Seja $(X_1, \dots, X_n)$ a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}$.

Um intervalo de confiança com coeficiente de confiança γ é um intervalo aleatório que satisfaz:

$$P_\theta(I_1(\boldsymbol{X}_n) \leq \theta \leq I_2(\boldsymbol{X}_n)) = \gamma \forall \theta \in \Theta$$

em que $\boldsymbol{X}_n = (X_1 \dots, X_n)$

Observação:

Se o intervalo aleatório $[I_1(\boldsymbol{X}_n), I_2(\boldsymbol{X}_n)$ satisfaz

$$P_\theta(I_1(\boldsymbol{X}_n) \leq \theta \leq I_2(\boldsymbol{X}_n)) \geq \gamma \forall \theta \in \Theta$$

Então o intervalo aleatório é um intervalo de confiança com pelo menos coeficiente de confiança γ

Observação: $I_1(\boldsymbol{X}_n)$ e $I_2(\boldsymbol{X}_n)$ são estatísticas e são tais que $I_1(\boldsymbol{X}_n) \leq I_2(\boldsymbol{X}_n)$

Observação:
Quando substituímos $(\boldsymbol{X}_n)$ pela amostra observada $(\boldsymbol{x}_n)$ temos que

$$P_\theta(I_1(\boldsymbol{x}_n) \leq \theta \leq I_2(\boldsymbol{x}_n)) = \begin{cases} 1, & \text{se } \theta \in [I_1(\boldsymbol{x}_n), I_2(\boldsymbol{x}_n)] \\ 0, & \text{caso contrário} \end{cases}$$

IC sob normalidade para μ com σ^2 conhecido

Em uma distribuição normal (θ, σ^2), por exemplo, conseguimos de forma genérica para qualquer $\gamma \in [0,1]$ encontrar pela tabela um valor de c_γ que satisfaça $P(-c_\gamma \leq Z \leq c_\gamma)$. Assim,

$$-c_\gamma \leq Z \leq c_\gamma \Leftrightarrow -c_\gamma \leq \frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\sigma^2}} \leq c_\gamma \Leftrightarrow \bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}$$

Portanto,

$$P_\theta\left(\bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}\right) = \gamma \forall \theta \in \Theta$$

Dessa forma,

$\bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}$ é um intervalo de confiança cujo coeficiente de confiança é γ.

Exemplo.

Considere que a amostra observada foi $1, 2.2, 3, 3.5$ de uma distribuição $N(\theta, 3)$.

Encontre o IC observado com coeficiente de confiança de 99%.

Primeiro encontramos a média: $\bar{x} = 2.42$.

Dessa forma, $c_{99\%} = 2.58$ e nosso intervalo de confiança é

$$\left[2.42 - 2.58\frac{\sqrt{3}}{\sqrt{4}}, 2.42 + 2.58\frac{\sqrt{3}}{\sqrt{4}}\right] = [0.18, 4.65]$$

Dessa forma, se repetirmos o experimento N vezes, esperamos que $\gamma = 99\%$ dos ICs observados contenham a quantidade de interesse.

Notação.

$$\mathrm{IC}(\theta, \gamma) = [I_1(\boldsymbol{X}_n), I_2(\boldsymbol{X}_n)]$$

Denotará o intervalo de confiança teórico

$$\mathrm{IC}_{\mathrm{Obs}}(\theta, \gamma) = [I_1(\boldsymbol{x}_n), I_2(\boldsymbol{x}_n)]$$

IC sob normalidade para μ quando σ^2 é desconhecido

Seja uma a.a de $X \sim N(\mu, \sigma^2)$ em que μ, σ^2 são desconhecidos. Então, o IC para μ com coeficiente de confiança γ é dado por

$$\mathrm{IC}(\mu, \gamma) = \left[\bar{X} - t_{\gamma, n-1}\sqrt{\frac{s^2(\boldsymbol{X}_n)}{n}}, \bar{X} + t_{\gamma, n-1}\sqrt{\frac{s^2(\boldsymbol{X}_n)}{n}}\right]$$

Em que $s^2(\boldsymbol{x}_n) = \frac{1}{n-1}\sum_{i=1}^{k}(X_i - \bar{x})^2$ e $t_{y,(n-1)}$ deve ser calculado da tabela $P(-t_{\gamma,n-1} \leq T_{n-1} \leq t_{\gamma,n-1}) = \gamma$, $T_{n-1} \sim \mathrm{t-Student}(n-1)$

Se a variância for desconhecida, substituímos σ^2 pelo estimador

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

e o valor de c_γ obtido de uma t-Student com $n-1$ graus de liberdade.

Justificativa:

Se uma a.a. de $X \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$.

1-) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
2-) $\sqrt{n}\frac{\bar{X}-\mu}{\sqrt{\sigma^2}} \sim N(0,1)$
3-) $\sum\limits_{i=1}^{n} \frac{(X_i-\bar{X})^2}{\sqrt{\sigma^2}} \sim \chi^2_{n-1}$
4-) $\sqrt{n}\frac{\bar{X}-\mu}{\sqrt{s^2}} = \sqrt{n}\frac{\bar{X}-\mu}{\sqrt{s^2}} \cdot \frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} = \frac{\sqrt{n}\frac{\bar{X}-\mu}{\sqrt{\sigma^2}}}{\sqrt{\frac{\sum(X_i-\bar{X})^2}{\sigma^2} \cdot \frac{1}{n-1}}}$
5-) Se $Z \sim N(0,1)$ e $W \sim \chi^2_k$, então $t = \frac{Z}{\sqrt{\frac{W}{k}}} \sim t_k$
6-) $\sqrt{n}\frac{(\bar{X}-\mu)}{\sqrt{s^2}} \sim t_{(n-1)}$

Exercício

Considere uma a.a de $X \sim N(\mu, \sigma^2)$, ambos parâmetros desconhecidos, em que X é o retorno de um ativo específico. Considere que a seguinte amostra foi observada:

$$(-1.3\%, 0.4\%, -1.7\%, 3.2\%, 0.7\%, -1.6\%, 1.0\%, 1.5\%, 1.2\%, -0.6\%)$$

Encontre o IC para μ, com coeficiente de confiança $\gamma = 99\%$

Temos que $s^2(\boldsymbol{x}_n) = 2.48$ e $\bar{x} = 0.28$. Da tabela, $t_{99\%,9} = 3.25$

Sendo assim,

$$\text{IC}_{\text{Obs}}(\mu, 99\%) = \left[0.28 - 3.25\sqrt{\frac{2.48}{10}}, 0.28 + 3.25\sqrt{\frac{2.48}{10}}\right] = [-1.34, 1.9]$$

IC sob normalidade para σ^2 com μ desconhecido

Seja uma a.a. de $X \sim N(\mu, \sigma^2)$ em que μ, σ^2 são desconhecidos.

O intervalo de confiança para σ^2 com coeficiente de confiança γ é dado por:

$$\mathrm{IC}(\sigma^2, \gamma) = \left[\frac{(n-1)s^2(\boldsymbol{X}_n)}{q^{(2)}_{\gamma,n-1}}, \frac{(n-1)s^2(\boldsymbol{X}_n)}{q^{(1)}_{\gamma,n-1}} \right]$$
$$\mathrm{IC}_{\mathrm{Obs}}(\sigma^2, \gamma) = \left[\frac{(n-1)s^2(\boldsymbol{x}_n)}{q^{(2)}_{\gamma,n-1}}, \frac{(n-1)s^2(\boldsymbol{x}_n)}{q^{(1)}_{\gamma,n-1}} \right]$$

em que $s^2(\boldsymbol{X}_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ e $s^2(\boldsymbol{x}_n) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$
E $q^{(1)}_{\gamma,n-1}, q^{(2)}_{\gamma,n-1}$ são obtidos calculando $P(q^{(1)}_{\gamma,n-1} \leq W \leq q^{(2)}_{\gamma,n-1}) = \gamma$ no qual $W \sim \chi^2_{n-1}$ e $P(\chi^2_{n-1} \leq q^{(1)}_{\gamma,n-1}) = \frac{1-\gamma}{2} = P(\chi^2_{n-1} \geq q^{(2)}_{\gamma,n-1})$

Demonstração:
Vamos construir um IC para a variância.

Note que

$$\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sqrt{\sigma^2}} \sim \chi^2_{n-1}$$

Note que
$W \sim \chi^2_{n-1}$, ou seja, $P(q^{(1)}_{\gamma,n-1} \leq W \leq q^{(2)}_{\gamma,n-1}) = \gamma$. Assim, $W = \frac{(n-1)s^2}{\sigma^2}$.

Dessa forma,

$$\frac{1}{q^{(2)}_{\gamma,n-1}} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{q^{(1)}_{\gamma,n-1}} \Leftrightarrow \frac{(n-1)s^2}{q^{(2)}_{\gamma,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{q^{(1)}_{\gamma,n-1}}$$

em que $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

Portanto,

$$P_\theta\left(\frac{(n-1)s^2}{q^{(2)}_{\gamma,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{q^{(1)}_{\gamma,n-1}} \right) = \gamma \forall \theta \in \Theta$$

Exercício:

Considere uma a.a de $X \sim N(0, \theta), \theta > 0$ em que X é o retorno de um ativo específico. Considere que a seguinte amostra foi observada:

$$(-1.3\%, 0.4\%, -1.7\%, 3.2\%, 0.7\%, -1.6\%, 1.0\%, 1.5\%, 1.2\%, -0.6\%)$$

Construa um intervalo de confiança para a variância (populacional) θ
com coeficiente de confiança $\gamma = 95\%$.

O IC para a variância é

$$\mathrm{IC}(\theta,\gamma)=\left[\frac{(n-1)s^2(\boldsymbol{x}_n)}{q^2_{\gamma,n-1}},\frac{(n-1)s^2(\boldsymbol{x}_n)}{q^1_{\gamma,n-1}}\right]$$

em que $s^2(\boldsymbol{x}_n)=\frac{1}{n-1}\sum_{i=1}^{k}(X_i-\bar{x})^2$
e $q^1_{\gamma,n-1},q^2_{\gamma,n-1}$ satisfazem as fronteiras que delimitam uma área de γ em torno da média. Nesse caso, como $n=10$, $q^1_9=2.7,q^2_9=19.023$.
Dessa forma, $\bar{x}=0.28,\sum_i x_i^2=2.308\Rightarrow s^2(\boldsymbol{x}_n)=\frac{10}{9}(2.308-0.28^2)=2.48$

Sendo assim

$$\mathrm{IC}_{\mathrm{Obs}}(\theta,95\%)=\left[\frac{9\cdot 2.48}{19.023},\frac{9\cdot 2.48}{2.7}\right]=[1.17,8.27]$$

Podemos concluir que, com uma confiança de 95%, a variância está nesse intervalo.

Intervalo de Confiança para a proporção

Seja $(\boldsymbol{x}_n)$uma amostra aleatória de $X\sim Ber(\theta)$ em que $\theta\in[0,1]$.
Sabemos que, pelo Teorema do Limite Central

$$\bar{X}\overset{a}{\approx}N\left(\theta,\frac{\theta(1-\theta)}{n}\right)$$

Dessa forma,

$$\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\theta(1-\theta)}}\overset{a}{\approx}N(0,1)$$

Além disso, pelo Teorema de Slutsky

$$\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}}\overset{a}{\approx}N(0,1)$$

Observe que, se $Z\sim N(0,1)$ então $-c_\gamma\le Z\le c_\gamma\Leftrightarrow -c_\gamma\le\frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}}\le c_\gamma$

$$\Leftrightarrow\bar{X}-c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\le\theta\le\bar{X}+c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

Então, seja c_γ tal que

$$P(-c_\gamma \leq N(0,1) \leq c_\gamma) = \gamma$$
$$\Rightarrow P_\theta\left(\bar{X} - c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \theta \leq \bar{X} + c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \approx \gamma \; \forall\theta \in \Theta = [0,1]$$

que melhora conforme $n \to \infty$

Logo, um IC aproximado com coeficiente de confiança γ é dado por

$$IC(\theta,\gamma) = \left[\bar{X} - c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + c_\gamma\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \cap \Theta$$

Exemplo

Seja $(\boldsymbol{x}_n)$ uma amostra aleatória de $X \sim \text{Ber}(\theta)$ em que $\theta \in [0,1]$.

$$X(w) = \begin{cases} 1, \text{ se w disser que vota no candidato} \\ \\ 0, c.\,c \end{cases}$$

A amostra observada foi $(0,0,0,1,0,0,0,1)$.
Encontre o IC para a proporção de intenção de votos no candidato considerando $\gamma = 99\%$

$\bar{x} = \frac{1}{4}, \bar{x}(1-\bar{x}) = \frac{3}{16}, n = 8, c_\gamma = 2.58$

$$\begin{aligned} \text{IC}_{\text{Obs}}(\theta, 99\%) &= \left[0.25 - 2.58\frac{\sqrt{3}}{4\sqrt{8}}, 0.25 - 2.58\frac{\sqrt{3}}{4\sqrt{8}}\right] \cap [0,1] \\ &= [0.25 - 0.39, 0.25 + 0.39] \cap [0,1] \\ &= [0, 0.64] \end{aligned}$$

Intervalos conservadores e otimistas

Um intervalo de confiança de proporção é dito ser conservador quando o calculamos tomando θ ***Na variância*** $(\theta(1-\theta))$ como o valor mais alto possível. No exemplo Bernoulli, o valor máximo para θ (derivando $\theta(1-\theta)$ e igualando a 0, $1 - 2\theta = 0$) é $\theta = 0.5$. Dessa forma, o IC conservador é calculado usando $\theta = 0.5$.

Por sua vez, um IC otimista é calculado usando o valor de θ obtido através do EMV para θ, no caso Bernoulli, usaríamos essa variância como $\bar{X}(1-\bar{X})$

Como interpretar intervalos de confiança

Importante:

Na estatística clássica (frequentista), devemos interpretar um intervalo de confiança $[a, b]$ com $\gamma = 0.95$ da seguinte forma:

"Com 95% de confiança, o intervalo $[a, b]$ conterá o valor da quantidade de interesse".

Isso é importante para diferenciar a interpretação frequentista (Theta do espaço paramétrico) da Bayesiana (Theta como variável aleatória). Dessa forma, estaria ***incorreto*** na estatística clássica dizer que

"O intervalo $[a, b]$ conterá a quantidade de interesse com probabilidade 95%" ou "O intervalo $[a, b]$ conterá a quantidade de interesse 95% das vezes"