**Anexo B Determinación del tamaño de la muestra**

El tamaño de la muestra se estima de la siguiente manera

$$n = \frac{n'}{1 + n'/N}$$

Siendo $n' = \frac{s^2}{\sigma^2}$, donde:

$\sigma^2$ es la varianza de la población respecto a determinadas variables. $s^2$ es la varianza de la muestra, la cual podrá determinarse en términos de probabilidad como $s^2 = p(1 - p)$, $se$ es error estandar que está dado por la diferencia entre ($\mu - \bar{x}$) la media poblacional y la media muestral. $(se)^2$ es el error estandar al cuadrado, que nos servirá para determinar $\sigma^2$, por lo que $\sigma^2 = (se)^2$ es la varianza poblacional.

Si la población es finita, es decir conocemos el total de la población y deseásemos saber cuántos del total tendremos que estudiar la respuesta seria:

$$n = \frac{N * Z_\alpha^2 p * q}{d^2 * (N-1) + Z_\alpha^2 * p * q}$$

Donde:

N = Total de la población, $Z_\alpha^2 = 1.96^2$ (si la seguridad es del 95%)

p = proporción esperada (en este caso 5% = 0.05)

q = 1 – p (en este caso 1-0.05 = 0.95)

d = precisión (en este caso deseamos un 3%).

Según diferentes seguridades el coeficiente de $Z_\alpha$ varía, así: si la seguridad $Z_\alpha$ fuese del 90% el coeficiente sería 1.645. Si la seguridad $Z_\alpha$ fuese del 95% el coeficiente sería 1.96. Si la seguridad $Z_\alpha$ fuese del 97.5% el coeficiente sería 2.24. Si la seguridad $Z_\alpha$ fuese del 99% el coeficiente sería 2.576.

$$n = \frac{k\hat{}2\ P\ Q\ N}{\varepsilon\hat{}2\ (N-1) + k\hat{}2\ PQ}$$

Donde k es el valor de referencia de la distribución normal con el nivel de confianza que se quiere trabajar (P. e para el 95.5% de confianza k = 2), ε es el error de muestreo, P es la proporción de la población que se pretende estimar, Q = 100-P, N es el tamaño de la población total.

**Cuadro 1**
Determinación del tamaño de la muestra

| ***h*ij** | ***Nh*ij** | ***Xh*ij** | ***VARh*ij** | ***Wh*ij** | ***nh*ij** | ***nh*ij*(entero)*** |
|---|---|---|---|---|---|---|
| 1 | 149 | 21.69 | 7.31 | 0.076449461 | 23.31708568 | 24 |
| 2 | 325 | 30.8 | 7.98 | 0.166752181 | 50.85941508 | 51 |
| 3 | 577 | 44.19 | 32.43 | 0.296049256 | 90.29502309 | 91 |
| 4 | 185 | 63.08 | 36.72 | 0.094920472 | 28.95074397 | 29 |
| 5 | 140 | | | 0.071831709 | 21.90867111 | 22 |
| 6 | 65 | 21.53 | 8.43 | 0.033350436 | 10.17188302 | 11 |
| 7 | 113 | 30.22 | 8.12 | 0.05797845 | 17.6834274 | 18 |
| 8 | 228 | 44.28 | 28.83 | 0.116983068 | 35.67983581 | 36 |
| 9 | 70 | 62.5142 | 32.4853 | 0.035915854 | 10.95433556 | 11 |
| 10 | 97 | | | 0.049769112 | 15.17957927 | 16 |
| | **1951** | **40.69** | **174.0145214** | | 305 | 309 |

Donde h representa cada uno de los estratos construidos con las diferentes combinaciones de rango de edad y región donde participan las socias dentro del programa, tuvieron que construir el estrato 5 y 10 para las socias de la región 0 y 1 respectivamente para las cuales no se contaba con datos de la edad.

Los promedios de edad (Xh*ij*) en los rangos *i* de cada una de las regiones *j* es muy parecido y parece no haber diferencia en el promedio de la edad, sin embargo se debe analizar con cautela estos datos, ya que existen en total 237 socias, es decir mas del 12% de la población, que no están representadas en este promedio, por lo que no se sabe con exactitud la distribución de estos datos faltantes.

Se observa también la variabilidad de los datos, donde cada uno de los estratos, tiene una menor varianza (VARh*ij*) que el total de la población, lo cual era esperado. Posteriormente se calcularon los pesos (Wh*ij*) de cada estrato para determinar que proporción de cada estrato es necesario para una muestra representativa, Así nh*ij* representa el tamaño de muestra en el estrato *ij*. Dado que el numero de individuos en una muestra es una variable discreta se redondea por simplicidad al numero siguiente, por lo que el total de la muestra (sumatoria de las muestras de cada estrato) es de 309 socias. Por otra parte se opto por realizar un muestreo estratificado aleatorio (asignación proporcional) por encima de una asignación optima ya que los resultados de la evaluación impacto serán del Grupo de tratamiento Nh*ij* vs. Grupo de control Nh*ij*, para todo *ij*; por lo que el análisis Grupo de tratamiento nh*ij* vs. Grupo de control nh*ij*, no resulta de interés en este proyecto.