

Especialização em *Data Science* e Estatística Aplicada

Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

Goiânia, 2024

IME

INSTITUTO DE
MATEMÁTICA E
ESTATÍSTICA

FEN

FACULDADE DE
ENFERMAGEM



UFG

UNIVERSIDADE
FEDERAL DE GOIÁS



Aula 3 - Parte 2

1. Análise de aderência e associação

- Teste de homogeneidade - (*Test of homogeneity*)

2. Referências Bibliográficas

Teste de homogeneidade

- O teste de independência está relacionado à questão:
 - Os dois critérios de classificação são independentes?
- Já o teste de homogeneidade está relacionado à:
 - As amostras são retiradas de populações homogêneas (semelhantes) com relação a algum critério de classificação?
- Em um teste de homogeneidade, testamos a afirmativa de que populações diferentes têm a mesma proporção de alguma característica.

Teste de homogeneidade

- Considere uma tabela de contingência em que as categorias de coluna podem representar as diferentes populações das quais as amostras são extraídas e as categorias de linha são níveis da variável categórica.
- Se, por exemplo, três populações forem amostradas, elas podem ser designadas como populações 1, 2 e 3, caso em que esses rótulos podem servir como títulos de coluna.
- Se a variável de interesse tiver três categorias, digamos, A, B e C, esses rótulos podem servir como títulos para linhas.

Teste de homogeneidade

Tabela 1: Tabela de contingência para dados para um teste Qui-Quadrado de homogeneidade

Variável categórica	População			Total
	1	2	3	
A	n_{A1}	n_{A2}	n_{A3}	$n_{A.}$
B	n_{B1}	n_{B2}	n_{B3}	$n_{B.}$
C	n_{C1}	n_{C2}	n_{C3}	$n_{C.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

Teste de homogeneidade

- Se as populações forem, de fato, homogêneas, ou, equivalentemente, se as amostras forem todas retiradas da mesma população, com relação às categorias A, B e C, nossa melhor estimativa da proporção na população que pertence à categoria A é $n_{A.}/n$;
- Analogamente, temos que a melhor estimativa da proporção na população que pertence à categoria B é $n_{B.}/n$ e que pertence à categoria C é $n_{C.}/n$.

Teste de homogeneidade

Exemplo 5

- Narcolepsia é uma doença que envolve distúrbios do ciclo sono-vigília.
- Membros da Sociedade Alemã de Enxaqueca e Cefaleia estudaram a relação entre enxaquecas em 96 indivíduos diagnosticados com narcolepsia e 96 controles saudáveis.
- Os resultados são mostrados na Tabela 2.
- Desejamos saber se podemos concluir, com base nesses dados, que a população com narcolepsia e as populações saudáveis representadas pelas amostras não são homogêneas com relação à frequência de enxaqueca.

Teste de homogeneidade

Continuação do Exemplo 5

Tabela 2: Frequência de enxaquecas por estado de narcolepsia.

Indivíduos	Enxaquecas relatadas		Total
	Sim	Não	
Narcolépticos	21	75	96
Saudáveis	19	77	96
Total	40	152	192

Fonte: The DMG Study Group, Migraine and Idiopathic Narcolepsy—A Case-Control Study, *Cephalagia*, 23, 786–789, 2003.

Teste de Homogeneidade

Solução do Exemplo 5

1. **Suposição:** Assumimos que temos uma amostra aleatória simples de cada uma das duas populações de interesse.

2. **Hipóteses:**

H_0 : As duas populações são homogêneas em relação à frequência de enxaqueca.

H_1 : As duas populações não são homogêneas em relação à frequência de enxaqueca.

Teste de Homogeneidade

Solução do Exemplo 5

3. **Estatística de teste:** A estatística de teste é

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

4. **Distribuição da estatística de teste:** Quando H_0 é verdadeira, χ^2 tem distribuição qui-quadrado com 1 grau de liberdade, pois $(r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$.

Teste de Homogeneidade

Solução do Exemplo 5

5. Cálculo do valor da estatística de teste

Neste caso, vamos calcular as frequências esperadas:

$$e_{11} = \frac{96 \times 40}{192} = 20, \quad e_{12} = \frac{96 \times 152}{192} = 76,$$

$$e_{21} = \frac{96 \times 40}{192} = 20, \quad e_{22} = \frac{96 \times 152}{192} = 76.$$

Logo,

$$\chi^2 = \frac{(21 - 20)^2}{20} + \frac{(75 - 76)^2}{76} + \frac{(19 - 20)^2}{20} + \frac{(77 - 76)^2}{76} = 0,126.$$

Teste de Homogeneidade

Solução do Exemplo 5

6. Calculando o valor- p no R:

$\hat{\alpha} = \text{pchisq}(0.126, \text{df} = 1, \text{lower.tail} = \text{FALSE}) = 0,723$

Teste de Homogeneidade

Solução do Exemplo 5

7. **Conclusão:** Ao nível de 1% não rejeitamos a hipótese nula, pois $\hat{\alpha} > \alpha = 0,01$.

Portanto, concluímos que as duas populações podem ser homogêneas em relação à frequência de enxaqueca.

Em outras palavras, os indivíduos narcolépticos e saudáveis são homogêneos, ao nível de 1%, em relação à frequência de enxaqueca.

Teste de Homogeneidade

Continuação do Exemplo 5

- Exemplo 5 no R.

Teste de Homogeneidade - Aplicação à dados reais

Exemplo 6

- Consideremos os dados do vacinômetro que indicam o número de doses da vacina contra COVID-19 administradas em diferentes faixas etárias em três diferentes municípios do estado de Goiás, a saber: Niquelândia, Padre Bernardo e Posse.
- Queremos verificar se a distribuição das doses aplicadas entre as faixas etárias é semelhante em todos os três municípios, ou seja, queremos testar a homogeneidade da distribuição das vacinas.

Teste de Homogeneidade - Aplicação à dados reais

Continuação do Exemplo 6

- **Hipóteses**

H_0 : As distribuições das vacinas por faixa etária são homogêneas entre os municípios de Niquelândia, Padre Bernardo e Posse.

H_1 : As distribuições das vacinas por faixa etária diferem entre os municípios de Niquelândia, Padre Bernardo e Posse.

Teste de Homogeneidade - Aplicação à dados reais

Continuação do Exemplo 6

- Exemplo 6 no R.

Teste de Homogeneidade

Frequências esperadas (pequenas)

- As regras para pequenas frequências esperadas fornecidas no teste de independência são aplicáveis ao realizar um teste de homogeneidade.

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Independentes

- O teste qui-quadrado de homogeneidade para o caso de duas amostras fornece um método alternativo para testar a hipótese nula de que duas proporções populacionais independentes são iguais.
- Vimos anteriormente, que o valor observado da estatística de teste para testar $H_0 : p_1 = p_2$ contra $H_1 : p_1 \neq p_2$ é

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

onde $\bar{p} = (x_1 + x_2)/(n_1 + n_2)$, x_1 e x_2 são os números de sucessos observados nas duas amostras, n_1 e n_2 são os tamanhos das duas amostras.

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Independentes

Exemplo 7

Suponha, por exemplo, que em um teste de $H_0 : p_1 = p_2$ contra $H_1 : p_1 \neq p_2$, os dados da amostra foram os seguintes: $n_1 = 100$, $n_2 = 120$, $\hat{p}_1 = 0,60$ e $\hat{p}_2 = 0,40$. Logo,

$$\bar{p} = \frac{0,60 \times 100 + 0,40 \times 120}{100 + 120} = 0,4909,$$

Então,

$$z_{obs} = \frac{0,60 - 0,40}{\sqrt{0,4909 \times (1 - 0,4909) \left(\frac{1}{100} + \frac{1}{120} \right)}} = 2,95469.$$

valor- p : $\hat{\alpha} = 2 * \text{pnorm}(2.95469, \text{lower.tail} = \text{FALSE}) = 0.0031298$.

Logo, rejeitamos a hipótese nula, pois $\hat{\alpha} < \alpha = 1\%$.

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Independentes

Continuação do Exemplo 7

Se quisermos testar a mesma hipótese usando a abordagem qui-quadrado, nossa tabela de contingência será

Tabela 3: Frequências observadas (Frequências esperadas).

Amostra	Característica presente		Total
	Sim	Não	
1	60 (49,09)	40 (50,91)	100
2	48 (58,91)	72 (61,10)	120
Total	108	112	220

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Independentes

Continuação do Exemplo 7

O valor da estatística de teste qui-quadrado é

$$\chi_{obs}^2 = \frac{(60 - 49,09)^2}{49,09} + \frac{(40 - 50,91)^2}{50,91} + \frac{(48 - 58,91)^2}{58,91} + \frac{(72 - 61,10)^2}{61,10} = 8,73.$$

Valor- p : $\hat{\alpha} = \text{pchisq}(8.73, \text{df} = 1, \text{lower.tail} = \text{FALSE}) = 0.003130$.

Logo, rejeitamos a hipótese nula, pois $\hat{\alpha} < \alpha = 1\%$.

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Independentes

Continuação do Exemplo 7

- Vemos, portanto, que chegamos à mesma conclusão por ambos os métodos.
- Isso não é surpreendente porque, pois $\chi_{obs}^2 = z_{obs}^2$, ou seja, $8,73 = (2,95469)^2$.

Teste de Homogeneidade - $H_0 : p_1 = p_2$ para Amostras Dependentes

- É comum que tenhamos medidas do mesmo conjunto de indivíduos em duas ocasiões diferentes, e desejamos saber se houve uma alteração na proporção.
- Neste caso, podemos usar o teste de McNemar, que veremos na disciplina de Métodos não paramétricos.

Análise de aderência e associação

Resumindo,

- **teste de aderência:** compara dados observados com uma distribuição teórica conhecida;
- **teste de independência:** avalia se duas variáveis categóricas são relacionadas ou independentes;
- **teste de homogeneidade:** compara distribuições entre dois ou mais grupos para verificar se são semelhantes.

Referências bibliográficas

1. Ministério da Saúde - Vacinômetro COVID-19. https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html, último acesso: 09/09/2024.
2. Vieira, S. Introdução à Bioestatística, 5ª Edição, Elsevier, 2008.

Especialização em *Data Science* e Estatística Aplicada

Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

tmelo@ufg.br

IME

INSTITUTO DE
MATEMÁTICA E
ESTATÍSTICA

FEN

FACULDADE DE
ENFERMAGEM



UFG

UNIVERSIDADE
FEDERAL DE GOIÁS

