Intervalos Estatísticos para uma única Amostra - parte II

Intervalo de confiança para proporção

Marcos Oliveira Prates

2012/02

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Construir intervalos de confiança para proporção de uma população.
- Esses intervalos serão construídos usando a distribuição normal e o Teorema Central do Limite.

- Muitas vezes queremos estimar a proporção de uma determinada população.
- Exemplo: proporção de consumidores que preferem determinada marca de refrigerante.
- Uma amostra de tamanho n é retirada de uma população grande.
- X ($X \leq n$) dessas observações pertencem a uma determinada classe.
- Então
$$\hat{P} = \frac{X}{n}$$
é um estimador da proporção p que pertence a essa classe.
- Observe que $X \sim Bin(n, p)$ e queremos estimar p.

- Seja X_i uma variável binária

$$X_i = \begin{cases} 1 & \text{se a i-ésima observação pertence à classe de interes} \\ 0 & \text{caso contrário.} \end{cases}$$

- Temos que

$$X = \sum_{i=1}^{n} X_i \,.$$

- Então X é uma média de variáveis i.i.d. com distribuição Bernoulli(p).
- Pelo Teorema Central do Limite

$$Z = \frac{\sum_i X_i - E(\sum_i X_i)}{\sqrt{Var(\sum_i X_i)}} = \frac{\sum_i X_i - nE(X)}{\sqrt{nVar(X_i)}} \sim N(0,1)$$

ou seja

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1) \,.$$

Aproximação Normal para uma Proporção Binomial

- Se n for grande, a distribuição de

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

será aproximadamente uma normal padrão.

Observação: Essa aproximação é boa desde que $np > 5$ e $n(1-p) > 5$.

- Para construírmos um intervalo para p com $100(1-\alpha)\%$ de confiança precisamos que

$$P(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

ou seja

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha\,.$$

- Isolando p ficamos com

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

- Porém não sabemos o valor de p para construir o intervalo.
- Então precisamos estimar esse valor por $\hat{P}$.
- O intervalo fica

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

Intervalo de Confiança para uma Proporção Binomial

- Considere uma amostra aleatória de tamanho n.
- Uma proporção $\hat{p}$ dessa amostra pertence a uma classe de interesse.
- Queremos construir um intervalo aproximado para a proporção p da população que pertence à classe.
- O intervalo com $100(1-\alpha)\%$ de confiança é dado por

$$\hat{P} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

onde $z_{1-\frac{\alpha}{2}}$ é tal que

$$P(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}\,.$$

Observações:

- Esse procedimento depende da adequação da aproximação da binomial pela normal.
- Quando a aproximação não é apropriada outros métodos devem ser usados.

Exemplo:

- Uma amostra de 85 mancais é selecionada.
- 10 deles tem acabamento de superfície mais rugoso do que as especificações permitidas.
- Uma estimativa pontual para proporção de mancais que excedem a rugosidade especificada é

$$\hat{P} = \frac{x}{n} = \frac{10}{85} = 0,12\,.$$

- Um intervalo com 95% de confiança para p é dado por

$$\hat{P} - z_{0,975}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + z_{0,975}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\,.$$

Exemplo: (continuação)

- Temos que $\alpha = 0,05$ e

$$P(Z \leq 1,96) = 0,975 \Rightarrow z_{0,975} = 1,96\,.$$

- Os dados são

$$\hat{p} = 0,12 \quad z_{0,975} = 1,96 \quad n = 85$$

logo o intervalo fica

$$0,12 - 1,96\sqrt{\frac{(0,12)(1-0,12)}{85}} \leq p \leq 0,12 + 1,96\sqrt{\frac{(0,12)(1-0,12)}{85}}$$

ou seja

$$0,05 \leq p \leq 0,19\,.$$

Determinação do Tamanho de Amostra

- $\hat{P}$ é estimador de p, logo o erro de estimação é

$$E = |p - \hat{P}| \, .$$

- Para um intervalor com $100(1-\alpha)\%$ de confiança o erro é no máximo é

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \, .$$

- No exemplo anterior o erro máximo era

$$1,96\sqrt{\frac{(0,12)(1-0,12)}{n}} = 0,07 \, .$$

- Podemos escolher n:
  - fixando o nível de confiança $100(1-\alpha)\%$;
  - e o erro máximo permitido E.
- Isolando n na expressão

$$E = z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\,.$$

temos que

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{E}\right)^2 p(1-p)\,.$$

Tamanho da amostra em uma Distribuição Binomial

- Fixado um nível de confiança $100(1-\alpha)\%$ e um erro E temos que

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{E}\right)^2 p(1-p)\,.$$

- Uma estimativa de p é necessária para calcular o valor de n.
- Existem algumas possibilidades.

Como estimar p para calcular n?

- Podemos usar uma estimativa $\hat{p}$ de uma amostra anterior.
- Uma amostra preliminar (amostra piloto) pode ser retirada e o valor $\hat{p}$ é calculado.
- Podemo encontrar p tal que $p(1-p)$ é máximo:
  - esse valor é $p = 0,5$;
  - então $p(1-p) = 0,25$ e

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{E}\right)^2 (0,25) .$$

Exemplo:

- Considere o exemplo dos mancais.
- Queremos construir um intervalo com 95% de confiança.
- O erro máximo cometido é 0,05.
- Se usarmos $\hat{p} = 0,12$ como estimativa de p temos que

$$n = \left(\frac{z_{0,975}}{E}\right)^2 \hat{P}(1-\hat{P}) = \left(\frac{1,96}{0,05}\right)^2 (0,12)(0,88) \approx 163.$$

- Se não quisermos usar $\hat{p}$ como estimativa de p temos que

$$n = \left(\frac{z_{0,975}}{E}\right)^2 (0,25) = \left(\frac{1,96}{0,05}\right)^2 (0,25) \approx 385\,.$$

- Se tivermos uma informação sobre p (de uma amostra passada ou de uma amsotra piloto)
  - podemos usar uma amostra de tamanho menor.

Limites unilaterais para proporção binomial

- Os limites aproximados inferior e superior de confiança $100(1-\alpha)\%$ são

$$\hat{P} - z_{1-\alpha}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p$$

$$p \leq \hat{P} + z_{1-\alpha}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}.$$