VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

GRADO EN ESTADÍSTICA
TRABAJO FIN DE GRADO

LA VIOLENCIA CIBERNÉTICA: UNA COMPARACIÓN DE DIFERENTES ESTRATEGIAS DE MUESTREO

Alumno/a: Yinan Qin
Tutor/a: María Teresa Cabero Morán

SALAMANCA, 2023

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

LA VIOLENCIA CIBERNÉTICA: UNA COMPARACIÓN DE DIFERENTES ESTRATEGIAS DE MUESTREO

Alumna: Yinan Qin

Tutora: María Teresa Cabero Morán

SALAMANCA, 2023

FIRMA DE LA AUTORA

Yinan Qin

FIRMA DE LA TUTORA

Certificado del/los tutor/es TFG

D./Dña. _María Teresa Morán_____profesor/a____del Departamento de__Estadística_____ de la Universidad de Salamanca,

HACE/N CONSTAR:

Que el trabajo titulado "La violencia cibernética: una comparación de diferentes estrategias de muestreo_____", que se presenta, ha sido realizado por _____Yinan Qin_____, con DNI ****54569 y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Grado en ___Estadística_____ en esta Universidad.

Salamanca, _28_ de _06____ de 20_23___

Fdo.: _____

Índice general

1. Introducción **1**
- 1.1. Estado de la cuestión . . . 1
- 1.2. Estructura del trabajo . . . 1
- 1.3. Objetivos . . . 2
- 1.4. Resultados fundamentales . . . 2

2. Muestreo Estadístico **3**
- 2.1. Muestreo Aleatorio Simple . . . 3
  - 2.1.1. Muestreo Aleatorio Simple con Reposición . . . 5
  - 2.1.2. Muestreo Aleatorio Simple sin Reposición . . . 6
  - 2.1.3. Errores de Muestreo . . . 8
  - 2.1.4. Cálculo del tamaño de la muestra . . . 10
- 2.2. Muestreo Estratificado . . . 11
  - 2.2.1. Estimaciones del Total Poblacional τ . . . 12
  - 2.2.2. Estimaciones de Media Poblacional μ . . . 13
  - 2.2.3. Estimaciones de la Proporción p . . . 13
  - 2.2.4. Cáculo del tamaño de muestra . . . 14
- 2.3. Muestreo por Conglomerados . . . 16
  - 2.3.1. Muestreo por conglomerados en una etapa . . . 17
  - 2.3.2. Muestreo por conglomerados en dos etapas . . . 18
- 2.4. Muestreo por cuotas . . . 22
  - 2.4.1. Pasos a realizar en el muestreo por cuotas . . . 22
  - 2.4.2. Ventajas y Desventajas . . . 23

3. Violencia Cibernética **25**
- 3.1. ¿Qué es la violencia cibernética? . . . 25
- 3.2. Casos reales de violencia cibernética . . . 26

4. Encuestas **27**
- 4.1. Datos Básicos . . . 27
- 4.2. Limpieza de datos de la encuesta . . . 28

5. Metodología **29**

6. Resultado **33**
- 6.1. Muestreo Aleatorio Simple . . . 33
- 6.2. Muestreo Estratificado . . . 34
- 6.3. Muestreo por Conglomerados en una etapa . . . 38
- 6.4. Muestreo por Conglomerados en dos etapas . . . 39
- 6.5. Muestreo por Cuotas . . . 42

7. Conclusiones **44**

1 | Introducción

1.1 Estado de la cuestión

Se ha decidido realizar la encuesta sobre el tema de violencia cibernética por darme en cuenta de que está pasando cada vez más grave la violencia digital. Según los datos publicados en Módulo sobre Ciberacoso 2020 en Instituto Nacional de Estadística y Geografía (INEGI)(10) $21\,\%$ de la población usuaria de 12 años o más de Internet fue víctima de ciberacoso entre octubre de 2019 y noviembre de 2020. Conforme con la publicación de la organización "Bullying Sin Fronteras"(9): El bullying y el ciberbullying son asesinos silenciosos que cada año matan 200,000 niños y jóvenes en todo el mundo. Por tanto considero que habrá que dar importancia a este tema para que a la gente le llama la atención.

Para la colección de datos se realizó una encuesta mediante Google formularios que fue enviado a través de las redes sociales incluyendo Whatsapp, Instagram y Wechat (red social china) a los participantes. Antes de pasarla a los participantes, se explicó en qué consistía la encuesta, para qué se utilizaba y el consentimiento informado de acuerdo con la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. La encuesta se hizo en dos idiomas: español y chino considerando en algunos de los participantes que no sabían alguno de los dos. La encuesta fue abierta desde 25 de julio de 2021 hasta 29 de septiembre de 2021 con participantes asiáticos, europeos, norteamericanos y latinoamericanos basados en dos preguntas dicotómicas sobre violencia cibernética: “¿Ha sufrido alguna vez la violencia cibernética?” y “¿Ha cometido alguna vez violencia cibernética contra otras personas?”, en las que se incluyía solo la violencia cibernética en el sector social y variables criterios que se formaban por sexo, edad, nacionalidad y estado civil.

1.2 Estructura del trabajo

El trabajo consiste en seis partes: introducción, parte teórica, encuesta, metodología, resultado y conclusión.

En la introducción incluye el estado de la cuestión, estructura del trabajo, objetivos y resultados fundamentales en los que se explica brevemente cómo se ha planteado las preguntas de la encuesta, cómo se ha conseguido los datos necesarios, cuál es el objetivo principal del dicho estudio y los resultados obtenidos posteriormente. A la siguiente parte le llamamos parte teórica en la que contiene dos partes: las técnicas de muestreo estadístico y la violencia cibernética. En cuanto a las técnicas, se introducen las definiciones y demostraciones de las técnicas de muestreo estadístico incluyendo muestreo aleatorio simple, muestreo estratificado, muestreo por conglomerados y muestreo por cuotas para que se tenga en cuenta de cómo son estas técnicas, cómo se emplean adecuadamente y las estimaciones necesarias en cada una de ellas. A continuación coincide con el título de este trabajo, explica qué es violencia cibernética y los casos reales que pasaron en unos paises.

Después, se escribe, primero se presentan los datos básicos de la encuesta y la limpieza de los datos, con el fin de que sean adecuados los datos para llevarse a cabo en el estudio, ya

que después de recoger los datos, se han detectado con defectos de colocación y faltas de ortografía como mayúsculas y minúsculas.

Segundo, en el apartado de metodología se basa en nueve puntos: diseño del estudio, sujetos de estudio, estimación del tamaño de la muestra, variables, recogida de datos, fases del estudio, análisis de datos, funciones utilizadas en EXCEL y limitaciones del estudio. Generalmente contienen los pasos de plantear el estudio, las características de variables y los softwares utilizados para realizar las estimaciones.

Tercero, en cuanto a resultados se realizan ejercicios prácticos de cada técnica de muestreo y se estiman los tamaños muestrales, parámetros, varianzas de proporción, errores y sus intervalos dependiendo de las características de las técnicas se utilizan diferentes formas de estimación. Finalmente se realizan comparaciones entre los diferentes métodos de muestreo y se obtendrá una conclusión general.

1.3 Objetivos

Se quiere hacer una comparación entre los diferentes métodos de muestreo estadístico mediante los resultados obtenidos a partir de los ejercicios propuestos, para conocer mejor los métodos, y en este estudio, cuáles de ellos son más adecuados para aplicar.

1.4 Resultados fundamentales

Los tamaños muestrales estimados, los errores y los intervalos de proporción serán los siguientes:

Método de Muestreo	n	e	I_p
Aleatorio Simple	202	0.0426	(19,5 % 28,02 %)
Estratificado	203	0.0455	(27,47 % 36,57 %)
Por Conglomerados en una etapa	7	0.0399	(21,64 % 29,62 %)
Por Conglomerados en dos etapa	10	0.05	(34,03 % 44,03 %)
Por Cuotas	203	0.0459	(38,88 % 48,06 %)

Cuadro 1.1: Resultados de cada método

⋆ Los métodos que tienen menor error serán muestreo aleatorio simple y muestreo por conglomerados en una etapa.

⋆ Aunque el muestreo aleatorio simple no tiene error mínimo, se entiende con facilidad y con resultados proyectables. Pero no se puede decir que el resultado tiene mucha precisión.

⋆ El muestreo estratificado y el muestreo por cuotas son métodos muy parecidos pero el primer método tiene más precisión.

⋆ En este estudio no es bueno aplicarse el muestreo por conglomerados en dos etapas por ser muy costoso y tener menos precisión.

⋆ Se ha realizado este estudio mediante los datos obtenidos de la encuesta, por lo tanto cuando se han aplicado los métodos que necesitaban estratos o conglomerados, se han limitado mucho.

⋆ Tal y como se ha podido deducir que según los resultados de muestreo por conglomerados en una etapa entre 406 individuos encuestados hay de 21.64 % a 29.62 % de ellos que han sufrido alguna vez violencia cibernética.

2 | Muestreo Estadístico

2.1 Muestreo Aleatorio Simple

Muestreo aleatorio simple (M.A.S) es un muestreo probabilístico, una de las técnicas más populares en muestreo estadístico. En este muestreo, todas las muestras posibles de un tamaño fijo tienen la misma probabilidad de ser seleccionadas. Para la selección de cada elemento se ha de hacer con un sorteo completamente aleatorio y estas unidades tienen que ser señaladas de una forma identificada.

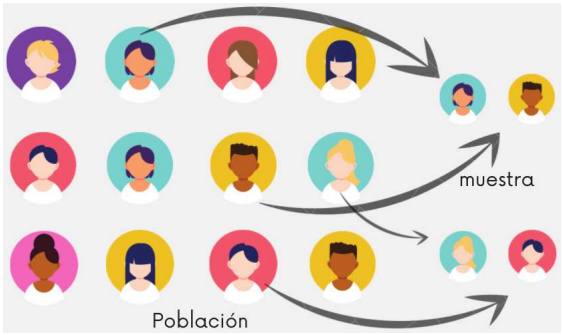

Figura 2.1: Selección de la muestra

Dependiendo de si las unidades de la población pueden ser seleccionadas más de una vez en la muestra o no, se considera muestreo aleatorio simple con reposición o sin reposición. En cuanto a çon reposición", una unidad que entra formar parte de la muestra, puede volver a ser elegida y esto es equivalente a realizar muestreo con poblaciones infinitas. Al contrario, cuando una unidad no vuelve a ser elegida, o bien no vuelve a formar parte de la población, le llamamos sin reposición y es equivalente a muestreo con poblaciones finitas.

Para seleccionar una muestra aleatoria simple, se consideran los siguientes pasos:

- Definir la población mediante la aplicación de una variable en el universo.
- Identificar un tamaño muestral.
- Evaluar el marco de muestreo y hacer ajustes necesarios.
- Asignar un número único a cada unidad.
- Determinar el tamaño muestral.
- Seleccionar al azar el número específico de unidades de la población.

En este caso, si tenemos una población de tamaño N y queremos seleccionar una muestra de tamaño n, el número de muestras posibles se determinará mediante la siguiente fórmula:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

cuando las variables son dependientes o sin reposición.

Al trabajar con la muestra, ha de obtener algunas mediciones de las características de esta muestra de la cual proviene de la población, es decir los resultados que obtengamos de esta

muestra son estimaciones de cómo se comporta la población. Si se extrae una muestra de ella se obtiene la media $\overline{x}$ y esta media es una estimación de la media poblacional μ.También desde esta muestra, se obtiene la cuasivarianza muestral sc^2 que es una estimación de la de la población σ^2.

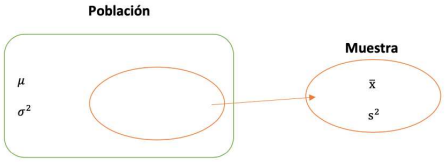

Figura 2.2: Población y Muestra

Las mediciones de características de la muestra se denominan los estimados, o bien estimaciones; Y los valores de las mediciones obtenidas a partir de la población se denominan los parámetros. Por consiguiente, mediante las aproximaciones si sabemos cómo se comporta la muestra, podemos inferir cómo se comporta la población.

Antes de realizar la búsqueda de los estimadores, se ha de saber los parámetros y estadísticos utilizados en muestreo, desde este punto de vista se consideran variables cuantitativas y cualitativas.

Parámetros utilizados en muestreo para variables cuantitativas

- Total poblacional: $\tau = x_1 + ... + x_N = \sum_{i=1}^{N} x_i$
- Media poblacional:$\mu = \frac{x_1+...+x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\tau}{N}$

Parámetros utilizados en muestreo para variables cualitativas

- Total poblacional: $\tau_A = a_1 + ... + a_N = \sum_{i=1}^{N} a_i$
- Media poblacional:$P_A = \frac{a_1+...+a_N}{N} = \frac{\sum_{i=1}^{N} a_i}{N} = \frac{\tau_A}{N}$

Estadísticos utilizados en muestreo para variables cuantitativas

- Media muestral: $\overline{X} = \frac{X_1+...+X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$
- Varianza muestral:$s^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}$
- Cuasivarianza muestral:$s_c^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1} = \frac{n}{n-1}s^2$
- Cuasivarianza muestral corregida:$s_{cN}^2 = \frac{N-1}{N}s_c^2$

Estadísticos utilizados en muestreo para variables cualitativas

- Proporción muestral: $\hat{p}_A = \frac{\sum_{i=1}^{n} A_i}{n}$
- Varianza muestral:$s_A^2 = \hat{p}_A(1-\hat{p}_A) = \hat{p}_A\hat{q}_A$

Búsqueda de los estimadores de τ, μ, τ_A y P_A

Sea $X_1...X_n$ o $A_1...A_n$ una muestra aleatoria simple de la población definida por X o A. Y un parámetro θ de P, sea $Z_1...Z_n$ también una muestra aleatoria simple en la que se puede considerar las variables tanto cualitativas como cuantitativas. Así, tenemos:

- Si $\theta = \mu \rightarrow W_i = \frac{1}{N} \ \ \forall i$
- Si $\theta = \tau \rightarrow W_i = 1 \ \ \forall i$

Ahora buscamos $\hat{\theta}$ estimador de θ. Cuando se cumple que la esperanza de la estimación es igual al parámetro, se dice que es un estimador insesgado. Por ejemplo, una media muestral es un estimador centrado. Por tanto, el estimador cetrado de $\hat{\theta}$ es $E[\hat{\theta}] = \theta$.

Demostración: $\hat{\theta} = \beta_1 Z_1 + ... + \beta_n Z_n \quad \beta_i \in \mathbb{R} \quad \forall i$

tal que $E[\hat{\theta}] = \theta$

donde $E[\hat{\theta}] = E(\beta_1 Z_1 + ... + \beta_n Z_n) = \beta_1 E[Z_1] + ... + \beta_n E[Z_n] = \beta_1 \mu + ... + \beta_n \mu = \mu \sum_{i=1}^{n} \beta_i = \theta$

Así que, $\sum_{i=1}^{n} \beta_i = \frac{\theta}{\mu} \rightarrow \beta_i = \frac{\theta}{n\mu} \quad \forall i$ es decir, β_i son iguales para todos i.

$$\hat{\theta} = \sum_{i=1}^{n} \beta_i Z_i = \sum_{i=1}^{n} \frac{\theta}{n\mu} Z_i = \frac{\theta}{\mu} \sum_{i=1}^{n} Z_i = \begin{cases} \text{cuantitativa:} \hat{\theta} = \frac{\theta}{\mu} \bar{Z} \\ \text{cualitativa:} \hat{\theta} = \frac{\theta}{P_A} \hat{P}_A \end{cases}$$

En este caso $\bar{Z}$ presenta como $\bar{X}$. Así, se obtienen los parámetros correspondientes:

$\tau : \hat{\tau} = \frac{\tau}{\mu} \bar{X} = \frac{\tau}{\frac{\tau}{N}} \bar{X} = N\bar{X}$

$\tau_A : \hat{\tau}_A = \frac{\tau_A}{P_A} \hat{P}_A = \frac{\tau_A}{\tau_A N} \hat{P}_A = N\hat{P}_A$

$\mu : \hat{\mu} = \frac{\mu}{\mu} \bar{X} = \bar{X}$

$P_A : \hat{P}_A = \frac{P_A}{P_A} \hat{P}_A = \hat{P}_A$ □

Además de estimar estos parámetros y estadísticos poblacionales, ha de fijarse un límite sobre el error de estimación. Para obtener el límite, primero se tiene que calcular la varianza del estimador mediante ciertos cálculos para cada uno de los estimadores anteriores. Desde este punto de vista, se consideran M.A.S con reposición y sin reposición.

2.1.1 Muestreo Aleatorio Simple con Reposición

En este tipo de muestreo, la probabilidad de seleccionar las unidades viene dada por $Pi = \frac{1}{n}$, y además todas las unidades tienen la misma probabilidad de ser elegidas, es decir, $P(primer\ extracto) = P(segundo\ extracto) = ... = P(n\ extracto) = \frac{1}{n}$

Primero, ha de conocer la varianza de $\hat{\theta}$:
$Var\hat{\theta} = Var(\frac{\theta}{\mu}\overline{z}) = Var(\frac{\theta}{\mu}\frac{\sum_{i=1}^{n} z_i}{n}) = \frac{\theta^2}{\mu^2 n^2} Var(\sum_{i=1}^{n} z_i) = \frac{\theta^2}{\mu^2 n^2} \sum_{i=1}^{n} Var(z_i) = \frac{\theta^2}{\mu^2 n^2} n\sigma^2 = \frac{\theta^2}{\mu^2} \frac{s_c^2}{n}$

Se fija que z_i son independientes.

- $\tau : Var(\hat{\tau}) = \frac{\tau^2}{\mu^2} \frac{s_c^2}{n} = \frac{\tau^2}{\frac{\tau^2}{N^2}} \frac{s_c^2}{n} = N^2 \frac{s_c^2}{n} \rightarrow \hat{\tau} = N\hat{\mu} \rightarrow Var(\hat{\tau}) = N^2 Var(\hat{\mu})$

- $\mu : Var(\hat{\mu}) = \frac{\mu^2}{\mu^2}\frac{s_c^2}{n} = \frac{s_c^2}{n}$
- $\tau_A : Var(\hat{\tau_A}) = \frac{\tau_A^2}{p_A^2}\frac{p_A q_A}{n} = \frac{\tau_A^2}{\frac{\tau_A^2}{N^2}} = N^2\frac{p_A q_A}{n}$
- $p_A : Var(\hat{p_A}) = \frac{p_A^2}{p_A^2}\frac{p_A q_A}{n} = \frac{p_A q_A}{n}$

2.1.2 Muestreo Aleatorio Simple sin Reposición

Llamaremos sin reposición a las unidades extraídas de la población no vuelve a ser elegidas y que no regresa a la base del muestreo. Cuando se considera este caso, $z_1 ... z_n$ son dependientes, así se fija la varianza:

Demostración: $Var(\hat{\theta}) = \frac{\theta^2}{\mu^2 n^2}\sum_{i=1}^n Var(z_i)$

donde $\sum_{i=1}^n Var(z_i) = \sum_{i=1}^n Var(z_i) + 2\sum_{i<j}^n Cov(z_i, z_j)$

$Cov(z_i, z_j) = \frac{\sum_{i,j}^n (z_i-\mu)(z_j-\mu)}{N} = E(z_i, z_j) - E(z_i)E(z_j)$ Ahora se desarrolla $E(z_i, z_j)$:

$$E(z_i, z_j) = \frac{1}{N(N-1)}\sum_{r=1}^N\sum_{s=1}^N z_r z_s = \frac{1}{N(N-1)}(N^2\mu^2 - \sum_{i=1}^N zr^2 \quad i \neq j, \quad r \neq s$$

$$= \frac{1}{N(-1)}[\sum_{r=1}^N z_r \sum_{s=1}^N z_s - \sum_{r=1}^N z_r^2]$$

$$= \frac{1}{N(N-1)}[\tau\tau - \sum_{r=1}^N z_r^2]$$

Para calcular $\sum_{r=1}^N z_r^2$, hay que saber su varianza:

$$s_c^2 = \frac{\sum_{i=1}^N (z_r-\mu)^2}{N} = \frac{\sum_{i=1}^N z_r^2}{N} - \mu^2$$

$$s_c^2 + \mu^2 = \frac{\sum_{i=1}^N z_r^2}{N}$$

$$\sum_{i=1}^N z_r^2 = N s_c^2 + N\mu^2$$

Así que, se ha obtenido $E(z_i, z_j)$, $Cov(z_i, z_j)$ y $Var(\sum_{i=1}^n z_i)$ con la siguiente forma:

$$E(z_i, z_j) = \frac{1}{N(N-1)}[N\mu^2(N-1) - N s_c^2] = \mu^2 - \frac{s_c^2}{N-1}$$

$$Cov(z_i, z_j) = \mu^2 - \frac{s_c^2}{N-1} - E[z_i]E[z_j] = \mu^2 - \frac{s_c^2}{N-1} - \mu^2 = -\frac{s_c^2}{N-1}$$

$$Var(\sum_{i=1}^n z_i) = \sum_{i=1}^n s_c^2 + 2\frac{n!}{(n-2)!2}(\frac{1}{N(N-1)}(N^2\mu^2 - N\mu^2 - N s_c^2) - \mu^2)$$

$$= n s_c^2 + n(n-1)(\frac{N}{N(N-1)}((N-1)\mu^2 - s_c^2) - \mu^2)$$

$$= n s_c^2 + n(n-1)(\mu^2 - \frac{s_c^2}{N-1} - \mu^2)$$

$$= ns_c^2 + n(n-1)\frac{-s_c^2}{N-1}$$

$$= \frac{n(N-n)-n(n-1)}{N-1}s_c^2$$

$$= \frac{n(N-n)}{N-1}s_c^2$$

$$= \frac{N-n}{N-1}ns_c^2$$

Así, hemos obtenido $Var(\hat{\theta})$:

$$Var(\hat{\theta}) = \frac{\theta^2}{n^2\mu^2}\frac{N-n}{N-1}ns_c^2 = \frac{N-n}{N-1}\frac{\theta^2}{\mu^2}\frac{s_c^2}{n}$$

□

A $\frac{n}{N}$ se le llama fracción de muestreo al cociente entre tamaño de la muestra y el de la población; a $\frac{N-n}{N-1}$ fracción de corrección por poblaciones finitas.

Es decir: $\lim_{N\to\infty}\frac{N-n}{N-1} = \lim_{N\to\infty}\frac{1-\frac{n}{N}}{1-\frac{1}{N}} = \lim_{N\to\infty}\frac{1-f}{\frac{1}{N}} = 1 \qquad f = \frac{n}{N}$

ya que se considera f y $\frac{1}{N}$ como 0, así se resulta que cuando el tamaño de población es muy grande, o bien cuando f $< 0{,}01$, $Var(\hat{\theta})$ con reposición $= Var(\hat{\theta})$ sin reposición.

Los parámetros de muestreo aleatorio simple sin reposición a partir de la varianza teórica serán:

- τ : $Var(\hat{\tau}) = \frac{N-n}{N-1}\frac{\tau^2}{\mu^2}\frac{s_c^2}{n} = \frac{N-n}{N-1}\frac{N^2\mu^2}{\mu^2}\frac{s_c^2}{n} = \frac{N-n}{N-1}N^2\frac{s_c^2}{n}$
- μ : $Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{\mu^2}{\mu^2}\frac{s_c^2}{n} = \frac{N-n}{N-1}\frac{s_c^2}{n}$
- τ_A : $Var(\hat{\tau_A}) = \frac{N-n}{N-1}\frac{\tau_A^2}{p_A}\frac{p_Aq_A}{n} = \frac{N-n}{N-1}N^2\frac{p_Aq_A}{n}$
- p_A : $Var(\hat{p_A}) = \frac{N-n}{N-1}\frac{p_A^2}{p_A^2}\frac{p_Aq_A}{n} = \frac{N-n}{N-1}\frac{p_Aq_A}{n}$

Cálculo de Las Varianzas Estimadas con Estimador Centrado

Se denomina sesgo de un estimador a la diferencia entre el valor esperado del estimador y el verdadero valor del parámetro a estimar. Es deseable que un estimador sea insesgado o centrado, es decir, que el sesgo sea nulo para que la esperanza del estimador sea igual al valor de parámetro que se desea estimar.
Por ejemplo, si se desea estimar la media de una población, la media aritmética de la muestra es un estimador insesgado de la misma.

Con Reposición

- τ : $V\hat{a}r(\hat{\tau}) = N^2\frac{\sigma^2}{n} = N^2\frac{s_c^2}{n}$
- μ : $V\hat{a}r(\hat{\mu}) = \frac{\sigma^2}{n} = \frac{s_c^2}{n}$
- τ_A : $V\hat{a}r(\hat{\tau_A}) = N^2\frac{\hat{p}_A\hat{q}_A}{n} = N^2\frac{\frac{n}{n-1}\hat{p}_A\hat{q}_A}{n} = \frac{N^2\hat{p}_A\hat{q}_A}{n-1} \quad \hat{p}_A\hat{q}_A \approx s^2 \quad s_c^2 = \frac{n}{n-1}s^2$

- $p_A : \hat{Var}(\hat{p_A}) = \frac{\hat{p}_A\hat{q}_A}{n} = \frac{\frac{n}{n-1}\hat{p}_A\hat{q}_A}{n} = \frac{\hat{p}_A\hat{q}_A}{n-1}$

Sin Reposición

- $\tau : \hat{Var}(\hat{\tau}) = \frac{N-n}{N-1}N^2\frac{s_c^2}{n} = \frac{N-n}{N-1}N^2\frac{s_c^2}{n}\frac{N-1}{N} = N^2(1-\frac{n}{N})\frac{s_c^2}{n} = N^2(1-f)\frac{s_c^2}{n}$
- $\mu : \hat{Var}(\hat{\mu}) = \frac{N-n}{N-1}N^2\frac{s_c^2}{n} = \frac{N-n}{N-1}\frac{s_c^2}{n} = (1-f)\frac{s_c^2}{n}$
- $\tau_A : \hat{Var}(\hat{\tau}_A) = \frac{N-n}{N-1}N^2\frac{\hat{p}_A\hat{q}_A}{n} = N^2\frac{N-n}{N}\frac{\hat{p}_A\hat{q}_A}{n-1} = N^2(1-f)\frac{\hat{p}_A\hat{q}_A}{n-1}$
- $p_A : \hat{Var}(\hat{p}_A) = \frac{N-n}{N-1}\frac{\hat{p}_A\hat{q}_A}{n} = (1-f)\frac{\hat{p}_A\hat{q}_A}{n}$

2.1.3 Errores de Muestreo

En este apartado, se desea estimar el error absoluto y el relativo de muestreo. Cuando se lean unas encuestas o cuestionarios se debe saber que los errores de muestreo influyen altamente en los datos y pueden sacarse conclusiones incorrectas. Por consiguiente, para que podamos sacar resultados válidos, ha de calcular error de muestreo y un intervalo de errores, así, para controlar que en nuestra investigación tenga menor error posible.
A fin de reducir un error de muestreo, se hacen en dos pasos:

- Aumentar el tamaño de la muestra: cuando sea mayor el tamaño de la muestra, se acercará más al tamaño real de la población, el resultado será más exacto.
- Dividir la población en grupos: prueba los grupos de acuerdo con su tamaño de la población, en vez de usar una muestra aleatoria.

Los cálculos de errores son sencillos, se utiliza distribución normal con el error de muestreo y k, cuantil de una distribución de probabilidad, ya que si la distribución es normal, los estimadores también lo son.

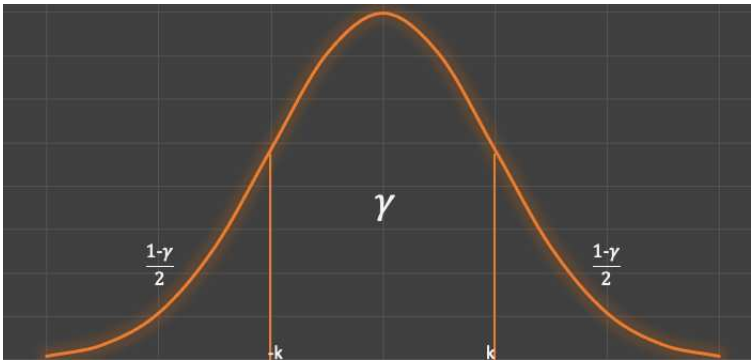

Figura 2.3: Distribución Normal

La relación entre el cuantil y γ la confianza será:

k	γ
1.96	95 %
2	95,45 %
3	99,74 %
4	99,99 %

Cuadro 2.1: cuantil y su nivel de confianza

Desigualdad de Chebyshev

La desigualdad de Chebyshev es una teorema que se utiliza para proporcionar una estimación conservadora, o bien intervalo de confianza de la probabilidad de que una variable aleatoria con su varianza correspondiente se sitúe en una cierta de su esperanza matemática, se expresa con la siguiente forma:

$$P(\mid X - \mu \mid < ks_c^2) \geq 1 - \frac{1}{k^2}$$

donde se supone que X es una variable aleatoria, $E(X) = \mu$, y su varianza $Var(X) = s_c^2$. Si X sigue una distribución normal, es decir $X \sim N(\mu, s_c^2)$, entonces:

$$P(\mid X - \mu \mid < ks_c) = 1 - \frac{1}{k^2}$$

Si X es igual a $\hat{\theta}$ centrado, así que $E(\hat{\theta}) = \theta$, se tiene que

$$P(\mid \hat{\theta} - \theta \mid < ks_c(\hat{\theta})) \geq 1 - \frac{1}{k^2}$$

En este caso, consideramos dos situaciones, si es una distribución normal, la confianza será $95,45\,\%$; Al contrario, ha de tener una confianza de $75\,\%$ como mínimo. Se define la relación entre cuantil y su nivel de confianza con la siguiente tabla:

k	$1 - \frac{1}{k^2} = \gamma$
2	75 %
3	88,9 %
4	93,75 %

Cuadro 2.2: Cuantil y su Nivel de Confianza (2)

Sin embargo, en cuanto a los errores de muestreo existen dos tipos: error absoluto y error relativo.

Error absoluto

Se define el error e como $e = k\sqrt{Var(\hat{\theta})}$ o $e = k\sqrt{V\hat{a}r(\hat{\theta})}$, este error absoluto se mide en las mismas unidades que la variable X, además se expresa siempre con un número distinto de cero, redondeándose siempre en exceso.

Error relativo

Para distinguir entre error absoluto y relativo, el relativo ya no tiene unidades, sino se mide en porcentajes, es decir, se obtendrá directamente el error en tanto por ciento, a la expresión anterior hay que multiplicarla por 100:

$$e = \frac{e_r}{|a|} \cdot 100\,\%$$

Cuando hablamos de error relativo de los estimadores μ y τ, consideramos que son iguales: $e_{r\mu} = e_{r\tau}$

$$e_{r\tau} = \frac{e_\tau}{|\hat{\tau}|} \cdot 100 = \frac{Ne_\mu}{|N\hat{\mu}|} \cdot 100 = \frac{e_\mu}{|\hat{\mu}|} \cdot 100 = e_{r\mu}$$

donde,

e_τ es un error absoluto
$\hat{\tau} = N\hat{\mu} \rightarrow Var(\hat{\tau}) = N^2 Var(\hat{\mu})$

entonces

$$e_\tau = k\sqrt{Var(\hat{\tau})} = k\sqrt{N^2 Var(\hat{\mu})} = kN\sqrt{Var(\hat{\mu})} = Ne_\mu$$

2.1.4 Cálculo del tamaño de la muestra

Antes de realizar la estimación del tamaño muestral, se ha de tener claro del parámetro que se quiere estimar, el error que se quiere meter y con qué confianza se considera. En los apartados anteriores, cuando hablamos de çon reposición.º "sin reposición", consideramos varianza estimada y teórica. No obstante, para calcular el tamaño de la muestra solo se utiliza varianza teórica por la necesidad de saber la varianza poblacional.
Se fija $\theta = \mu$, se supone un error e y una confianza γ, por lo tanto, su intervalo de confianza será de la siguiente forma:

$$I_\mu = \hat{\mu} \pm e = \hat{\mu} \pm k\sqrt{Var(\hat{\mu})}$$

Ahora a fin de calcular el tamaño muestral, teniendo en cuenta en el caso con reposición y sin reposición.

Con reposición

Se tiene que $Var(\hat{\mu}) = \frac{s_c^2}{n}$ con su error $e = k\sqrt{Var(\hat{\mu})} = k\sqrt{\frac{s_c^2}{n}}$. Se despeja n,

$$\frac{e^2}{k^2} = \frac{s_c^2}{n} \rightarrow n = \frac{k^2 s_c^2}{e^2}, \text{ se redondea por exceso}$$

A este n_∞ le llamamos $n_\infty = \frac{k^2 s_c^2}{e^2}$, recordándose de que la población es infinita en el caso de con reposición donde se puede volver a ser elegida una unidad de la muestra.

Sin reposición

Se tiene que $Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{s_c^2}{n}$ con su error $e = k\sqrt{Var(\hat{\mu})} = k\sqrt{\frac{N-n}{N-1}\frac{s_c^2}{n}}$. Se despeja n,

$$\frac{e^2}{k^2} = \frac{N-n}{N-1}\frac{s_c^2}{n}$$

$$\frac{e^2(N-1)}{k^2 s_c^2} = (N-n)\frac{1}{n}$$

$$\frac{e^2(N-1)}{k^2 s_c^2} n = N - n$$

$$(\frac{e^2(N-1)}{k^2 s_c^2} + 1)n = N$$

Entonces, se tiene n:

$$n = \frac{N}{\frac{e^2(N-1)}{k^2 s_c^2}+1} = \frac{N}{\frac{N-1}{n_\infty}+1} = \frac{n_\infty N}{N-1+n_\infty} = \frac{n_\infty}{1+\frac{n_\infty - 1}{N}}, \text{ se redondea n por exceso, pero } n_\infty \text{ no.}$$

Cuando la población total N es infinita, el tamaño muestral n también lo es, desde este punto de vista se permite utilizar tanto en con reposición como en sin reposición, ya que el tamaño muestral es lo mismo en los dos casos.

Caso más desfavorable en A

Se tiene A variable cualitativa y se quiere estimar una proporción, el tamaño muestral será $\frac{k^2 p_A q_A}{e^2}$, o n en el caso con reposición, donde $p_A = q_A = 0,5$. Se concluye que cuando $p_A q_A$ es mayor posible, n_∞ es lo mayor posible y n también lo es.

Influencia de N en el cálculo de n

Puesto que n=f(N), para que sea creciente, la derivada ha de ser mayor que 0.

1. f(N) es creciente.

f'(N)>0, f(N)=$\frac{Nn_\infty}{N+n_\infty-1}$, entonces

$$\text{f'(N)}=\frac{n_\infty(N+n_\infty-1)-Nn_\infty}{(N+n_\infty-1)^2}=\frac{n_\infty N+n_\infty^2-n_\infty-Nn_\infty}{(N+n_\infty-1)^2}=\frac{n_\infty(n_\infty-1)}{(N+n_\infty-1)^2}>0$$

Cuanto mayor es el valor de N, el valor de n también lo es.

2. Cuando f(N) tiene una asíntota en n_∞.

$\lim_{N\to\infty} f(N)=n_\infty$, es decir $\lim_{N\to\infty}\frac{n_\infty}{1+\frac{n_\infty-1}{N}}=n_\infty$

Se define: $\tilde{n}_\infty=\begin{cases}[n_\infty]+1 & n_\infty decimal\\ n_\infty & n_\infty entero\end{cases}$

donde $\tilde{n}_\infty\in\mathbb{Z}$, al final, se ha obtenido $\tilde{n}_\infty-1=\frac{n_\infty}{1+\frac{n_\infty-1}{\tilde{N}}}$. Se concluye que a partir de $\tilde{N}+1$, el tamaño de la muestra será n_∞. Además, se puede saber también la influencia de k, e y s_c^2 en el cálculo de n:

- Confianza k: a mayor confianza, n es mayor.
- Error e: a mayor error, n es menor.
- Varianza s_c^2: a mayor varianza o pq, n es mayor.

2.2 Muestreo Estratificado

Cada estrato se forma por miembros; por ejemplo si los estratos fueran personas, se podría pensar los miembros por sus edades, sexos, estado civil, etc. La población se divide en estratos, los estratos con grupos diferenciados por alguna variable que los hacen distintos; es decir, se llama L el número de estratos de la población. Se usa este muestreo para tener un objetivo de ser heterogéneo significativamente.
Para hacer un muestreo estratificado, se han de considerar en 6 pasos:

- Primero: Definir el objetivo principal.
- Segundo: Definir los estratos.
- Tercero: Realizar la estratificación.
- Cuarto: Dividir población en trozos.
- Quinto: Decidir el tamaño de la muestra de cada trozo.
- Sexto: Realizar un muestreo aleatorio simple de cada estrato.

Los estratos no siempre tienen el mismo número de unidades, es decir que pueden ser de tamaños diferentes, por consiguiente, la fracción de muestreo puede variar de un estrato a otro, esta fracción se representa como: $f_i=\frac{n_i}{N_i}$.
Comparando con muestreo aleatorio simple, el muestreo estratificado es una técnica más eficiente, pues se tiene una mejor calidad de las observaciones y se llega a toda la población. Las estimaciones pueden ser muy precisas, es decir se han de utilizar formas correctas para que el tamaño de la muestra de cada estrato sea bien seleccionado, por lo que se considera en tres métodos:

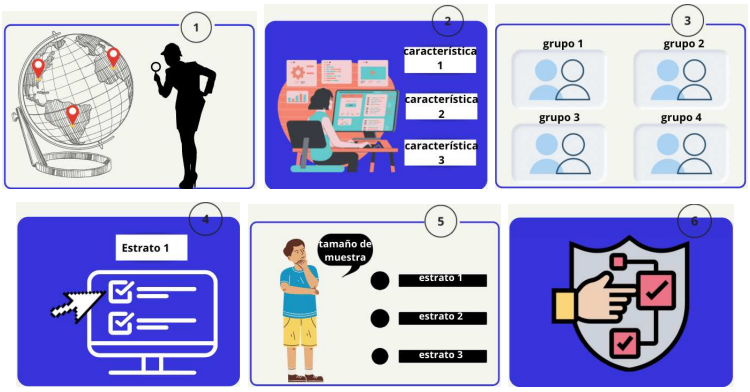

Figura 2.4: Pasos para hacer un muestreo estratificado

- **Muestra de igual tamaño o afijación constante:** se seleccionan el mismo número de unidad para cada estrato.
- **Muestra de asignación o afijación proporcional:** los tamaños de la muestra de cada estrato tienen un número de unidades en forma proporcional a las de los estratos de la población.
- **Muestra de asignación o afijación óptima:** ha de tener en cuenta los costes y el grado de variabilidad buscando que el error de estimación sea mínimo para un coste total.
- **Muestra de Neyman a costes constantes:** Se incorpora el factor coste, ya que uno de los objetivos del muestreo es recolectar la mayor cantidad de información, con mayor precisión y al menor costo. Es necesario conocer para cada estrato: El tamaño total, la desviación estándar y el costo de recolectar una encuesta en cada estrato.

2.2.1 Estimaciones del Total Poblacional τ

Para estimar el total poblacional τ: $\hat{\tau} = \hat{\tau}_1 + \hat{\tau}_2 + ... + \hat{\tau}_L = \sum_{i=1}^{L} \hat{\tau}_i$.

Además, τ es un estimador centrado, pues se tiene que tenemos:$E[\hat{\tau}] = E[\sum_{i=1}^{L} \hat{\tau}_i] = \sum_{i=1}^{L} E[\hat{\tau}_i] = \sum_{i=1}^{L} \tau_i = \tau$.

Estimador de total: $\hat{\tau} = \sum_{i=1}^{L} N_i \hat{\mu}_i$, ya que se tiene que: $\begin{cases} \hat{\tau}_i = N_i \hat{\mu}_i & \forall i = 1, 2, ..., L \\ \hat{\mu}_i = \overline{x}_i & \forall i = 1, 2, ..., L \end{cases}$

Una vez se obtiene el estimador del total se pueden calcular sus varianzas teóricas y estimadas tanto con reposición como sin reposición.

Varianza Teórica

Sin reposición: $Var(\hat{\tau}) = Var(\sum_{i=1}^{L} N_i \hat{\mu}_i) = \sum_{i=1}^{L} Var(N_i \hat{\mu}_i) = \sum_{i=1}^{L} N_i^2 Var(\hat{\mu}_i) = \sum_{i=1}^{L} N_i^2 \frac{N_i - n_i}{N_i - 1} \frac{s_{ci}^2}{n_i}$

Con reposición: $Var(\hat{\tau}) = \sum_{i=1}^{L} N_i^2 \frac{s_{ci}^2}{n_i}$

Varianza Estimada

Sin reposición: $V\hat{a}r(\hat{\tau}) = \sum_{i=1}^{L} N_i^2 V\hat{a}r(\hat{\mu}_i) = \sum_{i=1}^{L} N_i^2 (1 - \frac{n_i}{N_i}) \frac{s_{ci}^2}{n_i}$

Con reposición: $V\hat{a}r(\hat{\tau}) = \sum_{i=1}^{L} N_i^2 \frac{s_{ci}^2}{n_i}$

Su error correspondiente será: $e = k\sqrt{Var(\hat{\tau})}$ o $e = k\sqrt{V\hat{a}r(\hat{\tau})}$

2.2.2 Estimaciones de Media Poblacional μ

A fin de estimar la media poblacional, es necesario que se haga una estimación de la afijación; como se ha mencionado anteriormente existen tres tipos de afijación, pero en el dicho apartado se utiliza afijación proporcional. Sabemos que $\tau = N\mu \rightarrow \hat{\tau} = N\hat{\mu} \rightarrow \hat{\mu} = \frac{\hat{\tau}}{N}$. Así que, $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{\mu}_i = \sum_{i=1}^{L} \frac{N_i}{N}\hat{\mu}_i$, en el que llamamos a $W_i = \frac{N_i}{N}$ una fracción que corresponde al tamaño del estrato con respecto al tamaño de la población. De la misma forma, se tiene que $w_i = \frac{n_i}{n}$ es la fracción de muestra del estrato con respecto al tamaño entero de la muestra. A esta w_i se lo llamamos afijación.

Varianza Teórica
Sin reposición:
$Var(\hat{\mu}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 Var(\hat{\mu}_i) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{N_i-n_i}{N_i-1}\frac{s_{ci}^2}{n_i} = \sum_{i=1}^{L}(\frac{N_i}{N})^2 \frac{N_i-n_i}{N_i-1}\frac{s_{ci}^2}{n_i} = \sum_{i=1}^{L} W_i^2 \frac{N_i-n_i}{N_i-1}\frac{s_{ci}^2}{n_i}$

Con reposición: $Var(\hat{\mu}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{s_{ci}^2}{n_i}$

Varianza Estimada
Sin reposición: $V\hat{a}r(\hat{\mu}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 V\hat{a}r(\hat{\mu}_i) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(1-\frac{n_i}{N_i})\frac{s_{ci}^2}{n_i}$

Con reposición: $V\hat{a}r(\hat{\mu}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{S_{ci}^2}{n_i}$

Se considera el error en dos partes:
El error total:
$e = k\sqrt{Var(\hat{\mu})} = k\sqrt{\frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{N_i-n_i}{N_i-1}\frac{s_{ci}^2}{n_i}} = k\sqrt{\sum_{i=1}^{L}\frac{N_i^2}{N^2}Var(\hat{\mu}_i)} = k\sqrt{\sum_{i=1}^{L} W_i^2 Var(\hat{\mu}_i)}$

El error en cada estrato: $e_i = k\sqrt{Var(\hat{\mu}_i)} \quad i = 1,2,...,L \rightarrow Var(\hat{\mu}_i) = (\frac{e_i}{k})^2 = \frac{e_i^2}{k^2}$

Entonces, $e = k\sqrt{\sum_{i=1}^{L} W_i^2\frac{e_i^2}{k^2}} = k\sqrt{\frac{1}{k^2}\sum_{i=1}^{L} W_i^2 e_i^2} = \sqrt{\sum_{i=1}^{L} W_i^2 e_i^2}$. Sin embargo, se utiliza solo cuando se conoce el error en cada estrato de antemano.

2.2.3 Estimaciones de la Proporción p

$\hat{p}_i = \frac{X_i}{n_i} \quad i = 1,2,...,L \quad \hat{p} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{p}_i$

Varianza Teórica
Sin reposición: $Var(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 Var(\hat{p}_i) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{N_i-n_i}{N_i-1}\frac{p_i q_i}{n_i}$

Con reposición: $Var(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{p_i q_i}{n_i}$

Varianza Estimada
Sin reposición: $V\hat{a}r(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 V\hat{a}r(\hat{p}_i) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(1-\frac{n_i}{N_i})\frac{\hat{p}_i\hat{q}_i}{ni-1}$

Con reposición: $Var(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} \frac{N_i^2}{N^2}\frac{\hat{p}_i\hat{q}_i}{n_i-1}$

2.2.4 Cáculo del tamaño de muestra

Para que el resultado sea más precisa, se requiere conocer el tamaño de muestra, el de cada y sus costes correspondientes. En este caso se complica la situación, ya que hay que calcular $n_1, n_2, ..., n_L$ y $n = n_1 + n_2 + ... + n_L$. Se sabe que $w_i' = \frac{n_i}{n} \quad i = 1, .., L$, su afijación muestral será: $n_i = nw_i$ donde $0 \leq w_i \leq 1 \quad \forall i = 1, ..., L$, así se ha obtenido que $\sum_{i=1}^{L} w_i = 1$.
Según los diferentes tipos de afijación, se considera en cuatro puntos:

- **Constante:** $n_1 = ... = n_L \quad n_i = \frac{n}{L}$, entonces $w_i = \frac{1}{L} \quad \forall i = 1, ..., L$, es decir cuando tiene el mismo tamaño de la muestra cada estrato.

- **Proporcional:** Cada estrato tiene la misma proporción tanto en la población como en la muestra. $W_i = w_i \quad \frac{N_i}{N} = \frac{n_i}{n}$.

- **Óptima:** Si tiene el coste mínimo o error mínimo. $w_i = \frac{\frac{N_i s_{ci}}{\sqrt{c_i}}}{\sum_{i=1}^{L} \frac{N_i s_{ci}}{\sqrt{c_i}}} \quad \forall i = 1, ..., L$ donde c_i es el coste de extraer una unidad de la muestra en el estrato i.

- **Neyman:** Si el coste de cada estrato es constante: $c_1 = ... = c_L = c$ así que: $w_i = \frac{\frac{N_i s_{ci}}{\sqrt{c}}}{\frac{\sum_{i=1}^{L} N_i s_{ci}}{\sqrt{c}}} \rightarrow w_i = \frac{N_i s_{ci}}{\sum_{i=1}^{L} N_i s_{ci}}$.

En el caso de que conocer cuál ha de ser $n_1, ...n_L$, se debe considerar afijación óptima y el caso de coste fijo con un error mínimo. Sea C un coste de extraer por toda la muestra y c_i un coste de extraer por una observación en cada estrato donde $i = 1, ..., L$.
Así tenemos $C = n_1 c_1 + ... + n_L c_L = \sum_{i=1}^{L} n_i c_i$, según la afijación proporcional $n_i = w_i n$ y además $\sum_{i=1}^{L} w_i = 1$.
Con tal de obtener el tamaño de la muestra, se sustituye en:

$$C = \sum_{i=1}^{L} w_i n c_i = n \sum_{i=1}^{L} w_i c_i \rightarrow n = \frac{c}{\sum_{i=1}^{L} w_i c_i}$$

En el caso de que se calcula el tamaño, no solo se necesita conocer los costes, a no ser que se calcule también la afijación w_i. A partir de este punto de vista se debe llevar a cabo la estimación de varianza total, donde

$$\hat{\tau} = \sum_{i=1}^{L} \hat{\tau}_i \rightarrow Var(\hat{\tau}) = \sum_{i=1}^{L} Var(\hat{\tau}_i) = \sum_{i=1}^{L} N_i^2 \frac{N_i n_i}{N_i - 1} \frac{s_{ci}^2}{n_i}$$

Se sustituye por n_i:

$$Var(\hat{\tau}) = \sum_{i=1}^{L} N_i^2 \frac{N_i n_i}{N_i - 1} \frac{s_{ci}^2}{ni} = \sum_{i=1}^{L} N_i^2 \frac{N_i - w_i n}{N_i - 1} \frac{s_{ci}^2}{w_i n} = \sum_{i=1}^{L} \frac{N_i^3}{N_i - 1} \frac{s_{ci}^2}{w_i} \frac{1}{n} - \sum_{i=1}^{L} \frac{N_i^2 s_{ci}^2}{N_i - 1} \frac{w_i n}{w_i n}$$

Para que el error sea mínimo, ha de minimizar $Var(\hat{\tau})$ considerándose dos restricciones: $C = n \sum_{i=1}^{L} w_i c_i$ y $\sum_{i=1}^{L} w_i = 1$, se aplica mediante los multiplicadores de Lagrange a partir de la función construida:

$$F_(w_1, ..., w_L, n) = \sum_{i=1}^{L} \frac{N_i^3}{N_i - 1} \frac{s_{ci}^2}{w_i} \frac{1}{n} - \sum_{i=1}^{L} \frac{N_i s_{ci}^2}{N_i - 1} + \lambda_1 (n \sum_{i=1}^{L} w_i c_i - c) + \lambda_2 (\sum_{i=1}^{L} w_i - 1)$$

A partir de esta función, se permite considerar en dos partes:

- (1) $\frac{\partial F}{\partial n} = \frac{-1}{n^2} \sum_{i=1}^{L} \frac{N_i^3}{N_i - 1} \frac{s_{ci}^2}{w_i} + \lambda_1 \sum_{i=1}^{L} w_i c_i = 0$

- (2) $\frac{\partial F}{\partial w_i} = \frac{-1}{w_i^2} \frac{N_i^3}{N_i - 1} \frac{s_{ci}^2}{n} + \lambda_1 n c_i + \lambda_2 = 0$

Con la ecuación (2) se multiplica por w_i y se divide entre n:

$$-\frac{1}{w_i}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n^2}+\lambda_1 w_i c_i+\lambda_2\frac{w_i}{n}=0$$

Sumamos todas las ecuaciones obtenidas:

$$-\sum_{i=1}^{L}\frac{1}{w_i}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n^2}+\lambda_1\sum_{i=1}^{L}w_i c_i+\lambda_2\sum_{i=1}^{L}\frac{w_i}{n}=0$$

$$(1)+\lambda_2\sum_{i=1}^{L}\frac{w_i}{n}=0\rightarrow\frac{\lambda^2}{n}\sum_{i=1}^{L}w_i=0$$

Por las restricciones mencionadas anteriormente, ahora se sale:

$$\frac{\lambda^2}{n}1=0\rightarrow\lambda_2=0$$

se lo sustituye en la ecuación (2) y para despejar w_i:

$$-\frac{1}{w_i}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n}+\lambda_1 n c_i+0=0$$

$$\frac{1}{w_i}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n}=\lambda_1 n c_i$$

$$w_i^2=\frac{\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n}}{\lambda_1 n c_i}\rightarrow w_i=\sqrt{\frac{\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{n}}{\lambda_1 n c_i}}=\sqrt{\frac{1}{\lambda_1 n^2}}\frac{\sqrt{\frac{N_i^3}{N_i-1}s_{ci}^2}}{\sqrt{c_i}}$$

$$w_i=constante\frac{\sqrt{\frac{N_i^3}{N_i-1}s_{ci}^2}}{\sqrt{c_i}}$$

A efecto de saber cuánto vale w_i, hemos de calcular primero el valor del constante. Para llevarse a cabo es necesario sumar todos los w_i:

$$\sum_{i=1}^{L}w_i=\sum_{i=1}^{L}constante\frac{\sqrt{\frac{N_i^3}{N_i-1}s_{ci}^2}}{\sqrt{c_i}}=1\rightarrow 1=constante\sum_{i=1}^{L}\frac{\sqrt{\frac{N_i^3}{N_i-1}s_{ci}^2}}{\sqrt{c_i}}\rightarrow constante=\frac{1}{\sum_{i=1}^{L}\frac{\sqrt{\frac{N_i^3}{N_i-1}s_{ci}^2}}{\sqrt{c_i}}}$$

Por último, teniendo en cuenta de que en los campos de muestreo se tiene consideración a que N_i sea equivalente a $Ni-1$. Por consiguiente:

$$w_i=\frac{N_i\frac{s_{ci}}{\sqrt{c_i}}}{\sum_{i=1}^{L}N_i\frac{s_{ci}}{\sqrt{c_i}}}$$

A diferencia del tipo anterior (coste fijo y error mínimo), se considera también cuando el error sea fijo con un coste mínimo. Se sigue con el mismo proceso que el caso anterior, pero ahora se minimiza el coste. Se considera el error:

$$e=k\sqrt{Var(\hat{\mu})}=k\sqrt{\frac{1}{N^2}\sum_{i=1}^{L}N_i\frac{N_i-n_i}{N_i-1}\frac{s_{ci}^2}{n_i}}=k\sqrt{\frac{1}{N^2}\sum_{i=1}^{L}N_i^2\frac{N_i^2-w_i n}{N_i-1}\frac{s_{ci}^2}{w_i n}}.$$

Se despeja n:

$$(\frac{eN}{K})^2=(\sqrt{\sum_{i=1}^{L}N_i^2\frac{N_i^2-w_i n}{N_i-1}\frac{s_{ci}^2}{w_i n}})^2$$

$$(\frac{eN}{K})^2=\sum_{i=1}^{L}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{nw_i}-\sum_{i=1}^{L}\frac{N_i^2}{N_i-1}s_{ci}^2$$

$$(\frac{eN}{K})^2+\sum_{i=1}^{L}\frac{N_i^2}{N_i-1}s_{ci}^2=\frac{1}{n}\sum_{i=1}^{L}\frac{N_i^3}{N_i-1}\frac{s_{ci}^2}{w_i}$$

Entonces n será:

$$n = \frac{\sum_{i=1}^{L} \frac{N_i^3}{N_i - 1} \frac{s_{ci}^2}{w_i}}{(\frac{eN}{k})^2 + \sum_{i=1}^{L} \frac{N_i^2}{N_i - 1} s_{ci}^2} = \frac{\sum_{i=1}^{L} \frac{N_i^2 s_{ci}^2}{w_i}}{(\frac{eN}{k})^2 + \sum_{i=1}^{L} N_i s_{ci}^2}$$

Para que el error no sea sobrepasado y el coste sea mínimo, se redondea los tamaños de la muestra en el caso de coste fijo y error fijo respectivamente. Por consiguiente, se ha obtenido el siguiente resumen(Figura 5).

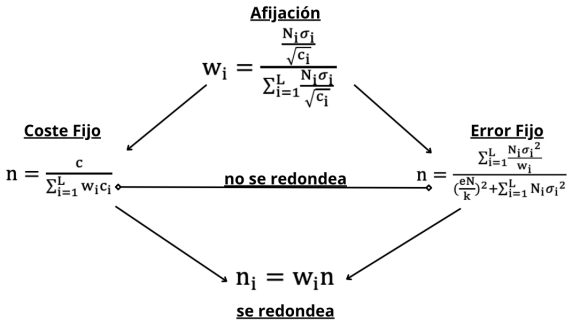

Figura 2.5: Esquema del cálculo de n en afijación óptima

2.3 Muestreo por Conglomerados

En cuanto a conglomerados, se divide la población en trozos que son una 'fotocopia' reducida de ella, es decir, tiene las mismas propiedades. A diferencia del muestreo estratificado, los individuos dentro de cada conglomerado son heterogéneos dentro de sí y homogéneos con los demás. Se extrae una muestra aleatoria simple de conglomerados y se estudian todos los individuos en cada conglomerado.

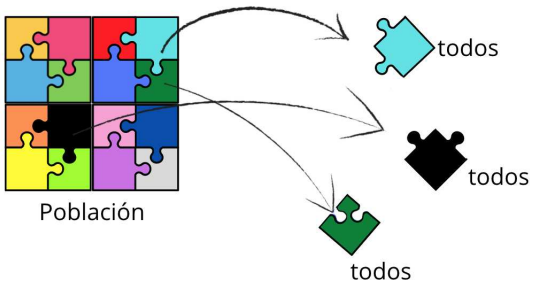

Figura 2.6: Muestreo por Conglomerados

Normalmente, cuando aplicamos esta técnica se definen los conglomerados geográficamente. Por ejemplo, si queremos estudiar qué proporción de la población española tiene cáncer de mama, se puede dividir el total de la población en comunidades autónomas y escoger algunas de ellas para ser estudiadas. Una vez se conocen los conglomerados, se permite ser seleccionados los conglomerados a estudiar mediante un muestreo aleatorio simple o sistemático. Ahora se puede estudiar todos los individuos que forman parte de los mismos grupos. Teniendo en cuenta de si se desea estudiar más sobre los individuos, estamos hablando de muestreo por conglomerados en dos etapas, es decir, se selecciona primero los trozos que se quiere estudiar, dentro de estos trozos se escoge de nuevo los individuos que nos interesen.

2.3.1 Muestreo por conglomerados en una etapa

Notaciones utilizadas en Muestreo por conglomerados

- N=número total de conglomerados en la población
- n=número de conglomerados en la muestra
- M_i=número de individuos en el conglomerados, $i = 1, ..., N$
- M=número total de individuos en la población, donde $M = \sum_{i=1}^{N} M_i$
- $\bar{M} = \frac{M}{N}$=número medio de individuos por conglomerados
- m=la suma del número de los individuos en la muestra
- $\hat{\bar{M}} = \bar{m} = \frac{\sum_{i=1}^{n} M_i}{n}$=número medio de individuos en conglomerados de la muestra

Estimaciones de μ y τ

Se considera un estimador de razón, donde 'x'=τ_i es el total de la variable en conglomerado i e 'y'=M_i es el tamaño de conglomerado i. El estimador de la media se escribe como $\hat{\mu}$ es un estimador de una razón: $\hat{\mu} = \frac{\sum_{i=1}^{n} \tau_i}{\sum_{i=1}^{n} M_i}$. Se define razón de muestreo cuando todos los objetos tienen las mismas probabilidades de ser elegidas en la muestra: $r = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$, su varianza teórica será:$Var(r) = \frac{N-n}{N-1}\frac{1}{\mu y^2}\frac{s_{cR}^2}{n}$ y su varianza estimada: $\hat{Var}(r) = (1 - \frac{n}{N})\frac{1}{\mu y^2}\frac{s_r^2}{n}$.

donde $\begin{cases} s_{cR}^2 = \frac{\sum_{i=1}^{N}(x_i - Ry_i)^2}{n} \\ s_r^2 = \frac{\sum_{i=1}^{n}(\tau_i - \mu M_i)^2}{n-1} \end{cases}$

Se sustituye en el estimador de la razón obteniéndose la varianza teórica y estimada de μ respectivamente:

$$Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{1}{\bar{M}^2}\frac{s_{c\tau}^2}{n},\ \hat{Var}(\hat{\mu}) = (1 - \frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_\tau^2}{n}$$

En este caso, si $\bar{M}$ es desconocida, lo estimamos por el valor de la muestra $\hat{\bar{M}} = \bar{m}$.
Respecto al total, se tiene $\hat{\tau} = M\hat{\mu}$, así se permite obtener la varianza teórica y estimada del total teniendo en cuenta de que M es un constante:

$$Var(\hat{\tau}) = N^2\frac{N-n}{N-1}\frac{s_{c\tau}^2}{n}$$

$$\hat{Var}(\hat{\tau}) = M^2(1 - \frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_\tau^2}{n} = M^2(1 - \frac{n}{N})\frac{1}{\frac{M^2}{N^2}}\frac{s_\tau^2}{n} = N^2(1 - \frac{n}{N})\frac{s_\tau^2}{n}$$

Estimador de p

En muestreo por conglomerados es equivalente a estimador de la media. Se supone un a_i total de los individuos que cumplen una condición propuesta. Se tiene que $\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i}$ y su varianza será:

$$\hat{Var}(\hat{p}) = (1 - \frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_a^2}{n},\ \text{donde } s_a^2 = \frac{\sum_{i=1}^{n}(a_i - \hat{p}M_i)^2}{n-1}$$

El tamaño de la muestra

Cuando se calcula el tamaño de la muestra para muestreo por conglomerados, es igual que calcularlo para una razón:

$$n_\infty = \frac{k^2 s_{c\tau}^2}{\bar{M}^2 e^2};\ n = \frac{n_\infty}{1 + \frac{n_\infty - 1}{N}}$$

2.3.2 Muestreo por conglomerados en dos etapas

En este apartado se realiza un primer muestreo de conglomerados, unidades de primera etapa y un segundo muestreo de unidades de segunda etapa o finales en este caso. Esto es, una vez muestreados los conglomerados, solo tomamos algunos elementos dentro de cada conglomerado. La notación de esta técnica ha de considerarse también en dos etapas:

- i=1,...,n o N, contador de conglomerados.
- j=1,...,m_i o M_i, contador de unidades del conglomerado i.
- x_{ij}=valor de la unidad j del conglomerado i.
- $\bar{X}_i = \frac{\sum_{j=1}^{n} x_{ij}}{m_i}$=media muestral para el i-ésimo conglomerado.
- $f_1 = \frac{n}{N}$ y $f_2 = \frac{m_i}{M_i}$=fracciones de muestreo.

Se utiliza esta técnica cuando un conglomerado contiene bastantes unidades para obtener una medición de cada uno, mientras que estos elementos son semejantes a la medición que una parte de ellos proporciona de la información del conglomerado total. Por otro lado, en beneficio de que la investigación no sea muy costosa económicamente o que sea imposible construir una lista completa de unidades poblacionales, este método tiene en cuenta a grandes poblaciones, por tanto aplicar cualquier otro método podría ser muy costoso.
No obstante, tras implementar este método, ha de tener claro de que la precisión sea menor y se compliquen las operaciones algebraicas a la hora de calcular varianzas.

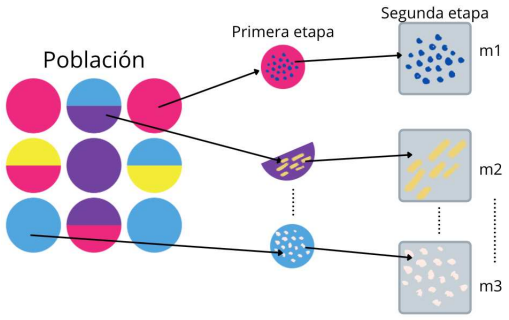

Figura 2.7: Muestreo por Conglomerados en dos etapas

Estimaciones de Medidas

Se tiene que distinguir en dos casos:
(1) Estimador insesgado: Cuando se conoce M
Estimaciones de media

$$\hat{\mu} = \frac{N}{M} \frac{\sum_{i=1}^{n} M_i \bar{X}_i}{n}$$

$$V\hat{a}r(\hat{\mu}) = (1 - \frac{n}{N}) \frac{1}{\bar{M}^2} \frac{s_b^2}{n} + \frac{1}{nN\bar{M}^2} \sum_{i=1}^{n} M_i^2 (1 - \frac{m_i}{M_i}) \frac{s_{ci}^2}{m_i}$$

$$s_b^2 = \frac{\sum_{i=1}^{n} (M_i \bar{X}_i - \bar{M}\hat{\mu})^2}{n-1}$$

donde s_{ci}^2 es la cuasivarianza dentro del conglomerado i.

El primer término de la varianza estimada de la media es debido a las diferencias que existe entre conglomerados, que llamaremos variabilidad inter: $(1-\frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_b^2}{n}$. Si todos los conglomerados fueran iguales, este sumando sería cero. El segundo término recoge las diferencias entre las unidades de segunda etapa o variabilidad intra: $\frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2(1-\frac{m_i}{M_i})\frac{s_{ci}^2}{m_i}$.

Estimaciones de proporciones

Al igual que en el muestreo por conglomerados en una etapa la estimación de proporciones, cuando se realizan dos etapas es muy parecida a lo visto para medias.

$$\hat{p} = \frac{N}{M}\frac{\sum_{i=1}^{n} M_i\hat{p}_i}{n}$$

Cuya su varianza viene dada por:

$$V\hat{a}r(\hat{p}) = (1-\frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_b^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2(1-\frac{m_i}{M_i})\frac{\hat{p}_i\hat{q}_i}{m_i-1}$$

$$s_b^2 = \frac{\sum_{i=1}^{n}(M_i\hat{p}_i-\bar{M}\hat{p})^2}{n-1}$$

Comparando con las estimaciones para medias, hemos sustituido medias por proporciones y la varianza s_b^2, dentro de los conglomerados por $\hat{p}_i\hat{q}_i$.

(2) Estimador sesgado: Cuando no se conoce M

Estimaciones de medias

En este caso se utiliza estimador de razón:

$$\hat{\mu}_r = \frac{\sum_{i=1}^{n} M_i\bar{X}_i}{\sum_{i=1}^{n} M_i}$$

$$V\hat{a}r(\hat{\mu}) = (1-\frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_r^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2(1-\frac{m_i}{M_i})\frac{s_{ci}^2}{m_i}$$

$$s_r^2 = \frac{\sum_{i=1}^{n}(M_i\bar{X}_i-\bar{M}\hat{\mu}_r)^2}{n-1}$$

Como M es desconocida, han de considerar los siguientes parámetros:

$$\hat{\bar{M}} = \frac{\sum_{i=1}^{n} M_i}{n} \text{ y } \hat{M} = \frac{\sum_{i=1}^{n} M_i}{n}N$$

Estimaciones de proporciones

$$\hat{p} = \frac{\sum_{i=1}^{n} M_i\hat{p}_i}{\sum_{i=1}^{n} M_i}$$

Cuya su varianza viene dada por:

$$V\hat{a}r(\hat{p}) = (1-\frac{n}{N})\frac{1}{\bar{M}^2}\frac{s_r^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2(1-\frac{m_i}{M_i})\frac{\hat{p}_i\hat{q}_i}{m_i-1}$$

$$s_r^2 = \frac{\sum_{i=1}^{n} M_i^2(\hat{p}_i-\hat{p})^2}{n-1}$$

Respecto a la variabilidad inter e intra, se resulta si la inter es pequeña compara con la intra, se podría escoger pocos conglomerados y muchas unidades dentro de cada conglomerado.

Cálculo del tamaño de la muestra

Se debe a ser dos etapas, es necesario encontrar n y m_i dependiendo de dos fuentes de variación entre (inter) conglomerados y dentro (intra) de conglomerados. Cuando la variación inter es grande y la intra es pequeña, significará que n es grande y m_i es pequeña o al revés. En algunas ocasiones, los conglomerados son del mismo tamaño, por ejemplo un gran envio de artículos en contenedores de 10 cajas (conglomerados de primera etapa) con 100 unidades elementales, estas se considera como segunda etapa. Bajo estas condiciones se cumplirá:

$$M_1 = M_2 = ... = M_N = \bar{M} = constante$$
$$m_1 = m_2 = ... = m_N = \bar{m} = constante$$

Los dos estimadores puntuales propuestos coinciden para estas condiciones, así como la varianza de los mismos. La estimación de la media poblacional se reduce a una media aritmética simple de los conglomerados muestreados.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\bar{X}_i$$

Y la varianza teórica de la media se puede escribir como:

$$Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{1}{\bar{M}^2}\frac{s_{cb}^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\frac{M_i-m_i}{M_i-1}\frac{s_{ci}^2}{m_i}$$

donde se ha considerado:

$$s_{cb}^2 = \frac{\bar{M}^2\sum_{i=1}^{N}(\mu_i-\mu)^2}{N} \rightarrow \hat{s}_{cb}^2 = s_b^2 = \frac{\bar{M}^2\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2}{n-1}$$

$$s_{ci}^2 = \frac{\sum_{j=1}^{M_i}(x_{ij}-\mu_i)^2}{M_i} \rightarrow \hat{s}_{ci}^2 = s_{ci}^2 = \frac{\sum_{j=1}^{m_i}(X_{ij}-\bar{X}_i)^2}{m_i-1}$$

de esta forma, se ha obtenido:

$$\begin{matrix} s_{c1}^2 = \frac{\sum_{i=1}^{N}(\mu_i-\mu)^2}{N} & \hat{s}_{c1}^2 = s_1^2 = \frac{\sum_{i=1}^{n}(\bar{X}_i-\hat{\mu})^2}{n-1} \\ s_{cw}^2 = \frac{\sum_{i=1}^{N} s_{ci}^2}{N} & \hat{s}_{cw}^2 = s_w^2\frac{\sum_{i=1}^{n} s_{ci}^2}{n} \end{matrix} \left\{ \hat{s}_{c1}^2 = s_1^2 - \frac{s_w^2}{\bar{m}} \right.$$

Así que nos permita simplificar la varianza teórica de esta forma:

$$Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{s_{c1}^2}{n} + \frac{\bar{M}-\bar{m}}{\bar{M}-1}\frac{s_{cw}^2}{n\bar{m}}$$

Optimización del tamaño de la muestra

La afijación óptima en muestreo por conglomerados en dos etapas trata de encontrar la pareja de valores n, número de conglomerados, y $\bar{m}$, valor medio del número de unidades a muestrear en la segunda etapa, que minimicen el coste del muestreo o el límite de error del mismo. "¿Qué nos resulta más eficiente, muestrear más conglomerados en primera etapa o más unidades dentro de cada conglomerados?". Se presentará las dos estrategias posibles, la de coste fijo donde conocemos el coste y minimizamos el error, o bien, nos fijamos el error y minimizamos el coste.
En primer lugar, se considera el caso más sencillo:

$$f_1 = f_2 = 0, \text{ fracciones muestreales despreciables.}$$

En estas condiciones la varianza teórica de la media puede simplificarse:

$$Var(\hat{\mu}) = \frac{s_{c1}^2}{n} + \frac{s_{cw}^2}{n\bar{m}}$$

Estrategia de error mínimo y coste fijo

$$C = nc_1 + n\bar{m}c_2,\ \bar{m} = \sqrt{\frac{s_{cw}^2 c_1}{s_{c1}^2 c_2}} \text{ (se despeja n de C)}$$

llamamos a c_1 el coste de extraer un coglomerado y a c_2 el de extraer un elemento del conglomerado. En cuanto a las relaciones entre s_{c1}^2, s_{cw}^2 y $\bar{m}$, se puede pensar en si s_{c1}^2 aumenta, $\bar{m}$ disminuye; Sin embargo, si s_{cw}^2 aumenta, $\bar{m}$ también.

Estrategia de coste mínimo y error fijo

$$e = k\sqrt{Var(\hat{\mu})},\ \bar{m} = \sqrt{\frac{s_{cw}^2 c_1}{s_{c1}^2 c_2}} \text{ (se despeja n de e)}$$

Si cumple que las fracciones muestrales son distintas de cero: $f_1 \neq 0$ y $f_2 \neq 0$. Se dice que son fracciones de muestreo o despreciables. El cálculo del tamaño medio óptimo del conglomerado y el número de estos sigue siendo como el caso despreciable. La varianza teórica es, en este caso:

$$Var(\hat{\mu}) = \frac{N-n}{N-1}\frac{s_{c1}^2}{n} + \frac{\bar{M}-\bar{m}}{\bar{M}-1}\frac{s_{cw}^2}{n\bar{m}}$$

Estrategia de error mínimo y coste fijo

$$C = nc_1 + n\bar{m}c_2,\ \bar{m} = \sqrt{\frac{c_1}{c_2}\frac{1}{\frac{s_{c1}^2}{s_{cw}^2} - \frac{1}{\bar{M}}}} \text{ (se despeja n de C)}$$

Estrategia de coste mínimo y error fijo

$$e = k\sqrt{Var(\hat{\mu})},\ \bar{m} = \sqrt{\frac{c_1}{c_2}\frac{1}{\frac{s_{c1}^2}{s_{cw}^2} - \frac{1}{\bar{M}}}} \text{ (se despeja n de e)}$$

Cálculo de tamaños de muestra para proporciones

Todo lo expuesto en el apartado anterior es de aplicación cuando lo que se pretende es estimar una proporción o porcentaje poblacional, tanto para fracciones muestrales despreciables, como para cuando no lo son. Hay también, como en el caso de estimación de medias, las dos estrategias para minimizar el error o el coste. Los esquemas para los cálculos necesarios en cada caso son idénticos a lo ya presentados, Con la particularidad de la varianzas inter e intra, vienen dadas por:

$$s_{cb}^2 = \frac{\bar{M}^2\sum_{i=1}^{N}(p_i-p)^2}{N} \rightarrow \hat{s}_{cb}^2 = s_b^2 = \frac{\bar{M}^2\sum_{i=1}^{n}(\hat{p}_i-\hat{p})^2}{n-1}$$

$$s_{ci}^2 = \frac{\sum_{j=1}^{M_i}(a_{ij}-p_i)^2}{M_i} \rightarrow \hat{s}_{ci}^2 = s_{ci}^2 = \frac{\sum_{j=1}^{m_i}(A_{ij}-\hat{p}_i)^2}{m_i-1}$$

Su estimación de la proporción y varianza teórica son respectivamente:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_i$$

$$Var(\hat{p}) = \frac{N-n}{N-1}\frac{1}{\bar{M}^2}\frac{s_{cb}^2}{n} + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n}M_i^2\frac{M_i-m_i}{M_i-1}\frac{p_i q_i}{m_i}$$

2.4 Muestreo por cuotas

El muestreo por cuotas es una de las técnicas pertenecidas a muestreo no probabilístico. El objetivo principal de esta ténica será investigar una característica de un subgrupo a que le interesa estudiar el investigador, mientras que se permite estudiar la relación entre los subgrupos de una población total. Una vez que se decidir aplicar esta técnica, lo fundamental es preparar adecuadamente la selección de los entrevistadores, la elección del método de encuesta y la elaboración de cuotas. Un plan de cuotas podría ser el número de entrevistas que quieren realizar, las características de los participantes y sus proporciones correspondientes. Las proporciones de la población reflejan las de la muestra.
Existen formas diversas para realizar el muestreo de cuotas, las más comunes se conocen como encuestas *online*, telefónicas o *in situ*. Para que los participantes tengan las mismas características que las de por cuotas, las primeras preguntas de la entrevista deben ser sencillas personales, por ejemplo, el sexo, la edad, el estado civil, etc. Tras este primer paso el entrevistador ya podría hacer preguntas reales sobre el estudio o la investigación que lo desea.
Antes de que se aplique esta técnica en un caso real, hemos de quedar claro en qué situación se pueda utilizar un muestreo por cuotas. Primero, cuando un investigador tiene criterios específicos para llevar a cabo la investigación en los que pueden ser el filtro para la formación de subgrupos o cuotas. Segundo, cuando le interesa al investigador un análisis comparativo entre dos grupos.

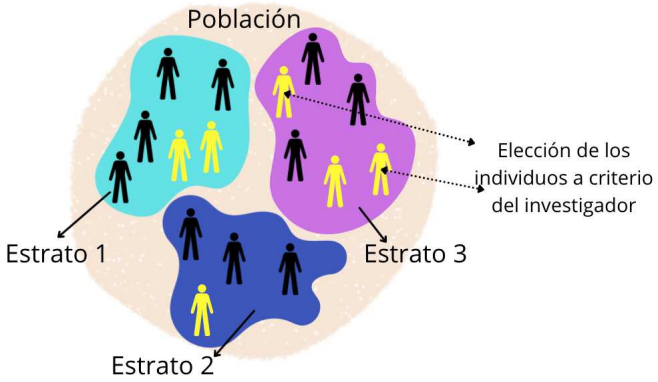

Figura 2.8: Muestreo por cuotas

2.4.1 Pasos a realizar en el muestreo por cuotas

- Se divide la población en k estratos con alguna característica común en los que tienen N_i elementos, tales que: $N = N_1 + N_2 + ... + N_k$
- Determina qué porcentaje respecto a la población total representa cada uno de los estratos.
- Se selecciona número de sujetos n_i en cada grupo, el total de elementos $n = n_1 + n_2 + ... + n_k$
- Calcula el número de elementos por cada cuota, de modo que sean proporcionales al porcentaje que representa cada uno de ellos respecto de la población total.
- Se toma el número de los elementos en cada estrato hasta completar la cuota correspondiente a cada uno sin ningún criterio.

2.4.2 Ventajas y Desventajas

Es una técnica sencilla y rápida con un costo bajo, se puede ayudar a los investigadores a estudiar una población usando cuotas específicas sin perder utilidad de los resultados. Sin embargo, a diferencia de otras técnicas probabilísticas, es imposible que acote el error que se está cometiendo al usar este tipo de muestreo. Por otro lado, para una cuota relevante se tiene el riesgo de obviarlo en un estudio; No obstante, tampoco se puede dejar de incluirla para el estudio. Eso ocasionará que los resultados estén fuertemente distorsionados.

3 | Violencia Cibernética

3.1 ¿Qué es la violencia cibernética?

Las redes sociales son plataformas digitales que facilitan las conexiones entre millones de usuarios de todas las partes del planeta. En ella, se puede compartir cualquier tipo de contenido. En los últimos años, estas plataformas han supuesto un cambio en nuestra forma de comunicarnos. No es difícil imaginar estar desconectados de estas redes ya que nos han facilitado en muchos aspectos nuestra forma de vida; Se cita un dicho de Marcos McKinnon: “la tecnología y las redes sociales han traído el poder a la gente”; Es decir, nos han permitido que nos comuniquemos de una forma más fluida incluso internacionalmente.

Sin embargo, el uso inadecuado puede tener consecuencias negativas. Por ejemplo, se difunden mentiras o publican fotografías vergonzosas; se envían mensajes hirientes o amenazas; se hacen pasar por la víctima y envían mensajes agresivos en nombre de dicha persona. En este sentido, le llamamos delitos cibernéticos; no obstante, existen muchos tipos de delitos cibernéticos como difusión de abuso sexual infantil, *hacking*, suplantación de identidad, etc. Dicho estudio, se enfoca a violencia cibernética, o bien ciberbulling del sector social: amenazas, difamación y coacción, donde la gente emprende acaso o intimidación a través de las tecnologías digitales como las redes sociales, las plataformas de juegos, de mensajería o los teléfonos móviles.(7) Este comportamiento se repite y se busca aterrorizar, enfadar o humillar a otras personas pudiendo tener consecuencias sobre efectos psicológicos, físicos o sexuales.

Con relación a detener dichos comportamientos, se publicaron leyes, aunque no son suficientemente completas. El ciberacoso lleva tipificado en el código penal en España desde 2013, según el artículo 131 de la Ley 26.904 y el artículo 183 ter de la Ley Orgánica 10/1995 de 23 de noviembre(8), se penalizan especialmente el acaso sexual a menos de edades por medios cibernéticos. Se nota que estas leyes se enfocan más cuando haya una consecuencia sexual. Sin embargo, en China la violencia cibernética se incluye en las leyes civiles, es decir, aunque se perjudica a otras personas físicamente o psicológicamente, no será penado con prisión, sino que se pagará una indemnización, o que se cierre su cuenta de redes sociales permanentemente con la que se comete la violencia. Por faltar completar dichas leyes, el público aún no ha dado prioridad a los perjuicios causados.

Según la investigación de una universidad en Beijing, China, el 88,44 % de los encuestados considera que la violencia cibernética suele ser conducta irracional de un grupo de personas, el 50,02 % de los que creen que será una forma para sentimentalismo, el 20,82 % de la gente considera que con la mayor posibilidad se podría influir por la violencia cibernética, incluye el estado de ánimo, la forma de hablar, etc., el 56,70 % solo creen que se influye parcialmente, por ejemplo sus discernimiento sobre un asunto, y el 22,48 % de los que creen que no se influyen nada.

3.2 Casos reales de violencia cibernética

Un joven chino, Liu Xuezhou, de diecisiete años llevaba varios meses buscando sus padres biológicos después de haber sido vendido por ellos cuando era pequeño. El joven los encontró y expresaba en las redes sociales que había sido un reencuenntro feliz. Sin embargo, la situación cambió después de que indicó a sus padres biológicos que necesitaba ayuda económica. Desde el 14 de diciembre de 2021 hasta el 24 de enero de 2022, estalló en conflictos con sus padres biológicos, y fue cuestionado por usuarios de Internet si su intención verdadera fuera pedir dinero. Finalmente, tras susfrir una gran violencia cibernética se suicidó tomando medicamentos al final de enero justo antes del año nuevo chino. Desgraciadamente, su suicidio no atrajo mucha atención oficial, los funcionarios chinos seguían ignorando la solicitud pública de legislación de que los abusadores de Internet deben asumir la responsabilidad penal, ya que ninguno de ellos fue considerado responsable legalmente salvo que sus cuentas fueron prohibidas.

El 6 de abril de 2010, una chica irlandesa de 15 años, Phoebe Prince, se suicidó ahorcándose tras sufrir insultos varios meses, siendo maltratada físicamente y psicológicamente a través de los móviles e Internet por algunos compañeros suyos del instituto en Massachusetts, EE.UU. A diferencia del primer caso, los acusadores fueron juzgados como adultos y podrían pasar el resto de su vida en la cárcel. Los responsables del instituto siguieron desempañado sus funciones. Este caso fue el primero que llevaba adelante la Fiscalía del Estado.

El 14 de octubre de 2019, la cantante de Corea del Sur Chol Jin-ri se suicidó en su casa. Según los reportajes ella sufrió años de depresión y maltrato físico por su empresa y violencia cibernética por el público. Durante las sesiones del tribunal se proclamó solamente que unos trabajadores de esa empresa habían sido culpable sin mencionar nada sobre violencia cibernética. A pesar de la causa principal de su muerte no fue cyberbullying, se influyó indirectamente.

4 | Encuestas

4.1 Datos Básicos

Se consideran los datos del número de personas que han realizado la encuesta sobre violencia cibernética mediante diferentes plataformas de redes sociales, con un total de 406 personas, entre ellos, 172 fueron hombres y 234 fueron mujeres. Se tiene consideración de las variables criterios como edad, nacionalidad y estado civil además de sexo. Se ha obtenido un rango de edad suficientemente grande, de 12 años a 60 años con 12 nacionalidades diferentes: china, española, venezolana, portuguesa, mexicana, británica, brasileña, estadounidense, francesa, ecuatoriana, italiana y coreana del sur; entre ellos, la mayoría de los participantes son de China y España. En cuanto al estado civil, se ha adquirido 293 solteros, 109 casados o con pareja, 3 divorciados y 1 viudo. No es difícil darse cuenta de que la mayoría de los participantes están en su época joven y adulta.

Se debe tener en cuenta también en dicho estudio que el objetivo es hacer comparaciones sobre diferentes técnicas de muestreo en vez de analizar las causas de sufrir o cometer violencia cibernética; por consiguiente, el planteamiento de pregunta ha sido sencillo: “¿Ha sufrido alguna vez la violencia cibernética ?” y “¿Ha cometido alguna vez violencia cibernética contra otras personas?” con respuestas dicotómicas, es decir, con sí o no. En este caso, se ha conseguido a partir de la primera pregunta 302 respuestas que no y 104 que sí. En la segunda pregunta había 369 de los participantes eligieron no y solo 37 de ellos eligieron sí.

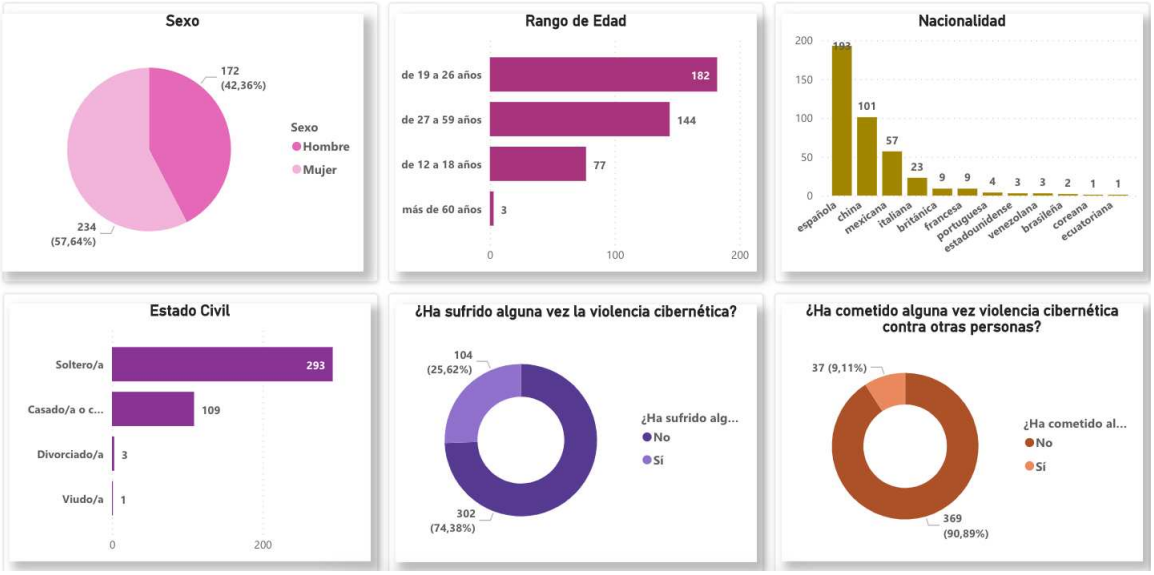

Figura 4.1: Visualización de datos

4.2 Limpieza de datos de la encuesta

Tras coleccionar todas las respuestas, debía de hacer limpieza y modificación a los datos debido a siguientes problemas:

- **Idioma**

Atendiendo a que los participantes fueran de diferentes países, era necesario preparar la encuesta en diferentes idiomas para evitar los problemas de malentendido de las preguntas, en dicho caso, había español y chino.

- **Letras mayúsculas y minúsculas**

En dicha encuesta, no todas las preguntas fueron dicotómicas, los participantes tenían que poner su nacionalidad con respuesta corta. Así que era obligatorio modificar las letras para que se permitía hacer cálculos en Excel.

- **Edad**

Además de nacionalidad, la variable “edad” también se hizo con respuesta corta y se podía considerar en que todas fueron adecuadas menos dos: “27 primaveras” y “52 años”, ya que solo se necesitaban los números.

Además se expresa esta variable por números mediante una pregunta corta, para facilitar los análisis posteriores, las separamos en 4 grupos dependiendo de adolescencia, juventud, adultez y persona mayor:

- **De 12 a 18 años:** 77 participantes
- **De 19 a 26 años:** 182 participantes
- **De 27 a 59 años:** 144 participantes
- **Más de 60 años:** 3 participantes

- **Formas de expresar nacionalidad**

Según las respuestas obtenidas, se han mostrado varias formas de poner la nacionalidad siendo pregunta corta, por ejemplo, sobre la nacionalidad española se han puesto en tres formas: “España”, “española” o “español”. A fin de que se mantuviera la misma forma, se utilizaba al final “española” tanto en nacionalidad española como en las demás.

5 | Metodología

Diseño del estudio

El objetivo principal de realizar este estudio es poder hacer una comparación entre diferentes técnicas de muestreo probabilístico y no probabilístico y predecir brevemente la influencia de violencia cibernética a los seres humanos.
En cuanto al fin de este estudio, se plantearán ejercicios adecuados de las diferentes técnicas del muestreo para poder obtener los porcentajes y sus intervalos.

Sujetos de estudio

La población se puede definir como el conjunto de los individuos que han realizado la encuesta, en este caso serían las personas que decidieron realizarla tras haber dado su consentimiento, siendo esta nuestra población accesible.

Tamaño de la Muestra

Para determinar el tamaño máximo de la muestra, se utilizará la siguiente fórmula:

$$n_{\infty} = \frac{k^2 p_A q_A}{e^2}$$

Siendo:

- k=cuantil de una distribución de probabilidad. Si se toma un valor de confianza $\gamma = 95\,\%$, así que el valor de k=1,96.
- $p_A = q_A = 0,5$, se considera el caso más desfavorable, que hace mayor el tamaño de la muestra y se prevé el error que se va a cometer.
- se toma un error de $5\,\% = 0,05$

Tras hacer el cálculo mediante esta fórmula, se ha logrado saber que el tamaño de la muestra tiene que ser de 400 sujetos.
Depediendo de las características de las técnicas habrá que estimar de nuevo el tamaño de la muestra con formas distintas.

Variables

Conforme a la encuesta, las variables son cualitativas nominales, como por ejemplo estado civil, y edad. Hay otras a las que llamamos nominales dicotómicas en las que solo hay dos categorías con sí o no. Seguidamente, se describen las variables utilizadas en dicho estudio:

- Sexo: con selección entre hombre y mujer.
- Edad: con respuestas cortas donde pone las edades de participantes en forma numérica. Se distribuirán en cuatro grupos: 12-18 años, 19-26 años, 27-59 años y más de 60 años.
- Nacionalidad: con respuestas cortas que se incluyen los países europeos, asiáticos y americanos.
- Estado civil: con selección entre soltero/a, casado/a o con pareja, divorciado/a y viudo/a.
- Dos preguntas dicotómicas constan de selecciones de sí o no sobre preguntas de violencia cibernética. Pregunta 1: “¿Ha sufrido alguna vez la violencia cibernética?” Pregunta 2: “¿Ha cometido alguna vez violencia cibernética contra otras personas?”

Todas estas variables se han recogido por medio de la encuesta realizada desde fin de junio hasta fin de septiembre enviadas por los redes sociales.

Variables	Tipos de variables	Categorías
Sexo	cualitativa dicotómica	hombre, mujer
Edad	respuesta corta	forma numérica
Nacionalidad	respuesta corta	nombre de nacionalidad
Estado Civil	cualitativa no dicotómica	soltero/a, casado/a con pareja, divorciado/a, viudo/a
Pregunta 1	cualitativa dicotómica	sí, no
Pregunta 2	cualitativa dicotómica	sí, no

Cuadro 5.1: Resumen de Variables

Recogida de datos

La recogida de datos se realizará mediante un formulario sencillo que preparé con el consentimiento informado y su posterior cumplimentación. Se realizó entre estudiantes universitarios de la Universidad de Salamanca y si es posible, se lo pasarían a los miembros de sus familias o sus amigos.
El formulario consta de 6 preguntas, por lo que su realización se hizo de forma simple y rápida. Para la realización del estudio se necesitaba únicamente el consentimiento informado y el formulario.

Fases del Estudio

-Fase 1: Revisión Bibliográfica
En esta fase, se llevaba a cabo una búsqueda de bibliografía para obtener todas las informaciones necesarias.

-Fase 2:Captación y Recogida de datos

Se recogieron los datos mediante la encuesta propuesta desde el 25 de julio hasta el 29 de septiembre en 2021 con una duración de tres meses, así como de conocer a la población con la que queremos trabajar.

-Fase 3: Análisis de los Datos
Una vez se obtuvieron los datos, se hizo modificaciones y limpieza para que se quedara en una forma adecuada y que puediera llevarse a cabo su posterior análisis y cálculos.

-Fase 4: Presentación de resultados
Una vez que ya se han terminado de modificar los datos, se podía comenzar con los cálculos de las técnicas de muestreo, posteriormente analizar los resultados de cada técnica y hacer la conclusión.

Análisis de Datos

A fin de realizar el análisis de estos datos obtenidos por la encuesta, se utiliza primero PowerBI, que es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial en el que se permite trabajar con orígenes muy complejos de datos con lenguaje de programación y otras opciones que están más cargadas por expertos de tecnologías de la información y que cualquier usuario con conocimientos medios o avanzados de Excel sea capaz de manejar la herramienta importar datos de Excel o crear informes. En dicho trabajo, se utiliza solamente para la visualización de datos.

Por otro lado, la herramienta principal que se utiliza en este trabajo es Excel que es una hoja de cálculo de Microsoft cuenta con cálculo, gráficos, tablas y un lenguaje de programación macro llamado "Visual Basic"para aplicaciones. En este caso, se realizan los cálculos mediante las fórmulas establecidas para poder comparar los resultados y asegurar de que no hay problemas de cálculos. Para conocer mejor los datos, se obtendrán parámetros como la media, el total, la varianza o las proporciones a partir de los datos adquiridos con anterioridad. Para así saber, por ejemplo, cuál es el porcentaje de las personas encuestadas que han sufrido alguna vez la violencia cibernética en España.

Al final, una vez obtenidos los parámetros y resultados anteriores, se procederá a la comparación entre diferentes técnicas de muestreo.

Funciones utilizadas de EXCEL

- ALEATORIO.ENTRE: función matemática y trigonométrica que devuelve un número mayor o igual a límite inferior y menor o igual a límite superior. =ALEATORIO.ENTRE(límite inferior, límite superior).

- BUSCARV: cuando se necesita encontrar elementos en una tabla o un rango por fila. =BUSCARV(Lo que desea buscar; dónde quiere buscarlo; el número de columna en el rango que contiene el valor a devolver; devuelve una coincidencia exacta/aproximada, 1/TRUE o 0/FALSE), donde una coincidencia exacta=0 y una aproximada=1.

- CONTAR.SI: función estadística para contar el número de celdas que cumple un criterio.=CONTAR.SI(celdas que quiere realizar la búsqueda; criterio)

Limitaciones del estudio

- En este estudio se ha recogido respuestas de personas de 12 países diferentes; sin embargo, existe una gran diferencia de datos entre ellos: la mayoría de la gente es de China, España o México en la que se ocupa aproximadamete un $86\,\%$ del total. El resto de los 9 paises ocupa el $14\,\%$; es decir, que hay una o dos personas.

6 | Resultado

El objetivo principal de dicho estudio es estimar proporciones mediante diferentes técnicas de muestreo y realizar una comparación entre ellas. Para todas de ellas se considera la población todos los individuos encuestados con un error del $5\,\%$.

6.1 Muestreo Aleatorio Simple

Se desea saber la proporción de los individuos que han sufrido alguna vez violencia cibernética sin considerarse el resto de las variables. En este caso, se sabe que $N = 406$ y $e = 5\,\%$.
Primero, se ha de estimar el tamaño muestral y se utiliza el caso más desfavorable, es decir, $p = 0,5$ y $q = 1 - p = 1 - 0,5 = 0,5$.

$$n_\infty = \frac{k^2 pq}{e^2} = \frac{2^2 \cdot 0,5 \cdot 0,5}{5\,\%^2} = 400$$

$$n = \frac{n_\infty}{1 + \frac{n_\infty + 1}{N}} = \frac{400}{1 + \frac{400+1}{406}} = 201,239$$

Se redondea n por exceso, para que no aumente el error; por tanto se ha obtenido el tamaño muestral igual que 202 individuos.
Con el objeto de generar individuos aleatoriamente, se utilizan dos funciones en Excel: Primero, la función =ALEATORIO.ENTRE devuelve un numero aleatorio entero que se encontrará entre el límite inferior y el superior especificados. En nuestro caso, el inferior=1 y el superior=406, así con esta función se generan 202 individuos aleatoriamente.

Después de conocer los individuos que se va a escoger, según el objetivo de esta técnica hemos de saber las respuestras correspondientes. Se utiliza la función BUSCARV con la intención de saber cuáles son las respuestas que coinciden con los individuos aleatorios. Esta función es una de las de búsqueda y referencia, cuando se quiere buscar elementos en una tabla o un rango por fila; =BUSCARV(lo que desea buscar; dónde quiere buscarlo; número de columna en el rango que contiene el valor a devolver; devuelve una coincidenncia exacta/aproximada, 1=TRUE o 0=FALSE). Según nuestros datos, se escribe con la siguiente forma:

=BUSCARV(individuos aleatorios obtenidos; numeración de individuos de 1 a 406; la sexta columna:¿ha sufrido alguna vez violencia cibernética; FALSE)

Desde esta función, se han obtenido las respuestas correspondientes con las que vamos a trabajar.

1	216	31	232	61	118	91	81	121	14	151	285	181	354
2	152	32	268	62	232	92	232	122	61	152	331	182	251
3	330	33	230	63	287	93	226	123	350	153	249	183	342
4	49	34	226	64	302	94	245	124	135	154	79	184	92
5	258	35	30	65	168	95	362	125	346	155	213	185	279
6	154	36	147	66	214	96	252	126	60	156	333	186	125
7	397	37	256	67	195	97	117	127	342	157	276	187	331
8	102	38	198	68	329	98	124	128	24	158	224	188	2
9	259	39	170	69	114	99	115	129	287	159	331	189	191
10	364	40	238	70	104	100	40	130	364	160	267	190	46
11	315	41	124	71	226	101	105	131	309	161	183	191	181
12	38	42	168	72	34	102	47	132	307	162	89	192	242
13	216	43	324	73	327	103	8	133	171	163	351	193	290
14	246	44	390	74	130	104	310	134	152	164	389	194	39
15	46	45	12	75	69	105	34	135	347	165	22	195	119
16	380	46	347	76	28	106	101	136	278	166	222	196	110
17	298	47	153	77	335	107	276	137	24	167	291	197	77
18	177	48	313	78	374	108	255	138	65	168	73	198	331
19	49	49	363	79	329	109	154	139	355	169	137	199	230
20	18	50	297	80	260	110	330	140	276	170	173	200	149
21	187	51	226	81	286	111	144	141	375	171	79	201	95
22	321	52	241	82	398	112	17	142	348	172	179	202	55
23	55	53	389	83	333	113	244	143	263	173	381		
24	199	54	38	84	155	114	169	144	205	174	405		
25	106	55	210	85	258	115	37	145	221	175	149		
26	403	56	366	86	270	116	141	146	395	176	77		
27	305	57	193	87	72	117	39	147	113	177	319		
28	348	58	314	88	153	118	57	148	61	178	280		
29	203	59	235	89	195	119	387	149	302	179	127		
30	251	60	164	90	29	120	395	150	306	180	213		

Figura 6.1: individuos aleatorios

En estos 202 individuos, hay 48 de ellos que han contestado que sí. Ahora se han obtenido los datos básicos para llevarse a cabo la estimación:

n=202, N=406, x=48

$$\hat{p} = \frac{x}{n} = \frac{48}{202} \approx 0{,}2376 = 23{,}76\,\%$$

q=1-p=1-0.2376=0.7624

$\gamma = 95{,}45\,\%$ es decir se considera k=2

Cálculo de varianza estimada por proporción:

$$V\hat{a}r(\hat{p}) = (1 - \frac{n}{N})\frac{\hat{p}\hat{q}}{n-1} = (1 - \frac{202}{406})\frac{0{,}2376 \cdot 0{,}7624}{202-1} \approx 4{,}5283 \times 10^{-4}$$

Cálculo de error:

$$e = k\sqrt{V\hat{a}r(\hat{p})} = 2\sqrt{4{,}5283 \times 10^{-4}} \approx 0{,}0426$$

Cálculo de intervalo de proporción:

$$I_p = \hat{p} \pm e = 0{,}2376 \pm 0{,}0426 = (0{,}195, 0{,}2802) = (19{,}5\,\%, 28{,}02\,\%)$$

Se concluye que dentro de estos 406 individuos encuestados habría de $19{,}5\,\%$ a $28{,}02\,\%$ de ellos que ha sufrido alguna vez la violencia cibernética.

6.2 Muestreo Estratificado

En cuanto a muestreo estratificado, ha de considerarse primero el coste y el error, en este caso será el error fijo y coste mínimo, pero aquí no se habla del coste por ser constante, ya

que se han conseguido los datos mediante la encuesta. Se calcula el tamaño que debe tener una muestra estratificada para la población de 406 individuos con afijación óptima con el caso más desfavorable: p=q=0.5.
Se quiere estudiar en tres estratos dependiendo de los estados: Asia(China y Corea del Sur), Europa(España, Portugal, Inglaterra, Francia e Italia) y América(Venezuela, México, Brasil, Los Estados Unidos y Ecuador). Así que se han obtenido los datos básicos:

$$N_1(\text{Asia})=102,\ N_2(\text{Europa})=238,\ N_3(\text{Amércia})=66,\ \text{N}=406$$

Se estiman las afijaciones óptimas mediante la siguiente fórmula considerándose el caso más desfavorable y coste constante:

$$w_i = \frac{N_i s_{ci}}{\sum_{i=1}^{L} N_i s_{ci}} = \frac{N i \sqrt{p_i q_i}}{\sum_{i=1}^{L} N_i \sqrt{p_i q_i}}$$

$$w_1 = \frac{102\sqrt{0{,}5\cdot 0{,}5}}{102\sqrt{0{,}5\cdot 0{,}5}+238\sqrt{0{,}5\cdot 0{,}5}+66\sqrt{0{,}5\cdot 0{,}5}} = 0{,}2512$$

$$w_2 = \frac{238\sqrt{0{,}5\cdot 0{,}5}}{102\sqrt{0{,}5\cdot 0{,}5}+238\sqrt{0{,}5\cdot 0{,}5}+66\sqrt{0{,}5\cdot 0{,}5}} = 0{,}5862$$

$$w_3 = 1 - w_1 - w_2 = 1 - 0{,}2512 - 0{,}5862 = 0{,}1626$$

En cuanto al tamaño de muestra, se calcula con p=q=0.5 y el error=5 %:

$$n = \frac{\frac{\sum_{i=1}^{L} N_i^2 p_i q_i}{w_i}}{\frac{N^2 e^2}{k^2} + \sum_{i=1}^{L} N_i p_i q_i}$$

$$n = \frac{(102^2\cdot 0{,}5\cdot 0{,}5)/0{,}2512+(238^2\cdot 0{,}5\cdot 0{,}5)/0{,}5862+(66^2\cdot 0{,}5\cdot 0{,}5)/0{,}1626}{\frac{406^2\cdot 0{,}05^2}{2^2}+102\cdot 0{,}5\cdot 0{,}5+238\cdot 0{,}5\cdot 0{,}5+66\cdot 0{,}5\cdot 0{,}5} = 201{,}4814 \text{(sin redondear)}$$

Con $n_i = nw_i$ ya se puede calcular el tamaño en cada estrato y nos permite encontrar un resultado óptimo.

$$n_1 = 201{,}4814 \cdot 0{,}2512 = 50{,}6121$$

$$n_2 = 201{,}4814 \cdot 0{,}5862 = 118{,}1084$$

$$n_3 = 201{,}4814 \cdot 0{,}1626 = 32{,}7609$$

Ahora hay que redondear n_i por defecto para encontrar una solución óptima, desde este punto de vista, ha de calcular el error para que no supere al 5 %. A fin de obtener el error, calculamos primero la varianza:

Caso 1: $n_1 = 50$, $n_2 = 118$, $n_3 = 32$, $n = 200$

$$Var(\hat{p}) = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \frac{N_i - n_i}{N_i - 1} \frac{p_i q_i}{n_i}$$

$$Var(\hat{p}) = \frac{1}{406^2}\left(102^2 \frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2 \frac{238-118}{238-1}\frac{0{,}5^2}{118} + 66^2 \frac{66-32}{66-1}\frac{0{,}5^2}{32}\right) = 6{,}3910 \times 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}3910 \times 10^{-4}} = 0{,}0506$$

Caso 2: $n_1 = 51$, $n_2 = 118$, $n_3 = 32$, $n = 201$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-51}{102-1}\frac{0{,}5^2}{51} + 238^2\frac{238-118}{238-1}\frac{0{,}5^2}{118} + 66^2\frac{66-32}{66-1}\frac{0{,}5^2}{32}) = 6{,}3286 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}3286 \cdot 10^{-4}} = 0{,}0503$$

Caso 3: $n_1 = 50$, $n_2 = 119$, $n_3 = 32$, $n = 201$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2\frac{238-119}{238-1}\frac{0{,}5^2}{119} + 66^2\frac{66-32}{66-1}\frac{0{,}5^2}{32}) = 6{,}3296 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}3296 \cdot 10^{-4}} = 0{,}0503$$

Caso 4: $n_1 = 50$, $n_2 = 118$, $n_3 = 33$, $n = 201$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2\frac{238-118}{238-1}\frac{0{,}5^2}{118} + 66^2\frac{66-33}{66-1}\frac{0{,}5^2}{33}) = 6{,}3275 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}3275 \cdot 10^{-4}} = 0{,}0503$$

Caso 5: $n_1 = 52$, $n_2 = 118$, $n_3 = 32$, $n = 202$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-52}{102-1}\frac{0{,}5^2}{52} + 238^2\frac{238-118}{238-1}\frac{0{,}5^2}{118} + 66^2\frac{66-32}{66-1}\frac{0{,}5^2}{32}) = 6{,}2685 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2685 \cdot 10^{-4}} = 0{,}0501$$

Caso 6: $n_1 = 51$, $n_2 = 119$, $n_3 = 32$, $n = 202$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-51}{102-1}\frac{0{,}5^2}{51} + 238^2\frac{238-119}{238-1}\frac{0{,}5^2}{119} + 66^2\frac{66-32}{66-1}\frac{0{,}5^2}{32}) = 6{,}2671 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2671 \cdot 10^{-4}} = 0{,}0501$$

Caso 7: $n_1 = 50$, $n_2 = 120$, $n_3 = 32$, $n = 202$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2\frac{238-120}{238-1}\frac{0{,}5^2}{120} + 66^2\frac{66-32}{66-1}\frac{0{,}5^2}{32}) = 6{,}2692 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2692 \cdot 10^{-4}} = 0{,}0501$$

Caso 8: $n_1 = 50$, $n_2 = 119$, $n_3 = 33$, $n = 202$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2\frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2\frac{238-119}{238-1}\frac{0{,}5^2}{119} + 66^2\frac{66-33}{66-1}\frac{0{,}5^2}{33}) = 6{,}2661 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2661 \cdot 10^{-4}} = 0{,}0501$$

Caso 9: $n_1 = 50$, $n_2 = 118$, $n_3 = 34$, $n = 202$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2 \frac{102-50}{102-1}\frac{0{,}5^2}{50} + 238^2 \frac{238-118}{238-1}\frac{0{,}5^2}{118} + 66^2 \frac{66-34}{66-1}\frac{0{,}5^2}{34}) = 6{,}2677 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2677 \cdot 10^{-4}} = 0{,}0501$$

Caso 10: $n_1 = 51$**,** $n_2 = 119$**,** $n_3 = 33$**,** $n = 203$

$$Var(\hat{p}) = \frac{1}{406^2}(102^2 \frac{102-51}{102-1}\frac{0{,}5^2}{51} + 238^2 \frac{238-119}{238-1}\frac{0{,}5^2}{119} + 66^2 \frac{66-33}{66-1}\frac{0{,}5^2}{33}) = 6{,}2036 \cdot 10^{-4}$$

$$e = k\sqrt{Var(\hat{p})} = 2\sqrt{6{,}2677 \cdot 10^{-4}} = 0{,}0498 \leq 0{,}05$$

Se concluye que, con un tamaño muestral de 203 individuos y 51, 119 y 33 tamaños muestrales en cada estrato, puede sacar la conclusión óptima con un error menor o igual que 0.05.
A continuación, aplicamos las mismas funciones utilizadas en Aleatorio Simple para encontrar cuántos individuos que han sufrido alguna vez la violencia cibernética con los tamaños muestrales estimados. Se ha conseguido los datos como en la tabla (figura 6.2):

Estrato	Ni	ni	Sí
Asia	102	51	25
Europa	238	119	33
América	66	33	7

Figura 6.2: Datos obtenidos

Posteriormente se realizan las estimaciones de proporciones necesarias.
Proporción en cada estrato y total:

$$\hat{p}_i = \frac{x_i}{n_i}$$

$$\hat{p_1} = \frac{25}{51} = 0{,}4902,\ \hat{p_2} = \frac{33}{119} = 0{,}2773,\ \hat{p_3} = \frac{7}{33} = 0{,}2121$$

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{p}_i = \frac{1}{406}(102 \cdot 0{,}4902 + 238 \cdot 0{,}2773 + 66 \cdot 0{,}2121) = 0{,}3202$$

Para estimar las varianzas también hay que considerarlas en cada estrato y el total:

$$\hat{Var}(\hat{p}_i) = (1 - \frac{n_i}{N_i})\frac{\hat{p}_i\hat{q}_i}{n_i - 1}$$

$$\hat{Var}(\hat{p_1}) = (1 - \frac{51}{102})\frac{0{,}4902(1-0{,}4902)}{51-1} = 2{,}499 \cdot 10^{-3}$$

$$\hat{Var}(\hat{p_2}) = (1 - \frac{119}{238})\frac{0{,}2773(1-0{,}2773)}{119-1} = 8{,}4917 \cdot 10^{-4}$$

$$\hat{Var}(\hat{p_3}) = (1 - \frac{33}{66})\frac{0{,}2121(1-0{,}2121)}{33-1} = 2{,}6111 \cdot 10^{-3}$$

Entonces, la varianza total de la proporción será:

$$\hat{Var}(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(1 - \frac{n_i}{N_i})\frac{\hat{p}_i\hat{q}_i}{n_i - 1}$$

$$\hat{Var}(\hat{p}) = \frac{1}{406^2}(102^2 \cdot 2{,}499 \cdot 10^{-3} + 238^2 \cdot 8{,}4917 \cdot 10^{-4} + 66^2 \cdot 2{,}6111 \cdot 10^{-3}) = 5{,}1854 \cdot 10^{-4}$$

Al final de todo, su error e intervalo correpondiente serán:

$$e = k\sqrt{V\hat{a}r(\hat{p})} = 2\sqrt{5{,}1854 \times 10^{-4}} = 0{,}0455$$

$$I_p = 0{,}3202 \pm 0{,}0455 = (0{,}2747, 0{,}3657) = (27{,}47\,\%, 36{,}57\,\%)$$

En conclusión, después de aplicar el muestreo estratificado entre un $27{,}47\,\%$ y un $36{,}57\,\%$, de los individuos encuestados han sufrido alguna vez la violencia cibernética.

6.3 Muestreo por Conglomerados en una etapa

Se sabe que la población total es 406 individuos con 12 países distintos basándose en los datos obtenidos. En este apartado vamos a utilizar muestreo por conglomerados en una etapa, de ser así, se consideran los 12 países distintos como número de conglomerados en la población y los 406 individuos como número total de individuos en la población. De esta forma, se tiene que N=12 y M=406.

Igual que las técnicas anteriores, habrá que estimar el número de conglomerados en la muestra, n, con un $5\,\%$ de error y el caso más desfavorable. A diferencia de los casos que preceden, hemos de calcular el número medio de individuos en la población $\bar{M}$:

$$\bar{M} = \frac{M}{N} = \frac{406}{12} = 33{,}8333$$

Ahora se estima n mediante n_∞:

$$n_\infty = \frac{k^2 pq}{\bar{M}e^2} = \frac{2^2 \cdot 0{,}5 \cdot 0{,}5}{33{,}8333 \cdot 0{,}05^2} = 11{,}8227$$

$$n = \frac{n_\infty}{1+\frac{n_\infty - 1}{N}} = \frac{11{,}8227}{1+\frac{11{,}8227-1}{12}} = 6{,}2163$$

Se redondea por exceso, por tanto n=7

Desde aquí, se sabe que se tiene que escoger 7 conglomerados en 12 de ellos. Por consiguiente, utilizamos las funciones de muestreo aleatorio simple: ALEATORIO.ENTRE y buscamos los valores correspondientes posteriormente.

Para que no se ocurran errores innecesarios, se numeran las nacionalidades de 1 a 12. Los 7 conglomerados obtenidos serán: 7=brasileña, 2=española, 6=británica, 9=francesa, 1=china, 11=italiana y 5=mexicana. A partir de estos datos ya nos permite realizar las estimaciones. Por tener el objetivo de estimar proporciones, se considera a_i el total de individuos que cumple una condición donde en nuestro caso serán los que han sufrido alguna vez violencia cibernética.

Ahora se ha obtenido los datos necesarios para llevarse a cado las estimaciones.

$$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i} = \frac{2+44+2+...+12}{2+193+9+...=57} = \frac{101}{394} = 0{,}2563 = 25{,}63\,\%$$

En estos siete conglomerados hay un $25{,}63\,\%$ de ellos han sufrido alguna vez la violencia cibernética. Ahora se calcula el número medio de individuos en conglomerados de la muestra.

Nacionalidad	M_i	a_i
brasileña	2	2
española	193	44
británica	9	2
francesa	9	0
china	101	35
italiana	23	6
mexicana	57	12

Cuadro 6.1: M_i y a_i de 7 conglomerados

$$\hat{\bar{M}} = \bar{m} = \frac{\sum_{i=1}^{n} M_i}{n} = \frac{394}{7} = 56{,}2857$$

En cuanto a la estimación de varianza, hemos de calcular primero s_a^2.

$$s_a^2 = \frac{\sum_{i=1}^{n} (a_i - \hat{p} M_i)^2}{n-1} = \frac{127{,}3813}{7-1} = 21{,}2302$$

Se sustituye estos valores obtenidos a la fórmula de calcular varianza de proporción.

$$V\hat{a}r(\hat{p}) = (1 - \frac{n}{N})\frac{1}{\bar{m}^2}\frac{s_a^2}{n} = (1 - \frac{7}{12})\frac{1}{56{,}2857^2}\frac{21{,}2302}{7} = 3{,}9889 \cdot 10^{-4}$$

Su error e intervalo correspondiente serán:

$$e = k\sqrt{V\hat{a}r(\hat{p})} = 2\sqrt{3{,}9889 \cdot 10^{-4}} = 0{,}0399$$

$$I_p = \hat{p} \pm e = 0{,}2563 \pm 0{,}0399 = (0{,}2164, 0{,}2962) = (21{,}64\,\%, 29{,}62\,\%)$$

En conclusión, aproximadamente de $21{,}64\,\%$ a $29{,}62\,\%$ de los individuos han sufrido alguna vez la violencia cibernética con un error de 0.0399.

6.4 Muestreo por Conglomerados en dos etapas

Aprovechamos los resultados de muestreo por conglomerados en una etapa y utilizamos el número de conglomerados en la muestra n=7 siendo una muestra piloto. En cada uno de ellos, planteamos que se elige una muestra del $40\,\%$ de los individuos. A continuación, se utiliza la función =ALEATORIO.ENTRE para obtener la segunda etapa es decir los individuos que han sufrido alguna vez la violencia cibernética. Los resultados son los siguientes:

Es decir, se tiene $N = 12$, $n = 7$, $M = 406$ y se puede estimar el tamaño del conglomerado promedio: $\bar{M} = \frac{M}{N} = \frac{406}{12} = 33{,}8333$
Ahora calculamos las estimaciones necesarias: $\hat{p}_i$ con su fórmula $\frac{X_i}{m_i}$, su $\hat{q}_i = 1 - \hat{p}_i$ correspondiente, y $\hat{p}_i\hat{q}_i$. Los resultados obtenidos serán los siguientes:

Tras obtener las proporciones de cada conglomerado, ha de estimar la proporción total $\hat{p}$ para el cálculo del tamaño de muestra.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \hat{p}_i$$

Nacionalidad	M_i	m_i	X_i
brasileña	2	2	1
española	193	77	30
británica	9	4	1
francesa	9	4	0
china	101	40	32
italiana	23	9	4
mexicana	57	23	8

Cuadro 6.2: M_i, m_i y X_i de 7 conglomerados

Nacionalidad	$\hat{p}_i$	$\hat{q}_i$	$\hat{p}_i\hat{q}_i$
brasileña	0.5	0.5	0.25
española	0.3896	0.6104	0.2409
británica	0.25	0.75	0.1875
francesa	0	1	0
china	0.8	0.2	0.16
italiana	0.4444	0.5556	0.2469
mexicana	0.3478	0.6522	0.2268
SUMA	2.732	4.268	1.3091

Cuadro 6.3: Estimaciones necesarias

Por lo tanto:

$$\hat{p} = \frac{1}{7} \cdot 2{,}732 = 0{,}3903 = 39{,}03\,\%$$

de las personas que han sufrido alguna vez la violencia cibernética.

En cuanto al cálculo del tamaño de la muestra, primero hemos de calcular las cuasivarianzas $\hat{s}^2_{c1}$ y $\hat{s}^2_{cw}$ y su mejor ajuste. A fin de calcular $\hat{s}^2_{c1}$ se tiene que conocer primero $(\hat{p}_i - \hat{p})^2$ y su sumatorio.

Nacionalidad	$(\hat{p}_i - \hat{p})^2$
brasileña	0.012
española	$4.756{\cdot}10^{-7}$
británica	0.0197
francesa	0.1523
china	0.1679
italiana	0.003
mexicana	0.0018
SUMA	0.3566

Cuadro 6.4: $(\hat{p}_i - \hat{p})^2$ y su sumatorio

Entonces, se tiene que:

$$\hat{s}^2_{c1} = s^2_1 = \frac{\sum_{i=1}^{n}(\hat{p}_i - \hat{p})^2}{n-1} = \frac{0{,}3566}{7-1} = 0{,}0594$$

$$\hat{s}^2_{cw} = s^2_w = \frac{\sum_{i=1}^{n}\hat{p}_i\hat{q}_i}{n} = \frac{1{,}3091}{7} = 0{,}187$$

Para calcular su mejor ajuste ha de saber que el tamaño de la muestra promedio
$\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} = \frac{159}{7} = 22{,}7143$
Así tenemos su mejor ajuste:

$$\hat{s}_{c1}^2 = s_1^2 - \frac{s_w^2}{\bar{m}} = 0{,}0594 - \frac{0{,}187}{22{,}7143} = 0{,}0512$$

A continuación, se tiene que estimar $\bar{m}$ para fracciones de muestreo no despreciable considerádose el error fijo y el coste mínimo para llevarse a cabo despejar n posteriormente. Se supone que el coste de extraer un conglomerado c_1 será mil euros más grande que el de extraer un elemento del conglomerado c_2; en este caso, se elige el dicho coste para mantener un error mínimo, por lo tanto:

$$\bar{m} = \sqrt{\frac{c_1}{c_2}\frac{1}{\frac{s_{c1}^2}{s_{cw}^2} - \frac{1}{\bar{M}}}} = \sqrt{\frac{1000}{1}\frac{1}{\frac{0{,}0512}{0{,}187} - \frac{1}{33{,}8333}}} = \sqrt{4094{,}3316} = 63{,}987$$

Se sabe que el error es fijo, es decir se considerar un error de $5\,\%$ y $e = k\sqrt{Var(\hat{p})}$, nos permite despejar n mediante las siguientes fórmulas:

$$Var(\hat{p}) = (\frac{e}{k})^2$$

$$Var(\hat{p}) = \frac{N-n}{N-1}\frac{s_{c1}^2}{n} + \frac{\bar{M}-\bar{m}}{\bar{M}-1}\frac{s_{cw}^2}{n\bar{m}}$$

Entonces:

$$(\frac{0{,}05}{2})^2 = \frac{12-n}{12-1}\frac{0{,}0512}{n} + \frac{33{,}8333-63{,}987}{33{,}8333-1}\frac{0{,}187}{63{,}987n}$$

$$6{,}25 \cdot 10^{-4} = \frac{12-n}{11}\frac{0{,}0512}{n} - \frac{2{,}684 \cdot 10^{-3}}{n}$$

$$6{,}25 \cdot 10^{-4} = \frac{0{,}6144-0{,}0512n}{11n} - \frac{2{,}684 \cdot 10^{-3}}{n}$$

$$6{,}25 \cdot 10^{-4} = \frac{0{,}5849-0{,}0512n}{11n}$$

$$6{,}25 \cdot 10^{-4} \cdot 11n = 0{,}5849 - 0{,}0512n$$

$$0{,}0581n = 0{,}5849$$

$$n = 10{,}0671$$

Se redondea por defecto para no pasarnos del coste, se tiene que n=10. En conclusión, se puede estimar su intervalo de proporción con un error de 0.05:

$$I_p = (0{,}3903 \pm 0{,}05) = (0{,}3403,\ 0{,}4403) = (34{,}03\,\%,\ 44{,}03\,\%)$$

Se concluye que si extraemos 10 coglomerados en la muestra con 1000 euros de extraer un conglomerado y un error de 0.05, entre un $34{,}03\,\%$ y un $44{,}03\,\%$ de los individuos encuestados han sufrido alguna vez la violencia cibernética.

6.5 Muestreo por Cuotas

Se sabe que el muestreo por cuotas tiene una característica similar a la de muestreo estratificado, salvo que se determinan los porcentajes de cada estrato de antemano. Por esta razón, se utilizan los mismos datos de muestreo estratificado.
Se supone que se desea estimar proporciones de si han sufrido alguna vez la violencia cibernética con respecto a los individuos encuestados. Se tiene una población de 406 personas dividiéndose en 3 cuotas: Asia, Europa y América con 102, 238 y 66 personas respectivamente. Según las características de muestreo por cuotas, es un muestreo no aleatorio, por consiguiente, los individuos no se tienen una forma fija de ser seleccionados. Nos permite calcular los porcentajes que vamos a considerar en cada cuota.

$$\text{Asia: } 102/406 = 0{,}2512 = 25{,}12\,\% \simeq 25\,\%$$

$$\text{Europa: } 238/406 = 0{,}5862 = 58{,}62\,\% \simeq 59\,\%$$

$$\text{América: } 66/406 = 0{,}1626 = 16{,}26\,\% \simeq 16\,\%$$

Aprovechando los resultados de Muestreo Estratificado con un tamaño de muestra de 203 individuos. Falta ahora determinar la cuota por cada segmento, el cual debe ser proporcional al tamaño de la muestra y el porcentaje por cuota. Por consiguiente, la cuota de número de individuos queda así:

$$\text{Ásia: } 203 \cdot 25{,}12\,\% = 50{,}9936 \simeq 51$$

$$\text{Europa: } 203 \cdot 58{,}62\,\% = 118{,}9986 \simeq 119$$

$$\text{América: } 203 \cdot 16{,}26\,\% = 33{,}0078 \simeq 33$$

Note que la suma de las cuotas tiene que ser igual al tamaño de la muestra: 51+119+33=203. A continuación, estas cuotas se establecen con información basada en las estadísticas de la población, es decir se puede considerar que este método es aproximadamente probabilístico. Aprovechamos las estimaciones de muestreo estratificado para llevarse a cabo su estimación de varianza y la de error.

$$\hat{p_1} = 25\,\%,\ \hat{p_2} = 59\,\%,\ \hat{p_3} = 16\,\%$$

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{p}_i = \frac{1}{406}(102 \cdot 25\,\% + 238 \cdot 59\,\% + 66 \cdot 16\,\%) = 0{,}4347$$

Para estimar las varianzas también hay que considerarse las en cada cuota y el total:

$$\hat{Var}(\hat{p}_i) = (1 - \frac{n_i}{N_i})\frac{\hat{p}_i\hat{q}_i}{n_i - 1}$$

$$\hat{Var}(\hat{p_1}) = (1 - \frac{51}{102})\frac{0{,}25(1-0{,}25)}{51-1} = 1{,}875 \cdot 10^{-3}$$

$$\hat{Var}(\hat{p_2}) = (1 - \frac{119}{238})\frac{0{,}59(1-0{,}59)}{119-1} = 1{,}025 \cdot 10^{-3}$$

$$\hat{Var}(\hat{p_3}) = (1 - \frac{33}{66})\frac{0{,}16(1-0{,}16)}{33-1} = 2{,}1 \cdot 10^{-3}$$

Entonces la varianza total de proporción será:

$$V\hat{a}r(\hat{p}) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2(1-\frac{n_i}{N_i})\frac{\hat{p}_i\hat{q}_i}{n_i-1}$$

$$V\hat{a}r(\hat{p}) = \frac{1}{406^2}(102^2 \cdot 1{,}875 \cdot 10^{-3} + 238^2 \cdot 1{,}025 \cdot 10^{-3} + 66^2 \cdot 2{,}1 \cdot 10^{-3}) = 5{,}2607 \cdot 10^{-4}$$

Al final de todo, su error e intervalo correpondiente serán:

$$e = k\sqrt{V\hat{a}r(\hat{p})} = 2\sqrt{5{,}2607 \cdot 10^{-4}} = 0{,}0459$$

$$I_p = 0{,}4347 \pm 0{,}0459 = (0{,}3888, 0{,}4806) = (38{,}88\,\%, 48{,}06\,\%)$$

En conclusión, después de aplicar muestreo estratificado entre un $38{,}88\,\%$ y un $48{,}06\,\%$ de los individuos encuestados han sufrido alguna vez violencia cibernética.

7 | Conclusiones

Utilizamos las técnicas de muestreo para poder obtener información fiable de la población a partir de una muestra con un margen de error medido en términos de probabilidades. De esta forma, con el fin de estudiar bien el comportamiento y las opiniones de la población, sería mejor tener un error pequeño para que los resultados sean fiables y que tengan más validez. En dicho estudio se consideran los 406 individuos como la población. Teniendo en cuenta de que con los métodos como muestreo estratificado y por conglomerados era necesario dividirse la población en estratos o conglomerados. Ahora vemos los resultados en cada técnica y una conclusión general.

Muestreo Aleatorio Simple

Se ha considerado un error de 0,05 para la estimación del tamaño muestral y hemos escogido 202 individuos. El proceso de escoger los individuos fue completamente aleatorio y se ha obtenido aproximadamente entre un $19{,}5\,\%$ y un $28{,}02\,\%$ de los individuos encuestados han sufrido alguna vez violencia cibernética con un error de 0,0426. Se entiende este método con facilidad y con resultados proyectables, pero se puede decir que el resultado tiene menos precisión.

Muestreo Estratificado

En este método, se ha estimado el tamaño muestra mediante afijación óptima, con el fin de obtener los tamaños más precisos hemos de probar diferentes números redondeados hasta que se salieran tamaños con un error menor o igual que 0,05 que es lo que planteamos al principio. Al final, en los tres estratos tenemos 51, 119 y 33 individuos respectivamente, y el tamaño total de la muestra fue 203. A través de estos tamaños, se ha llevado a cabo el resto de las estimaciones, se ha dado un error de 0.0455 con aproximadamente entre un $27{,}47\,\%$ y un $36{,}57\,\%$ de individuos encuestados que han sufrido la violencia cibernética.

Muestreo por conglomerados en una etapa

En cuanto a este método, se cosideran los 12 países como número total de conglomerados de la población y hemos estimado el número de conglomerados en la muestra que fue 7 países, y se han seleccionado aleatoriamente estos siete. Se ha obtenido aproximadamente que entre un $21{,}64\,\%$ y un $29{,}62\,\%$ de los individuos han sufrido alguna vez la violencia cibernética con un error de 0,0399.

Muestreo por conglomerados en dos etapas

Se ha utilizado la muestra de muestreo por conglomerados en una etapa como muestra piloto para estimar un nuevo tamaño muestral. Sin embargo, con este método hemos de considerar un coste de extraer un conglomerado, cuando aumenta este coste, disminuye el número de conglomerados de la muestra. Desde este punto de vista, se observa que este método es muy costoso para aplicar en una investigación. En otras palabras, si se desea obtener mayor precisión de resultados de una investigación, mayor coste será. Al final se ha obtenido un intervalo de proporción $(34{,}03\,\%0{,}1cm44{,}03\,\%)$ con un error de 0,05.

Muestreo por cuotas

A diferencia del muestreo estratificado, es necesario definir los porcentajes de cada cuota antes de realizar las estimaciones. Se ha escogido las cuotas iguales que los estratos de muestreo estratificado y con los mismos tamaños muestrales. Posteriormente, se ha aplicado muestreo estratificado para calcular sus estimaciones en las que el error fue 0,0459 con un intervalo de proporciones $(38{,}88\,\%\ ,\ 48{,}06\,\%)$.

En conclusión, el error fue el mínimo en muestreo por conglomerados en una etapa, es decir se puede concluir que en dicho estudio hay aproximadamente de $21{,}64\,\%$ a $29{,}62\,\%$ de los individuos encuestados que han sufrido alguna vez la violencia cibernética. Con muestreo aleatorio simple, también se ha obtenido un margen de error pequeño, el procedimiento de este método es simple y eficaz, pero tiene un inconveniente, ya que no se puede utilizar si el universo es muy grande. Desde esta imagen, se podría concluir que es fiable para nuestro estudio con un universo pequeño. En cuanto a muestreo aleatorio estratificado, el resultado no ha llegado a uno expectante, la causa de esto puede ser que en cada estrato el número de individuos tendría demasiadas diferencias, aquí también es la limitación de dicho estudio. Por ejemplo, en Europa había 238 individuos pero en América había menos de 100.

Según la definición de los métodos de muestreo estadístico, el muestreo por conglomerados es el menos fiable. Esto se nota en muestreo por conglomerados en dos etapas. Se ha contado con una población pequeña, lo que podría ser la causa de no tener resultados esperados. Además, en el procedimiento del cálculo de tamaño de muestra, se han encontrado problemas de tener el tamaño muestral más grande que el poblacional, a no ser que se hayan planteado un coste de extraer un coglomerado. Y cuanto mayor coste se planteó, menor error se obtuvo. Por consiguiente, se concluye que este método es muy costoso y no es el más apropiado para este estudio. Sin embargo, con estos datos recogidos se ha conseguido el menor error en muestreo por conglomerados en una etapa. Comparado con los otros métodos, el resultado es más preciso por tener más grupos. Cuando se tiene una población pequeña, se puede concluir que el resultado es tan viable como confiable. Si se tuviera una población más grande y se realizaran varias fases de muestreo sucesivas, se obtendría mejor resultado.

Conclusión general

En cuanto a violencia cibernética, se ha obtenido una conclusión breve mediante el resultado de muestreo por conglomerados en una etapa. Se observa que en cada 10 personas habría un mínimo de dos que han sufrido violencia cibernética. Es decir, se puede estimar que en España con una población total de 47 615 034 habitantes, habría un mínimo de 9 523 006 personas que han sufrido una vez violencia digital en su vida, y en China 288 millones de ellos la han sufrido por lo menos una vez.

En el dicho estudio, solo se han incluido tres casos reales de violencia cibernética; sin embargo, en cualquier sitio del mundo estarán pasando casos similares incluso ahora mismo. Por consiguiente, no es difícil observar cuánta importancia habría que dar a este tema. Las autoridades tendrían que establecer más leyes estrictas para que no pasen más suicidios por violencia digital y restricciones de edad al utilizar ciertas plataformas.

Como dice Marcos McKinnon: “la tecnología y las redes sociales han traído el poder a la gente”; es decir con un gran poder viene una gran responsabilidad. Hoy en día vivimos en un mundo digital, la protección de los datos personales y el mantenimiento de un ambiente sano sería imprescindible si quisieramos evitar más tragedias.

Bibliografía

[1] Otzen, T., & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. International journal of morphology, 35(1), 227-232.

[2] Bencardino, C. M. (2012). Estadística y muestreo. Colombia: Ecoe ediciones.

[3] Betin Vasquez, E. P., & González Ramos, I. J. (2004). Estudio de la eficiencia de dos métodos de muestreo estadístico (MAS-MS), mediante simulación.

[4] Rojas, H. A. G. (2016). Estrategias de muestreo: diseño de encuestas y estimación de parámetros. Ediciones de la U.

[5] Chile, Contraloría General de la República. (2012). Guía práctica para la construcción de muestras.

[6] https://bookdown.org/aquintela/EBE/muestreo-aleatorio-simple.html

[7] https://bolivia.unfpa.org/sites/default/files/pub-pdf/1-legislacion-violenciadigital-roxanaperezdelcastillo-presentacion-dmp-2020.pdf

[8] https://ciberintocables.com/ciberacoso-codigo-penal/

[9] https://bullyingsinfronteras.blogspot.com/2016/11/estadisticas-de-acoso-escolar-o.html

[10] https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/MOCIBA-2020.pdf

[11] López, C. P. (2010). Técnicas de muestreo estadístico. Ibergarceta.

[12] Martínez, C. (2012). Estadística y muestreo-13ra Edición. Ecoe ediciones.

[13] Manyoma, P. C., Klinger, R. A. (2006). El uso del muestreo estadístico en la medición del trabajo. Scientia et technica, 12(32), 363-368.

[14] Silva, L. A. E. (2020). Análisis de la violencia cibernética en las preparatorias pertenecientes a la Universidad Autónoma de Nuevo León del municipio de Monterrey. Ciencia Latina Revista Científica Multidisciplinar, 4(2), 324-344.

[15] Hernández Mejía, R. (2018). La violencia cibernética (ciberbullying) en la universidad pública: El caso de la Universidad Autónoma del Estado de México.

[16] Velázquez, A. P. (2017). Tipos de muestreo. México: Centrogeo.

[17] Tadeo Noble, A. E., De Los Santos Posadas, H. M., Ángeles Pérez, G., Torres Pérez, J. A. (2014). Muestreo por conglomerados para manejo forestal en el Ejido Noh Bec, Quintana Roo. Revista mexicana de ciencias forestales, 5(25), 64-83.

ANEXO

	A	B	C	D	E	F	G
1	Marca temporal	Sexo	Edad	Nacionalidad	Estado Civil	¿Ha sufrido alguna vez la violencia cibernética?	¿Ha cometido alguna vez violencia cibernética contra otras personas?
2	6/25/2021 12:43:45	Mujer	22	China	Soltero/a	Sí	No
3	6/25/2021 12:47:44	Mujer	23	china	Soltero/a	Sí	No
4	6/25/2021 19:52:03	Hombre	21	España	Soltero/a	No	No
5	7/13/2021 13:27:57	Mujer	20	China	Soltero/a	Sí	No
6	7/13/2021 15:32:46	Mujer	16	España	Soltero/a	No	No
7	7/13/2021 15:33:20	Mujer	16	España	Soltero/a	Sí	No
8	7/13/2021 16:03:49	Hombre	28	España	Soltero/a	Sí	Sí
9	7/13/2021 16:23:24	Mujer	15	España	Soltero/a	No	No
10	7/13/2021 16:24:17	Mujer	43	España	Casado/a o con pareja	No	No
11	7/13/2021 16:25:43	Mujer	15	España	Soltero/a	No	No
12	7/13/2021 16:34:37	Mujer	25	españa	Soltero/a	No	No
378	9/29/2021 23:03:43	Mujer	26	italiana	Soltero/a	No	No
379	9/29/2021 23:04:17	Mujer	25	china	Casado/a o con pareja	Sí	No
380	9/29/2021 23:04:36	Hombre	28	española	Soltero/a	No	No
381	9/29/2021 23:04:58	Mujer	40	mexicana	Soltero/a	No	No
382	9/29/2021 23:05:17	Mujer	25	mexicana	Soltero/a	Sí	Sí
383	9/29/2021 23:05:32	Hombre	37	española	Casado/a o con pareja	No	No
384	9/29/2021 23:05:53	Hombre	20	española	Soltero/a	No	No
385	9/29/2021 23:06:05	Hombre	42	española	Soltero/a	No	No
386	9/29/2021 23:06:19	Hombre	16	china	Soltero/a	No	No
387	9/29/2021 23:06:34	Hombre	27	italiana	Soltero/a	Sí	No
388	9/29/2021 23:06:48	Hombre	28	española	Soltero/a	No	No
389	9/29/2021 23:07:03	Hombre	26	española	Soltero/a	No	No
390	9/29/2021 23:07:16	Mujer	20	china	Soltero/a	Sí	No
391	9/29/2021 23:07:29	Mujer	23	china	Soltero/a	No	No
392	9/29/2021 23:07:44	Mujer	16	española	Soltero/a	No	No
393	9/29/2021 23:07:57	Mujer	20	italiana	Soltero/a	No	No
394	9/29/2021 23:08:12	Hombre	24	española	Soltero/a	No	No
395	9/29/2021 23:08:25	Hombre	28	española	Soltero/a	No	No
396	9/29/2021 23:08:38	Hombre	23	china	Casado/a o con pareja	Sí	No
397	9/29/2021 23:14:06	Mujer	20	china	Soltero/a	No	No
398	6/25/2021 13:46:11	男	20	中国	单身	否	否
399	7/14/2021 11:52:23	女	21	中国	单身	是	否
400	7/14/2021 11:57:11	男	21	中国	单身	否	否
401	7/14/2021 12:07:31	女	21	中国	单身	是	否
402	7/14/2021 12:08:08	女	22	中国	单身	是	是
403	7/14/2021 12:37:28	男	29	中国	已婚或有男/女朋友	否	否
404	7/14/2021 12:56:18	女	20	中国	单身	否	是
405	7/14/2021 16:40:39	女	21	中国	单身	是	是
406	7/15/2021 20:15:30	男	27	中国	已婚或有男/女朋友	否	否
407	7/17/2021 14:32:17	女	23	中国	单身	否	否

Figura 7.1: Datos originales

	A	B	C	D	E	F
1	Sexo	Edad	Nacionalidad	Estado Civil	¿Ha sufrido alguna vez la violencia cibernética?	¿Ha cometido alguna vez violencia cibernética contra otras personas?
2	Mujer	22	china	Soltero/a	Sí	No
3	Mujer	23	china	Soltero/a	Sí	No
4	Hombre	21	española	Soltero/a	No	No
5	Mujer	20	china	Soltero/a	Sí	No
6	Mujer	16	española	Soltero/a	No	No
7	Mujer	16	española	Soltero/a	Sí	No
8	Hombre	28	española	Soltero/a	Sí	Sí
9	Mujer	15	española	Soltero/a	No	No
389	Hombre	26	española	Soltero/a	No	No
390	Mujer	20	china	Soltero/a	Sí	No
391	Mujer	23	china	Soltero/a	No	No
392	Mujer	16	española	Soltero/a	No	No
393	Mujer	20	italiana	Soltero/a	No	No
394	Hombre	24	española	Soltero/a	No	No
395	Hombre	28	española	Soltero/a	No	No
396	Hombre	23	china	Casado/a o con pareja	Sí	No
397	Mujer	20	china	Soltero/a	No	No
398	Hombre	20	china	Soltero/a	No	No
399	Mujer	21	china	Soltero/a	Sí	No
400	Hombre	21	china	Soltero/a	No	No
401	Mujer	21	china	Soltero/a	Sí	No
402	Mujer	22	china	Soltero/a	Sí	Sí
403	Hombre	29	china	Casado/a o con pareja	No	No
404	Mujer	20	china	Soltero/a	No	Sí
405	Mujer	21	china	Soltero/a	Sí	Sí
406	Hombre	27	china	Casado/a o con pareja	No	No
407	Mujer	23	china	Soltero/a	No	No

Figura 7.2: Datos modificados

A	B	C
	Nacionalidad	individuos
1	china	101
2	española	193
3	venezolana	3
4	portuguesa	4
5	mexicana	57
6	británica	9
7	brasileña	2
8	estadounidense	3
9	francesa	9
10	ecuatoriana	1
11	italiana	23
12	coreana (sur)	1

K	L	M
	numeración de las nacionalidades	número de individuos
1	7	2
2	2	193
3	6	9
4	9	9
5	1	101
6	11	23
7	5	57

=ALEATORIO.ENTRE(1;12)
=BUSCARV(L4;$A$1:$C$13;3;FALSO)

Figura 7.3: número de conglomerados en la muestra y el de individuos

Encuesta para TFG: Violencia cibernética en el sector social

CONSENTIMIENTO INFORMADO
A través de este cuestionario queremos conocer si se ha pasado alguna vez violencia cibernética o si se ha cometido violencia cibernética contra otras personas.

La encuesta está realizada en colaboración con la Universidad de Salamanca. El único fin de la encuesta es de carácter académico, investigación y/o publicación. La encuesta es TOTALMENTE ANÓNIMA.

Para que pueda recoger estos datos es necesario contar con su consentimiento informado. Este se dará por aceptado con la mera cumplimentación y posterior envío de la encuesta.

Se han adoptado las medidas necesarias para garantizar la completa confidencialidad de los datos personales de todos aquellos sujetos que participen en la experimentación en este estudio, de acuerdo con la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Podrán ejercerse los derechos de acceso, rectificación, suspensión, olvido y portabilidad de los datos, cuando procedan, a la dirección de correo electrónico.

*Obligatorio

Datos Personales

1. Sexo *

   Marca solo un óvalo.

   - ◯ Mujer
   - ◯ Hombre

2. Edad *

   _____

3. Nacionalidad *

   _____

4. Estado Civil *

Marca solo un óvalo.

- Soltero/a
- Casado/a o con pareja
- Divorciado/a
- Viudo/a

Preguntas Dicotómicas

La violencia cibernética (o violencia digital/ciberbullying) consiste en utilizar la tecnología para amenazar, avergonzar, intimidar o criticar a otra persona. Amenazas en línea, textos groseros, agresivos o despectivos enviados por Twitter, comentarios publicados en Internet o mensajes: todo cuenta. Y también cuenta el hecho de colgar en Internet información, fotografías o vídeos de carácter personal para herir o avergonzar a otra persona.

5. ¿Ha sufrido alguna vez la violencia cibernética? *

Marca solo un óvalo.

- Sí
- No

6. ¿Ha cometido alguna vez violencia cibernética contra otras personas? *

Marca solo un óvalo.

- Sí
- No

¡Muchas Gracias!

Ha completado satisfactoriamente el cuestionario de violencia cibernética.
¡Su colaboración servirá de gran ayuda!

毕业论文调查问卷：社会性网络暴力

通过这份调查问卷我们想知道您是否曾经经历过网络暴力，或者您是否对他人实施过网络暴力。

该调查是与萨拉曼卡大学合作进行的，调查的唯一目的是进行学术研究和/或出版。该调查完全匿名。

为了收集这些数据，我们有必要获得您的知情同意。只要完成并发送此问卷，将被视为同意将您的答案用于本次学术研究。

根据2021年1月1日《中华人民共和国民法典》（隐私权和个人信息保护章节），本次调查将保证所有参与本研究实验的受试者的个人信息完全保密。

在适当的时候数据的访问、更正、暂停、遗失和可移植的权利，可能会被运用到电子邮件地址。

*Obligatorio

个人信息

1. 性别 *

   Marca solo un óvalo.

   - ◯ 女
   - ◯ 男

2. 年龄 *

   _____

3. 国籍 *

   _____

4. 婚姻状况 *

Marca solo un óvalo.

- 单身
- 已婚或有男/女朋友
- 离异
- 丧偶

关于网络暴力的问题

网络暴力（或数字暴力/网络欺凌）包括使用技术来威胁、羞辱、恐吓或批评他人。在线威胁、通过微信/微博等社交软件发送的粗鲁、攻击性或贬损性文本、在网络上发布的评论或消息。并且还包括在互联网上发布个人信息、照片或视频以达到伤害或使他人难堪的目的。

5. 您曾经是否经受过网络暴力？ *

Marca solo un óvalo.

- 是
- 否

6. 您是否曾经对他人实施过网络暴力？ *

Marca solo un óvalo.

- 是
- 否

非常感谢

您已成功完成此问卷。
您的合作会有很大帮助！