

Estimación puntual y por intervalos

Estadística Inferencial

El tiempo que tarda un estudiante en resolver un parcial de cierta materia sigue una distribución normal, con una desviación estándar $\sigma = 20$ minutos. Sin embargo no se conoce μ , la media del tiempo necesario para resolver el parcial.

¿Cómo se podría estimar μ ?

Objetivo

Tomar decisiones o sacar conclusiones respecto a la población basándose en la información contenida en una muestra (observaciones).

Generalmente se hacen inferencias respecto a los parámetros que describen la distribución de la población, a la forma de la distribución, a la independencia entre variables, a la normalidad (gaussianidad) de las observaciones.

Tipos de inferencias que veremos:

- ▶ **Estimación Puntual:** Se da un solo valor o punto como estimación del parámetro poblacional de interés
- ▶ **Estimación por Intervalos:** Se especifica un intervalo de posibles valores del parámetro en estudio
- ▶ **Test de hipótesis:** Se hacen conclusiones acerca de una hipótesis sobre los parámetros o distribución de la población. (próxima unidad)

Situación 1

El tiempo que tarda un estudiante en resolver un parcial de cierta materia sigue una distribución normal, con una desviación estándar $\sigma = 20$ minutos. Sin embargo no se conoce μ , la media del tiempo necesario para resolver el parcial.

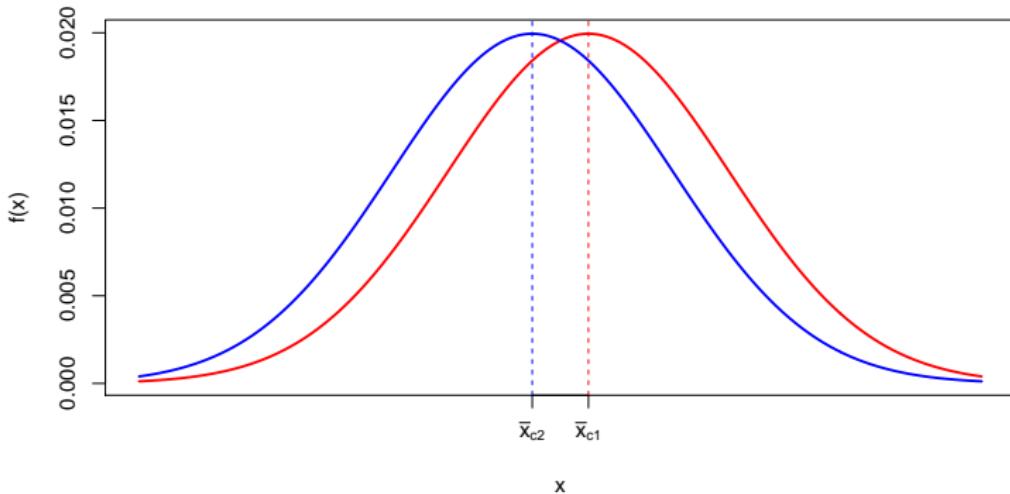
¿Como se podría estimar μ ?

Propuesta 1: Estimación Puntual

- ▶ El profesor a cargo de la comisión 1 registra los tiempos x_1, \dots, x_{30} en que cada uno de sus 30 alumnos entrega el parcial luego calcula el promedio de esas observaciones. Obtiene $\bar{x}_{c1} = 93.5$ m. La mediana de sus observaciones es $Me_{c1} = 92.4$ m
- ▶ El profesor a cargo de la comisión 2 realiza el mismo procedimiento con su grupo de 30 alumnos y obtiene $\bar{x}_{c2} = 86.1$ m y $Me_{c2} = 86.2$ m

¿Podemos tomar alguna conclusión? ¿Podemos afirmar con total certeza cual es el “verdadero” valor de μ ?

Sabiendo que *el tiempo que tarda un estudiante en resolver un parcial de cierta materia sigue una distribución normal, con una desviación estándar $\sigma = 20$* , los profesores graficaron la distribución del tiempo, usando como estimación de μ , los promedios obtenidos (\bar{x}_{c1} y \bar{x}_{c2})



Propuesta 2: Estimación por Intervalos

- ▶ Con los tiempos registrados el prof. a cargo de la comisión 1 calcula un intervalo del 95% de confianza para μ . Obtiene (86.3, 100.6)
- ▶ El profesor a cargo de la comisión 2 realiza el mismo procedimiento y obtiene (78.9, 93.2)

Observaciones

- ▶ Ambos profesores afirman que la probabilidad que μ esté en el intervalo que ellos calcularon, es de 0.95
- ▶ Notar que la amplitud de los intervalos es la misma:
 1. $100.6 - 86.3 = 14.3$
 2. $93.2 - 78.9 = 14.3$

confidencia: el verdadero valor de μ utilizado en la generación de estas muestras "aleatorias" de tiempo, fue $\mu = 90$

Supongamos que la variable aleatoria X tiene función de densidad $f(x)$ caracterizada por el parámetro θ .

Sea X_1, X_2, \dots, X_n es una muestra aleatoria de tamaño n de X , un **estimador puntual** del parámetro poblacional θ se obtiene seleccionando un estadístico apropiado $g(X_1, \dots, X_n)$. Al calcular su valor utilizando los datos muestrales x_1, \dots, x_n se obtiene $\hat{\theta}$, una **estimación puntual** de θ .

El **estimador** es una variable aleatoria. La **estimación** es un vector de números (de igual dimensión que θ).

Notación: Denotaremos por $\hat{\Theta}$ el estimador de θ , por $\hat{\theta}$ la estimación de θ .

Ejemplo

Se quiere determinar el tiempo de respuesta de un analgésico. Se asume que la variable aleatoria que mide el tiempo de respuesta del analgésico sigue una distribución normal con media μ y desviación estándar σ . Se administra el analgésico a 10 personas. Denotemos por X_1, \dots, X_{10} la muestra aleatoria que representan los tiempos de respuesta del analgésico en 10 pacientes, y por x_1, \dots, x_{10} los tiempos observados. Si:

$$\begin{array}{lllll} x_1 = 23.0 & x_2 = 28.4 & x_3 = 26.3 & x_4 = 35.1 & x_5 = 31.6 \\ x_6 = 30.9 & x_7 = 25.2 & x_8 = 28.0 & x_9 = 27.3 & x_{10} = 29.2 \end{array}$$

¿Cómo se podría estimar μ ?

Posibles estimadores de μ son:

1. \bar{X} (media muestral), con estimación
 $\hat{\mu} = \bar{x} = \sum_{i=1}^{10} x_i / 10 = 28.5$
2. $\tilde{X} = X_{Me}$ (mediana muestral), con estimación
 $\hat{\mu} = x_{Me} = 28.20$
3. $\frac{\min X_i + \max X_i}{2}$ (promedio de los tiempos de respuesta), con estimación $\hat{\mu} = \frac{23.0 + 35.1}{2} = 29.05$

¿Cuál de ellos está más próximo al verdadero valor de μ ?

Estimadores Insesgados

En el mejor de los casos, $\hat{\theta} = \theta$, pero esto es imposible ya que $\hat{\theta}$ es resultado de una variable aleatoria. Así:

$$\hat{\theta} = \theta + \epsilon$$

donde ϵ es un error de estimación. Un estimador “preciso” será el que haga ese error pequeño.

Definición: Se dice que el estimador $\hat{\Theta}$ es un **estimador insesgado** de θ si

$$E(\hat{\Theta}) = \theta$$

Si $\hat{\Theta}$ no es insesgado, entonces $E(\hat{\Theta} - \theta) = \text{sesgo de } \hat{\Theta}$

Sea X v.a. con media μ y varianza σ^2 . Sea X_1, \dots, X_n muestra aleatoria de tamaño n de X .

Un estimador de μ es \bar{X}

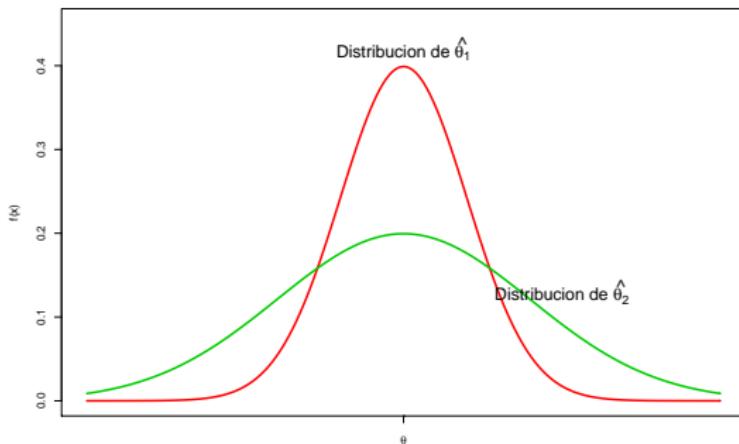
$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n\mu}{n} = \mu$$

Por lo tanto la media muestral \bar{X} es un *estimador insesgado* de la media poblacional μ .

Un estimador insesgado de σ^2 es $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$

Estimadores Puntuales Insesgados - EIMV

Puede ocurrir que varios haya varios estimadores de un parámetro poblacional θ que sean insesgados... ¿Cuál elegimos?



Elegimos el
Estimador
Insesgado de
Mínima Varianza

Otras propiedades de un *buen* estimador

- ▶ **Consistencia:** decimos que un estimador $\hat{\Theta}$ de θ es **consistente** si $\hat{\Theta}$ converge en probabilidad a θ cuando $n \rightarrow \infty$, es decir,
 $\forall \epsilon > 0 \quad P(|\hat{\Theta} - \theta| \leq \epsilon) \rightarrow 1$ cuando $n \rightarrow \infty$.
- ▶ **Eficiencia:** se dice que un estimador es **Eficiente** u óptimo cuando posee varianza mínima, o en términos relativos, cuando presenta menor varianza que otro.
- ▶ **Suficiencia:** Un estimador $\hat{\Theta}$ de θ es **suficiente** si $f_{\hat{\Theta}}(x_1, x_2, \dots, x_n)$ no depende de θ .

En resumen: Recordemos que los estimadores son **variables aleatorias**

- ▶ tienen una distribución de probabilidad (distribuciones muestrales)
- ▶ su distribución le confiere una serie de propiedades estadísticas: ser insesgado, consistente, eficiente, etc. Por ello se puede comparar la *calidad* de un estimador respecto a la de otros estimadores (del mismo parámetro).
- ▶ ningún estimador es perfecto.

Cuando se reporta el valor de un estimador puntual debe reportarse también alguna indicación de su precisión, por ejemplo

$$\sigma_{\hat{\theta}} = \sqrt{Var\hat{\theta}}$$

Si μ = media poblacional, X_1, \dots, X_n muestra aleatoria,
 p = parámetro binomial $Y \sim \mathcal{B}(n, p)$ entonces:

Parámetro	Tamaño muestra	Estimador Puntual	$E(\hat{\theta})$	$\sigma_{\hat{\theta}}$
μ	n	\bar{X}	μ	σ^2/n
p	n	Y/n	p	pq/n

Métodos de Estimación Puntual

Método de los momentos

Definiciones:

Sea X_1, \dots, X_n muestra aleatoria de la v.a. X (discreta o continua) con función de densidad (o función de probabilidad de masa) $f(x)$

- ▶ El *k*-ésimo momento de una v.a. X respecto al origen es $\mu_k = E(X^k)$.
- ▶ El *k*-ésimo momento de la muestra es $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$

- ▶ Los momentos caracterizan una distribución de probabilidad
- ▶ Si dos variables aleatorias tienen los mismos momentos, entonces estas variables tienen la misma función de densidad.
- ▶ Intuitivamente, los momentos poblacionales μ_k se parecerán a los momentos muestrales m_k

Ejemplos

Método de los momentos: Se eligen como estimaciones de los parámetros a aquellos valores de los mismos que son soluciones de las ecuaciones

$$m_k = \mu_k, k = 1, \dots, t$$

donde t es el número de parámetros de interés.

1. Sea X_1, \dots, X_n muestra aleatoria de una v.a. con distribución uniforme en el intervalo $(0, \theta)$, con θ desconocido. Estimar θ por método de los momentos.

Ayuda: Recordar que $E(X) = \theta/2$

Ejemplos

2. Se quieren estimar los parámetros μ y σ^2 de una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$.

¿Cómo lo haría utilizando el método de los momentos?

Ayuda: En este caso $t = 2$. Recordar que $E(X) = \mu$,
 $\sigma^2 = E(X^2) - E(X)^2$.

Método de Máxima Verosimilitud

Definiciones:

Sean x_1, \dots, x_n realizaciones de X_1, \dots, X_n , muestra aleatoria de una v.a. con función de densidad (o probabilidad de masa) $f(x)$ que involucra a algún parámetro θ (la denotaremos $f(x; \theta)$).

1. Se define la función de verosimilitud

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

2. El **estimador de máxima verosimilitud** del parámetro θ es el valor de θ que maximiza la función $L(x_1, \dots, x_n; \theta)$

Ejemplo

Sean x_1, \dots, x_n observaciones de X_1, \dots, X_n muestra aleatoria de una v.a. con distribución normal $N(\mu, \sigma^2)$. Suponiendo que σ^2 es conocida, encontrar el EMV de μ .

Recordar que $f(x_i; \mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$

Entonces:

$$\begin{aligned} L(x_1, \dots, x_n; \mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-(1/2\sigma^2) \sum_{i=1}^n (x_i - \mu)^2\right) \end{aligned}$$

Tomando logaritmo:

$$\begin{aligned}\ln(L(x_1, \dots, x_n; \mu)) &= \ln\left(\frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left(-(1/2\sigma^2) \sum_{i=1}^n (x_i - \mu)^2\right)\right) \\ &= -(n/2) \ln(2\pi\sigma^2) - (2\sigma^2)^{-1} \sum_{i=1}^n (x_i - \mu)^2\end{aligned}$$

Buscando el máximo de esa función resulta:

$$\frac{d \ln L(x_1, \dots, x_n; \mu)}{d\mu} = \sigma^{-2} \sum_{i=1}^n (x_i - \mu)$$

igualando a 0 y resolviendo para μ : $\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$

Continuación ejercicio 3:

b) ¿Cómo calcularía los estimadores de Máxima Verosimilitud de μ y σ^2 de una variable aleatoria X con distribución normal $\mathcal{N}(\mu, \sigma^2)$?

Ej 1, Guía TP 10

Se sabe que el número de máquinas a reparar de un Call Center, por mes, tiene distribución de Poisson. El encargado del Call Center desea conocer el número esperado de máquinas a reparar en un mes cualquiera para anticipar los gastos por reparación. Para ello, cuenta con los datos que corresponden al número de máquinas enviadas al taller en el último año, expuestos en la siguiente tabla:

E	F	M	A	M	J	J	A	S	O	N	D
2	4	2	3	6	6	2	2	1	3	3	2

1. Obtenga la expresión del estimador de momentos para una muestra aleatoria de v.a.i.i.d con distribución Poisson de parámetro λ .
2. Obtenga la expresión del estimador de máxima verosimilitud para una muestra aleatoria de v.a.i.i.d con distribución Poisson de parámetro λ .
3. Obtenga una estimación para el número esperado de máquinas del Call Center a reparar en un mes cualquiera.