MAT02025 - Amostragem 1

A estimativa do tamanho da amostra: formulações

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2022

A fórmula para n na amostragem para proporções

n na amostragem para proporções

- **Relembrando:** as unidades são classificadas em duas classes, C e C'.
- Admite-se uma certa margem de erro d na proporção estimada p de unidades na classe C, e há um pequeno risco α de que estejamos dispostos a incorrer de que o erro real seja maior do que d; ou seja, nós queremos[1]

$$\Pr(|p - P| \geq d) = \alpha.$$

[1] Um erro $(p - P)$ maior que d ocorre com probabilidade α.

n na amostragem para proporções

- A **amostragem aleatória simples** é assumida e p é considerado como normalmente distribuído. Da expressão **(2)** da aula 17 **(**Teo. 17.2)},

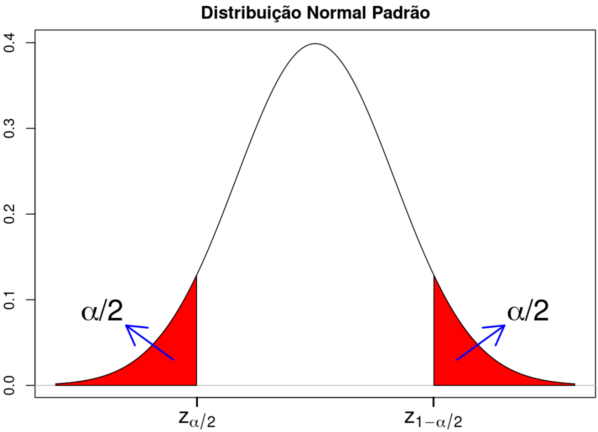
$$\sigma_p = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}.$$

- Portanto, a fórmula que conecta n com o grau de precisão desejado é

$$d = z_{\alpha}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}},$$

n na amostragem para proporções

em que z_α é a abscissa da curva normal que define uma área α nas caudas.



n na amostragem para proporções

- Resolvendo para n, encontramos

$$n = \frac{\frac{z_\alpha^2 PQ}{d^2}}{1 + \frac{1}{N}\left(\frac{z_\alpha^2 PQ}{d^2} - 1\right)}. \quad (1)$$

- Para uso prático, uma **estimativa antecipada** p de P é substituída nesta fórmula.
- Se N for grande, uma primeira aproximação é

$$n_0 = \frac{z_\alpha^2 pq}{d^2} = \frac{pq}{V}, \quad (2)$$

em que $V = (d^2/z_\alpha^2) = pq/n_0 =$ variância desejada da prporção amostral.

n na amostragem para proporções

- Na prática, primeiro calculamos n_0.
  - Se n_0/N for desprezível, n_0 é uma aproximação satisfatória para n de (1).
  - Se não, é aparente na comparação de (1) e (2) que n é obtido como

$$n = \frac{n_0}{1 + (n_0 - 1)/N} \approx \frac{n_0}{1 + (n_0/N)}. \tag{3}$$

Exemplo

- No **exemplo hipotético de grupos sanguíneos**, tivemos

$$d = 0,05, \quad p = 0,5, \quad \alpha = 0,05, \quad z_{\alpha} \approx 2.$$

Consequentemente,

$$n_0 = \frac{(4)(0,5)(0,5)}{0,0025} = 400.$$

Exemplo

- Suponhamos que haja apenas 3.200 pessoas na ilha. A cpf é necessária e encontramos

$$n = \frac{n_0}{1 + (n_0 - 1)/N} = \frac{400}{1 + 399/3200} = 356.$$

Exemplo

Pacote PracTools

- No R, o pacote PracTools possui funções para a estimativa do tamanho de amostra sob AAS.
  - A função nProp calcula o tamanho de amostra para a proporção conforme as expressões (2) e (3).

```
# install.packages("PracTools")
library(PracTools)

nProp(V0 = (0.05/2)^2, N = 3200, pU = 0.5)
```

```
## [1] 355.6543
```

```
# diferença ao usar z 'arrendondado'
nProp(V0 = (0.05/1.96)^2, N = 3200, pU = 0.5)
```

```
## [1] 343.0804
```

Comentários

- A fórmula para n_0 também é válida se d, p e q forem expressos como **porcentagens** em vez de proporções.
- Como o produto pq aumenta à medida que p se move em direção a $1/2$, ou 50%, uma estimativa conservadora de n é obtida escolhendo para p o valor mais próximo de $1/2$ no intervalo em que se pensa que p provavelmente estará.
- Se p parece estar entre 5 e 9%, por exemplo, assumimos 9% para a estimativa de n.

Comentários

- Às vezes, particularmente ao estimar o número total NP de unidades na classe C, desejamos controlar o **erro relativo** r em vez do erro absoluto em Np.
  - Por exemplo, podemos desejar estimar NP com um erro não superior a 10%. Ou seja, nós queremos

$$\Pr\left(\frac{|Np - NP|}{NP} \geq r\right) = \Pr(|p - P| \geq rP) = \alpha.$$

- Para esta especificação, substituímos rP ou rp para d nas fórmulas (1) e (2). De (2) obtemos

$$n_0 = \frac{z_\alpha^2 pq}{r^2 p^2} = \frac{z_\alpha^2}{r^2}\frac{q}{p}.$$

- A fórmula (3) permanece inalterada.

A fórmula para n com dados contínuos

n com dados contínuos

- Mais comumente, desejamos controlar o **erro relativo** r na estimativa total ou média da população.
- Com uma amostra aleatória simples tendo média $\overline{y}$, queremos

$$\Pr\left(\left|\frac{\overline{y}-\overline{Y}}{\overline{Y}}\right| \geq r\right) = \Pr\left(\left|\frac{N\overline{y}-N\overline{Y}}{N\overline{Y}}\right| \geq r\right) = \Pr(|\overline{y}-\overline{Y}| \geq r\overline{Y}) = \alpha,$$

em que α é uma probabilidade pequena.

- Assumimos que $\overline{y}$ é normalmente distribuído: do **Corolário 10.1** (aula 10), seu erro padrão é

$$\sigma_{\overline{y}} = \sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}.$$

n com dados contínuos

Portanto,

$$r\overline{Y} = z_{\alpha}\sigma_{\overline{y}} = z_{\alpha}\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

- Resolvendo para n, encontramos

$$n = \left(\frac{z_{\alpha}S}{r\overline{Y}}\right)^2 \Big/ \left[1 + \frac{1}{N}\left(\frac{z_{\alpha}S}{r\overline{Y}}\right)^2\right].$$

- Observe que a característica da população da qual n depende é seu coeficiente de variação $S/\overline{Y}$.
- Isso geralmente é mais estável e fácil de "estimar" antecipadamente que o próprio S.

n com dados contínuos

- Como uma primeira aproximação, tomamos

$$n_0 = \left(\frac{z_\alpha S}{r\overline{Y}}\right)^2 = \frac{1}{C}\left(\frac{S}{\overline{Y}}\right)^2 \qquad (4)$$

substituindo uma estimativa antecipada de $(S/\overline{Y})$. A quantidade C é o $(cv)^2$ desejado da estimativa amostral.

n com dados contínuos

- Se n_0/N é não desprezível, calculamos n como em (3)

$$n = \frac{n_0}{1 + (n_0/N)}.$$

- Se em vez do **erro relativo** r quisermos controlar o **erro absoluto** d em $\overline{y}$, tomamos $n_0 = z_\alpha^2 S^2/d^2 = S^2/V$, em que V é a variância desejada de $\overline{y}$.

Exemplo

- Em viveiros que produzem árvores jovens para venda, é aconselhável estimar, no final do inverno ou início da primavera, quantas árvores jovens saudáveis podem estar disponíveis, uma vez que isso determina a política de solicitação e aceitação de pedidos.
- Um estudo de métodos de amostragem para a estimativa do número total de mudas foi realizado.
- Os dados a seguir foram obtidos de uma camada de mudas de bordo prateado com 1 pé de largura e 430 pés de comprimento.
- A unidade de amostragem foi de 1 pé do comprimento do leito, de modo que $N = 430$.
- Pela enumeração completa do leito se descobriu que $\overline{Y} = 19$, $S^2 = 85,6$, sendo esses os valores reais da população.

Exemplo

- Com a amostragem aleatória simples, quantas unidades devem ser tomadas para estimar $\overline{Y}$ em 10%, além de uma chance de 1 em 20? De (4) obtemos

$$n_0 = \frac{z_\alpha^2 S^2}{r^2 \overline{Y}^2} = \frac{(4)(85,6)}{(1,9)^2} = 95$$

- Uma vez que n_0/N não é desprezível, tomamos

$$n = \frac{95}{1 + \frac{95}{430}} = 78.$$

- Quase 20% do leito deve ser contado para se atingir a precisão desejada.

Exemplo

Pacote PracTools

- A função `nCont` calcula o tamanho de amostra com dados contínuos visto anteriormente.

```r
# install.packages("PracTools")
library(PracTools)

nCont(S2 = 85.6, ybarU = 19, N = 430, CV0 = (0.10/2))
```

```
## [1] 77.70729
```

```r
# diferença ao usar z 'arrendondado'
nCont(S2 = 85.6, ybarU = 19, N = 430, CV0 = (0.10/1.96))
```

```
## [1] 75.168
```

Comentários

- As fórmulas para n fornecidas aqui se aplicam apenas à amostragem aleatória simples em que a média da amostra é usada como a estimativa de $\overline{Y}$.
- As fórmulas apropriadas para outros métodos de amostragem e estimativa são apresentadas com a discussão dessas técnicas **(Amostragem 2)**.

Para casa

- Revisar os tópicos discutidos nesta aula.
- Consultar e estudar a página:

https://sites.google.com/hcpa.edu.br/bioestatistica/softwares-e-aplicativos/pss-health?authuser=0

Próxima aula

- Tamanho de amostra considerando a abordagem do poder (de um teste de hipóteses).

Por hoje é só!

Bons estudos!