Distribuição Qui-quadrado e Beta

Prof. José Roberto Silva dos Santos

Depto. de Estatística e Matemática Aplicada - UFC

Fortaleza, 26 de outubro 2022

Sumário

Sumário

Distribuição Qui-quadrado

Definição:

Um caso particular muito importante da distribuição gama será obtido, se fizermos $\alpha = n/2$ e $\lambda = 1/2$, em que n é um inteiro positivo. Nesse caso,

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \underset{(0,\infty)}{I}(x)$$

Distribuição Qui-quadrado

Definição:

Um caso particular muito importante da distribuição gama será obtido, se fizermos $\alpha = n/2$ e $\lambda = 1/2$, em que n é um inteiro positivo. Nesse caso,

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \underset{(0,\infty)}{I}(x)$$

Nesse caso, dizemos que X possui distribuição *qui-quadrado* com n graus de liberdade (Notação: $X \sim \chi^2_{(n)}$)

Sumário

Teste de Associação

- Considere duas características designadas por A e B e suponha que existem r categorias $A_1, A_2, \ldots, A_r$ para A e c categorias $B_1, B_2, \ldots, B_c$ para B.

Teste de Associação

- Considere duas características designadas por A e B e suponha que existem r categorias $A_1, A_2, \ldots, A_r$ para A e c categorias $B_1, B_2, \ldots, B_c$ para B.
- Suponha que uma amostra de tamanho n é classificada e distribuída nas caselas da tabela produzindo uma tabela de frequência em que:
  n_{ij} = frequência de observações com as características A_i e B_j conjuntamente.

Teste de Associação

- Considere duas características designadas por A e B e suponha que existem r categorias $A_1, A_2, \ldots, A_r$ para A e c categorias $B_1, B_2, \ldots, B_c$ para B.
- Suponha que uma amostra de tamanho n é classificada e distribuída nas caselas da tabela produzindo uma tabela de frequência em que:
  n_{ij} = frequência de observações com as características A_i e B_j conjuntamente.
  n_{i0} = total da i-ésima linha, ou frequência de A_i.

Teste de Associação

- Considere duas características designadas por A e B e suponha que existem r categorias $A_1, A_2, \ldots, A_r$ para A e c categorias $B_1, B_2, \ldots, B_c$ para B.
- Suponha que uma amostra de tamanho n é classificada e distribuída nas caselas da tabela produzindo uma tabela de frequência em que:
  $n_{ij} =$ frequência de observações com as características A_i e B_j conjuntamente.
  $n_{i0} =$ total da i-ésima linha, ou frequência de A_i.
  $n_{0j} =$ total da j-ésima coluna, ou frequência de B_j.

Teste de Associação

Tabela de contingência $r \times c$

	B_1	B_2	$\cdots$	B_c	total da linha
A_1	n_{11}	n_{12}	$\cdots$	n_{1c}	n_{10}
A_2	n_{21}	n_{22}	$\cdots$	n_{2c}	n_{20}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	n_{r0}
total da coluna	n_{01}	n_{02}	$\cdots$	n_{0c}	n

Teste de Associação

- Queremos avaliar se existe associação entre as variáveis. Podemos o coeficiente de associação de Pearson

$$\chi^2 = \sum_{rc \text{ componentes}} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}.$$

com

$$E_{ij} = n\hat{p}_{ij} = \frac{n_{i0}n_{0j}}{n}.$$

Teste de Associação

- Queremos avaliar se existe associação entre as variáveis. Podemos o coeficiente de associação de Pearson

$$\chi^2 = \sum_{rc \text{ componentes}} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}.$$

com

$$E_{ij} = n\hat{p}_{ij} = \frac{n_{i0}n_{0j}}{n}.$$

- Pode-se mostrar que esse coeficiente tem distribuição χ^2 com $(r-1)\times(c-1)$ graus de liberdade, para n grande.

Exemplo 4

Uma amostra aleatória de 500 pessoas responde um questionário sobre filiação partidária e atitude mediante um programa de racionamento de energia.

Exemplo 4

Uma amostra aleatória de 500 pessoas responde um questionário sobre filiação partidária e atitude mediante um programa de racionamento de energia.

Tabela de contingência: filiação partidária e opinião sobre o racionamento de energia

	favorável	indiferente	contrário	Total
A	138	83	64	285
B	64	67	84	215
Total	202	150	148	500

Exemplo 4

Antes de apresentar uma análise formal estatística consideramos a tabela de um ponto de vista descritivo, transformando as contagens em proporções, primeiramente por linhas, depois por casela.

Exemplo 4

Antes de apresentar uma análise formal estatística consideramos a tabela de um ponto de vista descritivo, transformando as contagens em proporções, primeiramente por linhas, depois por casela.

a) Proporções por linhas

	favorável	indiferente	contrário	total
A	0,48	0,30	0,22	1
B	0,30	0,31	0,39	1

Exemplo 4

Antes de apresentar uma análise formal estatística consideramos a tabela de um ponto de vista descritivo, transformando as contagens em proporções, primeiramente por linhas, depois por casela.

a) Proporções por linhas

	favorável	indiferente	contrário	total
A	0,48	0,30	0,22	1
B	0,30	0,31	0,39	1

b) Proporções por casela

	favorável	indiferente	contrário	total
A	0,276	0,166	0,128	0,570
B	0,128	0,134	0,168	0,430
Total	0,404	0,300	0,296	1

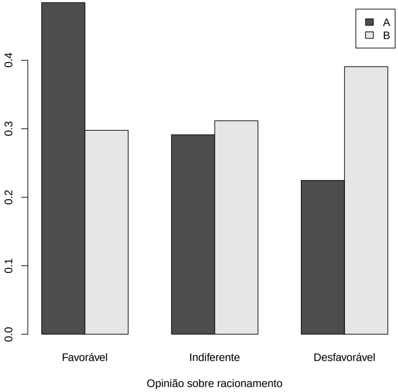

Gráfico de barras: frequências relativas por linha.

Exemplo 4

Usando os dados sobre filiação partidária, a tabela abaixo mostra as frequências observadas e as esperadas.

Tabela de contingência para filiação partidária e opinião sobre racionamento

	favorável	indiferente	contrário	total
A	138	83	64	285
	(115,14)	(85,50)	(84,36)	
B	64	67	84	215
	(86,86)	(64,50)	(63,64)	
Total	202	150	148	500

Exemplo 4

A estatística χ^2 tem o valor observado de
$\chi^2 = 4,539 + 0,073 + 4,914 + 6,016 + 0,097 + 6,514 = 22,153$ com $(2-1)(3-1) = 2$ g.l.

Exemplo 4

A estatística χ^2 tem o valor observado de
$\chi^2 = 4,539 + 0,073 + 4,914 + 6,016 + 0,097 + 6,514 = 22,153$ com
$(2-1)(3-1) = 2$ g.l.
Usando o nível de significância $\alpha = 0,05$, o χ^2 tabulado é $5,991$ que é menor do que o observado e daí a hipótese nula de indepêndencia é rejeitada.

Relação com a distribuição normal

- Se $X \sim N(0,1)$, qual a distribuição de $Y = X^2$?

Sumário

Distribuição Beta

Definição:

Seja X uma variável aleatória contínua, que tome somente valores no intervalo $[0, 1]$. Diremos que X possui distribuição beta, se sua fdp for dada por

Distribuição Beta

Definição:

Seja X uma variável aleatória contínua, que tome somente valores no intervalo $[0, 1]$. Diremos que X possui distribuição beta, se sua fdp for dada por

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \underset{(0,1)}{I}(x),$$

Definição:

Seja X uma variável aleatória contínua, que tome somente valores no intervalo $[0, 1]$. Diremos que X possui distribuição beta, se sua fdp for dada por

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \underset{(0,1)}{I}(x),$$

para todo $a > 0$ e $b > 0$. Temos que, $B(a, b)$ é a função beta definida por

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} d_x$$

- Pode-se mostrar a seguinte relação:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

- Pode-se mostrar a seguinte relação:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

em que $\Gamma(.)$ é a função gama.

- Pode-se mostrar a seguinte relação:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

  em que $\Gamma(.)$ é a função gama.

- **Notação:** $X \sim \text{Beta}(a,b)$.

Comportamento da fdp da distribuição beta para diferentes valores dos parâmetros

Distribuição Beta

- Se $a = b = 1$ então $X \sim \text{Beta}(1, 1) \equiv U(0, 1)$.

- Se $a = b = 1$ então $X \sim \text{Beta}(1, 1) \equiv U(0, 1)$.
- O valor esperado de X é dado por

$$\mathbb{E}(X) = \frac{a}{a + b}.$$

Distribuição Beta

- Se $a = b = 1$ então $X \sim \text{Beta}(1,1) \equiv U(0,1)$.
- O valor esperado de X é dado por

$$\mathbb{E}(X) = \frac{a}{a+b}.$$

- A variância de X é dada por

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

- Se $a = b = 1$ então $X \sim \text{Beta}(1, 1) \equiv U(0, 1)$.
- O valor esperado de X é dado por

$$\mathbb{E}(X) = \frac{a}{a+b}.$$

- A variância de X é dada por

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

- O k-ésimo momento de X é dado por

$$\mathbb{E}(X^k) = \frac{\Gamma(a+b)\Gamma(k+a)}{\Gamma(a)\Gamma(k+a+b)};\ k = 1, 2, \ldots$$

- Uma loja de comércio eletrônico envia *e*-mails com ofertas especiais a seus clientes cadastrados. Suponha que, após o recebimento de uma mensagem, a proporção de clientes que efetivam uma compra é uma variável aleatória com densidade dada por

$$f(x) = cx(1-x)^5 \underset{(0,1)}{I}(x).$$

(a) Encontre o valor de c.
(b) Calcule a probabilidade de que um *e*-mail resulte em alguma compra para mais de 50% dos seus destinatários.