Departamento de Ingeniería Industrial
Facultad de Cs. Físicas y Matemáticas
Universidad de Chile

IN3401-01: Estadística para la Economía y Gestión
Profesores: Marcelo Henriquez, Sebastiàn Maldonado, Manuel Reyes
Auxiliar: José Miguel Carrasco

Auxiliar –Tema: Muestreo

Pregunta 1

La media de las estaturas de una muestra aleatoria de 400 personas de una ciudad es 1,75m. Se sabe que la estatura de las personas de esa ciudad es una variable aleatoria que sigue una distribución normal con varianza $\sigma^2 = 0{,}16$ m^2.

a. Construya un intervalo, de un 95% de confianza, para la media de las estaturas de la población.

b. ¿Cuál sería el mínimo tamaño muestral necesario para que pueda decirse que la verdadera media de las estaturas está a menos de 2 cm de la media muestral, con un nivel de confianza del 90%?

Soluciòn

La media de las estaturas de una muestra aleatoria de 400 personas de una ciudad es 1,75 m. Se sabe que la estatura de las personas de esa ciudad es una variable aleatoria que sigue una distribucion normal con varianza $\sigma^2 = 0{,}16m^2$

a) Construye un intervalo, de un 95% de confianza, para la media de las estaturas de la población.

$$n=400 \quad x=1.75 \quad \sigma=0.4$$
$$1-\alpha=0.95 \quad z\,\alpha/2=1.96$$
$$(1.75 \pm 1.96 \cdot 0.4/20) \rightarrow (1.7108, 1.7892)$$

b)

$$1.65 \cdot \frac{0.4}{\sqrt{n}} < 0.02 \qquad \frac{\sqrt{n}}{0.658} > \frac{1}{0.02}$$

$$\sqrt{n} > 32.9 \qquad n > 1\,082.41$$

La muestra debe tener al menos 1083 personas.

Pregunta 2 (Control Nº1 IN540- Otoño 2010)

En un amplio estudio sobre hábitos de consumo de medios en el Gran Santiago, uno de los objetivos es conocer la proporción de personas que ven diariamente 2 o más horas de TV. Justificando cada pregunta responda:

a. Para realizar la estimación a un 95% de confianza y con un error máximo de 0.01. ¿Cuál es el tamaño muestral necesario en un MAS?

Considerando la aproximación normal para la distribución de la proporción muestra P_n, vimos que el error muestral se expresa como: $\varepsilon = z\alpha/_2\sqrt{V(P_n)}$. Ahora bien, dado el tamaño de la población del Gran Santiago, se puede suponer que $\frac{(N-n)}{N} = 1$, y como la proporción poblacional p no es conocida, conviene utilizar el supuesto de varianza máxima, es decir la varianza que se consigue maximizando p(1-p), que se alcanza con $\hat{p} = 0.5$. Luego, usando las indicaciones,

$$0.01 = \varepsilon = 1.96\sqrt{\frac{0.5\times(1-0.5)}{n}} = \frac{1.96\times 0.5}{\sqrt{n}}$$

De donde $n = \left(\frac{1.96\times 0.5}{0.01}\right)^2 = 9604$ es el tamaño muestral requerido.

b. Si dicho tamaño muestral se distribuye de forma directamente proporcional a la población por comunas y si Santiago y Providencia representan respectivamente el 17.41% y el 15.81% del Gran Santiago, ¿Qué precisión tendrán los resultados parciales de esas comunas?

Utilizando el tamaño obtenido en a), y dado que la muestra es proporcional a la población en las comunas del Gran Santiago, si n_1 es el tamaño de la submuestra de la comuna de Santiago y n_2 es el tamaño de la submuestra de la comuna de Providencia, se tiene $n_1 = 0.1741 \times 9604 = 1672$ y $n_2 = 0.1581 \times 9604 = 1519$. Reemplazando estos tamaños en la expresión del error, con el mismo nivel de significación, se tienen los siguientes errores:

- *Para Santiago: $\varepsilon_1 = \frac{1.96\times 0.5}{\sqrt{1672}} = 0.024$, es decir 2,4%*
- *Para Providencia: $\varepsilon_2 = \frac{1.96\times 0.5}{\sqrt{1519}} = 0.025$, es decir 2,5%*

c. Si resultaron 890 casos de personas que ven 2 o más horas al día de TV, en la muestra de Santiago y 715 casos para Providencia, estime la diferencia de proporciones e interprete el resultado.

De acuerdo a los tamaños calculados en b), se tiene estimaciones de proporción:

- *Para Santiago: $P_1 = \frac{890}{1672} = 0.532$, es decir 53.2% en Santiago ve TV 2 o más horas al día.*
- *Para Providencia: $P_2 = \frac{715}{1519} = 0.471$, es decir 47.1% en Providencia ve TV 2 o más horas al día.*

Luego, la diferencia de proporciones se estima como $P_1 - P_2 = 0.532 - 0.471 = 0.061$.

Pregunta 3

El siguiente cuadro presenta información de una encuesta sobre el uso mensual de una tarjeta de crédito bancaria por parte de los clientes de cierto banco. La población esta estratificada por ingreso:

Estrato del ingreso (i)	Proporción del estrato (π_i)	Desviación estándar (σ_i)	Costo de la entrevista en M$ (c_i)	Media de la muestra i en M$
Bajo	0.3	10	2.5	80
Medio	0.5	20	2.5	250
Alto	0.2	40	10.0	400

Responda los siguientes, considerando la asignación de Neyman.

a. Suponga que se lleva a cabo un total de 1000 encuestas. ¿Cómo distribuye las entrevistas entre los tres estratos?

b. Si una muestra aleatoria simple de un tamaño de 1000 fuera obtenida de la población, ¿Cuál sería el tamaño de la muestra para el estrato alto?

c. ¿Por qué la recomendación de tamaño muestral para el estrato alto en a) difiere de b)?

d. Estime la media de la población

e. Calcule el costo total de la encuesta.

f. Si el costo de entrevista para el estrato alto fuera $2,500 .como aplicaría las 1000 entrevistas?

La asignación de Neyman para la muestra en cada estrato, se traduce en:

$$n_i = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum_i (\pi_i \sigma_i / \sqrt{c_i})} n$$

Donde:

n: El tamaño total de la muestra

π_i: La proporción de la población en el estrato *i*

σ_i: La desviación estándar de la población en el estrato *i*

c_i: El costo de una entrevista en el estrato *i*

n_i: El tamaño de la muestra para el estrato *i*

Utilizando la fórmula con $n = 1000$ y los datos del cuadro, calculamos n_i

Estrato del ingreso (i)	$\pi_i \sigma_i / \sqrt{c_i}$	n_i
Bajo	0.06	176.47 $\approx$ 177
Medio	0.20	588.23 $\approx$588
Alto	0.08	235.29 $\approx$ 235
$\sum$	0.34	1000

b) Aproximadamente 200 dado que el 20% de la poblacion es de ese estrato.
c) Por la desviación estandar que es la más alta en ese estrato.

d) Estimacion de la media de la poblacion = $\sum_i y_i * \pi_i = 229$ en miles de $

e) Costo total= $\sum_i c_i * n_i = 2{,}5 * 177 + 2{,}5 * 588 + 10 * 235 = 4262{,}5$ en M$

f) En este caso:

Estrato del ingreso (i)	Costo de la entrevista (c_i)	$\pi_i \sigma_i / \sqrt{c_i}$	n_i	Proporción (π_i)	Desviación estándar (σ_i)
Bajo	25	0.06	$142.85 \approx 143$	0.3	1
Medio	25	0.20	$476.19 \approx 476$	0.5	2
Alto	25	0.16	$380.95 \approx 381$	0.2	4
$\sum$		0.42	1000		

y el nuevo Costo total es obviamente 1000*2.5 = 2500 (enM$).