MAT02025 - Amostragem 1

AAS: estimação de proporções para classificações em mais de duas categorias

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, 2021

Proporções: duas ou mais categorias

- Até o momento foi tratado o caso em que temos uma variável dicotômica (ou dicotomizada), resultando na classificação da população em duas categorias.
- Muitas vezes temos a necessidade de definir mais de duas categorias.
  - Estudar a distribuição por faixas etárias de um grupo de pessoas.
  - Estudar a classificação econômica das empresas de determinado país.
  - Estimar a intenção de votos dos candidatos em uma eleição com mais de 2 candidatos, além das possibilidades de voto em branco ou nulo ou, ainda, eleitores indecisos.
- Nesses casos, há interesse de estimar a proporção de unidades em cada uma das possíveis categorias e respectiva precisão.

Proporções: duas ou mais categorias

Exemplo: seja uma escola com 1.000 alunos distribuídos entre as 9 etapas do ensino fundamental:

Etapa de ensino	Alunos	Proporção
1° ano	110	0,110
2° ano	108	0,108
3° ano	110	0,110
4° ano	115	0,115
5° ano	104	0,104
6° ano	119	0,119
7° ano	116	0,116
8° ano	107	0,107
9° ano	111	0,111
Total	**1.000**	**1,000**

Proporções: duas ou mais categorias

- Observe que, para calcular as proporções em cada uma das categorias, na verdade o que se faz é atribuir o valor 1 às unidades da categoria em questão e o valor 0 para as unidades pertencentes às demais categorias.
- Em outras palavras, se a variável tem m categorias é como se fossem m problemas com duas categorias.
- A proporção de unidades da população pertencentes à categoria $C \in (1, 2, .., m)$, é dada por:

$$P_C = \frac{A_C}{N},$$

em que A_C é o número de unidades na categoria C e N é o tamanho total da população.

Proporções: duas ou mais categorias

- Seja uma amostra aleatória simples (com ou sem reposição) de tamanho n e seja a variável indicadora Y_i definida como:

$$Y_i = \begin{cases} 1, & \text{se a unidade } i \text{ pertence à categoria } C \\ 0, & \text{se a unidade } i \text{ pertence a outra categoria} \end{cases}$$

- Com tal definição pode-se ver que o número de unidades da categoria C na amostra será dado por:

$$a_C = \sum_{i=1}^{n} Y_i.$$

- Um estimador para a proporção de unidades populacionais pertencentes à categoria C é dado por:

Proporções: duas ou mais categorias

$$p_C = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{a_C}{n}.$$

Proporções: duas ou mais categorias

- O problema foi reduzido ao caso de estimar proporções em variáveis com duas categorias.
- Pode-se obter, também, estimativas de precisão utilizando os mesmos resultados já apresentados nas aulas 17, 18 e 19.

Proporções: duas ou mais categorias

- Muitas vezes pode-se estar interessado em estimar proporções para agrupamentos das categorias originais.
- Voltando ao exemplo da escola do ensino fundamental, pode ser de interesse estudar a proporção de seus alunos que estão matriculados no primeiro segmento do ensino fundamental.
  - Nesse caso, seriam contabilizados como pertencentes à categoria C de interesse todos os alunos do 1º até o 5º ano, para os quais $Y_C = 1$, sendo $Y_C = 0$ para os demais alunos da escola.

Proporções: duas ou mais categorias

- Outro caso de interesse ocorre quando, na aplicação de um questionário, por exemplo, aparecem respondentes que se recusaram a responder ou, mesmo, disseram que não sabiam a resposta.
- Num caso como esse, pode-se estar interessado em estimar a proporção das pessoas que responderam determinada alternativa, entre as pessoas que efetivamente responderam a pesquisa escolhendo uma das alternativas válidas.
- Um exemplo prático seria uma pesquisa sobre a intenção de voto numa eleição com apenas dois candidatos.
  - Nesse caso, o entrevistado poderia responder que votará no canditato A, no candidato B, que votará **nulo** ou *em branco*, onde apenas as duas primeiras alternativas seriam consideradas como **votos válidos**.

Proporções: duas ou mais categorias

- Pode-se estimar a proporção para cada uma das quatro categorias iniciais ou apenas a proporção de votos válidos para cada um dos dois candidatos:

$$p_A = \frac{a_A}{a_A + a_B} \quad \text{e} \quad p_B = \frac{a_B}{a_A + a_B}.$$

- Vale notar que na expressão acima, tanto o numerador como o denominador do estimador da proporção são variáveis aleatórias, pois a população (eleitores que efetivamente vão votar num dos candidatos) é desconhecida.

Proporções: duas ou mais categorias

```
library(survey)
data(api)

# View(apisrs)
# ?apisrs

# O objeto design
api.des <- svydesign(id = ~1,
                     fpc = ~fpc,
                     data = apisrs)

# Estimativa da proporção de escolas
# do tipo "Elementary/Middle/High School"
svyciprop(formula = ~I(stype == "E"), design = api.des)
```

```
##                         2.5% 97.5%
## I(stype == "E") 0.710 0.644  0.77
```

```
svyciprop(formula = ~I(stype == "M"), design = api.des)
```

```
##                         2.5% 97.5%
## I(stype == "M") 0.165 0.120  0.22
```

Proporções: duas ou mais categorias

```
svyciprop(formula = ~I(stype == "H"), design = api.des)
```

```
##                           2.5% 97.5%
## I(stype == "H") 0.1250 0.0861  0.18
```

Para casa

- Revisar os tópicos discutidos nesta aula.
- Faça uma busca por um levantamento por amostragem que tenha apresentado (parte dos) seus resultados em termos de proporções. Investigue os procedimentos de amostragem e métodos de estimação. Compartilhe os seus achados no Fórum Geral do Moodle.

Próxima aula

- Proporções e totais das subpopulações.

Por hoje é só!

Bons estudos!