Estadística Inferencial

Clifford Torres

Facultad de Ciencias - UNI (Clase 1)

August 19, 2019

Introducción

- El propósito de la estadística es usar la información contenida en una muestra para hacer inferencias acerca de la población de la cual se toma la muestra.
- Debido a que las poblaciones están caracterizadas por medidas descriptivas numémricas llamadas parámetros (parámetros objetivos), el objetivo de muchas investigaciones estadísticas es calcular el valor de uno o más parámetros relevantes.
- Las distribuciones muestrales que se conocerán poco a poco, desempean un importante papel en el desarrollo de los procedimientos de estimación.

Estimación

Definición

Un estimador es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.

Ejemplo

La media muestral

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

es un posible estimador puntual de la media poblacional μ.

Observación

Muchos estimadores diferentes (reglas de estimación) pueden obtenerse para el mismo parámetro poblacional.

Sesgo y error cuadrático medio de estimadores puntuales

Motivación

- *La estimación puntual es similar, en muchos aspectos, a disparar a un blanco con un revólver.*
- *El estimador, que genera estimaciones, es análogo al revólver; una estimación particular es comparable a un tiro; y el parámetro de interés corresponde al centro del blanco o diana.*
- *Extraer una sola muestra de una población y usarla para calcular una estimación del valor del parámetro equivale a disparar un solo tiro al centro del blanco.*
- *Suponga que un hombre dispara un solo tiro y acierta en el centro del blanco. ¿Concluimos que es un excelente tirador? ¿Se atrevería a sostener el blanco cuando se haga un segundo tiro?*

Sesgo y error cuadrático medio de estimadores puntuales

FIGURA 1
Distribución de estimaciones

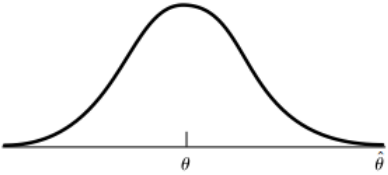
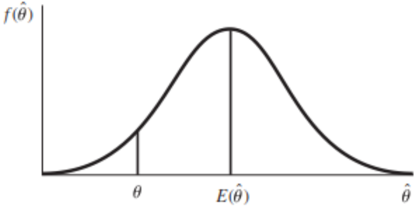


FIGURA 2
Distribución muestral para un estimador sesgado positivamente



Sesgo y error cuadrático medio de estimadores puntuales

Definición

Si $\hat{\theta}$ *es un estimador puntual de un parámetro* θ*, entonces* $\hat{\theta}$ *es un estimador insesgado si* $E(\hat{\theta}) = \theta$*. Si* $E(\hat{\theta}) \neq \theta$*, se dice que* $\hat{\theta}$ *está sesgado.*

Definición

El sesgo de un estimador puntual $\hat{\theta}$ *está dado por* $B(\hat{\theta}) = E(\hat{\theta}) - \theta$*.*

Sesgo y error cuadrático medio de estimadores puntuales

¿Qué tipo de distribución preferiríamos para nuestro estimador?

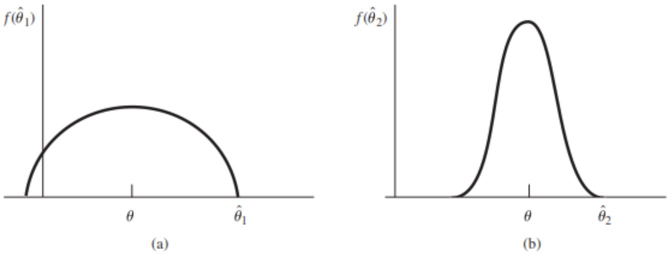


FIGURA 3 Distribuciones muestrales para dos estimadores insesgados: (a) estimador con variación grande; (b) estimador con variación pequeña

Sesgo y error cuadrático medio de estimadores puntuales

- Por consiguiente, además de preferir un estimador insesgado, necesitamos que la varianza de la distribución del estimador $V(\hat{\theta})$ sea lo más pequeña posible.
- Dados dos estimadores insesgados de un parámetro θ seleccionaríamos el estimador con la menor varianza mientras, todo lo demás permanece igual.
- Más que usar el sesgo y la varianza de un estimador puntual $\hat{\theta}$ para caracterizar su bondad, emplearemos...

Sesgo y error cuadrático medio de estimadores puntuales

Definición

El error cuadrático medio de un estimador puntual $\hat{\theta}$ *es*

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Ejemplo

El error cuadrático medio de un estimador u, MSE (u), es una función de su varianza y su sesgo.

$$MSE(\hat{\theta}) = V(\hat{\theta}) - [B(\hat{\theta})]^2.$$

Observación

A menudo buscamos estimadores insesgados con varianzas relativamente pequeñas.

Algunos estimadores puntuales insesgados comunes

- Nos concentramos en algunos estimadores que ameritan consideración con base en la intuición. Por ejemplo,
  - Parece natural usar la media muestral $\bar{Y}$ para estimar la media poblacional μ;
  - usar la proporción muestral $\hat{p} = \frac{Y}{n}$ para estimar un parámetro binomial p;
  - Si una inferencia está basada en muestras aleatorias independientes de n_1 y n_2 observaciones seleccionadas de dos poblaciones diferentes, ¿cómo estimaríamos la diferencia entre medias $(\mu_1 - \mu_2)$ o la diferencia en dos parámetros binomiales, $(p_1 - p_2)$?

Algunos estimadores puntuales insesgados comunes

- Como los cuatro estimadores $\bar{Y}$, $\hat{p}$, $(\bar{Y}_1 - \bar{Y}_2)$ y $(\hat{p}_1 - \hat{p}_2)$ son funciones de las variables aleatorias observadas en muestras, podemos hallar sus valores y varianzas esperados.

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	$\bar{Y}$	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	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}^{*\dagger}$
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}^{\dagger}$

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

† Se supone que las dos muestras son independientes.

Algunos estimadores puntuales insesgados comunes

Ejercicio

Justifique la insegadez de la varianza poblacional y muestral.

Pueden hacerse dos comentarios finales respecto a los estimadores puntuales de la Tabla 1.

1. Los valores esperados y los errores estándar para $\bar{Y}$ y $\bar{Y}_1 - \bar{Y}_2$ dados en la tabla son válidos cualquiera que sea la distribución de la(s) población(es) de donde se tome(n) la(s) muestra(s).
2. Los cuatro estimadores poseen distribuciones de probabilidad que son aproximadamente normales para muestras grandes (El teorema del límite central justifica este enunciado para $\bar{Y}$ y $\hat{p}$, y teoremas similares para funciones de medias muestrales justifican la afirmación para $\bar{Y}_1 - \bar{Y}_2$ y $(\hat{p}_1 - \hat{p}_2)$. ¿Qué tan grande es 'grande'?

Algunos estimadores puntuales insesgados comunes

- Sabemos que $\bar{Y}$, $\hat{p}$, $(\bar{Y}_1 - \bar{Y}_2)$ y $(\hat{p}_1 - \hat{p}_2)$ son insesgados con distribuciones muestrales casi normales (al menos con forma de campana) para muestras de tamaño moderado; ahora utilizaremos esta información para responder algunas preguntas prácticas.
- Si usamos un estimador una sola vez y obtenemos una sola estimación, ¿qué tan buena será ésta? ¿Cuánto podemos confiar en la validez de nuestra inferencia?