Probabilidades y Estadística (C) – 2019

Intervalos de Confianza – Parte 3 (7/11)

Resumen bajo normalidad: $X_i \sim \mathcal{N}(\mu, \sigma^2)$

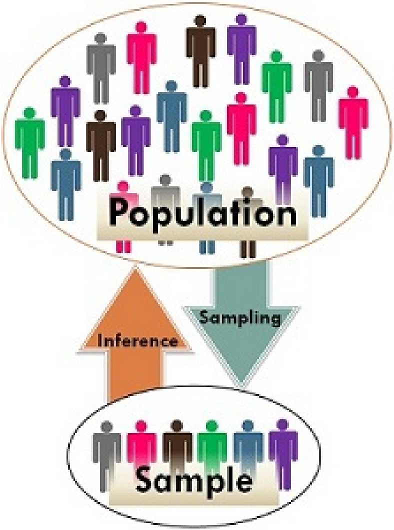
Parámetro	Supuesto	Pivot
$\mu = E(X)$	Normales σ^2 conocido	$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$
$\mu = E(X)$	Normales σ^2 DESconocido	$\sqrt{n}\frac{\overline{X}_n - \mu}{S} \sim T_{n-1}$
$\sigma^2 = V(X)$	Normales μ conocido	$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$
$\sigma^2 = V(X)$	Normales μ DESconocido	$\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi_{n-1}^2$

Resumen bajo normalidad: $X_i \sim \mathcal{N}(\mu, \sigma^2)$

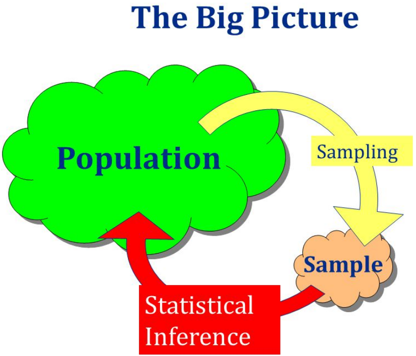
Parámetro	Supuesto	Pivot
$\mu = E(X)$	Normales σ^2 conocido	$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$
$\mu = E(X)$	Normales σ^2 DESconocido	$\sqrt{n}\frac{\overline{X}_n - \mu}{S} \sim T_{n-1}$
$\sigma^2 = V(X)$	Normales μ conocido	$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$
$\sigma^2 = V(X)$	Normales μ DESconocido	$\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2_{n-1}$

- Intervalo estimado nivel 0.95 para μ con $\overline{X}_{\text{obs}} = 7.34$, $\sigma^2 = 2$ y $n = 5$.
- Intervalo estimado nivel 0.95 para μ con $\overline{X}_{\text{obs}} = 7.34$, $S^2_{\text{obs}} = 2$ y $n = 5$.
- $X_i \sim \mathcal{N}(\mu, 1)$. Intevalo reportado: $(2.18, 3, 41)$, con $n = 10$. Hallar $\overline{x}$ y el nivel $1 - \alpha$.

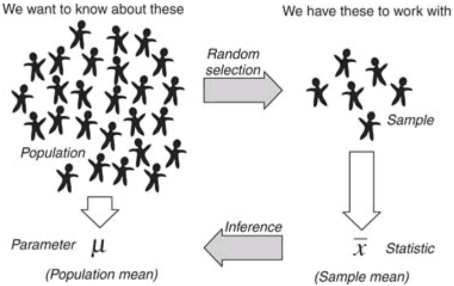
Población vs. Muestra



Población vs. Muestra - En forma abstracta...



Población vs. Muestra -



	Valor Poblacional	Valor Muestral
Media	μ	$\overline{X}$
Desvio Standar	σ	S
Proporción	p	$\widehat{p}$
Acumulada	F	$\widehat{F}$ LA empírica (o $F(\cdot\,,\widehat{\theta}_n)$)
Densidad	f	Histograma (o $f(\cdot\,,\widehat{\theta}_n)$)

Intervalos de confianza ASINTÓTICOS

- Intervalo de confinanza de nivel asintótico $1-\alpha$ para el parámetro θ:

$$\lim_{n\to\infty} P\left(a(X_1,\ldots,X_n) \leq \theta \ < \ b(X_1,\ldots,X_n)\right) = 1-\alpha \ .$$

- Pivot (asintótico):

$H(X_1,\ldots,X_n,\theta) \longrightarrow^{d}$ Distribución *tabulada*, cuando $n \to \infty$

- Encerramos al pivot entre los percentiles de su distribución límite y despejamos θ.

Intervalos de confianza ASINTÓTICOS

- Intervalo de confinanza de nivel asintótico $1-\alpha$ para el parámetro θ:

$$\lim_{n\to\infty} P\left(a(X_1,\ldots,X_n) \leq \theta \ < \ b(X_1,\ldots,X_n)\right) = 1-\alpha\ .$$

- Pivot (asintótico):

$H(X_1,\ldots,X_n,\theta) \longrightarrow^d$ Distribución *tabulada*, cuando $n\to\infty$

- Encerramos al pivot entre los percentiles de su distribución límite y despejamos θ.

En adelante, utilizaremos $W_n \approx W$ o $W_n \overset{(a)}{\sim} W$ para denotar que (W_n) converge en distribución a W:

$W_n \overset{(a)}{\sim} W \quad W_n \approx W$ es equivalente a decir que $\quad W_n \longrightarrow^d W$

Intervalo de confianza asintótico para $\mu = E(X)$ - σ^2 conocido

- TCL:

$$\frac{(\overline{X}_n - \mu)}{\sqrt{\sigma^2/n}} \overset{(a)}{\sim} \mathcal{N}(0,1)$$

- z_β: $P(Z \geq z_\beta) = \beta$, entonces

$$P\left(-z_{\alpha/2} \ < \ \frac{(\overline{X}_n - \mu)}{\sqrt{\sigma^2/n}} \ < \ z_{\alpha/2}\right) \approx 1 - \alpha$$

- luego

$$\left(\overline{X}_n - \frac{\sqrt{\sigma^2}\, z_{\alpha/2}}{\sqrt{n}} \quad , \quad \overline{X}_n + \frac{\sqrt{\sigma^2}\, z_{\alpha/2}}{\sqrt{n}}\right)$$

  es un IC de nivel asintótico $1 - \alpha$ para $\mu = E(X)$.

- ¿Qué hacemos si σ^2 es desconocido?

Teorema de Slutzky (sin demostrar)

Ejemplo:

$$\frac{\left(\overline{X}_n - \mu\right)}{\sqrt{\sigma^2/n}} \longrightarrow^d \mathcal{N}(0,1)\ , \qquad \frac{\sqrt{\sigma^2/n}}{\sqrt{S^2/n}} \longrightarrow^P 1$$

Entonces, por Slutzky

$$\frac{\sqrt{\sigma^2/n}}{\sqrt{S^2/n}} \frac{\left(\overline{X}_n - \mu\right)}{\sqrt{\sigma^2/n}} = \frac{\left(\overline{X}_n - \mu\right)}{\sqrt{S^2/n}} \longrightarrow^d \mathcal{N}(0,1)$$

Teorema de Slutzky: Si $W_n \longrightarrow^d W$ y $A_n \longrightarrow^P$ cte., entonces $A_n W_n \longrightarrow^d$ cte.W.

Intervalo para $\mu = E(X)$ con σ^2 Desconocido

- TCL

$$\frac{(\overline{X}_n - \mu)}{\sqrt{\sigma^2/n}} \overset{(a)}{\sim} \mathcal{N}(0,1)$$

- Además, $\frac{\sqrt{\sigma^2/n}}{\sqrt{S^2/n}} \longrightarrow^P 1$ y por Slutzky vale que

$$\frac{\sqrt{\sigma^2/n}}{\sqrt{S^2/n}} \frac{(\overline{X}_n - \mu)}{\sqrt{\sigma^2/n}} = \frac{(\overline{X}_n - \mu)}{\sqrt{S^2/n}} \approx \mathcal{N}(0,1)$$

- z_β: con $P(Z \geq z_\beta) = \beta$, entonces

$$P\left(-z_{\alpha/2} \ < \ \frac{(\overline{X}_n - \mu)}{\sqrt{S^2/n}} \ < \ z_{\alpha/2}\right) \approx 1 - \alpha$$

- luego,

$$\left(\overline{X}_n - \frac{\sqrt{S^2}\, z_{\alpha/2}}{\sqrt{n}} \quad , \quad \overline{X}_n + \frac{\sqrt{S^2}\, z_{\alpha/2}}{\sqrt{n}}\right)$$

es un IC de nivel asintótico $1 - \alpha$ para $\mu = E(X)$.

Mundo asintótico

- Intervalo estimado nivel 0.95 para μ con $\overline{X}_{\text{obs}} = 7.34$, $\sigma^2 = 2$ y $n = 120$.

- Intervalo estimado nivel 0.95 para μ con $\overline{X}_{\text{obs}} = 7.34$, $S^2_{\text{obs}} = 2$ y $n = 120$.

$X_i \sim B(1,p)$: Intervalo de confianza asintótico para p

- TCL

$$\frac{\sqrt{n}\left(\overline{X}_n - p\right)}{\sqrt{p(1-p)}} \overset{(a)}{\sim} \mathcal{N}(0,1)$$

- Además, $\widehat{p} \to p$ y por Slutzky

$$\frac{\sqrt{p(1-p)}}{\sqrt{\widehat{p}(1-\widehat{p})}} \frac{\sqrt{n}\,(\widehat{p}-p)}{\sqrt{p(1-p)}} = \frac{\sqrt{n}\,(\widehat{p}-p)}{\sqrt{\widehat{p}(1-\widehat{p})}} \approx \mathcal{N}(0,1)$$

- Por consiguiente,

$$\left(\widehat{p} - \frac{\sqrt{\widehat{p}(1-\widehat{p})}\, z_{\alpha/2}}{\sqrt{n}} \quad , \quad \widehat{p} + \frac{\sqrt{\widehat{p}(1-\widehat{p})}\, z_{\alpha/2}}{\sqrt{n}}\right)$$

es un intervalo de confianza de nivel asintótico $1-\alpha$ para p.

Algunos ejemplos

Para estimar la proporción de personas a favor de cierta ley, se realiza una encuesta y su optiene que el 38% de los encuestados está a favor.

- ¿Cuál es el parámetro de interés?
- ¿Cuál es su estimador?
- ¿Cuál es su estimación en base a los datos obtenidos ?
- Hallar el intervalo de confianza estimado de nivel asintótico 0.90 para la proporción de personas a favor de la ley.

Algunos ejemplos

Para estimar la proporción de personas a favor de cierta ley, se realiza una encuesta y su optiene que el 38% de los encuestados está a favor.

- ¿Cuál es el parámetro de interés?
- ¿Cuál es su estimador?
- ¿Cuál es su estimación en base a los datos obtenidos ?
- Hallar el intervalo de confianza estimado de nivel asintótico 0.90 para la proporción de personas a favor de la ley. ¿qué información le está faltando?

Algunos ejemplos

Para estimar la proporción de personas a favor de cierta ley, se realiza una encuesta y su optiene que el 38% de los encuestados está a favor.

- ¿Cuál es el parámetro de interés?
- ¿Cuál es su estimador?
- ¿Cuál es su estimación en base a los datos obtenidos ?
- Hallar el intervalo de confianza estimado de nivel asintótico 0.90 para la proporción de personas a favor de la ley. ¿qué información le está faltando?

El tamaño n de la muestra SI importa!

EL TCL y la distribución de algunos estimadores

$$\widehat{\mu}_n \overset{(a)}{\sim} \mathcal{N}(\mu, \mathsf{se}^2) \ , \qquad \mathsf{se}^2 = \sigma^2/n \ , \ \ \widehat{\mathsf{se}} = \sqrt{S^2/n} \ ,$$
$$\mathsf{IC} = (\widehat{\mu_n} - z_{\alpha/2}\widehat{\mathsf{se}} \ , \ \widehat{\mu_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

EL TCL y la distribución de algunos estimadores

$$\widehat{\mu}_n \overset{(a)}{\sim} \mathcal{N}(\mu, \mathsf{se}^2)\ , \quad \mathsf{se}^2 = \sigma^2/n\ ,\ \ \widehat{\mathsf{se}} = \sqrt{S^2/n}\ ,$$
$$\mathsf{IC} = (\widehat{\mu_n} - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{\mu_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

$$\widehat{p}_n \overset{(a)}{\sim} \mathcal{N}(p, \mathsf{se}^2)\ ,\ \ \mathsf{se}^2 = p(1-p)/n\ ,\ \widehat{\mathsf{se}} = \sqrt{\widehat{p}_n(1-\widehat{p}_n)/n}\ ,$$
$$\mathsf{IC} = (\widehat{p_n} - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{p_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

EL TCL y la distribución de algunos estimadores

$$\widehat{\mu}_n \overset{(a)}{\sim} \mathcal{N}(\mu, \mathsf{se}^2) \;, \quad \mathsf{se}^2 = \sigma^2/n \;, \;\; \widehat{\mathsf{se}} = \sqrt{S^2/n} \;,$$
$$\mathsf{IC} = (\widehat{\mu_n} - z_{\alpha/2}\widehat{\mathsf{se}} \;, \; \widehat{\mu_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

$$\widehat{p}_n \overset{(a)}{\sim} \mathcal{N}(p, \mathsf{se}^2) \;, \;\; \mathsf{se}^2 = p(1-p)/n \;, \; \widehat{\mathsf{se}} = \sqrt{\widehat{p}_n(1-\widehat{p}_n)/n} \;,$$
$$\mathsf{IC} = (\widehat{p_n} - z_{\alpha/2}\widehat{\mathsf{se}} \;, \; \widehat{p_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

$$X_i \sim \mathcal{P}(\lambda) \;, \quad \widehat{\lambda}_n \overset{(a)}{\sim} \mathcal{N}(\lambda, \mathsf{se}^2) \;, \;\; \mathsf{se}^2 = \lambda/n \;, \; \widehat{\mathsf{se}} = \sqrt{\widehat{\lambda_n}/n} \;,$$
$$\mathsf{IC} = (\widehat{\lambda_n} - z_{\alpha/2}\widehat{\mathsf{se}} \;, \; \widehat{\lambda_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

En muchos casos,

$$\widehat{\theta}_n \overset{(a)}{\sim} \mathcal{N}(\theta, \mathsf{se}^2) \quad \equiv \quad \frac{\widehat{\theta}_n - \theta}{\mathsf{se}} \overset{(a)}{\sim} \mathcal{N}(0,1)$$

EL TCL y la distribución de algunos estimadores

$$\widehat{\mu}_n \overset{(a)}{\sim} \mathcal{N}(\mu, \mathsf{se}^2)\ , \quad \mathsf{se}^2 = \sigma^2/n\ ,\ \ \widehat{\mathsf{se}} = \sqrt{S^2/n}\ ,$$
$$\mathsf{IC} = (\widehat{\mu_n} - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{\mu_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

$$\widehat{p}_n \overset{(a)}{\sim} \mathcal{N}(p, \mathsf{se}^2)\ ,\ \ \mathsf{se}^2 = p(1-p)/n\ ,\ \widehat{\mathsf{se}} = \sqrt{\widehat{p}_n(1-\widehat{p}_n)/n}\ ,$$
$$\mathsf{IC} = (\widehat{p_n} - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{p_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

$$X_i \sim \mathcal{P}(\lambda)\ , \quad \widehat{\lambda}_n \overset{(a)}{\sim} \mathcal{N}(\lambda, \mathsf{se}^2)\ ,\ \ \mathsf{se}^2 = \lambda/n\ ,\ \widehat{\mathsf{se}} = \sqrt{\widehat{\lambda_n}/n}\ ,$$
$$\mathsf{IC} = (\widehat{\lambda_n} - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{\lambda_n} + z_{\alpha/2}\widehat{\mathsf{se}})$$

En muchos casos,

$$\widehat{\theta}_n \overset{(a)}{\sim} \mathcal{N}(\theta, \mathsf{se}^2) \quad \equiv \quad \frac{\widehat{\theta}_n - \theta}{\mathsf{se}} \overset{(a)}{\sim} \mathcal{N}(0,1)$$

Si podemos estimar se con $\widehat{\mathsf{se}}$ tenemos que

$(\widehat{\theta}_n - z_{\alpha/2}\widehat{\mathsf{se}}\ ,\ \widehat{\theta}_n + z_{\alpha/2}\widehat{\mathsf{se}})$ es IC de nivel asintótica $1-\alpha$.

Cómo se informa: $\pm$

On the other hand, the invariance of retrograde speeds distribution suggests that the endogenous and KHC576-mCherry-Pex26 motors contribute similarly to retrograde transport. In addition, KHC576-mCherry-Pex26 did not affect the median values of plus and minus-end directed run lengths (Table 1) and the number of reversions in the trajectories (1.8 $\pm$ 0.1 reversions/trajectory).

- Se informa poniendo el valor de la estimación y **”su error”**.
- ¿A qué llamamos ahora **”error”**?

Error de estimación

Definición: llamamos error de una estimación a la estimación del desvío (exacto o aproximado) del estimador con el cual estimamos.

- Estimador: $\widehat{\theta}_n$
- Estimación: $\widehat{\theta}_{n,\mathsf{obs}}$
- $\mathsf{se} = \sqrt{V(\widehat{\theta}_n)}$ o $\mathsf{se} \approx \sqrt{V(\widehat{\theta}_n)}$
- Error de estimación: $\widehat{\mathsf{se}}_{\mathsf{obs}}$

Diferentes culturas

- Intervalo estimado

$$\left(\widehat{\theta}_{n,\mathsf{obs}} - z_{\alpha/2}\,\widehat{\mathsf{se}}_{\mathsf{obs}} \quad , \quad \widehat{\theta}_{n,\mathsf{obs}} + z_{\alpha/2}\,\widehat{\mathsf{se}}_{\mathsf{obs}}\right)$$

- Otra manera de informar

$$\widehat{\theta}_{n,\mathsf{obs}} \quad (\pm\widehat{\mathsf{se}}_{\mathsf{obs}})$$

$$(\widehat{\theta}_{n,\mathsf{obs}} \quad \pm\,\widehat{\mathsf{se}}_{\mathsf{obs}})$$

Errores de estimación

$$\widehat{\mu}_n \ , \qquad \mathsf{se}^2 = V(\widehat{\mu}_n) = \sigma^2/n \ , \ \ \widehat{\mathsf{se}} = \sqrt{S^2/n} \ ,$$

Errores de estimación

$$\widehat{\mu}_n \ , \quad \mathsf{se}^2 = V(\widehat{\mu}_n) = \sigma^2/n \ , \ \ \widehat{\mathsf{se}} = \sqrt{S^2/n} \ ,$$

$$\widehat{p}_n \ , \ \ \mathsf{se}^2 = V(\widehat{p}_n) = p(1-p)/n \ , \ \widehat{\mathsf{se}} = \sqrt{\widehat{p}_n(1-\widehat{p}_n)/n} \ ,$$

Error de estimación para la mediana?

- Desvio del Estimador:

$$\mathsf{se} = \mathsf{se}(\mathsf{med}(X_1, \ldots, X_n)) = \sqrt{V\{\mathsf{med}(X_1, \ldots, X_n)\}} = ???$$

- $\widehat{\mathsf{se}} = ??$

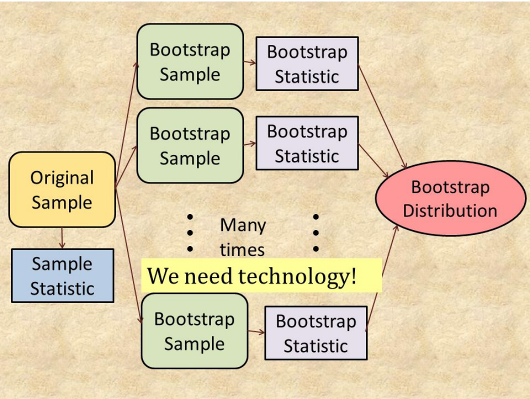
Error de estimación para la mediana?

- Desvio del Estimador:

$$\mathsf{se} = \mathsf{se}(\mathsf{med}(X_1, \ldots, X_n)) = \sqrt{V\{\mathsf{med}(X_1, \ldots, X_n)\}} = ???$$

- $\widehat{\mathsf{se}} = ??$
- Bootstrap! $\widehat{\mathsf{se}}_{boot}$

Esquema Bootstrap



Esquema Bootstrap. All of Statistics. Wasserman. Cap 8.

Bootstrap Variance Estimation

1. Draw $X_1^*, \ldots, X_n^* \sim \widehat{F}_n$.
2. Compute $T_n^* = g(X_1^*, \ldots, X_n^*)$.
3. Repeat steps 1 and 2, B times, to get $T_{n,1}^*, \ldots, T_{n,B}^*$.
4. Let

$$v_{\text{boot}} = \frac{1}{B} \sum_{b=1}^{B} \left(T_{n,b}^* - \frac{1}{B} \sum_{r=1}^{B} T_{n,r}^* \right)^2. \tag{8.1}$$

Bootstrap for The Median

```
Given data X = (X(1), ..., X(n)):

T <- median(X)
Tboot <- vector of length B
for(i in 1:B){
    Xstar <- sample of size n from X (with replacement)
    Tboot[i] <- median(Xstar)
    }
se <- sqrt(variance(Tboot))
```

Intevalos Bootstrap Normal

- $\widehat{\theta}_n$ asintóticamente normal si

$$\frac{\widehat{\theta}_n - \theta}{\mathsf{se}} \approx \mathcal{N}(0,1)$$

- Sea $\widehat{\mathsf{se}}_{\mathsf{boot}}$ el estimador bootstrap de se

- Intervalo boot normal nivel $1-\alpha$:

$$(\widehat{\theta}_n - z_{\alpha/2}\, \widehat{\mathsf{se}}_{\mathsf{boot}}\ ,\ \widehat{\theta}_n + z_{\alpha/2}\, \widehat{\mathsf{se}}_{\mathsf{boot}})$$