

Testes Qui-Quadrado - Teste de Aderência

- Consideremos uma tabela de frequências com k frequências, $k \geq 2$
 - k : total de categorias
 - frequências observadas: O_1, \dots, O_k
 - seja $p_1 = p_{01}, \dots, p_k = p_{0k}$ as probabilidades especificadas e associadas as k categorias
 - $\sum_{i=1}^n O_i = n$
 - frequências esperadas: $E_1, \dots, E_k, E_i = np_{0i}$
 - $\Rightarrow \sum_{i=1}^n E_i = n$

Testes Qui-Quadrado - Teste de Aderência

- $H_0 : p_1 = p_{01}, \dots, p_k = p_{0k}$ vs $H_1 : \text{ao menos uma é diferente}$
- Estatística do teste: $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$
- Resultado: assumindo H_0 como verdadeira, se as k categorias são mutuamente exclusivas e as E_i são suficientemente grandes, então χ^2 tem distribuição Qui-Quadrado com $k - 1$ graus de liberdade.
- Rejeição do teste: se χ^2 assumir valores grandes $\Rightarrow O_i$ é muito diferente de E_i , logo H_0 não é verdadeira

$$R_c = \{\chi^2 \geq c\} \text{ onde } c \text{ depende do nível de significância } \alpha$$

Testes Qui-Quadrado - Teste de Homogeneidade

- Suponha r subpopulações S_1, \dots, S_r . De cada subpopulação é extraída uma amostra de n_i elementos, $i = 1, \dots, r$, com $\sum_{i=1}^r n_i = n$. Em seguida, os n_i elementos de S_i são distribuídos segundo c categorias A_1, \dots, A_C

| | A_1 | ... | A_C | Total |
|----------|---------------|----------|---------------|--------------|
| S_1 | n_{11} | ... | n_{1C} | $n_{1\cdot}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| S_r | n_{r1} | ... | n_{rC} | $n_{r\cdot}$ |
| Total | $n_{\cdot 1}$ | ... | $n_{\cdot C}$ | n |

Testes Qui-Quadrado - Teste de Homogeneidade

- Objetivo: verificar se as distribuições de probabilidade das categorias A_1, \dots, A_C são as mesmas para as r subpopulações
- $H_0 : P_1(A_1) = \dots = P_r(A_1), \dots, P_1(A_C) = \dots = P_r(A_C)$ vs H_1 :
ao menos uma é diferente
- $P_i(A_j)$ = probabilidade de um elemento da subpopulação i ser classificado na categoria A_j
- Estatística do teste:
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Testes Qui-Quadrado - Teste de Homogeneidade

- O_{ij} : frequência observada na subpopulação i , na categoria j
- E_{ij} : frequência esperada na subpopulação i , na categoria j
- sob H_0 : $E_{ij} = \frac{n_i \cdot n_j}{n}$
- Região crítica do teste: $R_c = \{\chi^2 \geq c\}$
- χ^2 tem uma distribuição Qui-Quadrado com $(r - 1)(C - 1)$ graus de liberdade

Testes Qui-Quadrado - Exemplo 1

- Teste de Aderência: Conforme a herança mendeliana, a descendência de certo cruzamento deveria ser só vermelho, preta ou branca na seguinte proporção: $p_{01} = \frac{9}{16}$, $p_{02} = \frac{3}{16}$ e $p_{03} = \frac{4}{16}$. Se um experimento mostrou $O_1 = 74$, $O_2 = 32$ e $O_3 = 38$ descendentes nessas categorias respectivamente, a teoria está afirmada?
 - $O_1 + O_2 + O_3 = n = 144$
 - $E_1 = np_{01} = 81$, $E_2 = np_{02} = 27$ e $E_3 = np_{03} = 36$
 - $\chi^2 = 0.60 + 0.93 + 0.11 = 1.64$
 - se $\alpha = 0.05$, o valor c da região de rejeição é $c = 5.9915$ (2 graus de liberdade) \Rightarrow não rejeita H_0

Testes Qui-Quadrado - Exemplo 2

- Teste de Homogeneidade: considere 2 escolas diferentes, e seus estudantes são submetidos a um mesmo exame, em que A, B, C, D e E são as notas por eles obtidas

| | A | B | C | D | E | Total |
|----------|----|----|-----|----|----|-------|
| escola 1 | 18 | 39 | 129 | 48 | 66 | 300 |
| escola 2 | 18 | 26 | 41 | 6 | 9 | 100 |
| Total | 36 | 65 | 170 | 54 | 75 | 400 |

Testes Qui-Quadrado - Exemplo 2

- A distribuição das notas obtidas pelos alunos é a mesma nas duas escolas?
 - $r = 2$
 - $C = 5$
 - $E_{11} = 27, E_{12} = 48.75, E_{13} = 127.5, E_{14} = 40.5, E_{15} = 56.25$
 - $E_{21} = 9, E_{22} = 16.25, E_{23} = 42.5, E_{24} = 13.5, E_{25} = 18.75$
 - $\chi^2 = 32.186$
- se $\alpha = 0.05$, com $(r - 1)(C - 1) = 4$ graus de liberdade, $c = 9.4877$
 \Rightarrow rejeita H_0

Testes Qui-Quadrado - Teste de Independência

- Sejam n indivíduos selecionados aleatoriamente de uma população. Vamos classificar cada indivíduo segundo 2 variáveis A e B.
 - A tem r categorias
 - B tem c categorias

| | A_1 | ... | A_r | Total |
|----------|----------|----------|----------|----------|
| B_1 | n_{11} | ... | n_{1r} | $n_{1.}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| B_C | n_{C1} | ... | n_{Cr} | $n_{C.}$ |
| Total | $n_{.1}$ | ... | $n_{.r}$ | n |

Testes Qui-Quadrado - Teste de Independência

- Objetivo: testar se A e B são independentes
- $H_0 : P(A_i \text{ e } B_j) = P(A_i)P(B_j)$ vs H_1 : ao menos uma é diferente
- Observação: diferença entre os testes de homogeneidade e independência
 - Teste de homogeneidade: selecionamos uma amostra de elementos de cada uma das r subpopulações e distribuímos os elementos de cada uma dessas amostras segundo C categorias
 - Teste de independência: distribuímos uma amostra de n elementos de "uma" população segundo as categorias da variável A e as categorias da variável B.