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.
¿Como se podría estimar μ?

Estadística Inferencial

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.

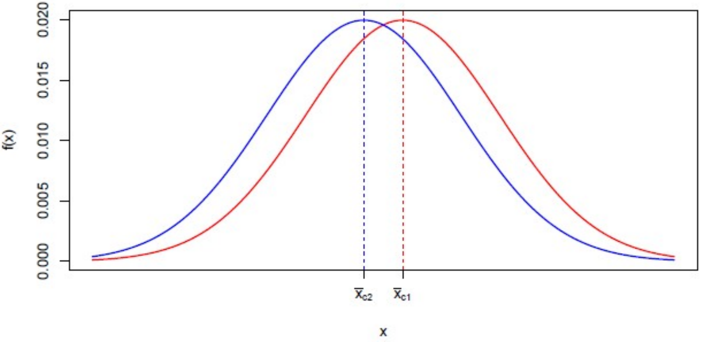
¿Como se podría estimar μ?

Propuesta 1: Estimación Puntual

- El profesor a cargo de la comisión 1 registra los tiempos $x_1, \ldots, x_{30}$ 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 $Me_{c1} = 92.4m$
- 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 $Me_{c2} = 86.2m$

¿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 $\mu = 90$

Estimación Puntual

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

Sea $X_1, X_2, \cdots, X_n$ es una muestra aleatoria de tamaño n de X, un estimador puntual del parámetro poblacional θ se obtiene seleccionando un estadístico apropiado $g(X_1, \cdots, X_n)$. Al calcular su valor utilizando los datos muestrales $x_1, \ldots, x_n$ 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 $X_1, \cdots, X_{10}$ la muestra aleatoria que representan los tiempos de respuesta del analgésico en 10 pacientes, y por $x_1, \cdots, x_{10}$ los tiempos observados. Si:

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

¿Como 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. $\frac{minX_i + maxX_i}{2}$ (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.

<u>Definición</u>: 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 $E(\hat{\Theta} - \theta) =$ sesgo de $\hat{\Theta}$

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

Un estimador de μ es $\bar{X}$

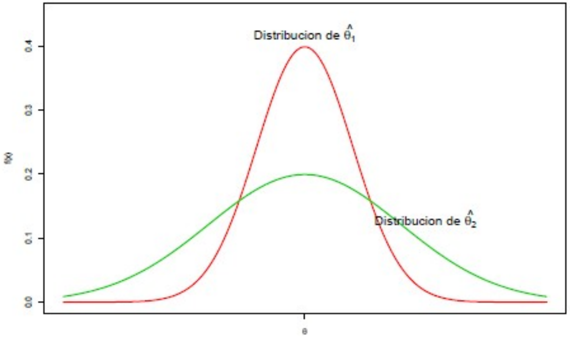
$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{n\mu}{n} = \mu$$

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 $n \to \infty$, es decir, $\forall \epsilon > 0$ $P(|\hat{\Theta} - \theta| \leq \epsilon) \to 1$ cuando $n \to \infty$.
- 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 $f_{\hat{\Theta}}(x_1, x_2, \ldots, x_n)$ no depende 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 $\mu =$ media poblacional, $X_1, \cdots, X_n$ 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 $X_1, \cdots, X_n$ 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

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

donde t es el número de parámetros de interés.

1. Sea $X_1, \cdots, X_n$ 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 $x_1, \cdots, x_n$ realizaciones de $X_1, \cdots, X_n$, 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 $L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$
2. El estimador de máxima verosimilitud del parámetro θ es el valor de θ que maximiza la función $L(x_1, \ldots, x_n; \theta)$

Ejemplo

Sean $x_1, \cdots, x_n$ observaciones de $X_1, \cdots, X_n$ 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:

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

Tomando logaritmo:

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

Buscando el máximo de esa función resulta:

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

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

Estimación por Intervalos

Estimación por Intervalos

Cualquier parámetro que quiera estimarse tiene como conjunto de valores posibles un intervalo de n^{o}s reales.
Ejemplos:

1. Tiempo de vida útil (promedio) de cierta lámpara de bajo consumo puede ser cualquier número, digamos, entre 900 y 1100 hrs.
2. Cantidad de clientes que utilizan cierto cajero automático entre las 10 y las 11 de la mañana, puede ser cualquier número natural entre (por ejemplo) 1 y 60.

- Muestras aleatorias $\Rightarrow$ altamente probable que $\hat{\theta} \neq \theta$.
- El estimador puntual no da información de cuán próximo está $\hat{\theta}$ de θ.

Objetivo:Encontrar un intervalo de números reales que con alta probabilidad contenga al verdadero valor del parámetro en estudio.

Intervalo de Confianza para un parámetro poblacional

Intervalo de valores posibles que puede tomar dicho parámetro. Sus extremos inferior y superior son denominados límites de confianza inferior y superior, respectivamente, y son cantidades aleatorias que *dependen de la muestra*.

Se pretende que el intervalo estimado:

- Contenga al verdadero valor del parámetro con una alta probabilidad (nivel de confianza)
- Sea relativamente angosto

Intervalos de confianza

Sean θ_I, θ_S los límites inferior y superior del intervalo de confianza para el parámetro θ ($\theta \in \mathbb{R}$).

Se busca encontrar estimadores $\hat{\Theta}_I, \hat{\Theta}_S$ tales que

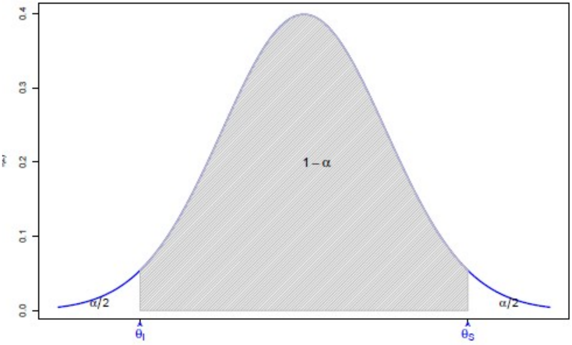
$$P(\hat{\Theta}_I < \theta < \hat{\Theta}_S) = 1 - \alpha$$

$0 < \alpha < 1$

$(1 - \alpha)$ es el nivel de confianza del intervalo.

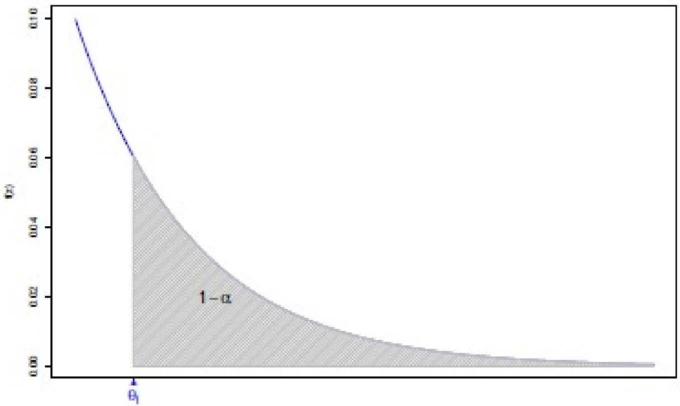
En tal caso el intervalo $(\hat{\Theta}_I, \hat{\Theta}_S)$ es un intervalo de confianza del $100 \cdot (1 - \alpha)\%$ (bilateral)

Intervalo de confianza para media poblacional de un distribución simétrica



Intervalos de Confianza Unilaterales

$P(\hat{\Theta}_I < \theta) = 1 - \alpha \Rightarrow (\hat{\Theta}_I, \infty)$ $\qquad$ $P(\theta < \hat{\Theta}_S) = 1 - \alpha \Rightarrow (-\infty, \hat{\Theta}_S)$



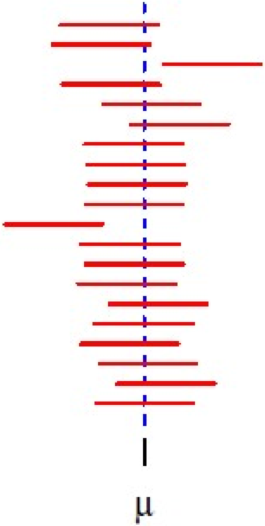
Intervalos de Confianza - Interpretación

Observaciones

- $\hat{\Theta}_I$, $\hat{\Theta}_S$ son estadísticos (funciones de la muestra).
- $P(\hat{\Theta}_I \leq \theta \leq \hat{\Theta}_S) = 1 - \alpha$ se lee como *la probabilidad que el intervalo* $(\hat{\Theta}_I\ ,\hat{\Theta}_S)$ *contenga al verdadero valor de* θ *es* $1 - \alpha$
- Una vez observada la muestra se calcula $(\hat{\theta}_I, \hat{\theta}_S)$, que es el intervalo "observado". El verdadero valor de θ puede, o no, estar contenido en ese intervalo.
- Si se toman muestras aleatorias de una misma población y con cada una de ellas se calcula un int. de confianza para $\theta \Rightarrow (1 - \alpha)\%$ de las veces dicho intervalo contendrá el verdadero valor de θ

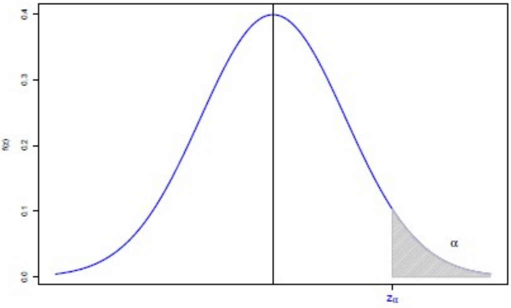
Interpretación

- Objetivo: Estimar μ, media poblacional de una v.a. $X \sim f(x;\mu)$.
- Calcular intervalo de 90% de confianza para μ.
- Se obtuvieron 20 muestras aleatorias de $X \sim f(x;\mu)$ y con ellas se calcularon los 20 intervalos de confianza



Sea $Z \sim N(0,1)$. Se define z_α a aquel valor que satisface

$$P(Z \geq z_\alpha) = \alpha$$



$1-\alpha$	0.90	0.95	0.99
$\alpha/2$	0.05	0.025	0.005
$z_{\alpha/2}$	1.64	1.96	2.57

Intervalos de confianza para media poblacional, μ

En esta materia derivaremos intervalos de confianza (IC) para la media poblacional ($E(X) = \mu$) en las siguientes situaciones:

1. Población normal, con varianza conocida. $(X \sim \mathcal{N}(\mu, \sigma^2))$
2. Población normal, varianza desconocida, tamaño de muestra n pequeño. ($(X \sim \mathcal{N}(\mu, \sigma^2)$, $n < 30$)
3. Muestras grandes ($n \geq 30$, independientemente de la distribución de la variable)
   - 3.1 varianza conocida
   - 3.2 varianza desconocida

1. IC - Población normal, σ^2 conocida

Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a. con distribución $N(\mu, \sigma^2)$, con σ conocida.
Si se quiere encontrar un intervalo del $100 \cdot (1 - \alpha)\%$ de confianza para μ, se buscan $\hat{\Theta}_I, \hat{\Theta}_S$ tales que $P(\hat{\Theta}_I < \mu < \hat{\Theta}_S) = 1 - \alpha$.

$$X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, (\sigma/\sqrt{n})^2) \text{ y } \mathrm{Z} = \frac{\overline{\mathrm{X}} - \mu}{\sigma/\sqrt{\mathrm{n}}} \sim \mathrm{N}(0,1)$$

1. IC media - Población normal, σ^2 conocida

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2} - \bar{X}\right) \\
&= P\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < \frac{\sigma}{\sqrt{n}}\cdot z_{\alpha/2} + \bar{X}\right)
\end{aligned}
$$

1. IC media - Población normal, σ^2 conocida

$$(\hat{\Theta}_I, \hat{\Theta}_S) = (\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, +\bar{X} + \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2})$$

es un intervalo del $100.(1-\alpha)\%$ de confianza para μ.

- Es un intervalo aleatorio, centrado en $\bar{X}$ y se extiende $z_{\alpha/2} \cdot \sigma/\sqrt{n}$ a cada lado de $\bar{X}$.
- La longitud del intervalo es $2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ que NO es aleatoria.

Cálculo IC media - Población normal, σ^2 conocida

Si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución $N(\mu, \sigma^2)$ con σ conocida, y $x_1, x_2, \ldots, x_n$ es una realización de dicha muestra. Un intervalo de confianza del $100.(1-\alpha)\%$ para μ está dado por

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Ejemplo 1

La duración de la batería de cierto modelo de teléfono movil se puede aproximar por una distribución normal con una desviación típica de 5 meses. Se toma una muestra aleatoria simple de 10 baterías y se obtienen las siguientes duraciones (en meses)

$$33, 34, 26, 37, 30, 39, 26, 31, 36, 19$$

Encontrar un intervalo de confianza del 95% para la duración media de ese modelo de batería.

Elección del tamaño de muestra

- El error de estimación , $\mathrm{E} = |\bar{X} - \mu| \leq z_{\alpha/2}\sigma/\sqrt{n}$, con un nivel de confianza del $100 \cdot (1 - \alpha)\%$
- Si la situación (experimento) lo permite, fijado $(1 - \alpha)$, puede elegirse n, el tamaño de la muestra, de manera tal que el error en la estimación no exceda cierta cota E(fijada de antemano). En tal caso

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

- La amplitud del intervalo de confianza es $2 \cdot z_{\alpha/2}\sigma/\sqrt{n}$
- Si se quiere estimar un intervalo de amplitud menor o igual a A, entonces debe tomarse una muestra de tamaño, $n \geq \left(\frac{2z_{\alpha/2}\sigma}{A}\right)^2$

Ejemplo 2

El tiempo de conexión a internet de los alumnos de cierta universidad sigue una distribución normal con desviación estándar igual a 15 minutos. Para estimar la media de la población se quiere calcular un intervalo de confianza que tenga una amplitud menor o igual que 6 minutos, con un nivel de confianza del 95%. Determinar cuál es el tamaño mínimo de la muestra que es necesario observar.

Ejemplo 3

Se estima que el tiempo de reacción de un conductor ante un obstáculo imprevisto tiene una distribución normal con desviación estándar de 0,05 segundos. Si se quiere conseguir que el error de estimación de la media no supere los 0,01 segundos con un nivel de confianza del 99%,
¿qué tamaño mínimo ha de tener la muestra de tiempos de reacción?

2. IC media - población normal, σ desconocida

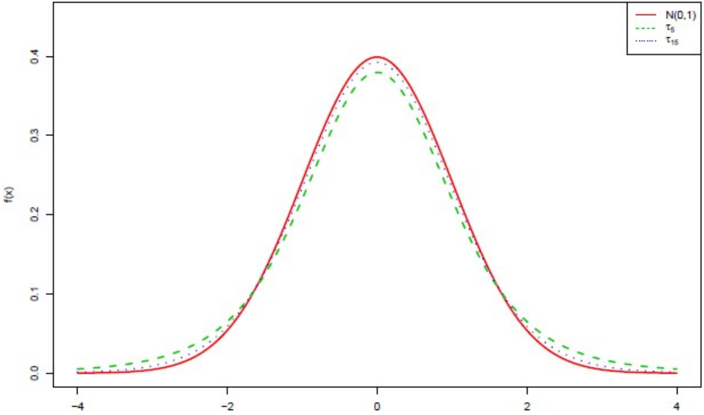
Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a normal con media μ y desviación estándar σ desconocidas.

Entonces la v.a.

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

tiene distribución t de Student con $n-1$ grados de libertad.$\bar{X}$ y S son la media muestral y la desviación estándar muestral, respectivamente.

Distribución t

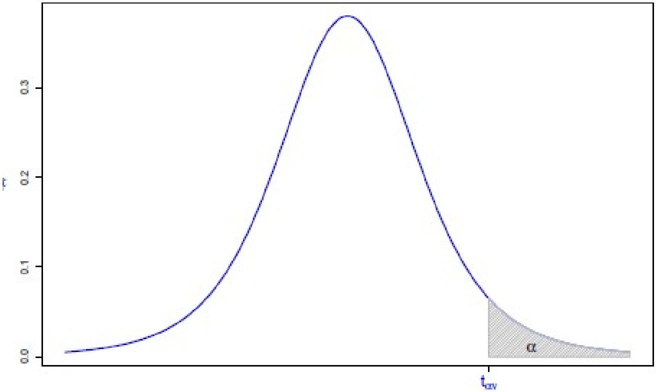


2. IC media - población normal, σ desconocida

De manera similar a lo hecho anteriormente, resulta que un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para la media μ de una población normal, cuando no se conoce σ es

$$\left(\bar{x} - t_{\frac{\alpha}{2},(n-1)} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}/,(n-1)} \cdot \frac{s}{\sqrt{n}} \right)$$

donde $\bar{x}, s$ son la media y desviación estándar muestrales, respectivamente, y $t_{\frac{\alpha}{2},\nu}$ es el valor crítico tal que $P(T > t_{\frac{\alpha}{2},\nu}) = \alpha/2$


0.0
0.1
0.2
0.3
α

Ejemplo 4

El número de errores diarios que se cometen al intentar conectar con una determinada red informática se distribuye normalmente con media desconocida. Para intentar conocer dicha media se realiza un M.A.S. de tamaño 10 días; resultando en

$$2;\ 3;\ 4;\ 5;\ 4;\ 3;\ 5;\ 0;\ 0;\ 1$$

errores.

a) Obtenga un intervalo de confianza para la media de errores cometidos diariamente con un nivel de significancia del 1%.

3a). IC media - muestras grandes, σ conocida

Sea $X_1, \ldots, X_n$ muestra aleatoria con media μ y desviación estándar $\sigma < \infty$.

Si n es grande ($n \geq 30$) $\Rightarrow$ Teo. Central del Límite dice que $\bar{X}$ es aproximadamente normal y que $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ es aproximadamente $N(0,1)$. Así

$$P(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha$$

Similarmente a lo hecho anteriormente,

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

es un intervalo de confianza del $100.(1-\alpha)\%$ para la media μ cuando el tamaño de la muestra es grande y σ es conocida.

3b. IC media - muestras grandes, σ desconocida

Sea $X_1, \ldots, X_n$ muestra aleatoria con media μ y desviación estándar σ desconocida.

Se usa S como estimación de σ, si n es grande ($n \geq 30$)
$\Rightarrow Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ es aproximadamente $N(0,1)$ y

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

es un intervalo de confianza del $100.(1-\alpha)\%$ para la media μ cuando el tamaño de la muestra es grande y σ NO es conocida.

Ejemplo 5

Un método para resolver la carencia de energía eléctrica requiere de la construcción de plantas eléctricas flotantes unas millas mar adentro. Se necesita una estimación de la densidad del tráfico naval en el área para evitar posibles colisiones de un barco con la planta flotante (aunque anclada).

a) Sea X: número de barcos que pasan por día en un radio de 10 kilómetros de una cierta planta flotante. ¿Qué distribución propondría para X?
b) Si quiere estimar el número promedio de barcos que pasan por día en un radio de 10km de esa planta, ¿cómo lo haría?

Continuación ejemplo 5

c) El número de barcos que pasan en un radio de 10 kilómetros de la ubicación propuesta de la planta eléctrica registrado en 60 días de verano tuvo una media de $\bar{x} = 7.2$. Suponiendo $\sigma = 2$ conocida, calcular un intervalo de confianza del 95% para el número promedio de barcos que pasan dentro de un radio de 10 kilómetros de localización de la planta eléctrica.

d) De una muestra de 90 observaciones diarias de barcos en los meses de junio a agosto se obtuvieron los siguientes datos: $\bar{x} = 4.7, s^2 = 4.9$. Obtener un intervalo de confianza del 95% para para el número promedio de barcos que pasan dentro de un radio de 10 kilómetros de localización de la planta eléctrica en los meses de invierno.

Ejemplo 6

Se hicieron pruebas para estimar el número promedio μ de horas por mes que pasa un niño en edad escolar frente al televisor, obteniéndose dos intervalos de confianza:

$$36.7 \leq \mu \leq 38.5 \quad ; \quad 34.8 \leq \mu \leq 40.4$$

a) ¿Cuál es el valor del promedio muestral de horas mensuales que pasaron los niños frente al televisor?

b) Uno de los intervalos fue calculado con un nivel de confianza del 90% y el otro con el 99%. Si se usó la misma muestra para ambos intervalos, ¿cuál de ellos se calculó con un 90% de confianza? Justificar

Intervalos de Confianza para σ^2 Dist. Normal

Sea $X_1, \ldots, X_n$ muestra aleatoria de una v.a normal con media μ y desviación estándar σ desconocidas.

Entonces la v.a.

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2, \text{ con n} - 1 \text{ g.l}$$

Y un intervalo del $100 \cdot (1-\alpha)\%$ de confianza para σ^2 de una distribución normal es:

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,(n-1)}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,(n-1)}} \right)$$

donde s^2 es la varianza muestral.

Ejemplo 7

El número de errores diarios que se cometen al intentar conectar con una determinada red informática se distribuye normalmente con media desconocida. Para intentar conocer dicha media se realiza un M.A.S. de tamaño 10 días; resultando en

$$2;\ 3;\ 4;\ 5;\ 4;\ 3;\ 5;\ 0;\ 0;\ 1$$

errores.

Halle un intervalo de confianza de 99% para la desviación estándar de la población muestreada

Intervalos de confianza para proporciones

Se toma una muestra aleatoria de tamaño n de una población

- Sea X = n^o de individuos de la muestra que pertencen a una clase de interés (solo dos clases)
- $\hat{P} = \frac{X}{n}$ estimador puntual de la proporción p de la población que pertenece a dicha clase.
- Si n es grande ($n \geq 30$) por Teo. DeMoivre -Laplace, $Z = \frac{X-np}{\sqrt{np(1-p)}} = \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$ es aproximadamente normal estándar
- $\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$ es un intervalo del $100\cdot(1-\alpha)\%$ de confianza para la proporción poblacional p.

Ejemplo 8

Una muestra de 100 votantes elegidos al azar entre todos los de un distrito dado, indicó que 55 de ellos estaban a favor de un determinado candidato.
a) Halle un intervalo de confianza para la proporción de todos los votantes que estaban a favor de ese candidato, con un nivel de significación del 5%.

Situaciones:

1. Un veterinario desea comparar el efecto de dos dietas de engorde en terneros
2. Se desea estudiar los efectos de dos marcas distintas de combustible en la eficiencia operativa de motores de automóviles (k/l).
3. Se quiere comparar la vida útil de pilas Duracell y pilas Eveready
4. Un médico desea comparar el efecto de dos concentraciones diferentes de una droga antidepresiva en la capacidad de atención.

Preguntas:

- ¿Qué similitudes hay con los casos vistos hasta ahora?
- ¿Qué diferencias hay con los casos vistos hasta ahora?

IC diferencia de medias

En algunas aplicaciones interesa estimar la diferencia entre dos parámetros poblacionales, como por ejemplo, la diferencia entre las medias de dos poblaciones distintas.

Suposiciones

- $X_{11}, X_{12}, \ldots X_{1n_1}$, es una muestra aleatoria de tamaño n_1 de la población 1 (X_1).
- $X_{21}, X_{22}, \ldots X_{2n_2}$, es una muestra aleatoria de tamaño n_2 de la población 2 (X_2).
- Las dos poblaciones, X_1 y X_2, son independientes.

IC diferencia de medias - Poblaciones normales, varianzas conocidas

Si $X_{11}, X_{12}, \ldots X_{1n_1}$ y $X_{21}, X_{22}, \ldots X_{2n_2}$ son muestras aleatorias de tamaños n_1 y n_2 de dos poblaciones normales $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ con σ_1 y σ_2 conocidas, entonces

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

y un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $\mu_1 - \mu_2$ es

$$(\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

donde

$\bar{x}_i$, es la media muestral de la población i, $i = 1, 2$.

IC diferencia de medias- Poblaciones normales, varianzas desconocidas

Si $X_{11}, X_{12}, \ldots X_{1n_1}$ y $X_{21}, X_{22}, \ldots X_{2n_2}$ son muestras aleatorias de tamaños n_1 y n_2 de dos poblaciones normales $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ con σ_1 ,σ_2 desconocidas pero iguales $(\sigma_1 = \sigma_2)$, entonces

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$$

Un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $\mu_1 - \mu_2$ es

$(\bar{x_1} - \bar{x_2} - t_{\alpha/2,(n_1+n_2-2)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x_1} - \bar{x_2} + t_{\alpha/2,(n_1+n_2-2)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$

donde

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$\bar{x}_i$, s_i^2 son la media y varianza muestrales, respectivamente, de la población i, $i = 1, 2$.

IC Diferencia de Proporciones

Si se toman dos muestras aleatorias de tamaño n_1 y n_2 de dos poblaciones X_1 y X_2, donde X_i cuenta el número de observaciones que pertenecen a la clase de interés en la muestra i, y siempre que n_1 y n_2 sean lo suficientemente grandes, entonces:

$$\hat{P}_1 = X_1/n_1; \quad \hat{P}_2 = X_2/n_2$$

tienen aproximadamente distribución normal, y un intervalo bilateral del $100 \cdot (1-\alpha)\%$ de confianza para $p_1 - p_2$ es

$$\left(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

donde
$\hat{p}_i = x_i/n_i$, es la proporción muestral para la población i, $i = 1, 2$.

Ejemplo 9

Se usan dos máquinas diferentes para fabricar tapas de plástico para botellas de dicho material. Una tapa se considera "defectuosa" si está excesivamente deformada o descolorida. Se seleccionan dos muestras aleatorias, cada una de tamaño 300. Se encuentran 15 tapas defectuosas fabricadas por la máquina 1 y 8 defectuosas fabricadas por la máquina 2.
Encuentre un intervalo del 95% de confianza para la diferencia de proporciones de tapas defectuosas.

Ejercicios

1. Se tomó una muestra de 50 cascos para motociclistas a los que se les realizó una prueba de impactos, resultando 18 de ellos con algún tipo de daño. Encontrar un intervalo del 95% de confianza para la proporción de cascos que podrían resultar dañados en este tipo de pruebas.

2. Las tensiones de rotura (en Kp.) de 5 cables de un determinado metal fueron 660, 460, 540, 580, 550. Suponiendo normalidad para las tensiones, estimar la tensión media de rotura mediante un intervalo de confianza del 95%

3. Una máquina produce piezas mecánicas de forma cilíndrica. Se toma una muestra de piezas y los diámetros son: 1.01, 0.97, 1.04, 0.99, 0.98, 0.99, 1.01 y 1.03 cm. Encontrar un intervalo del 99% de confianza para el diámetro medio de las piezas de la máquina. ¿Hay que hacer alguna suposición?

4. Se hicieron pruebas para estimar el número promedio μ de horas por mes que pasa un niño en edad escolar frente al televisor, obteniéndose dos intervalos de confianza:

$$36.7 \leq \mu \leq 38.5 \quad ; \quad 34.8 \leq \mu \leq 40.4$$

a) ¿Cuál es el valor del promedio muestral de horas mensuales que pasaron los niños frente al televisor?

b) Uno de los intervalos fue calculado con un nivel de confianza del 90% y el otro con el 99%. Si se usó la misma muestra para ambos intervalos, ¿cuál de ellos se calculó con un 90% de confianza? Justificar