Tabelas Bidimensionais

Unidade I

Parte 3

Estatística Qui-quadrado de Pearson

- *EXEMPLO:*

*A linhagem produzida pelo cruzamento de entre dois tipos de planta pode ter qualquer um de três genótipos designados por **A**, **B** e **C**. Um modelo teórico de herança genética sugere que a linhagem dos tipos **A**, **B** e **C** deve estar na razão de 1 : 2 : 1. Para verificação experimental, 90 plantas foram geradas pelo cruzamento dos dois tipo de plantas. Sua classificações genéticas estão registradas na tabela a seguir.*

Genótipo	Nº de Plantas
A	18
B	44
C	28
Total	**90**

Estes dados confirmam ou contradizem o modelo genético ?

Estatística Qui-quadrado de Pearson

A hipótese nula do modelo genético corresponde a $\pi_1 = 0{,}25, \pi_2 = 0{,}5$ *e* $\pi_3 = 0{,}25$

onde π_i *é a probabilidade de ocorrência de cada genótipo.*

Assim,

$$H_0\,)\ \pi_1 = 0{,}25, \pi_2 = 0{,}5, \pi_3 = 0{,}25$$

Se H_0 *é verdadeira , se espera observar cerca de* $1/4$ *de plantas do genétipo A, ou seja, a freqüência esperada do genótipo A é dada por:*

$$\mu_1 = n\pi_1 = 90 \times 0{,}25 = 22{,}5$$

Analogamente pode-se calcular as freqüências esperadas do genótipo B e do genótipo C.

A idéia é comparar as freqüências amostrais das células com as esperadas para decidir se os dados contradizem H_0 *. Quanto maior as diferenças , mais forte a evidência contra* H_0*.*

Estatística Qui-quadrado de Pearson

Deseja-se testar a hipótese nula (H_0 *) que as probabilidades das células de uma tabela de contingência são iguais a certos valores fixados* $\{\pi_{ij}\}$*.*

*Para uma amostra de tamanho **n** com freqüências das células* $\{n_{ij}\}$ *, os valores* $\{\mu_{ij} = n\pi_{ij}\}$ *são chamados **frequências esperadas** e representam os valores das expectâncias* $\{E(n_{ij})\}$ *quando* H_0 *é verdadeira.*

A idéia é comparar as frequências amostrais das células com as esperadas para decidir se os dados contradizem H_0 *. Quanto maior as diferenças* $\{n_{ij} - \mu_{ij}\}$ *mais forte a evidência contra* H_0*.*

*A **estatística Qui-quadrado de Pearson** para testar* H_0 *é:*

$$\chi^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

tem distribuição qui-quadrado para amostras "grandes"($\{\mu_{ij} \geq 5\}$*).*

Testes Qui-quadrado

- *Teste de Comparação de Proporções*

  → Teste Qui-quadrado de Homogeneidade

Em tabelas 2x2, por exemplo:

$$H_0)\pi_{11}=\pi_{21} \quad e \quad \pi_{12}=\pi_{22} \quad \Leftrightarrow H_0)\pi_{11}=\pi_{21}$$

Se H_0 é verdadeira :

$$\hat{\mu}_{ij} = n_{i+}\cdot p_{+j} = n_{i+}\cdot\left(\frac{n_{+j}}{n}\right) = \frac{n_{i+}\cdot n_{+j}}{n}$$

- *Teste Qui-quadrado de Independência*

$$H_0)\pi_{ij}=\pi_{i+}\pi_{+j}$$

Se H_0 é verdadeira :

$$\hat{\mu}_{ij} = n\cdot p_{ij} = n\cdot p_{i+}\cdot p_{+j} = n\cdot\frac{n_{i+}}{n}\cdot\frac{n_{+j}}{n} = \frac{n_{i+}\cdot +n_{+j}}{n}$$

Teste Qui-quadrado de Independência

- *Deseja-se estudar se existe associação entre gênero e identificação partidária.*

Na pesquisa General Social Survey –1991, duas das variáveis estudadas foram gênero e identificação partidária. Os entrevistados indicavam se eles se identificavam mais fortemente com o partido Democrático ou com o Republicano ou com o Independente. A tabela a seguir apresentada os resultados obtidos para esta variável bem como o gênero do entrevistado.

Gênero	Identificação Partidária			Total
	Democrático	Independente	Republicano	
Feminino	279	73	225	577
Masculino	165	47	191	403
Total	444	120	416	980

- *Determinando as frequências relativas com relação ao total das colunas temos o resultado apresentado na seguinte tabela e no gráfico a seguir:*

.

	Democrático	Independente	Republicano	
Fem	48,4	12,7	39,0	100
Masc	40,9	11,7	47,4	100
	45,3	12,2	42,4	100,0

Deseja-se testar se sexo (gênero) e identificação partidária são associados ou não. As hipóteses do teste são então:

***H_0)** Identificação partidária e gênero não estão associados (Independência);*

***H_1)** Identificação partidária e gênero estão associados.*

Frequências Esperadas

	Democrático	Independente	Republicano	
Fem	261,42	70,65	244,93	577
Masc	182,58	49,35	171,07	403
	444	120	416	980

Cálculo Qui-quadrado

	Democrático	Independente	Republicano	
Fem	1,18	0,08	1,62	2,882
Masc	1,69	0,11	2,32	4,127
	2,88	0,19	3,94	7,01

Estatística da Razão de Verossimilhança

Rejeita H_0 1 Aceita H_0

Uma estatística alternativa para testar $\boldsymbol{H_0}$ *resulta do método da razão de verossimilhança para testes de significância.*

O teste determina os valores dos parâmetros que maximizam a função de verosimilhança sob a suposição que $\boldsymbol{H_0}$ é verdadeira. Ele também determina o valor que maximiza a função de verossimilhança sob a condição mais geral de que $\boldsymbol{H_0}$ pode ou não ser verdadeira.

O teste se baseia na razão das funções de verossimilhança maximizadas,

$$\Lambda = \frac{\text{máximo da finção de verossimilhança quando parâmetros satisfazem } H_o}{\text{máximo da finção de verossimilhança quando parâmetros são irrestritos}}$$

A razão não pode exceder 1. Se a função de verossimilhança *maximizada* é muito maior quando os parâmetros não são forçados a satisfazer $\boldsymbol{H_0}$, então a razão Λ é bastante abaixo de 1 e existe forte evidência contra $\boldsymbol{H_0}$.

0 1

Rejeita H_0. **Não Rejeita H_0.**

Estatística da Razão de Verossimilhança

A ***estatística do teste para o Teste da Razão de Verossimilhança*** *é igual a*

$$-2\log(\Lambda)$$

tem distribuição aproximadamente qui-quadrado com $\boldsymbol{v}$ *graus de liberdade.*

v g.l. = nº parâmetros sob H_1 - nº parâmetros sob H_0

I J – 1 **(I-1)+(J-1)**

Este valor é não negativo e pequenos valores de Λ *produzem grande valores de* $-2\log(\Lambda)$.

0 $+\infty$

Não Rejeita H_0. **Rejeita H_0.**

Para tabelas de contingência bidimensionais, esta estatística pode ser simplificada para a fórmula:

$$G^2 = 2\sum n_{ij}\log\left(\frac{n_{ij}}{\mu_{ij}}\right)$$

Frequências Esperadas

	Democrático	Independente	Republicano	
Fem	261,42	70,65	244,93	577
Masc	182,58	49,35	171,07	403
	444	120	416	980

Cálculo G_2

	Democrático	Independnete	Republicano	
fem	18,16	2,39	-19,10	1,450734
masc	-16,71	-2,29	21,05	2,050088
				3,500822

$$G_2 = 2\times 3,500822 = 7,01644$$

Resíduos

A estatística do teste e seu *p-value* simplesmente descrevem a evidência contra a hipótese nula $\boldsymbol{H_0}$.

A comparação, célula por célula, da freqüência observada com a esperada ajuda a entender melhor a natureza desta evidência. Entretanto a diferença absoluta (bruta) é insuficiente.

Os resíduos úteis têm a forma

$$\frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-p_{i+})(1-p_{+j})}}$$

e são denominados ***resíduos ajustados***.

Quando $\boldsymbol{H_0}$, cada resíduo ajustado tem para grandes amostras, distribuição N(0,1). Um resíduo ajustado que seja maior que *2 ou 3* em valor absoluto indica falta de ajustamento de $\boldsymbol{H_0}$ nesta célula.

	Identificação Partidária			
Sexo	Democrático	Independente	Republicano	Total
Feminino	279 (2,29)	73 (0,46)	225 (-2,62)	577
Masculino	165 (-2,29)	47 / **49,3** (-0,46)	191 / 171,1 (2,62)	403
Total	444	120	416	980