

# Aula de Exercícios - Intervalo de Confiança para proporções

*Organização:* Airton Kist    *Digitação:* Guilherme Ludwig

# Intervalo de Confiança para proporções

## Exemplo

Suponha que  $p = 30\%$  dos estudantes de uma escola sejam mulheres. Colhemos uma amostra aleatória simples de  $n = 10$  estudantes e calculamos  $\hat{p} =$  proporção de mulheres na amostra. Qual a probabilidade de que  $\hat{p}$  difira de  $p$  em menos de 0,01?  
*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 276.*

## Intervalo de Confiança para proporções

Temos que a probabilidade é dada por

$$P(|\hat{p} - p| < 0.01) = P(-0.01 < \hat{p} - p < 0.01)$$

Mas  $\hat{p} - p$  pode ser aproximada por uma normal  $N(0, p(1 - p)/n)$ , e como  $p = 0.3$ , temos que

$$\text{Var}(\hat{p} - p) = \frac{0.3 \cdot 0.7}{10} = 0.021$$

e, portanto, a probabilidade pedida é igual a

$$P\left(\frac{-0.01}{\sqrt{0.021}} < Z < \frac{0.01}{\sqrt{0.021}}\right) = P(-0.07 < Z < 0.07) = 0.056$$

*Solução: Morettin & Bussab, Estatística Básica 5ª edição, pág 276.*

# Introdução

## Exemplo

Uma variável aleatória  $X$  tem distribuição normal, com média 100 e desvio padrão 10.

- (a) Qual a  $P(90 < X < 110)$ ?
- (b) Se  $\bar{X}$  for a média de uma amostra de 16 elementos retirados dessa população, calcule  $P(90 < \bar{X} < 110)$ .
- (c) Represente, num único gráfico, as distribuições de  $X$  e  $\bar{X}$ .
- (d) Que tamanho deveria ter a amostra para que  $P(90 < \bar{X} < 110) = 0.95$ ?

Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 274.

## Introdução

- (a) Devemos padronizar o evento, para comparar com a distribuição normal padrão.

$$P(90 < X < 110) = P\left(\frac{90 - 100}{10} < \frac{X - 100}{10} < \frac{110 - 100}{10}\right)$$

$$= P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = \Phi(1) - \Phi(-1)$$

Consultando a tabela disponível na página da disciplina<sup>1</sup>, vemos que  $\Phi(1) = 0.8413$ . Para encontrar  $\Phi(-1)$ , note que a distribuição normal é simétrica e portanto  $\Phi(-z) = 1 - \Phi(z)$ , daí  $\Phi(-1) = 0.1569$  e portanto  $\Phi(1) - \Phi(-1) = 0.6844$ .

---

<sup>1</sup><http://www.ime.unicamp.br/~veronica/Coordenadas1s/N.pdf>

# Introdução

(b) Se temos uma amostra e tiramos a média, note agora que

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{(1)}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \stackrel{(2)}{=} \frac{\text{Var}(X)}{n}$$

onde a igualdade (1) vale por independência, e a igualdade (2) vale por serem identicamente distribuídas. Consequentemente, o desvio padrão novo será  $\sigma/\sqrt{n}$ , ou 10/4. Temos então que

$$P(90 < \bar{X} < 110) = P\left(\frac{90 - 100}{10/4} < \frac{\bar{X} - 100}{10/4} < \frac{110 - 100}{10/4}\right)$$

# Introdução

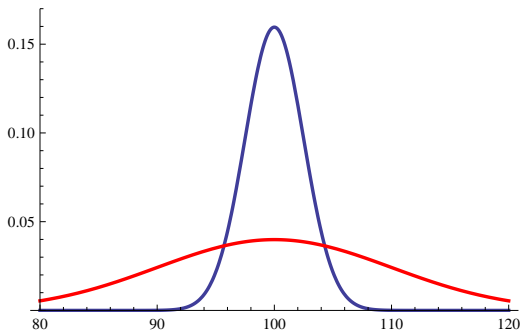
(b) Continuando,

$$= P(-4 < Z < 4) = P(Z < 4) - P(Z < -4) = \Phi(4) - \Phi(-4)$$

Se consultarmos a tabela agora, veremos que a probabilidade  $P(Z < 4)$  é tão grande nem está listada. Ela então pode ser considerada 1 (com a ajuda do pacote estatístico **R**, podemos verificar que  $\Phi(4) - \Phi(-4)$  é igual a 0.9999367).

# Introdução

(c) A distribuição de  $X$  está em vermelho, e a de  $\bar{X}$  em azul.





# Introdução

(d) Queremos resolver a seguinte equação:

$$P\left(\frac{90 - 100}{10/\sqrt{n}} < \frac{\bar{X} - 100}{10/\sqrt{n}} < \frac{110 - 100}{10/\sqrt{n}}\right) = 0.95$$

Note que, consultando a tabela, vemos que

$P(-q < Z < q) = 0.95$  se  $q = 1.96$ . Então a equação que queremos resolver pode ser reescrita como:

$$\frac{110 - 100}{10/\sqrt{n}} = 1.96 \Leftrightarrow \sqrt{n} \frac{110 - 100}{10} = 1.96 \Leftrightarrow n = 1.96^2$$

Portanto,  $n = 4$  é suficiente para obtermos a confiança desejada.

# Introdução

## Exemplo

Um procedimento de controle de qualidade foi planejado para garantir um máximo de 10% de itens defeituosos na produção. A cada 6 horas sorteia-se uma amostra de 20 peças e, havendo mais de 15% de defeituosas, encerra-se a produção para verificação do processo. Qual a probabilidade de uma parada desnecessária?

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 276.*

# Introdução

O procedimento é desnecessário se o processo está dentro da margem de 10% de itens defeituosos na produção. Suponha o pior caso, que corresponde a exatamente 10%.

Nesse caso, o número  $Y_{20}$  de peças defeituosas no sorteio de 20 peças tem distribuição binomial(20,0.1) e a proporção de peças defeituosas,  $Y_{20}/20$ , tem distribuição aproximadamente  $N(0.1, 0.1 \cdot 0.9/20)$ .

## Introdução

Queremos calcular  $P(Y_{20}/20 > 0.15)$  utilizando a aproximação Normal. Note que o desvio padrão é  $\sqrt{0.1 \cdot 0.9/20} \approx 0.067$ . Então

$$P\left(\frac{Y_{20}}{20} > 0.15\right) = P\left(\frac{Y_{20}/20 - 0.1}{0.067} > \frac{0.15 - 0.1}{0.067}\right) = P(Z > 0.7462)$$

Recorde-se que  $P(Z > 0.7462) = 1 - \Phi(0.7462)$ , e portanto, basta consultar a tabela para vermos que  $\Phi(0.7462) = 0.7734$  e consequentemente  $P(Z > 0.7462) = 0.2266$ . O procedimento produzirá uma parada desnecessária com 22.66% de probabilidade.

# Introdução

## Exemplo

Supondo que a produção do exemplo anterior esteja sob controle, isto é,  $p = 10\%$ , e que os itens sejam vendidos em caixas com 100 unidades, qual é a probabilidade que uma caixa:

- (a) tenha mais do que 10% de itens defeituosos?
- (b) não tenha itens defeituosos?

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 277.*

# Introdução

- (a) A proporção de defeituosos em uma caixa de 100 unidades tem distribuição  $N(0.1, 0.1 \cdot 0.9/100)$ . Então a probabilidade de uma proporção de mais de 0.1 defeituosos é dada por  $P(Z > 0) = 0.5$
- (b) Note que neste caso a aproximação Normal não é recomendável, pois o evento  $\{\bar{X} \leq 0\}$  não faz sentido, e  $\{\bar{X} = 0\}$  tem probabilidade zero. Felizmente é fácil calcular a probabilidade exata de um evento equivalente, que tem distribuição binomial:  
 $P(0 \text{ defeituosos em } 100) = 0.9^{100} \approx 2.65 \times 10^{-5}$ , ou quase zero.

# Introdução

## Exemplo

Um distribuidor de sementes determina, por meio de testes, que 5% das sementes não germinam. Ele vende pacotes com 200 sementes com garantia de 90% de germinação. Qual é a probabilidade de que um pacote não satisfaça à garantia?

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 284.*

# Introdução

Temos que a proporção  $\bar{X}$  de sementes que não germinam entre  $n$  sementes tem distribuição  $N(0.05, 0.05 \cdot 0.95/n)$ . Se  $n = 200$  e a promessa do vendedor é que não menos que 0.9 germinarão, então a probabilidade da promessa não ser cumprida é dada por

$$P(\bar{X} > 0.1) = P\left(\frac{\bar{X} - 0.05}{\sqrt{0.05 \cdot 0.95/n}} > \frac{0.1 - 0.05}{\sqrt{0.05 \cdot 0.95/n}}\right) =$$

$$P\left(Z > \sqrt{n} \frac{0.05}{\sqrt{0.0475}}\right) \stackrel{(n=200)}{=} P\left(Z > \sqrt{200} \cdot 0.2294\right) = 1 - \Phi(3.24)$$

Consultando a tabela, vemos que  $\Phi(3.24) = 0.9994$  e portanto a garantia do vendedor é falsa em somente 0.06% das vezes.



# Estimação

## Exemplo

Qual deve ser o tamanho de uma amostra cujo desvio-padrão é 10 para que a diferença da média amostral para a média da população, em valor absoluto, seja menor que 1, com coeficiente de confiança igual a:

- (a) 95%
- (b) 99%

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 308.*

# Estimação

- (a) Note que se  $X \sim N(\mu, \sigma^2)$ , então  $X - \mu \sim N(0, \sigma^2)$ . Sabemos que  $\sigma = 10$ , e que o desvio-padrão do estimador da média,  $\bar{X}$ , será  $10/\sqrt{n}$ . Queremos que  $P(|\bar{X} - \mu| < 1) = 0.95$ . Mas o evento é equivalente a

$$P(-\sqrt{n}/10 < Z < \sqrt{n}/10)$$

Como  $P(-1.96 < Z < 1.96) = 0.95$ , então  $\sqrt{n}/10 = 1.96$  ou  $n \approx 385$ .

- (b) De modo análogo, temos que  $P(-2.57 < Z < 2.57) = 0.99$ , então  $\sqrt{n}/10 = 2.57$  ou  $n \approx 665$ .

# Intervalo de Confiança

## Exemplo

Uma amostra aleatória de 625 donas de casa revela que 70% delas preferem a marca A de detergente. Construir um intervalo de confiança para  $p =$  proporção das donas de casa que preferem A com coeficiente de confiança  $\gamma = 90\%$ .

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 308.*

## Intervalo de Confiança

Temos que em nossa amostra aleatória  $\hat{p} = 0.7$ . Como  $\hat{p} \sim N(p, p(1-p)/n)$ , então o intervalo de confiança é dado por

$$\left( \hat{p} - z(\gamma)\sqrt{\hat{p}(1-\hat{p})/n} ; \hat{p} + z(\gamma)\sqrt{\hat{p}(1-\hat{p})/n} \right)$$

Temos que para  $\gamma = 0.90$ ,  $z(\gamma) = 1.68$  e portanto o intervalo de confiança para a proporção de donas de casa que preferem o detergente A é dado por

$$\left( 0.7 - 1.68\sqrt{0.7 \cdot 0.3/625} ; 0.7 + 1.68\sqrt{0.7 \cdot 0.3/625} \right)$$

$$(0.6692 ; 0.7308)$$

# Intervalo de Confiança

## Exercício

Suponha que estejamos interessados em estimar a porcentagem de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

- (a) O intervalo de confiança de  $p$ , com c.c. de 95%; interprete o resultado.
- (b) O tamanho da amostra para que o erro da estimativa não exceda 0.02 unidades com probabilidade de 95%; interprete o resultado.

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 309.*

## Intervalo de Confiança

- (a) O intervalo de confiança a 95% de confiabilidade é dado por:

$$IC(p; 0.95) = 0.333 \pm 1.96 \sqrt{\frac{0.333 \cdot 0.667}{300}} = 0.333 \pm 0.053$$

Ou simplesmente (0.280; 0.387). Se tomarmos o valor máximo possível para  $p(1 - p)$ , isto é,  $1/4$ , temos um IC conservador dado por:

$$IC_c(p; 0.95) = 0.333 \pm 1.96 \sqrt{\frac{1}{4 \cdot 300}} = 0.333 \pm 0.057$$

Ou simplesmente (0.277; 0.390). *Interpretação:* Se pudéssemos construir um grande número de intervalos aleatórios para  $p$ , todos baseados em amostras de tamanho  $n$ , 95% deles conteriam o parâmetro  $p$ .

## Intervalo de Confiança

- (b) Utilizando a estimativa da amostra observada ( $\hat{p} = 0.333$ ), temos que  $n$  é dado por

$$n = \left( \frac{1.96}{0.02} \right)^2 \times 0.333 \times 0.667 \cong 2134$$

Utilizando o valor máximo de  $p(1 - p)$ , obtemos

$$n = \left( \frac{1.96}{0.02} \right)^2 \times \frac{1}{4} \cong 2401$$

*Interpretação:* Utilizando o tamanho amostral encontrado, teremos uma probabilidade de 95% de que a proporção amostral não difira do verdadeiro valor de  $p$  em menos que 2%.

# Intervalo de Confiança

## Exemplo

Antes de uma eleição em que existiam dois candidatos, A e B, foi feita uma pesquisa de opinião com 400 eleitores escolhidos ao acaso, e verificou-se que 208 destes pretendiam votar no candidato A. Construa um intervalo de confiança, com coeficiente de confiança  $\gamma = 0.95$ , para a porcentagem de eleitores favoráveis ao candidato A na época das eleições.

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 318.*



## Intervalo de Confiança

A proporção estimada de votos no candidato A é de  $\bar{x}_A = 208/400 = 0.52$ . Para  $\gamma = 0.95$ , temos  $z(\gamma) = 1.96$ , e o intervalo de confiança é dado por:

$$\left( 0.52 - 1.96\sqrt{0.52 \cdot 0.48/400} ; 0.52 + 1.96\sqrt{0.52 \cdot 0.48/400} \right)$$
$$(0.471 ; 0.589)$$

# Teste de Hipóteses

## Exemplo

Estão sendo estudados dois processos para conservar alimentos, cuja principal variável de interesse é o tempo de duração destes. No processo A, o tempo  $X$  de duração segue a distribuição  $N(\mu_A, 100)$ , e no processo B o tempo  $Y$  obedece à distribuição  $N(\mu_B, 100)$ . Sorteiam-se duas amostras independentes: a de A, com 16 latas, apresentou tempo médio de duração igual a 50, e a de B, com 25 latas, duração média igual a 60.

(a) Construa um IC para  $\mu_A$  e  $\mu_B$ , separadamente.

# Teste de Hipóteses

## Exemplo

- (b) Para verificar se os dois processos podem ter o mesmo desempenho, decidiu-se construir um IC para a diferença  $\mu_A - \mu_B$ . Caso o zero pertença ao intervalo, pode-se concluir que existe evidência de igualdade dos processos. Qual seria sua resposta?

*Fonte: Morettin & Bussab, Estatística Básica 5ª edição, pág 318.*

## Teste de Hipóteses

- (a) Para o caso geral, o intervalo de confiança para  $\mu$ , com coeficiente de confiabilidade  $\gamma$ , é dado por

$$\left( \bar{X} - z(\gamma)\sqrt{\sigma/n} ; \bar{X} + z(\gamma)\sqrt{\sigma/n} \right)$$

Repare que  $\sigma_A = \sigma_B$ ; para o coeficiente de confiança  $\gamma = 0.95$ , por exemplo, temos  $z(\gamma) = 1.96$ , e os intervalos de confiança serão, respectivamente:

$$IC(\mu_A) = \left( 50 - 1.96\sqrt{10/16} ; 50 + 1.96\sqrt{10/16} \right)$$

$$IC(\mu_B) = \left( 60 - 1.96\sqrt{10/25} ; 60 + 1.96\sqrt{10/25} \right)$$

## Teste de Hipóteses

- (a) (cont.) Fazendo as contas, obtemos que

$$IC(\mu_A) = (48.45 ; 51.55)$$

$$IC(\mu_B) = (58.76 ; 61.24)$$

Observe que os intervalos não se interceptam; temos evidência para dizer que as durações médias serão diferentes, a 95% de confiança.

- (b) Temos aqui duas amostras diferentes; a diferença  $\bar{X}_A - \bar{X}_B$  tem distribuição Normal, com média  $\mu_A - \mu_B$  e variância  $\sigma_A^2/n_A + \sigma_B^2/n_B$  (pois são independentes).

## Teste de Hipóteses

(b) (cont.) Então o intervalo de confiança para  $\mu_A - \mu_B$  é dado por

$$\left( \bar{X}_A - \bar{X}_B - z(\gamma)\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B} ; \right. \\ \left. \bar{X}_A - \bar{X}_B + z(\gamma)\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B} \right)$$

Aplicando os valores conhecidos ou observados, e fixando a confiança em  $\gamma = 0.95$  temos:

$$\text{IC}(\mu_A - \mu_B) = \left( 50 - 60 - 1.96\sqrt{10/16 + 10/25} ; \right. \\ \left. 50 - 60 + 1.96\sqrt{10/16 + 10/25} \right)$$

## Teste de Hipóteses

(b) (cont.) Executando as contas, obtemos finalmente que

$$IC(\mu_A - \mu_B) = (-11.98 ; -8.02)$$

Em concordância com o item (a), vemos que 0 não está contido no intervalo e, portanto, rejeitamos a hipótese, a  $\gamma = 0.95$  de confiança, das médias  $\mu_A$  e  $\mu_B$  serem iguais.