UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

MAT02215 - Estatística Geral 2 - 2021/2

Plano Aula 23 e 24

Markus Stein

(continuação) Testes de Aderência e Associação

Nosso interesse agora será em teste para mais de duas proporções para **duas variáveis categóricas** de interesse.

- **Exemplo 1**: A quantidade de abstenções na última eleição para reitor na UFRGS está associada à categoria dos votantes?
- **Exemplo 2**: As proporções de votos em cada chapa pode estar associada à categoria dos votantes?

Teste Qui Quadrado de Associação/Independência (Bussab e Morettin - seção 14.4)

Para duas variáveis qualitativas podemos estar interessados em verificar se existe associação (relação) entre elas.

- O que significam frequências esperadas nesse caso?
- **. . . cont. exemplo 1**: Os dados observados de abstenções e total de votos por categorias foram

	Docentes	Técnicos	Estudantes	Total
Votos	2605	1828	11292	15725
Abstenções	327	717	28836	29880
Total habilitados	2932	2545	40128	45605

- **. . . cont. exemplo 2**: Os dados observados de votos em cada chapa por categorias foram

	Docentes	Técnicos	Estudantes	Total
Chapa 1	436	208	1216	1860
Chapa 2	1454	516	2713	4683
Chapa 3	679	1056	7212	8947
Total votos válidos	2605	1828	11292	15725

Como usar proporções/frequências observadas para testar as hipóteses acima?

Relembrando sobre independência em probabilidade:

- Duas variáveis aleatórias X e Y são ditas *independentes* se e somente se a probabilidade conjunta é igual ao produto das probabilidades marginais

$$P(X = i, Y = j) = P(X = i) \times P(Y = j),$$

para todo $i = 1, \ldots, k$ e $j = 1, \ldots, l$.

- No teste de associação, a hipótese nula é dada por

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j} \text{ ou } H_0 : P(X = i, Y = j) = P(X = i) \times P(Y = j).$$

Estatística qui-quadrado

Seja n_{ij} o número de observações pertencentes a categoria conjunta $(X = i, Y = j)$.

- *proporções* (frequências relativas) **observadas** $p_{ij} = n_{ij}/n$ ou *frequências* observadas $o_{ij} = n_{ij}$;
- *proporções* **esperadas** π_{ij} (sob H_0) ou *frequências* esperadas $e_{ij} = \frac{n_{i.}n_{.j}}{n}$. em que $n_{i.} = \sum_{i=1}^{k} n_{ij}$ é o número de elementos na categoria i de X e $n_{.j} = \sum_{j=1}^{l} n_{ij}$ na categoria j de Y.

Teorema (**Distribuição Qui-Quadrado 2, nossa versão 2**): Sob certas suposições

$$Q = \sum_{i,j} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2_{(\nu)}$$

em que $i = 1, \ldots, k$ é o índice da linha e $j = 1, \ldots, l$ das colunas e $\nu = (k-1) \times (l-1)$ são os graus de liberdade.

- Quais as suposições necessárias???
- Para tabelas 2×2 devemos usar a correção (de continuidade) de Yates!

Análise dos resíduos

O que as diferenças $o_{ij} - e_{ij}$ nos dizem sobre a associação entre X e Y?

- Os resíduos podem res vistos como o desvio de uma observação em relação ao seu valor esperado sob H_0.

$$res_{ij} = \frac{o_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$

Ler slides das aulas 23 e 24

Continuar os exercícios da lista 3-1 e fazer a lista 3-2

Fazer avaliação pontual 1 da área 3

REFERÊNCIA EXTRA

Página 'Probabilidade e Estatística (EaD)' da UFRGS

- Capítulo 6 - Inferência para dados categóricos
  - Seção - Testando a qualidade do ajuste usando a qui-quadrado