

Taller #2
Muestreo
Econometría 06169

Profesor: Julio César Alonso
Monitora: Stephanie Vergara

Notas:

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de enero.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

En el archivo anexo (T2-01-05.xls en la Hoja1) se reporta el ingreso mensual por hogar (medido en miles de pesos) y el consumo de Energía por hogar (en kilovatios hora) para el total de los 5000 hogares de un pequeño municipio. A partir de esta información responda:

1. Suponga que nuestra variable de interés es la demanda de energía promedio por hogar del municipio. De acuerdo a esta información responda (Suponga que usted no cuenta con la información de la población, sólo conoce el tamaño de esta!!):
 - a. Se desea poder contar con un intervalo de confianza para la demanda de energía promedio por hogar del municipio con un 99% de confianza y que este no sea más amplio que 3 kilovatios por hora. Determine el tamaño de la MAS sin reposición necesaria para garantizar esto.
 - b. Ahora calcule el tamaño de la MAS sin reposición necesaria para garantizar cada una de las siguientes combinaciones:

<u>Tamaño del Intervalo</u>	<u>α</u>	<u>Tamaño del Intervalo</u>	<u>α</u>
2	0.01	6	0.1
4	0.01	8	0.1
6	0.01	2	0.05
8	0.1	3	0.05
2	0.1	4	0.05
3	0.1	6	0.05
4	0.1	8	0.05

Además discuta los resultados obtenidos.

- c. Suponga que ahora se desea emplear una MAS con reposición. Calcule el tamaño adecuado de la muestra para cada una de las situaciones descritas en los apartados a) y b). Compare los resultados con los obtenidos anteriormente.

2. Ahora, a partir de una MAS sin reposición, que garantice un tamaño de intervalo de 6 y un nivel de confianza del 95%, encuentre:
 - a. La media muestral y compárela con la poblacional.
 - b. Calcule un intervalo de confianza para la media muestral. Compárelo con la media poblacional.
3. En el municipio bajo estudio se cree que existe una gran inequidad en la forma como se factura la energía. En la actualidad se cobra una única tarifa para todos los hogares. El gobierno municipal está interesado en adoptar un nuevo esquema tarifario que tenga en cuenta las diferencias en los ingresos. En especial, se cree que existen tres diferentes grupos de hogares si se tienen en cuenta los ingresos mensuales por hogar.
 - a. Calcule un histograma que describa la distribución poblacional de los ingresos mensuales de todos los hogares del municipio. Comente sus resultados.
 - b. Ahora consideré la posibilidad de que existan 3 diferentes grupos, de tal forma que la membresía a un grupo dependen del nivel de ingresos del hogar. Calcule, para cada uno de los 3 grupos, un histograma para los ingresos mensuales de los hogares de toda la población.
4. Continuando con la pregunta anterior, suponga que el gobierno municipal está planeando cobrar un impuesto al consumo de energía al 50% de los hogares con mayores ingresos. Suponga que no es fácil determinar el grupo al que pertenece cada hogar (al momento de efectuar la facturación pero al momento de la investigación si es posible), aunque determinar su ingreso si es mucho más sencillo. Suponga además que no es fácil acceder a la población, pero si conoce el número de hogares en cada grupo.
 - a. Determine un esquema por medio del cual el gobierno municipal pueda garantizar que el mayor número de hogares pague el impuesto. Explique claramente los supuestos necesarios para su esquema. (NO realice los cálculos aún)
 - b. Realice sus cálculos y provea un esquema que sea de fácil de aplicar. Sea lo mas explícito posible y explique todos los cálculos realizados.
 - c. Suponga que aplica el criterio que usted a diseñado en la parte a)y b), determine que tan exitoso será este esquema. (Para esto emplee la información de la población.
5. El Departamento de economía de La Universidad Estatal de un País Caribeño se encuentra interesado en conocer la opinión de los estudiantes de esta carrera sobre la reducción del tiempo de duración del programa, pasando de 5 años a 4 años. Según la observación de las discusiones y percepción inicial de los profesores, se cree que el 20% de los estudiantes prefiere que el programa dure 5 años, mientras que el 80% restantes preferirían que la carrera se redujera a 4 años.

Dado que la población de estudiantes de economía en esta universidad es muy amplia, resultaría muy tedioso realizar un censo para conocer la opinión que tienen los alumnos al respecto. Además se posee la información reportada en la Tabla 1.

Tabla 1. Población de estudiantes de economía por semestre.

Semestre	Número de estudiantes
1	150
2	49
3	94
4	45
5	50
6	46
7	63
8	46
9	66
10*	55
664	

* Este semestre no se tiene en cuenta para la encuesta pues los estudiantes se encuentran haciendo prácticas profesionales por fuera de la universidad.

Usted como un consultor muy renombrado, ha sido encargado para realizar una encuesta para tal fin y como paso inicial debe resolver las siguientes preguntas:

- a. ¿Cuál debe ser el tipo de muestreo que se debe utilizar?
- b. Determine el tamaño óptimo de la muestra si se desea tener un error de más o menos 10% y un grado de confiabilidad del 95%. Explique claramente como llega a su resultado.
- c. Determine el número de encuestas que se deberían realizar por cada conglomerado, estrato o muestra, según lo que halla decidido en la pregunta anterior.

Taller #2
Respuestas Sugeridas
Muestreo
Econometría 06169

Profesor: Julio César Alonso
Monitora: Stephanie Vergara

Notas:

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 31 de enero.

INSTRUCCIONES:

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

En el archivo anexo (T2-01-05.xls en la Hoja1) se reporta el ingreso mensual por hogar (medido en miles de pesos) y el consumo de Energía por hogar (en kilovatios hora) para el total de los 5000 hogares de un pequeño municipio. A partir de esta información responda:

1. Suponga que nuestra variable de interés es la demanda de energía promedio por hogar del municipio. De acuerdo a esta información responda (Suponga que usted no cuenta con la información de la población, sólo conoce el tamaño de esta!!):
 - a. Se desea poder contar con un intervalo de confianza para la demanda de energía promedio por hogar del municipio con un 99% de confianza y que este no sea más amplio que 3 kilovatios por hora. Determine el tamaño de la MAS sin reposición necesaria para garantizar esto.

Noten que en este caso tenemos que:

$$\begin{aligned} \delta &= 1.50 \\ \alpha &= 0.01 \\ \frac{z_{\alpha/2}}{2} &= 2.576 \\ S^2 &= 3112.79 \end{aligned}$$

(La varianza se calcula a partir de una muestra piloto de tamaño 20. Así los resultados pueden variar de un estudiante a otro)
 Así, tamaño de la MAS sin reposición necesaria para garantizar las condiciones requeridas es:

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left(\frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{\frac{2.576^2 \cdot 2922.609}{1.5^2}}{1 + \frac{1}{5000} \left(\frac{2.576^2 \cdot 2922.609}{1.5^2} \right)} = 3236,846454 \approx 3237$$

Es decir, el tamaño de la muestral debe ser de 3237.

- b. Ahora calcule el tamaño de la MAS sin reposición necesaria para garantizar cada una de las siguientes combinaciones:

Tamaño del Intervalo	α	Tamaño del Intervalo	α
2	0.01	6	0.1
4	0.01	8	0.1
6	0.01	2	0.05
8	0.1	3	0.05
2	0.1	4	0.05
3	0.1	6	0.05
4	0.1	8	0.05

Además discuta los resultados obtenidos.

Empleando la información anterior y la siguiente fórmula

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left(\frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)}$$

Obtenemos los resultados reportados en la Tabla 1. De estos resultados se puede observar que:

- o Un aumento del tamaño del intervalo (ceteris paribus) implica una disminución del tamaño de la muestra.
- o Un aumento del nivel de significancia (ceteris paribus) implica una disminución del tamaño de la muestra.

Tabla 1. Cálculo del tamaño muestral para diferentes casos planteados MAS sin reposición.

Tamaño del Intervalo	α	$Z_{\alpha/2}$	n
2	0,01	2,58	4026
3	0,01	2,58	3237
4	0,01	2,58	2541
6	0,01	2,58	1573
8	0,1	1,64	477
2	0,1	1,64	3138
3	0,1	1,64	2141
4	0,1	1,64	1482
6	0,1	1,64	789
8	0,1	1,64	477
2	0,05	1,96	3526
3	0,05	1,96	2577
4	0,05	1,96	1871
6	0,05	1,96	1050
8	0,05	1,96	651

- c. Suponga que ahora se desea emplear una MAS con reposición. Calcule el tamaño adecuado de la muestra para cada una de las situaciones descritas en los apartados a) y b). Compare los resultados con los obtenidos anteriormente.

En este caso la fórmula indicada para ser aplicada es:

$$n = \frac{z_{\alpha/2}^2 S^2}{\delta^2} \quad (1.1)$$

Los resultados para cada uno de los casos se reporta en la Tabla 2.

Tabla 2. Calculo del tamaño muestral para diferentes casos planteados MAS con reposición.

Tamaño del Intervalo	α	$Z_{\alpha/2}$	n (MAS con Reposición)	n (MAS sin Reposición)
2	0.01	2.58	20654	4026
3	0.01	2.58	9180	3237
4	0.01	2.58	5164	2541
6	0.01	2.58	2295	1573
8	0.01	2.58	1291	477
2	0.1	1.64	8422	3138
3	0.1	1.64	3744	2141
4	0.1	1.64	2106	1482
6	0.1	1.64	936	789
8	0.1	1.64	527	477
2	0.05	1.96	11958	3526
3	0.05	1.96	5315	2577
4	0.05	1.96	2990	1871
6	0.05	1.96	1329	1050
8	0.05	1.96	748	651

Noten que el tamaño de la muestra es más grande cuando existe reposición. De hecho, es importante resaltar que para el MAS con reposición existen seis casos (resaltados en rojo en la tabla anterior) para los cuales según la fórmula la muestra debería ser más grande que la población!. Obviamente esto no tiene ningún sentido, para esos casos es mejor emplear un censo, pues la precisión (tamaño del intervalo) y/o nivel de confianza son muy altos.

- 2. Ahora, a partir de una MAS sin reposición, que garantice un tamaño de intervalo de 6 y un nivel de confianza del 95%, encuentre:

- a. La media muestral y compárela con la poblacional.

En este caso es necesario implementar una muestra de tamaño 1050 como se puede observar en la Tabla 2. Para nuestro caso, tenemos que la media muestral es 823,7889 y la desviación estándar de ellas o error estándar corresponde a 51,69265 para una varianza de 2672,13. Igualmente, el valor de la media poblacional es $\mu = 824,7783$, por lo

que podemos concluir que como era de esperarse, el promedio del estimador de las medias (media muestral) es muy cercano al valor real.

Nota: Este resultado puede cambiar pues depende de la muestra aleatoria.

- b. Calcule un intervalo de confianza para la media muestral. Compárelo con la media poblacional.

Si nos basamos en la muestra utilizada en el punto anterior, tenemos que:

$$\bar{y} = 823,7889$$

$$S_{\bar{y}} = 51,69265$$

$$t_{0,025;1049} = 0,031345$$

Con estos datos ya tenemos la información suficiente para hallar el IC con un 95% de confianza.

$$\bar{y} \pm t_{\alpha/2} S_{\bar{y}}$$

$$823,7889 \pm 1,62033$$

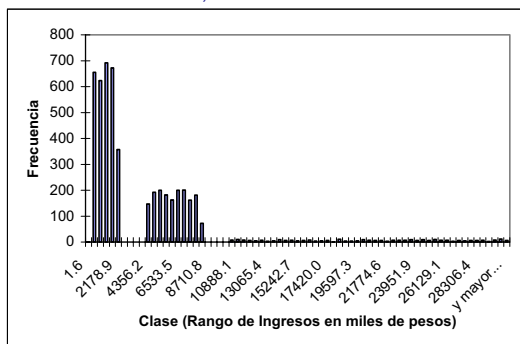
$$(823,158; 825,4093)$$

En este punto es importante tener en cuenta que la mitad de la longitud del IC es 1,125617 kilovatios/hora, un número muy aproximado al que queríamos (1.5 kilovatios por hora). Además, el valor de la media poblacional (824.77) se encuentra dentro de este intervalo, lo cual también es un punto importante para considerar.

- 3. En el municipio bajo estudio se cree que existe una gran inequidad en la forma como se factura la energía. En la actualidad se cobra una única tarifa para todos los hogares. El gobierno municipal está interesado en adoptar un nuevo esquema tarifario que tenga en cuenta las diferencias en los ingresos. En especial, se cree que existen tres diferentes grupos de hogares si se tienen en cuenta los ingresos mensuales por hogar.

- a. Calcule un histograma que describa la distribución poblacional de los ingresos mensuales de todos los hogares del municipio. Comente sus resultados.

Gráfico 1. Histograma de los Ingresos mensuales por hogar (Distribución Poblacional)



En el Gráfico 1 se reporta el histograma de los ingresos mensuales de cada uno de los 5000 hogares. Es de anotar que los datos no tienen una distribución simétrica, es decir, para algunos casos como en el nivel de menores ingresos, se encuentra una gran concentración de individuos, mientras que en el rango de mayores ingresos se ubican pocos individuos.

- b. Ahora consideré la posibilidad de que existan 3 diferentes grupos, de tal forma que la membresía a un grupo dependen del nivel de ingresos del hogar. Calcule, para cada uno de los 3 grupos, un histograma para los ingresos mensuales de los hogares de toda la población.

La distribución de las poblaciones para el Grupo 1, 2 y 3, se presentan en las Gráficas 2, 3 y 4, respectivamente. Observen que las distribuciones para cada uno de los grupos es uniforme y simétrica. Lo cual contrasta con lo observado para la población total. Es decir, existen claramente tres grupos muy heterogéneos en la población, y cada grupo parece ser muy homogéneo a su interior.

Gráfico 2. Histograma de los Ingresos mensuales para el Grupo 1.

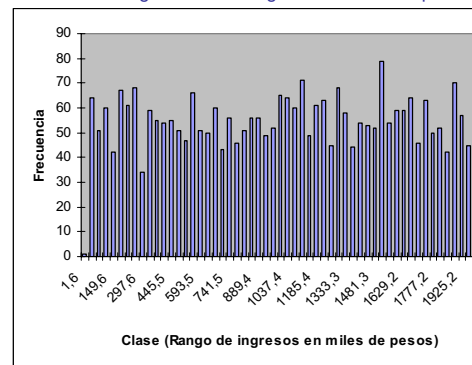


Gráfico 3. Histograma de los Ingresos mensuales para el Grupo 2.

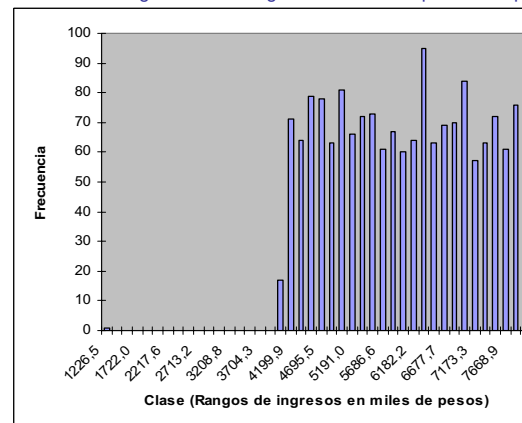
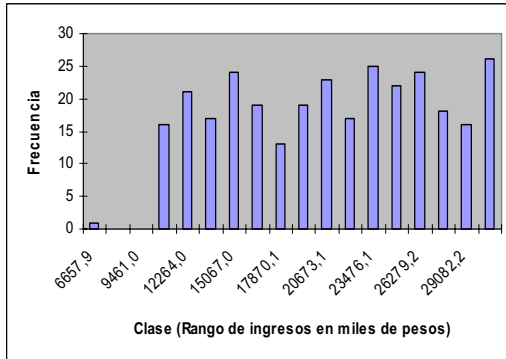


Gráfico 4. Histograma de los Ingresos mensuales para el Grupo 3.



4. Continuando con la pregunta anterior, suponga que el gobierno municipal está planeando cobrar un impuesto al consumo de energía al 50% de los hogares con mayores ingresos. Suponga que no es fácil determinar el grupo al que pertenece cada hogar (al momento de efectuar la facturación pero al momento de la investigación si es posible), aunque determinar su ingreso si es mucho más sencillo. Suponga además que no es fácil acceder a la población, pero si conoce el número de hogares en cada grupo.
 - a. Determine un esquema por medio del cual el gobierno municipal pueda garantizar que el mayor número de hogares pague el impuesto. Explique claramente los supuestos necesarios para su esquema. (NO realice los cálculos aún)

Para empezar, de la observación de nuestro histograma de la Gráfica 1, nos dimos cuenta de que la distribución de los ingresos de esta población no es simétrica. Adicionalmente, sabemos que tenemos 2999, 1700 y 301 individuos en los grupos 1, 2 y 3 respectivamente. Además, es claro que el grupo 1 corresponde a los hogares con menores ingresos. Así, todos los hogares del grupo 2 y 3 deberían pagar el impuesto pues esto corresponde a 2001 hogares del total de 5000 hogares.

Recuerden que deseamos diseñar un mecanismo que discrimine al 50% con mayores ingresos de la población. Así, será necesario discriminar 499 hogares del grupo 1 que posean los ingresos más altos. Es decir, nos debemos concentrar únicamente en el grupo 1. Es decir, consideraremos a ese grupo como nuestra población.

Además, podemos apelar a un resultado estadístico muy importante: la desigualdad de **Chebychev**. Según esta desigualdad, para cualquier variable aleatoria X y constante k se tiene que:

$$P(\mu - k\sigma \leq x \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2} \tag{1.2}$$

En este caso, necesitamos encontrar las 499 hogares de mayores ingresos de una subpoblación de 2999 individuos, es decir el 17% ($499/2999 = 0.1664$). Para aplicar la desigualdad (1.2), tenemos que:

$$0.5 = 1 - \frac{1}{k^2}$$

$$k = \sqrt{2} \approx 1.414$$

Es decir, de acuerdo a esta desigualdad tenemos que:

$$P(\mu - \sqrt{2}\sigma \leq x \leq \mu + \sqrt{2}\sigma) \geq 0.5$$

Dado que podemos esperar que la distribución de los hogares en el subgrupo 1 es simétrica, podríamos esperar que mas o menos el 25% de los individuos estén por encima del ingreso $\mu + \sqrt{2}\sigma$. Cómo desconocemos el valor de la media y de la varianza, será necesario estimarlos, pero antes será necesario determinar el tamaño de la muestra. Noten que en este caso un simple MAS sin reposición para el grupo 1 será suficiente, pues no es necesario considerar el resto de la población.

Así, lo que haremos será determinar el tamaño de la muestra adecuada para estimar la media del ingreso del grupo 1 y después a partir de la muestra calcularemos $\bar{x} + \sqrt{2}s$, cantidad que nos dará una cota que en principio nos permitirá dividir a la población total en dos grupos, los hogares que pagan y los que no pagan el impuesto.

- b. Realice sus cálculos y provea un esquema que sea de fácil de aplicar. Sea los mas explícito posible y explique todos los cálculos realizados.

El primer paso será hacer una prueba piloto que nos permita tener una idea sobre el tamaño de la varianza de los ingresos de los hogares del grupo 1. En este caso $s^2 = 239567.6729$. Ahora, escogiendo a

$$\delta = 100$$

$$\alpha = 0.05$$

$$z_{\frac{\alpha}{2}} = 1.96$$

Así, tamaño de la MAS sin reposición necesaria para garantizar las anteriores condiciones es:

$$n = \frac{z_{\alpha/2}^2 S^2}{\delta^2} = \frac{1.96^2 \cdot 239567.67}{100^2} = 327.87 \approx 328$$

$$n = \frac{1}{1 + \frac{1}{N} \left(\frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{1}{1 + \frac{1}{5000} \left(\frac{1.96^2 \cdot 239567.67}{100^2} \right)} = 327.87 \approx 328$$

A continuación se genera una muestra de tamaño 328 y se obtiene: $\bar{x} = 971.65$ y $s^2 = 331909.7$. Entonces de acuerdo a nuestro criterio escogido en el punto anterior, tenemos que $\bar{x} + \sqrt{2}s = 971.65 + \sqrt{2}(\sqrt{331909.7}) \approx 1786.4$

Así, el criterio para cobrar el impuesto será: cobrar el impuesto si el hogar tiene un ingreso mayor de 1786.4.

- c. Suponga que aplica el criterio que usted ha diseñado en la parte a) y b), determine que tan exitoso será este esquema. (Para esto emplee la información de la población).

En este caso, tenemos que la todos los hogares del grupo 2 y 3 poseen ingresos mayores de 1786.4. Además del grupo 1 tenemos que 301 hogares tiene ingresos mayor a 1786.4. Así, 2302 hogares pagarán el impuesto. En este caso el mecanismo es relativamente exitoso.

- 5. El Departamento de economía de La Universidad Estatal de un País Caribeño se encuentra interesado en conocer la opinión de los estudiantes de esta carrera sobre la reducción del tiempo de duración del programa, pasando de 5 años a 4 años. Según la observación de las discusiones y percepción inicial de los profesores, se cree que el 20% de los estudiantes prefiere que el programa dure 5 años, mientras que el 80% restantes preferirían que la carrera se redujera a 4 años.

Dado que la población de estudiantes de economía en esta universidad es muy amplia, resultaría muy tedioso realizar un censo para conocer la opinión que tienen los alumnos al respecto. Además se posee la información reportada en la Tabla 4.

Tabla 4. Población de estudiantes de economía por semestre.

Semestre	Número de estudiantes
1	150
2	49
3	94
4	45
5	50
6	46
7	63
8	46
9	66
10*	55
664	

* Este semestre no se tiene en cuenta para la encuesta pues los estudiantes se encuentran haciendo prácticas profesionales por fuera de la universidad.

Usted como un consultor muy renombrado, ha sido encargado para realizar una encuesta para tal fin y como paso inicial debe resolver las siguientes preguntas:

- a. ¿Cuál debe ser el tipo de muestreo que se debe utilizar?

El tipo de muestreo que resultaría mejor para este caso es el estratificado aleatorio, en donde cada semestre equivale a un estrato. Este método es el mejor pues en una muestra aleatoria simple sería complicado discriminar a los estudiantes que no sean de economía. Por otra parte, en un muestreo por conglomerados necesitaríamos estar seguros de que cada conglomerado es homogéneo respecto a los demás, pero heterogéneo en el interior, lo cual también es difícil de determinar. Sin embargo, es importante que tengan en cuenta que este tipo de muestreo solo se utiliza para

- b. Determine el tamaño óptimo de la muestra si se desea tener un error de más o menos 10% y un grado de confiabilidad del 95%. Explique claramente como llega a su resultado.

En primer lugar, es importante que tengan en cuenta que los estudiantes de décimo semestre no se incluyen pues no se encuentran en la Universidad, por lo que la Población (N) de estudio son solamente 609 estudiantes. Así mismo, como se espera que se tengan mas o menos la misma proporción de aceptación por estrato, tendremos que $P_h = \hat{P}$ y $Q_h = \hat{Q}$. Para determinar el tamaño de la muestra se podría emplear la siguiente fórmula:

$$n = \frac{z_{\alpha/2}^2 \left(\sum_{h=1}^H W_h S_h \right)^2}{\delta^2 + \frac{z_{\alpha/2}^2}{N} \left(\sum_{h=1}^H W_h S_h^2 \right)}$$

pero dado que $P_h = \hat{P}$ y $Q_h = \hat{Q}$, entonces $S_h = S$ y entonces la anterior fórmula se convierte en:

$$n = \frac{z_{\alpha/2}^2 \hat{P} \hat{Q}}{\delta^2 + \frac{z_{\alpha/2}^2}{N} \hat{P} \hat{Q}}$$

En este caso, podemos iniciar otorgándole a los valores \hat{P} y \hat{Q} , los porcentajes supuestos por los profesores, es decir, 20% y 80% respectivamente. De tal caso que obtendríamos:

$$n = \frac{\frac{Z_{\alpha}^2 \hat{P}\hat{Q}}{\delta^2}}{1 + \frac{1}{N} \left(\frac{Z_{\alpha}^2 \hat{P}\hat{Q}}{\delta^2} \right)} = \frac{(1.645)^2 (0.20 * 0.80)}{(0.05)^2} = 134.84 \approx 135$$

Por ende, la muestra debería ser de 135 personas.

- c. Determine el número de encuestas que se deberían realizar por cada conglomerado, estrato o muestra, según lo que halla decidido en la pregunta anterior.

El número de estratos viene dado por:

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^H W_h S_h}$$

dado que $S_h = S$ entonces tenemos que:

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^H W_h S_h} = n \frac{W_h S}{S \sum_{h=1}^H W_h} = n W_h$$

A partir del tamaño de muestra total tenemos los siguientes tamaños de muestra como se reporta en la Tabla 5.

Tabla 5. Numero de encuestas por semestre.

Semestre	#de Estudiantes	Peso (W)	Tamaño de la muestra	Tamaño Ajustado
1	150	0.25	33.25	33
2	49	0.08	10.86	11
3	94	0.15	20.84	21
4	45	0.07	9.98	10
5	50	0.08	11.08	11
6	46	0.08	10.20	10
7	63	0.10	13.97	14
8	46	0.08	10.20	10
9	66	0.11	14.63	15
Total	609	1	135	135