Tema 8

psicologia.isipedia.com

ESTIMACIÓN

Conceptos previos

Población y muestra:

Población se refiere al conjunto total de elementos que se quieren estudiar una o más características. Debe estar bien definida. Llamaremos ***N*** al número total de elementos de una población. También se suelen utilizar los términos ***individuos, sujetos y casos*** para referirnos a los elementos de la población.
Cuando se dispone de un **censo** (listado) de la población, se puede estudiar a todos ellos.
No siempre es factible estudiar a la totalidad de una población; por lo que se estudia un subconjunto de los elementos totales; es decir, un **muestra**. Llamaremos ***n*** al número de los elementos de una muestra.

Muestreo:

El muestreo es un proceso de selección con el fin de obtener una muestra lo más semejante posible a la población y así obtener estimaciones precisas. El tamaño es una característica esencial; ya que debe ser lo suficientemente amplia para representar adecuadamente las propiedades de la población y reducida para que pueda ser examinada en la práctica.
El muestreo ***probabilístico*** se conoce la probabilidad asociada a una muestra y cada elemento de la población tiene una probabilidad conocida de pertenecer a la muestra. El ***no-probabilístico*** se desconoce, o no se tiene en cuenta, la probabilidad asociada a cada muestra y se selecciona la que más le parezca representativa al investigador.

Una forma de obtener una muestra de una población homogénea es utilizar:

- El ***muestreo aleatorio simple***; por el cual se garantiza que cada elemento de la población tenga la misma probabilidad de formar parte de la muestra. Primero se asigna un número a cada elemento y después mediante algún medio (sorteo, papeletas,...) se elijen tantos elementos como sea necesario para la muestra.

- Cuando los elementos están ordenados o pueden ordenarse se utiliza el ***muestreo sistemático***. Se selecciona al azar entre los que ocupan los lugares $\frac{N}{n}$. Ejemplo: N = 100; n = 5; 100/5= 20; escogeríamos los elementos situados en las posiciones 20. El riesgo de este muestreo es la falta de representación; que se pudiese dar, del total de los elementos.
- Cuando topamos con una población heterogénea, utilizamos el ***muestreo estratificado***. Se emplea cuando disponemos de información suficiente sobre alguna característica y podemos elegir una muestra en función del número de elementos según estas características o estratos.
- Ante poblaciones desordenadas y conglomeradas en grupos, se emplea el ***muestreo por conglomerados***; donde se van seleccionando de todos los grupos, subgrupos, clases, … y finalmente de los elementos restantes la muestra.
- De la unión del estratificado y del conglomerado, surge otro ***muestreo el polietápico***.

En ocasiones el muestreo es muy costoso y se recurre a métodos no probabilísticos:

- El ***muestreo por cuotas*** (accidental) se basa en un buen conocimiento de los estratos o individuos más representativos para la investigación. Similar al estratificado pero carente del carácter aleatorio.
- El ***muestreo opinático*** (intencional) muestra el interés por incluir en la muestra a grupos supuestamente típicos.
- El ***causal*** (incidental) selección de los individuos de fácil acceso.
- ***Bola de nieve***; donde un elemento seleccionado lleva a otro y éste a otro y así sucesivamente hasta completar la muestra.

Una muestra es representativa si exhibe internamente el mismo grado de diversidad que la población y es aleatoria si los elementos han sido extraídos al azar de la población.

Inferencia estadística

El valor estadístico obtenido de una muestra (como media) no será igual al valor del parámetro de población. Para inferir un parámetro a partir de un estadístico hay que aplicar herramientas estadísticas de tipo inferencial como la ***estimación por intervalo*** (intervalos de confianza) o contraste de hipótesis.

Estimación de la media

La media muestral es una variable aleatoria que toma un valor según la muestra concreta que se obtenga. Se denomina ***distribución muestral de la media*** a su función de probabilidad.
La ***distribución muestral de un estadístico*** es un concepto central, tanto de la estimación como del contraste de hipótesis.

Distribución muestral de la media

Una función de probabilidad queda caracterizada por su forma, su media y su varianza. La media de la distribución muestral de la media $(\mu_{\bar{x}})$ es igual a la media de la población (μ). La varianza de la distribución muestral de la media es $\frac{\sigma}{n}$ y la desviación típica de la distribución muestral de la media, denominada error típico de la media, es $\sigma_{x} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n^2}}$.

La forma de la distribución original de la media se parece a una distribución normal aunque la distribución original de la variable en la población no es normal.
Dado el muestreo aleatorio simple:

- Si la distribución de X en la población es normal con media μ y desviación típica σ, entonces la distribución muestral de la $\bar{X}$ es normal $\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Si la distribución de X en la población no es normal con media μ y desviación típica σ, entonces la distribución muestral de la $\bar{X}$ tiende a la normal a medida que n crece (***Teorema Central del Límit***e), siendo la aproximación buena para $n \geq 30$.

Media, varianza y desviación típica de la variable cuantitativa X en la población y en la muestra, y de la distribución muestral de la media ($\bar{X}$).

	Población	**Muestra**	**Distribución muestral**

			de la media
Media	$\mu = \frac{\Sigma X}{N}$	$X = \frac{\Sigma X}{n}$	$\mu_{\bar{x}} = \mu$
Varianza	$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$	$S^2_{n-1} = \frac{\Sigma(X - \bar{X})^2}{n - 1}$ Cuasivarianza	$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$
Desviación típica	$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$	$S^2_{n-1} = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}$ Cuasidesviación típica	$\Sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$ Error tipico de la media

La media como estimador

Un estimador es un estadístico que se utiliza para estimar un parámetro. Por lo que la media de la muestra es un estimador de la media poblacional; y el valor del estimador en una muestra se denomina e***stimación*** o ***estimación puntual***.

La media muestral X es un estimador ***insesgado*** de la media poblacional (μ). El error típico de la media es un indicador de la precisión de la estimación de la media; cuanto menor es el error típico, mayor es la precisión. Dependiendo de la desviación típica de la población y del tamaño de la muestra.

Estimación de la proporción

Para la obtención de la distribución muestral de la proporción se puede hacer como la media.

Distribución muestral de la proporción

Sea X una variable que sólo toma valores 0 y 1, la proporción de la muestra P se define como:

$$P = \frac{\Sigma X}{n}$$

Dado el muestreo aleatorio simple, el estadístico proporción (P) se distribuye según una binomial:

$$\mu_p = \pi \;\; y \;\; \sigma^2_p = \frac{\pi (1 - \pi)}{n}$$

Como P es la media de los valores de X en la muestra, según el Teorema Central del Límite, a medida que el tamaño crece, la distribución muestral dela proporción tiende a la normal con media π y varianza $\frac{\pi (1 - \pi)}{n}$. Cuanto más alejado esté π de 0,5, más elementos debe tener la muestra para realizar la aproximación a la normal.

Media, varianza y desviación típica de la variable dicotómica o dicotomizada (X) en la población y en la muestra, y de la distribución muestral de la proporción (P):

	Población	**Muestra**	**Distribución muestral de la proporción (P)**
Media	$\pi = \frac{\Sigma X}{N}$ donde X: 0,1	$P = \frac{\Sigma X}{n}$ donde X: 0,1	$\mu_p = \mu$
Varianza	$\sigma^2 = \pi\,(1 - \pi)$	$S^2 = P\,(1 - P)$	$\sigma^2_p = \frac{\pi\,(1 - \pi)}{n}$
Desviación típica	$\sigma = \sqrt{\pi\,(1 - \pi)}$	$S = \sqrt{P\,(1 - P)}$	$\sigma_p = \sqrt{\frac{\pi\,(1 - \pi)}{n}}$ Error típico de la proporción

La proporción como estimador

La proporción muestral (p) es un estimador insesgado de la proporción poblacional (π).
El error típico de la proporción, es un indicador de la precisión de la estimación de la proporción. Cuanto menor es el error típico, mayor es la precisión.

Intervalos de confianza

Concepto

La finalidad de un intervalo de confianza es estimar un parámetro desconocido de una población a partir de una muestra. Al estimar la media de la población a partir de una muestra, podemos cometer un error de estimación $|\bar{X} - \mu|$.
La estimación por intervalo consiste en acotar el error con una alta probabilidad 1 – α (***nivel de confianza***) de forma que $|X - \mu|$ no sea superior a un estimado máximo ($E_{máx}$).
El error de estimación máximo ($E_{máx}$) es función de la variabilidad de la variable en la población, del nivel de confianza (n.c.) y del tamaño de la muestra:

$$E_{máx} = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

donde:

- $z_{1-\alpha/2}$ es función del n.c. = 1 – α y se obtiene en la tabla de la distribución normal tipificada (tabla IV).
- $\frac{\alpha}{\sqrt{n}}$ Es la desviación típica de la distribución muestral de la media, es decir, el error típico de la media.
- σ es la desviación típica de la población que es conocida.
- n es el tamaño de la muestra.

A partir de esta ecuación deducimos tanto el tamaño de la muestra como los límites del intervalo de confianza.

El tamaño de la muestra se obtiene despejando n de la ecuación:

$$n = \frac{z^2_{1-\alpha/2}\,\sigma^2}{E^2_{máx}}$$

vemos que n depende de:

- La desviación típica de la población.
- El nivel de confianza.
- El error de estimación máximo.

Los ***límites inferior (L_i) y superior (L_s)*** se obtienen a partir del $E_{máx}$:

$$L_i = X - E_{máx} \quad // \quad L_i = X - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$L_s = X + E_{máx} \quad // \quad L_s = X + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

El n.c. o ***probabilidad 1 – α*** significa que si extrajésemos todas las muestras posibles de una población, calculásemos la media en cada una de ellas y el intervalo de confianza, una proporción 1 – α de todos los intervalos de confianza contendrá la media poblacional y una proporción α no lo contendrá.

Tamaño de la muestra

Interesa que un intervalo sea lo más estrecho posible y con alta probabilidad. A mayor nivel de confianza mayor es el error de estimación máximo, por lo que más amplio será el intervalo y menos precisa será la estimación. Una forma de mantener y reducir el error de estimación máximo dado y aumentar el n.c., es aumentando n.

Otro factor que interviene es la variabilidad de la variable, cuanto mayor sea la desviación típica de la población, mayor debe ser n para alcanzar una misma precisión.
Para calcular el tamaño de la muestra desconociendo σ, hay que sustituir en la ecuación, la desviación típica por la cuasidesviación típica (S_{n-1}) y $z_{1-\alpha/2}$ por $t_{n-1,\ 1-\alpha/2}$ (tabla VI).

Aplicaciones

Los pasos para aplicar un intervalo de confianza son los siguientes:

- Establecer un error de estimación máximo para un nivel de confianza $1 - \alpha$.
- Obtener el tamaño de la muestra n para el error de estimación máximo especificado.
- Extraer una muestra aleatoria de tamaño n y medir la variable.
- Calcular el estadístico (es estimador del parámetro) con las medidas obtenidas.
- Calcular los límites del intervalo de confianza.

Intervalo de confianza para la media

Límites de los intervalos de confianza y supuestos para la estimación de la media:

Supuestos	Límites del intervalo de confianza para la media
• Muestreo aleatorio simple. • σ conocida. • Distribución normal o no normal con $n \geq 30$ (aprox. a la normal).	$L_i = \bar{X} - z_{1-\alpha/2}\sigma_{\bar{x}}$ $L_s = \bar{X} + z_{1-\alpha/2}\sigma_{\bar{x}}$ $z_{1-\alpha/2} \rightarrow$ Tabla IV $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
• Muestreo aleatorio simple. • σ desconocida. • Distribución normal. • $n < 30$[5].	$L_i = \bar{X} - t_{n-1;1-\alpha/2}\hat{S}_{\bar{x}}$ $L_s = \bar{X} + t_{n-1;1-\alpha/2}\hat{S}_{\bar{x}}$ $Tt_{n-1;1-\alpha/2} \rightarrow$ Tabla VI $\hat{S}_{x} = \frac{S_{n-1}}{\sqrt{n}}$
• Muestreo aleatorio simple. • σ desconocida. • Distribución normal o no normal con $n \geq 30$ (aprox. a la normal).	$L_i = \bar{X} - z_{1-\alpha/2}\hat{S}_{\bar{x}}$ $L_s = \bar{X} + z_{1-\alpha/2}\hat{S}_{\bar{x}}$ $z_{1-\alpha/2} \rightarrow$ Tabla IV $\hat{S}_{x} = \frac{S_{n-1}}{\sqrt{n}}$

S_{n-1} es la cuasidesviación típica calculada en la muestra.	

Intervalo de confianza para la proporción

El error de estimación máximo de la proporción es:

$$E_{máx} = z_{1-\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

donde:

- $z_{1-\alpha/2}$ es función del nivel de confianza $1-\alpha$ (tabla IV).
- $\sqrt{\frac{\pi(1-\pi)}{n}}$ es el error típico de la proporción: σ_p.
- π es la proporción de la población que no es conocida.
- n es el tamaño de la muestra y se debe cumplir $n\pi(1-\pi) \geq 5$ para la aproximación a la normal.

Los límites inferior y superior del intervalo de confianza se obtienen a partir del error de estimación máximo. Como desconocemos π, que es lo que precisamente queremos estimar, operamos con la proporción muestral P. Así, si en $E_{máx}$ sustituimos π por la proporción muestral P, los límites inferior y superior del intervalo de confianza son:

$$L_i = P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = P - E_{máx}$$

$$L_s = P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} = P + E_{máx}$$

Y la probabilidad de obtener un intervalo de confianza que contenga al parámetro π es:

$$P\left(P - z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} \leq \pi \leq P + z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}\right) = 1 - \alpha$$