```

```
5 25  50
```

```
> est3=read.table("est3.txt",T)
```

```
> est3
```

```
  x  ni
```

```
1 30 20
```

2	35	18
3	50	10
4	60	8
5	100	4

-Como conocemos las fórmulas para la media y la desviación estándar de la afijación de Neyman, definimos la función de la siguiente manera.

R Source

```
#función para la media y la desviación estándar con la afijación de Neyman

meanvar=function(x,f){

n=sum(f)

prom=sum(x*f)/n

dev=sqrt(sum((x-prom)^2*f)/(n-1))

r=data.frame(prom=prom,sd=dev)

r=data.frame(r,Nh=n)

return(r)}
```

-A continuación, calculamos el promedio y la desviación estándar para cada uno de los estratos.

R Source

```
d1=meanvar(est1$x,est1$ni)

d1
```

```

d2=meanvar(est2$x,est2$ni)

d2

d3=meanvar(est3$x,est3$ni)

d3

#Generamos otro data frame en forma de tabla y la modificamos para que nos aparezca el
nombre de cada estrato

#Utilizamos rbind para introducir filas y cbind para introducir columnas

tabla=rbind(d1,d2)

tabla=rbind(tabla,d3)

tabla=cbind(Grup=c("Est1","Est2","Est3"),tabla)

tabla

```

R Output

```

> tabla=cbind(Grup=c("Est1","Est2","Est3"),tabla)

> tabla

  Grup Grup  prome    sd    Nh
1 Est1 Est1 5.095238 1.904534  210
2 Est2 Est2 14.450000 4.566061  600
3 Est3 Est3 43.500000 18.555916   60

```

R Source

```
tabla$NhSh=tabla$Nh*tabla$sd
```

```
tabla$Wh=tabla$NhSh/sum(tabla$NhSh)
```

```
tabla
```

R Output

	Grup	prome	sd	Nh	NhSh	Wh
1	Est1	5.095238	1.904534	210	399.9522	0.09404125
2	Est2	14.450000	4.566061	600	2739.6367	0.64417420
3	Est3	43.500000	18.555916	60	1113.3550	0.26178455

-Por último calculamos el tamaño de la muestra.

R Source

```
#Cálculo del tamaño de la muestra
```

```
#utilizamos un margen de error de 0.03
```

```
e=0.03
```

```
n=sum((tabla$Wh*tabla$sd)^2)/(e^2+sum(tabla$Wh*tabla$sd^2)/sum(tabla$Nh))
```

```
(e^2+sum(tabla$Wh*tabla$sd^2)/sum(tabla$Nh))
```

```
#redondeamos a la parte entera, sin decimales
```

```
n=round(n,0)
```

```
n
```

R Output

```
[1] 268
```

-La distribución de este tamaño de muestra (268) será:

R Source
<pre>tabla\$nh=round(n*tabla\$Wh,0) tabla sum(tabla\$nh)</pre>
R Output
<pre>> tabla\$nh=round(n*tabla\$Wh,0) > tabla Grup prome sd Nh NhSh Wh nxest nh 1 Est1 5.095238 1.904534 210 399.9522 0.09404125 25.20305 25 2 Est2 14.450000 4.566061 600 2739.6367 0.64417420 172.63869 173 3 Est3 43.500000 18.555916 60 1113.3550 0.26178455 70.15826 70 > sum(tabla\$nh) [1] 268</pre>

-Podemos comprobar que el tamaño de muestra habiendo realizado la distribución coincide con el tamaño de muestra antes de realizar la distribución.

8. CONCLUSIONES

Como resultado de este TFG podemos extraer esta serie de conclusiones.

En primer lugar, obtener una visión general sobre el muestreo y el estudio de distintos tipos de muestreo. En relación a este apartado, se ha obtenido diferentes definiciones que hay que conocer antes de realizar cualquier tipo de muestreo.

En segundo lugar, conocer el tamaño de la muestra necesaria para realizar un muestreo correcto. En este caso, es importante destacar conforme aumente el tamaño de la muestra, menor será el error de muestreo y viceversa.

En tercer lugar, seleccionar el tipo de muestra. Hemos elaborado una serie de pasos previos a la realización de ésta con la finalidad de ahorrar tiempo y dinero.

En cuarto lugar se ha seleccionado un tipo de muestreo aleatorio simple y estratificado. Para estos apartados hemos realizado varios ejemplos en el programa estadístico R. Hemos de comentar que la realización de un muestreo aleatorio simple es más sencilla puesto que los cálculos son más fáciles. De la misma forma, la realización de este tipo de muestreo puede llegar a ser menos costosa que en el muestreo estratificado.

En quinto lugar, analizar los datos con ponderaciones y estimación de errores.

Por último, hemos realizado varios ejemplos a modo resumen de un muestreo aleatorio simple y de un muestreo estratificado.

9. BIBLIOGRAFÍA

INTRODUCCIÓN AL MUESTREO. Izcarra Palacios, Simón Pedro(2007), México D.F. : Editorial Miguel Ángel Porrúa.

MUESTREO Y PREPARACIÓN DE LA MUESTRA. Rodríguez Alonso, Juan José (2013), Barcelona : Cano Pina, Ediciones Ceysa.

MÉTODOS DE MUESTREO: CASOS PRÁCTICOS. Rodríguez Osuna, Jacinto (1993), Madrid: CIS.

MUESTREO ESTADÍSTICO. Alba Fernández, María Virtudes. ; e-libro, Corp.; Ruiz Fuentes, Nuria, aut. (2004), Oviedo : Septem Ediciones.

MUESTREO: DISEÑO Y ANÁLISIS. Lohr, Sharon L. (cop 2000), México ; Madrid: International Thompson.

METODOLOGÍA DE LA INVESTIGACIÓN (6ª ed., pp. 170-191). Hernández-Sampieri, R., Fernández-Collado, C. y Baptista-Lucio, P. (2014). Selección de la muestra. México: McGraw-Hill.

MUESTREO ESTADÍSTICO. DISEÑO Y APLICACIONES. Manuel Vivanco(2005), Editorial Universitaria.

ELEMENTOS DE MUESTREO. Richard L.Scheaffer, William Mendenhall, Lyman Ott (2006). Thomson Editores.

METODOLOGÍA DE LA INVESTIGACIÓN SOCIAL CUANTITATIVA. Disponible en: https://ddd.uab.cat/pub/caplli/2017/185163/metinvsoccua_cap2-4a2017.pdf

HOJA DE REFERENCIA PARA R MARKDOWN. Disponible en: <https://rstudio.com/wp-content/uploads/2015/03/rmarkdown-spanish.pdf>

SINTAXIS R MARKDOWN. Disponible en: <https://markdown.es/sintaxis-markdown/>

EL PAQUETE ESTADÍSTICO R. Jesús Bouso Freijo (2013), Cuadernos metodológicos.