Universidade Federal de Santa Maria
Departamento de Estatística

INTERVALO DE CONFIANÇA

Estatística Inferencial

Professora: Laís Helen Loose

Intervalo de confiança para média

σ conhecido

- Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição normal com média desconhecida μ e variância conhecida σ^2, isto é, $X_i \sim \mathrm{N}(\mu, \sigma^2)$.

- Na aula passado, vimos que:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1)$$

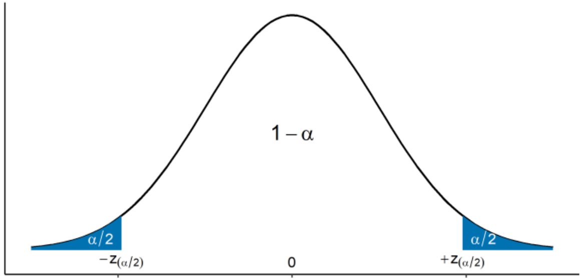
- Nesse caso, podemos encontrar dois valores a e b tais que a probabilidade dessa variável aleatória assumir um valor nesse intervalo seja igual a $1 - \alpha$.

$$P\left(a \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) = 1 - \alpha$$

σ conhecido

- Há vários pares (a, b) que satisfazem a equação acima. É desejável que o comprimento do intervalo (a, b) seja o menor possível.

- Como a distribuição normal padrão é simétrica em torno da origem, o intervalo (a, b) de comprimento mínimo também deve ser simétrico em torno da origem.



σ conhecido

- Assim,

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

  em que $z_{\alpha/2}$ é o quantil da distribuição normal padrão.

- Podemos isolar apenas o parâmetro μ no centro:

$$P\left(\overline{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

σ conhecido

- A partir da última equação, podemos construir um intervalo de confiança (com coeficiente de confiança $1-\alpha$) para o parâmetro μ:

$$IC(\mu, 1-\alpha) = \left[\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

em que:

  - $z_{\alpha/2}$ é chamado de **valor crítico.**
  - $1-\alpha$ é o **coeficiente de confiança ou nível de confiança do intervalo.**
  - A quantidade $e = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$ é chamada de **margem de erro** (também chamada de erro máximo da estimativa).

INTERVALO DE CONFIANÇA OBSERVADO

- Considere que a partir de uma amostra obtivemos um intervalo de confiança com $(1-\alpha)$ de confiança.

- Uma vez que o **intervalo observado** já não é mais aleatório e o parâmetro não é uma variável aleatório, **não é correto dizer** que a probabilidade do parâmetro μ pertencer ao intervalo observado é $(1-\alpha)$.

- Esse interpretação só é válida para o intervalo aleatório.

INTERVALO DE CONFIANÇA OBSERVADO

Interpretação do intervalo de confiança

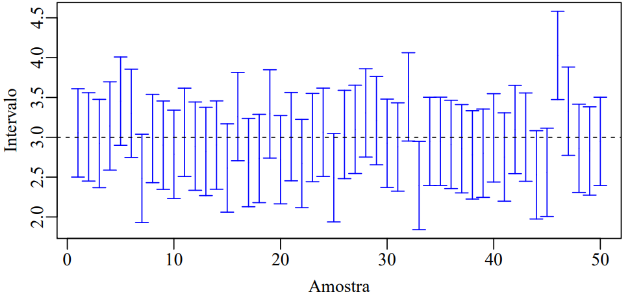
- O intervalo de confiança é calculado a partir de uma amostra aleatória. Isso significa que para cada amostra aleatória, um intervalo diferente será calculado.

- Se repetirmos o experimento várias vezes e construirmos intervalos de confiança com nível de confiança $1 - \alpha$ para cada um dos experimentos, **esperamos que em média $100 \cdot (1 - \alpha)\%$ dos intervalos de confiança contenham o verdadeiro valor do parâmetro populacional.**

INTERVALO DE CONFIANÇA OBSERVADO

Interpretação do intervalo de confiança

- Na prática apenas um experimento é realizado, portanto, teremos apenas um intervalo.
- **Interpretação:** Temos $100 \cdot (1 - \alpha)\%$ de confiança de que o intervalo observado contém o parâmetro populacional.



DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Exemplo

Vamos supor que a glicemia de jejum de uma população de pessoas não diabéticas siga a distribuição normal, com desvio padrão igual a 16 mg/dl, mas não conhecemos o valor da média dessa população. Também vamos supor que extraímos uma amostra aleatória de tamanho 36 dessa população e obtivemos a média amostral da glicemia de jejum igual a 92 mg/dl. Obtenha intervalos de confiança com 90%, 95% e 99% de confiança.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$\overline{x} = 92 \text{ mg/dl} \qquad \sigma = 16 \text{ mg/dl} \qquad n = 36$$

i. $(1-\alpha) = 0,90 \Longrightarrow \alpha = 0,10 \Longrightarrow z_{\alpha/2} = 1,645$

$$\begin{aligned} \text{IC}(\mu, 90\%) &= [92 - 1,645 \cdot 16/\sqrt{36};\ \ 92 + 1,645 \cdot 16/\sqrt{36}] \\ &= [87,61;\ \ 96,38] \end{aligned}$$

ii. $(1-\alpha) = 0,95 \Longrightarrow \alpha = 0,05 \Longrightarrow z_{\alpha/2} = 1,96$

$$\begin{aligned} \text{IC}(\mu, 95\%) &= [92 - 1,96 \cdot 16/\sqrt{36};\ \ 92 + 1,96 \cdot 16/\sqrt{36}] \\ &= [86,77;\ \ 97,22] \end{aligned}$$

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$\overline{x} = 92 \text{ mg/dl} \qquad \sigma = 16 \text{ mg/dl} \qquad n = 36$$

iii. $(1-\alpha) = 0,99 \Longrightarrow \alpha = 0,01 \Longrightarrow z_{\alpha/2} = 2,575$

$$\begin{aligned} \text{IC}(\mu, 99\%) &= [92 - 2,575 \cdot 16/\sqrt{36};\ 92 + 2,575 \cdot 16/\sqrt{36}] \\ &= [85,13;\ 98,86] \end{aligned}$$

Podemos observar que à medida que o nível de confiança aumenta, a amplitude do intervalo de confiança também aumenta.

σ desconhecido

- Considere uma amostra aleatória $X_1, \ldots, X_n$ de $X \sim \mathrm{N}(\mu, \sigma^2)$.
- É muito raro conhecermos o desvio padrão da população. O que acontece quando não conhecemos esse valor?
- Quando σ não é conhecido, usamos sua estimativa S para a construção do intervalo de confiança para média.
- Neste caso, temos que:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

em que t_{n-1} denota a distribuição t-Student com $n-1$ graus de liberdade.

σ desconhecido

- A partir desse resultado podemos construir um intervalo de confiança para o parâmetro μ de forma análoga ao caso em que σ é conhecido.

$$P\left(\overline{X} - t_{(n-1,\alpha/2)} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{(n-1,\alpha/2)} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

- O intervalo de confiança com coeficiente de confiança $1 - \alpha$ é dado por:

$$IC(\mu, 1-\alpha) = \left[\overline{X} - t_{(n-1,\alpha/2)} \cdot \frac{S}{\sqrt{n}}; \overline{X} + t_{(n-1,\alpha/2)} \cdot \frac{S}{\sqrt{n}}\right]$$

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Exemplo

Um estudo foi idealizado para estimar o tempo médio que os funcionários de uma empresa necessitam para realizar uma tarefa. Uma amostra de $n = 15$ funcionários forneceu uma média amostral de $21,39$ min e um desvio padrão de $5,38$ min. Construa um intervalo de confiança de 95% para o tempo médio de realização da tarefa.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$\overline{x} = 21,39 \text{ min} \qquad s = 5,38 \text{ min} \qquad n = 15$$

i. $(1-\alpha) = 0,95 \Longrightarrow \alpha = 0,05 \Longrightarrow t_{(\nu=15,\alpha/2)} = 2,145$

$$\begin{aligned} \text{IC}(\mu, 95\%) &= [21,39 - 2,145 \cdot 5,38/\sqrt{15};\ 21,39 + 2,145 \cdot 5,38/\sqrt{15}] \\ &= [18,41;\ 24,37] \end{aligned}$$

Intervalo de confiança para a proporção

Intervalo para proporção

- Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim$ Bernoulli(p), em que o parâmetro p é desconhecido.

- Pelo Teorema do Limite Central, a proporção amostral tem distribuição aproximadamente normal:

$$\widehat{p} \sim \mathrm{N}(p, \frac{p(1-p)}{n})$$

- Argumentos análogos ao caso da média levam a:

$$P\left(\widehat{p} - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{p} + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

- Note que p aparece na expressão da margem de erro, o que na prática impossibilita o uso dessa equação.

Intervalo para proporção

Estimativa otimista

Uma opção é substituir p por $\widehat{p} = \overline{X}$:

$$P\left(\widehat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right) = 1 - \alpha$$

$$IC(p, 1-\alpha) = \left[\overline{X} - z_{\alpha/2} \cdot \sqrt{\frac{\overline{X}(1-\overline{X})}{n}};\ \overline{X} + z_{\alpha/2} \cdot \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right] = 1 - \alpha$$

Intervalo para proporção

Estimativa conservadora

- Outra alternativa é usar $p = 0,5$. Quando $p = 0,5$, o termo $p(1-p)$ terá valor máximo, o que produz o intervalo de maior amplitude (intervalo conservador).

$$P\left(\widehat{p} - z_{\alpha/2} \cdot \sqrt{\frac{0,5 \cdot (1-0,5)}{n}} \leq p \leq \widehat{p} + z_{\alpha/2} \cdot \sqrt{\frac{0,5 \cdot (1-0,5)}{n}}\right) = 1 - \alpha$$

$$IC(p, 1-\alpha) = \left[\overline{X} - z_{\alpha/2} \cdot \frac{1}{2\sqrt{n}};\ \overline{X} + z_{\alpha/2} \cdot \frac{1}{2\sqrt{n}}\right]$$

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Exemplo

Vamos considerar que desejamos estimar a proporção de aparelhos de raios-X que estejam com defeito e produzam um excesso de radiação. Tomando-se uma amostra de 40 aparelhos, identificou-se que 12 estavam com defeito. Para um nível de confiança de 95%, calcule o intervalo de confiança para a verdadeira proporção de aparelhos que que possam estar com defeito, utilizando:

i. $p = \widehat{p}$

ii. $p = 0,5$

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$\widehat{p} = \overline{x} = 12/40 = 0,3 \qquad n = 40$$

i. $(1-\alpha) = 0,95 \implies \alpha = 0,05 \implies z_{\alpha/2} = 1,96$

$$\begin{aligned}\text{IC}(p, 95\%) &= \left[\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}};\ \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right] \\ &= \left[0,30 - 1,96 \cdot \sqrt{\frac{0,3 \cdot 0,7}{40}};\ 0,30 + 1,96 \cdot \sqrt{\frac{0,3 \cdot 0,7}{40}}\right] \\ &= [0,158;\ 0,442]\end{aligned}$$

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$\widehat{p} = \overline{x} = 12/40 = 0,3 \qquad n = 40$$

ii. $(1-\alpha) = 0,95 \Longrightarrow \alpha = 0,05 \Longrightarrow z_{\alpha/2} = 1,96$

$$\begin{aligned} \text{IC}(p, 95\%) &= \left[\overline{x} - z_{\alpha/2}\frac{1}{2\sqrt{n}};\ \overline{x} - z_{\alpha/2}\frac{1}{2\sqrt{n}}\right] \\ &= \left[0,3 - 1,96\frac{1}{2\sqrt{40}};\ 0,3 - 1,96\frac{1}{2\sqrt{40}}\right] \\ &= [0,145;\ 0,455] \end{aligned}$$

Intervalo de confiança para a variância

Intervalo para variância

- Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$, então temos que:

$$(n-1) \cdot \frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

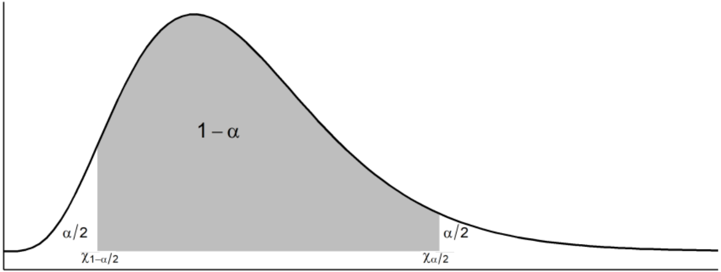
em que χ^2_{n-1} denota a distribuição qui-quadrado com $n-1$ graus de liberdade.

- Argumentos análogos ao caso da média, levam a;

$$IC(\sigma^2, 1-\alpha) = \left[\frac{(n-1)S^2}{\chi^2_{(\alpha/2, n-1)}}; \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2, n-1)}}\right]$$

em que $\chi^2_{(\alpha/2, n-1)}$ e $\chi^2_{(1-\alpha/2, n-1)}$ são os quantis da distribuição qui-quadrado com $n-1$ graus de liberdade.

Intervalo para variância



- Note que nesse caso o intervalo não é simétrico.

Intervalo para variância

Exemplo

Uma amostra de 6 cobaias é analisada para verificar a dosagem de um certo composto, obtendo-se a média amostral igual a $14,1$mg/dl e a variância igual a $2,1$ (mg/dl)^2. Obter intervalos de confiança, ao nível de 95%, para a média e a variância populacional, assumindo que os dados seguem uma distribuição normal.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$s^2 = 2,1 \qquad n = 6$$

ii. $(1-\alpha) = 0,95 \implies \alpha = 0,05 \implies \chi_{(\nu=5,\alpha/2)} = 12,833$ e $\chi_{(\nu=5,1-\alpha/2)} = 0,831$

$$\begin{aligned} \text{IC}(\sigma^2, 1-\alpha) &= \left[\frac{(n-1)s^2}{\chi^2_{(\alpha/2,n-1)}}; \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2,n-1)}}\right] \\ &= \left[\frac{(6-1)\cdot 2,1}{12,833}; \frac{(6-1)\cdot 2,1}{0,831}\right] \\ &= [0,818;\ 12,635] \end{aligned}$$

Intervalo de confiança para a diferença de médias

Intervalo para a diferença de médias

σ_x^2 e σ_y^2 conhecidas

- Sejam duas amostras aleatórias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ independentes, com $X_i \sim \mathsf{N}(\mu_x, \sigma_x^2)$ e $Y_i \sim \mathsf{N}(\mu_y, \sigma_y^2)$, onde as médias μ_x e μ_y são desconhecidas e as variâncias σ_x^2 e σ_y^2 são conhecidas. Então,

$$\overline{X} - \overline{Y} \sim \mathsf{N}(\mu_x - \mu_y, \sigma^2),$$

em que

$$\sigma^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}.$$

- Temos que

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sigma} \sim \mathsf{N}(0, 1).$$

.

Intervalo para a diferença de médias

σ_x^2 e σ_y^2 conhecidas

Dessa forma, o intervalo de confiança para $\mu_x - \mu_y$ (diferença entre as médias populacionais) com coeficiente de confiança $(1-\alpha)100\%$ é dado por

$$IC(\mu_x - \mu_y, 1-\alpha) = \left[\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\sigma^2}; \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\sigma^2}\right], \qquad (1)$$

em que

$$\sigma^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}.$$

Intervalo para diferenças de médias

σ_x^2 e σ_y^2 conhecidas

Exemplo

Testes de resistência à tensão foram feitos em dois tipos diferentes de estruturas de alumínio. Essas estruturas foram usadas na fabricação das asas de um avião comercial. De experiências passadas com o processo de fabricação dessas estruturas e com o procedimento de testes, os desvios-padrão das resistências à tensão são considerados conhecidos. Os dados obtidos são os seguintes: $n = 10$, $\overline{x} = 87{,}6$, $\sigma_x = 1$, $m = 12$, $\overline{y} = 74{,}5$, $\sigma_y = 1{,}5$. μ_x e μ_y denotam as resistências médias verdadeiras à tensão para os dois tipos da estrutura. Obtenha um intervalo de confiança de 90% para a diferença na resistência média $\mu_x - \mu_y$.

Intervalo para a diferença de médias

$\sigma_x^2 = \sigma_y^2$ desconhecidas

- Sejam duas amostras aleatórias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ independentes, com $X_i \sim \mathrm{N}(\mu_x, \sigma_x^2)$ e $Y_i \sim \mathrm{N}(\mu_y, \sigma_y^2)$, onde as médias μ_x e μ_y são desconhecidas e as variâncias σ_x^2 e σ_y^2 são **desconhecidas e supostamente iguais**, isto é, $\sigma_x^2 = \sigma_x^2 = \sigma^2$. Então,

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2},$$

em que

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-1}.$$

Intervalo para a diferença de médias

$\sigma_x^2 = \sigma_y^2$ desconhecidas

O intervalo de confiança com coeficiente de confiança $(1-\alpha)100\%$, para $\mu_x - \mu_y$, é dado por

$$IC(\mu_x - \mu_y, 1-\alpha) = \left[\overline{X} - \overline{Y} - t_{\alpha/2,\nu}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}; \overline{X} - \overline{Y} + t_{\alpha/2,\nu}\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}\right],$$

em que $\nu = n + m - 2$.

Exemplo

Realizou-se uma experiência a fim de comparar dois tipos de dieta, A e B, destinados à redução de peso. Dois grupos, cada um com 30 pessoas obesas, foram tratados, cada um usando uma dessas dietas, sendo cada pessoa associada a uma dieta aleatoriamente. Antes e depois de um período de 30 dias de dieta, foram anotados os pesos de todas as pessoas, obtendo-se os resultados a seguir, expressos em termos da redução de peso verificada:

Grupo	**Média** ($\overline{x}$)	**Desvio Padrão** (s)	**Amostra (n)**
Dieta A	$21,3$	$2,6$	30
Dieta B	$13,4$	$1,9$	30

Determine um intervalo de confiança de 95% para a diferença entre a média de perda de peso das pessoas que fizeram os dois tipos de dieta. Interprete o intervalo de confiança.

Suponha que as amostras provém de populações normais com variâncias iguais.

Intervalo para a diferença de médias

σ_x^2 e σ_y^2 desconhecidas e diferentes

Assumimos até aqui que as variâncias populacionais são desconhecidas mas iguais (ou pelo menos próximas). A violação desta suposição pode causar grandes problemas teóricos e práticos pois, não é trivial encontrar uma quantidade pivotal para $\mu_x - \mu_y$.

Na literatura encontramos diversos métodos para resolver este problema, mas nenhum deles é completamente satisfatório. Um método aproximado consiste em utilizar a quantidade pivotal

$$T = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim t_\nu,$$

em que $\nu = \frac{(w_x + w_y)^2}{\frac{w_x^2}{(n-1)} + \frac{w_y^2}{(m-1)}}$, $w_x = \frac{S_x^2}{n}$ e $w_y = \frac{S_y^2}{m}$.

Intervalo para a diferença de médias

σ_x^2 e σ_y^2 desconhecidas e diferentes

O intervalo de confiança com coeficiente de confiança $(1-\alpha)100\%$, para $\mu_x - \mu_y$, é dado por

$$IC(\mu_x - \mu_y, 1-\alpha) = \left[\overline{X} - \overline{Y} - t_{\alpha/2,\nu}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}; \overline{X} - \overline{Y} + t_{\alpha/2,\nu}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}\right],$$

em que $\nu = \frac{(w_x + w_y)^2}{\frac{w_x^2}{(n-1)} + \frac{w_y^2}{(m-1)}}$, $w_y = \frac{S_x^2}{n}$ e $w_y = \frac{S_y^2}{m}$.

Intervalo para diferenças de médias

σ_x^2 e σ_y^2 desconhecidas e diferentes

Exemplo

Uma pesquisadora está interessada em avaliar o comportamento de condução dos estudantes universitários do sexo masculino e feminino. Há uma série de maneiras que pode quantificar comportamentos de condução. Optou-se por avaliar a velocidade mais rápida já dirigida por um indivíduo. Foi realizado um levantamento de uma **amostra aleatória com 34 estudantes universitários do sexo masculino e 29 estudantes universitários do sexo feminino.** Os dados da pesquisa estão no quadro abaixo. Considere que as duas populações têm **distribuição normal** e que podemos assumir que $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$. Obtenha o intervalo de confiança para a diferença entre médias com 95% de confiança.

Grupo	Média ($\bar{x}$)	Desvio Padrão (s)	Amostra (n)
Masculino	$105,5$	$20,1$	34
Feminino	$90,9$	$12,2$	29

Intervalo para a diferença de proporções

Se $X_1, \ldots, X_n$ amostra aleatória de $X \sim$ Bernoulli(p_1) e $Y_1, \ldots, Y_m$ amostra aleatória de $Y \sim$ Bernoulli(p_2), independentes, o intervalo de confiança tradicional para $p_1 - p_2$ pode ser encontrado por meio da quantidade

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}} \sim \mathrm{N}(0,1).$$

O intervalo de confiança com coeficiente de confiança $(1-\alpha)100\%$, para $p_1 - p_2$, é dado por

$$\left[\widehat{p}_1 - \widehat{p}_2 - z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}; \widehat{p}_1 - \widehat{p}_2 - z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}\right].$$

Intervalo para diferenças de proporções

Exemplo

A tabela mostra o número de unidades vendidas e não vendidas de livros por dois vendedores. Utilize 90% de confiança e obtenha um intervalo de confiança para a diferença de proporções de unidades vendidas entre os dois vendedores.

	Unidades vendidas	**Unidades não vendidas**	**Total**
Vendedor A	216	51	267
Vendedor B	192	69	261
Total	408	120	528

Intervalo para a razão de variâncias

- Sejam duas amostras aleatórias $X_1, \cdots, X_n$ e $Y_1, \cdots, Y_m$ independentes, com $X_i \sim \mathsf{N}(\mu_x, \sigma_x^2)$ e $Y_i \sim \mathsf{N}(\mu_y, \sigma_y^2)$, onde as médias μ_x e μ_y e as variâncias σ_x^2 e σ_y^2 são desconhecidas. Então,

$$F = \frac{S_x^2}{S_y^2}\frac{\sigma_y^2}{\sigma_x^2} \sim \mathsf{F}(n-1, m-1)\cdot$$

- O intervalo de confiança para a razão $\frac{\sigma_x^2}{\sigma_y^2}$ com coeficiente de confiança $(1-\alpha)100\%$ é dado por

$$IC\left(\frac{\sigma_x^2}{\sigma_y^2}, 1-\alpha\right) = \left[\frac{1}{F_{(\frac{\alpha}{2};n-1;m-1)}}\frac{S_x^2}{S_y^2}; \frac{1}{F_{(1-\frac{\alpha}{2};n-1;m-1)}}\frac{S_x^2}{S_y^2}\right].$$

Intervalo para a razão de duas variâncias

Exemplo

Um pesquisador está interessado em estudar as variâncias do tempo que os clientes esperam antes de terem seus pedidos atendidos em duas empresas distintas, A e B. Uma **amostra aleatória de 10 clientes da empresa A**, resultou em uma **variância de 400 min.** Na **empresa B**, uma **amostra aleatória de 21** clientes resultou em uma **variância de 256 min.** Use o nível de confiança de 95% para obter o intervalo de confiança da razão entre as variâncias. Suponha que ambas as populações sejam normalmente distribuídas.

Tamanho da amostra para estimar a média

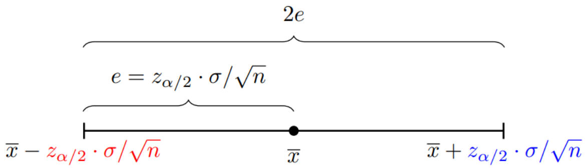
Tamanho da amostra

- Quando o desvio padrão é conhecido, já vimos que intervalo de confiança para média é dado por:

$$IC(\mu, 1-\alpha) = \left[\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

- A amplitude do intervalo de confiança é dada pela diferença entre o limite superior e inferior

$$A_{IC(\mu)} = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



Tamanho da amostra

- Observe que a amplitude do intervalo de confiança observado não é aleatório, assim podemos observar que:
  - A amplitude do intervalo diminui à medida que o tamanho da amostra aumenta, ou seja, quanto mais informações você tiver, mais preciso será o intervalo.
  - Se o nível de confiança $1-\alpha$ aumentar, então $z_{\alpha/2}$ será um valor maior, e como consequência, a amplitude do intervalo também vai crescer.
  - Se a dispersão dos dados for alta, ou seja, se o desvio padrão é grande, então a amplitude do intervalo tende a ser grande.

Tamanho da amostra

- O erro máximo da estimativa ou margem de erro é dado por:

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Assim, podemos isolar n a partir da equação anterior:

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

- O tamanho da amostra depende:
  - Do nível de confiança $(1 - \alpha)$
  - Do desvio-padrão (σ)
  - Do erro máximo admitido (e)
- O tamanho da amostra precisa ser um número inteiro, dessa forma, arredondamento para o primeiro maior número inteiro obtido $\lceil n \rceil$.

Tamanho da amostra

Exemplo

Uma pesquisa é elaborada para determinar as despesas médicas anuais das famílias dos empregados de uma grande empresa. A gerência da empresa deseja construir um intervalo de confiança com 90% de confiança e com uma margem de erro de $R\$\,50,00$. Um estudo piloto indica que o desvio padrão pode ser calculado como sendo igual $R\$\,100,00$.

a. Qual o tamanho de amostra necessário?

b. Se a gerência deseja uma margem de erro de $R\$\,30,00$ que tamanho de amostra será necessário?

c. Calcule novamente o tamanho de amostra do item a. considerando um nível de confiança de 99%.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$z_{0,05/2} = 1,64 \qquad z_{0,01/2} = 2,575 \qquad \sigma = 100 \qquad e = 50$$

i. $n = \left(\frac{z_{\alpha/2 \cdot \sigma}}{e}\right)^2 = \left(\frac{1,64 \cdot 100}{50}\right)^2 \approx 11$

ii. $n = \left(\frac{z_{\alpha/2 \cdot \sigma}}{e}\right)^2 = \left(\frac{1,64 \cdot 100}{30}\right)^2 \approx 30$

iii. $n = \left(\frac{z_{\alpha/2 \cdot \sigma}}{e}\right)^2 = \left(\frac{2,575 \cdot 100}{50}\right)^2 \approx 27$

Tamanho amostral: variável quantitativa - média

Um intervalo de confiança, com coeficiente de confiança $(1-\alpha)$100%, para a **média populacional** μ com correção para populações finitas, é dado por

$$IC(\mu, 1-\alpha) = \left[\overline{X} - z_{\alpha/2}\sqrt{\left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}}; \quad \overline{X} + z_{\alpha/2}\sqrt{\left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}}\right]$$

em que $Z \sim N(0,1)$ e $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.

$$e = z_{\alpha/2}\sqrt{\left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}}$$

corresponde ao erro na estimação de μ

Tamanho amostral: variável quantitativa - média

Para um coeficiente de confiança fixo $(1-\alpha)$, um erro fixado (e) e σ conhecido, n é obtido de:

$$n = \frac{N \cdot (z_{\alpha/2})^2 \cdot \sigma^2}{(N-1) \cdot e^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$$

Tamanho amostral: variável quantitativa - média

População infinita	$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{e}\right)^2$
População finita (< 10000)	$n = \frac{N \cdot (z_{\alpha/2})^2 \cdot \sigma^2}{(N-1) \cdot e^2 + \sigma^2 \cdot (z_{\alpha/2})^2}$

n: tamanho da amostra;

$z_{\alpha/2}$: valor crítico para o grau de confiança desejado, usualmente: 1.96 (95%, $\alpha = 0.05$);

σ: desvio padrão populacional da variável;

e: margem de erro, usualmente: $\pm$5% da média ($1.05\cdot$média);

N: tamanho da população (finita);

Note que para população finita temos um “ajuste"para o tamanho da amostra.

Tamanho da amostra para estimar a proporção

Tamanho da amostra

- Usando o mesmo raciocínio do tamanho de amostra para a média, a partir da equação da margem de erro para a distribuição amostral da proporção, temos:

$$e = z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

- Isolando n, temos que:

$$n = \left(\frac{z_{\alpha/2}}{e}\right)^2 \cdot p(1-p)$$

Tamanho da amostra

Exemplo

Considere que você está analisando uma campanha política e quer estimar, com 95% de confiança, a proporção dos eleitores que irão votar em um determinado candidato. Sua estimativa deve ter uma margem de erro de 2%. Encontre o número mínimo da amostra necessária se:

a. Não há nenhuma estimativa prévia;

b. Há uma estimativa prévia de $\widehat{p} = 0,3$.

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA

Solução

$$z_{0,05/2} = 1,96 \qquad e = 0,02$$

i. $n = \left(\frac{z_{\alpha/2}}{e}\right)^2 \cdot p(1-p) = \left(\frac{1,96}{0,02}\right)^2 \cdot 0,5 \cdot 0,5 \approx 2401$

ii. $n = \left(\frac{z_{\alpha/2}}{e}\right)^2 \cdot p(1-p) = \left(\frac{1,96}{0,02}\right)^2 \cdot 0,3 \cdot 0,7 \approx 2017$

Tamanho amostral: variável qualitativa - proporção

Um intervalo de confiança, com coeficiente de confiança $(1-\alpha)$100% com correção para populações finitas, para a para a proporção populacional p é dado por

$$IC(p, 1-\alpha) = \left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{N-n}{N-1}\right)\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right],$$

em que $Z \sim N(0,1)$ e $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.

Tamanho amostral: variável qualitativa - proporção

Um intervalo de confiança, com coeficiente de confiança $(1-\alpha)$100% com correção para populações finitas, para a para a proporção populacional p é dado por

$$IC(p, 1-\alpha) = \left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{N-n}{N-1}\right)\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right],$$

em que $Z \sim N(0,1)$ e $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.

$e = z_{\alpha/2} \cdot \sqrt{\left(\frac{N-n}{N-1}\right)\frac{\hat{p}\cdot(1-\hat{p})}{n}}$ corresponde ao erro na estimação de p.

Tamanho amostral: variável qualitativa - proporção

Um intervalo de confiança, com coeficiente de confiança $(1-\alpha)$100% com correção para populações finitas, para a para a proporção populacional p é dado por

$$IC(p, 1-\alpha) = \left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{N-n}{N-1}\right) \frac{\hat{p} \cdot (1-\hat{p})}{n}}\right],$$

em que $Z \sim N(0,1)$ e $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.

$e = z_{\alpha/2} \cdot \sqrt{\left(\frac{N-n}{N-1}\right) \frac{\hat{p} \cdot (1-\hat{p})}{n}}$ corresponde ao erro na estimação de p.

Para um coeficiente de confiança fixo $(1-\alpha)$, um erro fixado (e) e $\hat{p}$ conhecido, n é obtido de:

$$n = \frac{N \cdot p \cdot (1-p) \cdot (z_{\alpha/2})^2}{(N-1) \cdot e^2 + p \cdot (1-p) \cdot (z_{\alpha/2})^2}$$

Tamanho amostral: variável qualitativa - proporção

Fórmula para cálculo do tamanho amostral: variável qualitativa

População infinita $$n = \left(\frac{z_{\alpha/2} \cdot \sqrt{p \cdot (1-p)}}{e} \right)^2$$

População finita (< 10000) $$n = \frac{N \cdot p \cdot (1-p) \cdot (z_{\alpha/2})^2}{(N-1) \cdot e^2 + p \cdot (1-p) \cdot (z_{\alpha/2})^2}$$

n: tamanho da amostra;

$z_{\alpha/2}$: valor crítico para o grau de confiança desejado, usualmente: 1.96 (95%, $\alpha = 0.05$);

e: margem de erro, usualmente: $\pm$5% da proporção dos casos (precisão absoluta);

p: proporção de resultados favoráveis da variável na população;

N: tamanho da população (finita);

Tamanho amostral: variável qualitativa - proporção

Exemplo: O objetivo de um estudo é descrever a prevalência de insuficiência venosa nos membros inferiores, com um erro tolerável de $\pm 5\%$ e com 90% de confiança, na população de pacientes obesos mórbidos de um ambulatório específico de obesidade que possui um volume de 315 pacientes (630 membros). Um estudo anterior estimou a proporção de membros acometidos de 69.3%. Qual o tamanho de amostra necessário para o estudo?

Resolução:

Tamanho amostral: variável qualitativa - proporção

Exemplo: O objetivo de um estudo é descrever a prevalência de insuficiência venosa nos membros inferiores, com um erro tolerável de $\pm 5\%$ e com 90% de confiança, na população de pacientes obesos mórbidos de um ambulatório específico de obesidade que possui um volume de 315 pacientes (630 membros). Um estudo anterior estimou a proporção de membros acometidos de 69.3%. Qual o tamanho de amostra necessário para o estudo?

Resolução: O cálculo do tamanho amostral será baseado no slide anterior, fórmula para população finita. Consideremos agora 95% de confiança.

Note que $N = 630$, $e = 0.05$, $\hat{p} = 0.693$, $1 - \hat{p} = 0.307$, assim

$$n = \frac{630 \cdot 0.693 \cdot 0.307 \cdot (1.96)^2}{(630 - 1) \cdot (0.05)^2 + 0.693 \cdot 0.307 \cdot (1.96)^2} \approx 216 \text{ membros}$$

Tamanho amostral: variável qualitativa - proporção

Nota: caso não se tenha informação preliminar sobre a proporção, outra possibilidade é utilizar $p = 0.5$. Esse valor fornece a maior variância possível, consequentemente o maior tamanho amostral.

Tamanho amostral: variável qualitativa - proporção

Nota: caso não se tenha informação preliminar sobre a proporção, outra possibilidade é utilizar $p = 0.5$. Esse valor fornece a maior variância possível, consequentemente o maior tamanho amostral.

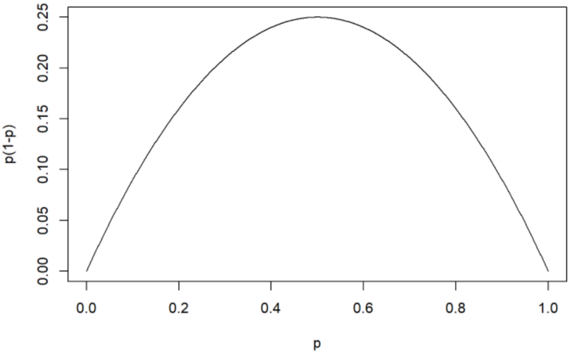


Figura 1: Gráfico da função $f(p) = p(1-p)$