Estadística Inferencial

José UGARTE - Clifford TORRES

Facultad de Ciencias - UNI (Clase 5)

September 9, 2019

Selección del tamaño muestral

- El diseño de un experimento es en esencia un plan para adquirir una cantidad de información.
- La investigación, ya sea científica o de otro tipo, se hace para obtener información.
    1. Obviamente, deberíamos buscar obtener información a un costo mínimo.
- El procedimiento de muestreo, o diseño experimental, como suele llamarse, afecta la canti-dad de información por medición.
- éste, junto con el tamaño muestral n controla la cantidad total de información relevante en una muestra.

Selección del tamaño muestral

- En esta etapa de nuestro estudio nos ocuparemos de la situación de muestreo más sencilla: muestreo aleatorio de una población relativamente grande.
- Dedicamos nuestra atención a la selección del tamaño muestral n.
- Una de las preguntas más frecuentes que se plantea un estadístico es ¿cuántas mediciones deben incluirse en la muestra?

  1. Desafortunadamente, el estadístico no puede contestar esta pregunta sin saber cuánta infor-mación desea obtener el experimentador.
  2. Si nos referimos específicamente a una estimación, nos gustaría saber qué tan precisa desea el experimentador que sea.
  3. El experimentador puede indicar la precisión deseada al especificar un límite en el error de estimación.

Ejemplo: Selección del tamaño muestral

Suponga que deseamos estimar el promedio diario de producción μ de un producto químico y deseamos que el error de estimación sea menor que 5 toneladas con probabilidad de $0,95$. Veamos,

- Debido a que aproximadamente 95% de las medias muestrales estarán a no más de $2\sigma_{\bar{Y}}$ de μ en muestreo repetido.

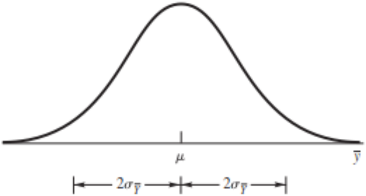

FIGURA 1 Distribución aproximada de $\bar{Y}$ para muestras grandes

- Se pide que $2\sigma_{\bar{Y}}$ sea igual a 5 toneladas, es decir,

$$\frac{2\sigma}{\sqrt{n}} = 5 \equiv n = \frac{4\sigma^2}{25}.$$

Ejemplo: Selección del tamaño muestral

- No podemos obtener un valor numérico exacto para n a menos que se conozca la desviación estándar poblacional σ
- Esto es exactamente lo que esperaríamos porque la variabilidad asociada con el estimador $\bar{Y}$ depende de la variabilidad exhibida en la población de la cual se sacó la muestra.
- A falta de un valor exacto para σ, usamos la mejor aproximación disponible, por ejemplo una estimación s obtenida de una muestra previa, o el conocimiento de la amplitud de las mediciones en la población.
- Como la amplitud es aproximadamente igual a 4σ (recuerde la regla empírica), un cuarto de la amplitud da un valor aproximado de σ.

Ejemplo: Selección del tamaño muestral

- Para nuestro ejemplo, suponga que se sabe que la amplitud de la producción es aproximadamente de 84 toneladas. Entonces

$$\sigma = \frac{84}{4} = 21$$

$$n = \frac{4\sigma^2}{25} \approx \frac{(4)(21)^2}{25} = 70,56.$$

- Usando un tamaño muestral $n = 71$, podemos tener cierta seguridad (con un coeficiente de confianza de alrededor de $0,95$) de que nuestro cálculo se encuentra a no más de 5 toneladas del verdadero promedio diario de producción.

Selección del tamaño muestral

- El método de seleccionar los tamaños muestrales para todos los procedimientos de estimación de muestra grande indicados en la Tabla 1 es análogo al que acabamos de describir.

Tabla 1 Valores esperados y errores estándar de algunos estimadores puntuales comunes

Parámetro objetivo θ	Tamaño(s) muestral(es)	Estimador puntual $\hat{\theta}$	$E(\hat{\theta})$	Error estándar $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\dfrac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \dfrac{Y}{n}$	p	$\sqrt{\dfrac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 y n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}^{*\dagger}$
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}^{\dagger}$

$^{*}\sigma_1^2$ y σ_2^2 son las varianzas de las poblaciones 1 y 2, respectivamente.

† Se supone que las dos muestras son independientes.

- El experimentador debe especificar un límite deseado en el error de estimación y un nivel de confianza asociado $1 - \alpha$.

Ejemplo: Selección del tamaño muestral

- Por ejemplo, si el parámetro es θ y el límite deseado es B, igualamos
$$z_{\frac{\alpha}{2}}\sigma_{\hat{\theta}} = B,$$
donde
$$P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}.$$

Ejemplo

- La reacción de un individuo a un estímulo en un experimento psicológico puede tomar una de dos formas, A o B. Si un experimentador desea estimar la probabilidad p de que una persona reaccione en una forma A, ¿cuántas personas deben incluirse en el experimento? Suponga que el experimentador estará satisfecho si el error de estimación es menor que $0,04$ con probabilidad igual a $0,90$. Suponga también que él espera que p se encuentre en algún punto cercano a $0,6$.

Ejemplo

- Un experimentador desea comparar la efectividad de dos métodos de capacitación para obreros que van a realizar una operación de ensamble. Los obreros seleccionados han de dividirse en dos grupos de igual tamaño, el primero para recibir el método 1 de capacitación y el segundo el método 2 de capacitación. Después de la capacitación cada obrero realizará la operación de ensamble y se registrará el tiempo que le tome hacerlo. El experimentador espera que las mediciones para ambos grupos tengan una amplitud de aproximadamente 8 minutos. Si la estimación de la diferencia en los tiempos promedio de ensamble debe ser correcta con una variación de no más de 1 minuto con probabilidad 0,95, ¿cuántos trabajadores deben incluirse en cada grupo de capacitación?