# Bioestatística

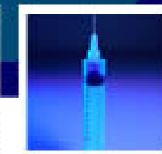
## Aula 8

## Testes para Dados Categorizados

Prof. Tiago A. E. Ferreira

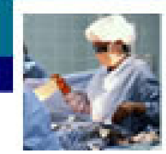
# Categorização

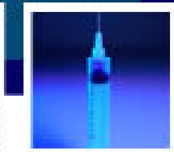




- Dados Categorizados:
  - Contagem de frequência de uma variável classificada ou subdividida em categorias.
  - Tipicamente procedimento aplicado à variáveis qualitativas.
      - » Ex.: Variável Sexo – Masculino e Feminino
  - Entretanto também é possível se aplicar à variáveis quantitativas.
    - Ex.: Variável Duração da Reação -. Até 2  horas; Mais de 2 horas.

# Tabelas de Contingência





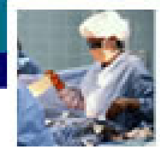
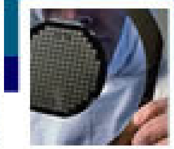
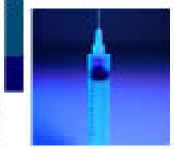
- Tipicamente, dados categorizados são apresentados em tabelas de contingência:

| | | **Fator Discriminado B** | | | | |
|---|---|---|---|---|---|---|
| | | **$B_1$** | **$B_2$** | **...** | **$B_s$** | **Totais A** |
| **Fator Discriminante A** | **$A_1$** | $O_{11}$ | $O_{12}$ | ... | $O_{1s}$ | **$T_{A1}$** |
| | **$A_2$** | $O_{21}$ | $O_{22}$ | ... | $O_{2s}$ | **$T_{A2}$** |
| | ⋮ | ⋮ | ⋮ | ... | ⋮ | ⋮ |
| | **$A_r$** | $O_{r1}$ | $O_{r2}$ | ... | $O_{rs}$ | **$T_{Ar}$** |
| | **Totais B** | **$T_{B1}$** | **$T_{B2}$** | **...** | **$T_{Bs}$** | T |

# Tabelas de Contingência

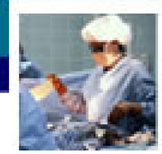
- A variável *A* foi denominada de *fator discriminante*
  - A partir das mudanças nas categorias de *A* deseja-se verificar o comportamento da variável *B*, chamada *fator discriminado.*
  - Assim,
    - Variável *A* é a variável independente.
    - Variável *B* é a variável dependente.

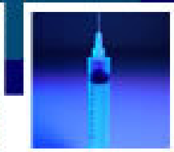
# Testes Categorizados

- O objetivo dos testes para dados categorizados é determinar se o *fator discriminante* exerce alguma influência sobre o *fator discriminado*.
  - **A hipótese nula:**
    - **H0 $\rightarrow$ $p_{A1} = p_{A2} = p_{Ar}$**
      - As categorias de *A* exercem a mesma influência sobre *B*
  - **Contra $H_1$:**
    - Pelo menos uma categoria apresenta diferençaem relação a *B*.

# Teste de Qui-Quadrado Clássico

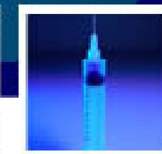




- A soma dos quadrados de variáveis normais tem uma distribuição chamada Qui-Quadrada.
  - Desta forma, o teste de Qui-Quadrado irá verificar se uma variável adere ou não a uma distribuição Qui-Quadrada.

# Teste de Qui-Quadrado Clássico





- Primeiro Passo:
  - Construir a matriz r×s de valores esperados, E, dada por:

$$E_{ij} = \frac{\sum_{l=1}^{s} O_{il} \cdot \sum_{k=1}^{r} O_{kj}}{\sum_{l=1}^{s} \sum_{k=1}^{r} O_{lk}} = \frac{T_{A_i} \cdot T_{B_j}}{T}$$

# Teste de Qui-Quadrado Clássico

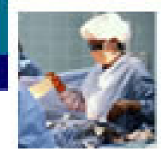
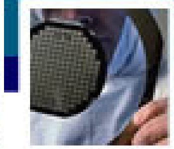
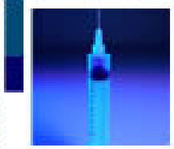
- Segundo Passo:
  - Calcula-se a estatística: $$\chi_c^2 = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

    - Esta é a expressão do **teste Qui-Quadrado Clássico.**
      - Esta expressão só pode ser utilizada se o número total de dados for maior que 40.
      - Caso o número de dados estiver entre 20 e 40, e o valor esperado das células for maior que 5, deve-se ser aplicada a correção de Yates:

$$\chi_c^2 = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{\left(\left|O_{ij} - E_{ij}\right| - 0{,}5\right)^2}{E_{ij}}$$

      - Caso nº menor que 20, aplica-se o teste exato de Fisher

# Teste de Qui-Quadrado Clássico

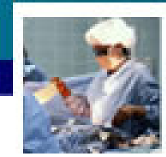




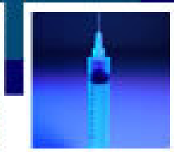
- O Valor calculado $\chi_c^2$ é comparado com o valor tabelado $\chi_t^2$ da distribuição Qui-Quadrado.
  - Este valor está associado a um nível de significância $\alpha$ e a uma quantidade de graus de liberdade $gl=(s-1)(r-1)$.
  - Aplica-se a Regra de Decisão:

$$Se \quad \chi_c^2 \leq \chi_t^2 \rightarrow H_0 \; deve \; ser \; aceita$$

$$Se \quad \chi_c^2 > \chi_t^2 \rightarrow H_0 \; deve \; ser \; rejeitada$$

# Exemplo



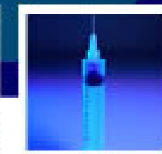


- Um grupo de 154 pacientes que apresentaram dor abdominal foram divididos em dois grupos:
  - Grupo de Tratamento (T)
  - Grupo de Controle (C)

| | **Permanência da Dor Abdominal** | | |
|---|---|---|---|
| **Grupo** | **Sim** | **Não** | **Total** |
| T | 6 | 57 | **63** |
| C | 30 | 61 | **91** |
| **Total** | **36** | **118** | **154** |

# Exemplo





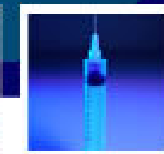
- Hipóteses:
  - Hipótese Nula:
    - $H_0 \rightarrow p_C = p_T$
      - Ou seja, os resultados do grupo de controle são iguais ao grupo de tratamento.
  - Hipótese Alternativa:
    - $H_1 \rightarrow p_C \neq p_T$

# Exemplo

- Construção da Matriz E
  - E11 = (63 . 36) / 154 $\cong$ 14,73
  - E12 = (63 . 118) / 154 $\cong$ 48,27
  - E21 = (91 . 36) / 154 $\cong$ 21,27
  - E22 = (91.118) / 154 $\cong$ 69,73

# Exemplo



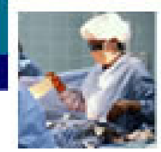


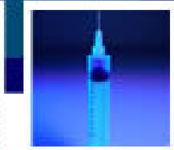
- Quantificação das diferenças entre as matrizes O e E:

$$\chi_c^2 = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} = \frac{(6-14{,}73)^2}{14{,}73} + \frac{(57-48{,}27)^2}{48{,}27} + \cdots + \frac{(61-69{,}73)^2}{69{,}73}$$

$$\chi_c^2 = 11{,}4290$$

# Exemplo





- O número de graus de liberbuiade será (2-1)(2-1) = 1.
  - Assim para este grau de liberdade, os valores da distribuição Qui-Quadrado será:

| $\alpha$ | $\chi_t^2$ |
|---|---|
| 10% | 2,7056 |
| 5% | 3,8416 |
| 1% | 6,6354 |
| 0,0724% | 11,4290 |

- Assim, aos níveis de 10%, 5% e 1% de significância H0 deve ser rejeitada!

# Distribuição F

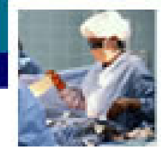
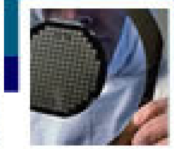
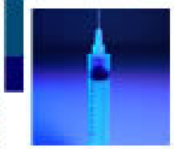
- A razão entre duas duas distribuições Qui-Quadradas tem uma distribuição F.

$$F(n,m) = \frac{\left(\chi^2(n)\big/ n\right)}{\left(\chi^2(m)\big/ m\right)}$$

  – Se cada uma das ditribuições qui-quadradas têm graus de liberdade n e m, então a distribuição F terá dois valores de graus de liberdade.

  – Ditribuição F – F(n,m)

# Teste F

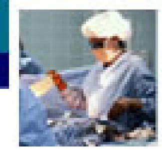
- Dado duas somas quadradas, SSi e SSj, com grau de liberdade vi e vj.
  - Assim, a hipótese de que SSi é menor ou igual a SSj é rejeitada com significância $\alpha$ se a razão

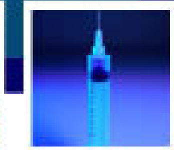
$$\frac{SSi/vi}{SSj/Vj}$$

  é maior que 1- $F_{[1-\alpha;vi,vj]}$

  - *Ou seja, SSi é* maior que SSj caso F computado seja maior que o F tabelado!

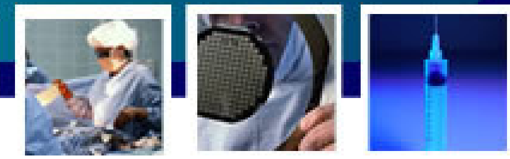
# Teste Exato de Fisher





- Este teste é utilizado para comparar dados categóricos em tabelas 2×2 quando o número total de casos é menor que 20.
  - Passo 1
    - A partir da tabela de contingência são construídas duas outras tabelas, X1 e X2:
      - X1 e X2 são elaboradas tomando-se o total dos elementos de uma das linhas, considerando que estes elementos estão primeiro na condição B1 (ou B2) e depois na condição B2 (ou B1).

# Teste Exato de Fisher



- Passo 2:
  – A partir da matriz original e as matrizes x1 e x2, calcula-se a estatística:

$$F = \frac{T_{A1}! \cdot T_{A2}! \cdot T_{B1}! \cdot T_{B2}!}{\left( \prod_{i=1}^{s} \prod_{j=1}^{r} O_{ij}! \right) \cdot T!}$$

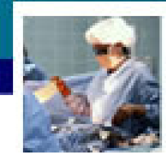
- Passo 3:
  – O Nível de significância é calculado somando os valores obtidos no Passo 2:

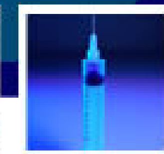
$$p = F_O + F_{X1} + F_{X2}$$

# Teste Exato de Fisher

- Passo 4:
  - Com o valor de *p* calculado toma-se a decisão estatística conveniente, aceitando-se ou rejeitando-se $H_0$.

# Exemplo





- Suponha um grupo de 16 ratos, divididos em dois grupos, Experimental e Normal.
  - O grupo Experimental é formado por 9 ratos geneticamente modificados.
  - Após um ano e meio o nº de ratos vivos do grupo Experimental e do Normal seja:

| | **Sobrevida +1,5 anos** | | |
|---|---|---|---|
| **Grupo** | Vivos | Mortos | **Total** |
| Normal | 5 | 2 | 7 |
| Experimental | 1 | 8 | 9 |
| **Total** | 6 | 10 | **16** |

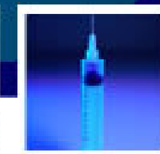
# Exemplo

- O Teste de Fisher consiste em elaborar, com base nos totais marginais do fator discriminante, duas tabelas, X1 e X2 (Matrizes Extremas)
  - X1 é a matriz em que todos os animais mortos pertencem ao grupo normal
    - Observe: são 10 mortos, mas o grupo normal só tem 7 ratos!

| | **Sobrevida +1,5 anos** | | |
|---|---|---|---|
| **Grupo** | Vivos | Mortos | **Total** |
| Normal | 0 | 7 | 7 |
| Experimental | 6 | 3 | 9 |
| **Total** | 6 | 10 | **16** |

# Exemplo



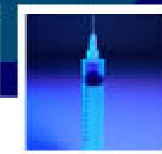


- X2 corresponde a uma matriz com os mesmos totais marginais, mas com todos os animais vivos no grupo normal

| | **Sobrevida +1,5 anos** | | |
|---|---|---|---|
| **Grupo** | Vivos | Mortos | **Total** |
| Normal | 6 | 1 | 7 |
| Experimental | 0 | 9 | 9 |
| **Total** | 6 | 10 | **16** |

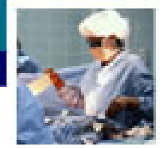
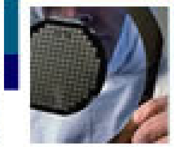
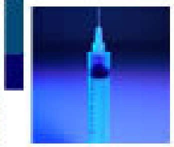
# Exemplo





- Observe que:
  - Foi escolhida a condição “Normal” ($A_1$) como referência
    - Para esta, foi considerados primeiramente todos os animais mortos (B2), gerando a matriz X1, e depois a condição vivos (B1) para a elaboração da matriz X2
- Caso a condição “Experimental” ($A_2$) tivesse sido escolhida como referência e o fator B1 tivesse sido usado para obter X1 e B2 para X2 os resultados seriam idênticos!

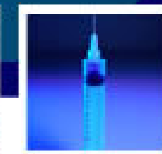
# Exemplo

- Calculando-se a estatística F:
  - Matriz Original: $$F_0 = \frac{7! \cdot 9! \cdot 10! \cdot 6!}{2! \cdot 5! \cdot 8! \cdot 1! \cdot 16!} = 0{,}02360$$
  - Matriz X1: $$F_{X1} = \frac{7! \cdot 9! \cdot 10! \cdot 6!}{0! \cdot 7! \cdot 6! \cdot 3! \cdot 16!} = 0{,}01050$$
  - Matriz X2: $$F_0 = \frac{7! \cdot 9! \cdot 10! \cdot 6!}{6! \cdot 1! \cdot 0! \cdot 9! \cdot 16!} = 0{,}0009$$

# Exemplo

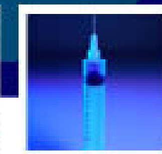




- Finalmente, o valor de p:
  - $p = F_0 + F_{X1} + F_{X2} = 0{,}0236 + 0{,}0105 + 0{,}0009 = 0{,}0350$ *ou 3,5%*
  - Assim, é possível afirmar que a sobrevivência dos ratos alterados é menor que a dos ratos normais envolve uma probabilidade de 3,5% de erro
    - Logo, ao nível de 3,5% de significância, rejeita-se $H_0$.

# Referências






- Livro Texto: Bioestatística. Teoriaia e Computacional (Héctor G. Arango). Guanaba
  - Capítulo 9
- Leituras Complementares:
  - Livro: Bioestatística. Princípios e Aplicações (Sidia M. Callegari-Jaqcques)
    - Capítulo 15.
  - Livro: Estatística Básica (Wilton de O. Bussab e Pedro A. Morettin). Saraiva
    - Capítulo: 7