

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Exercício 1.

(2,5 pontos). Uma pesquisa realizada entre estudantes da USP (Campus da Capital, Faculdade de Direito e Complexo de Saúde) foi conduzida no ano de 2009, sendo que um dos objetivos era verificar a existência de associação entre a área do curso e o uso de álcool. Para isso, uma amostra de 3715 alunos foi selecionada e os alunos responderam um questionário. Dos 958 alunos de cursos da área biológica (exceto medicina), 630 disseram ter utilizado álcool nos últimos 30 dias. Dentre os alunos de ciências exatas, 450 disseram não ter ingerido álcool nos últimos 30 dias e, finalmente, 970 dos 1373 alunos de cursos de ciências humanas disseram ter ingerido bebidas alcoólicas nos últimos 30 dias.

- (a) (0,5 pontos). Escreva as informações da pesquisa em uma tabela de distribuição conjunta de frequências.

Solução:

Área do curso	Utilizou álcool nos últimos 30 dias		
	Sim	Não	Total
Biológicas (exceto medicina)	630	328	958
Exatas	934	450	1384
Humanas	970	403	1373
Total	2534	1181	3715

□

- (b) (0,25 pontos). Qual é a proporção de alunos que disseram ter ingerido álcool nos últimos 30 dias? Dentre os alunos de ciências exatas, qual é essa proporção?

Solução:

A proporção de estudantes, sem levar em consideração a área de estudo, que utilizou álcool nos últimos 30 dias é $\frac{2534}{3715} = 0,6821$.

A proporção de alunos da área de exatas que consumiram álcool nos últimos 30 dias é $\frac{934}{1384} = 0,6749$. □

- (c) (0,25 pontos). Formule hipóteses H e A adequadas para a situação.

Solução:

Considere a hipótese **H: O consumo de álcool e a área de estudo são independentes.**

Considere a hipótese **A: O consumo de álcool e a área de estudo não são independentes.**

As seguintes hipóteses também são válidas: Considere a hipótese **H: Não existe associação entre o consumo de álcool e a área de estudo.**

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Considere a hipótese **A: Existe associação entre o consumo de álcool e a área de estudo.** □

- (d) (0,5 pontos). Se não há associação entre área do curso e uso de álcool, quantos alunos do curso de ciências humanas esperam-se que usem álcool?

Solução:

Se não há associação entre o consumo de álcool e a área de estudo, então o valor esperado de estudantes de ciências humanas que consumiram álcool nos últimos 30 dias é de $\frac{2534 \times 1373}{3715} = 936,5227$.

Assim a proporção de alunos de ciências humanas que consumiram álcool nos últimos 30 dias, considerando que não há associações entre as duas variáveis, é $\frac{936,5227}{1373} = 0,6821$. □

- (e) (1 ponto). Por meio do nível descritivo, conclua sobre suas hipóteses, adotando um nível de significância igual a 1%.

Solução:

A estatística de teste é da forma

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

onde r é o número de linhas e s é o número de colunas, portanto $(r - 1) \times (s - 1) = (3 - 1) \times (2 - 1) = 2$ Na tabela a baixo está apresentado os dados observados e os dados esperados.

Área do curso	Utilizou álcool nos últimos 30 dias				Total
	Sim		Não		
	Observado	Esperado	Observado	Esperado	
Biológicas (exceto medicina)	630	653,4514	328	304,5486	958
Exatas	934	944,0258	450	439,9742	1384
Humanas	970	936,5227	403	436,4773	1373
Total	2534		1181		3715

Assim,

$$\begin{aligned}\chi^2 &= \frac{(630 - 653,4514)^2}{653,4514} + \frac{(328 - 304,5486)^2}{304,5486} + \frac{(934 - 944,0258)^2}{944,0258} + \\ &+ \frac{(450 - 439,9742)^2}{439,9742} + \frac{(970 - 936,5227)^2}{936,5227} + \frac{(403 - 436,4773)^2}{436,4773} = \\ &= 0,8416 + 1,8058 + 0,1065 + 0,2285 + 1,1967 + 2,5667 \\ &= 6,7468.\end{aligned}$$

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Então, $\chi^2 \sim \chi_2^2$ e o nível descritivo é $P(\chi_2^2 \geq 6,7468) = 0,0343$. Com 1% de significância e um nível descritivo 3,43% decide-se por não rejeitar a hipótese **H**. Não há indícios que sustentem a associação do consumo de álcool com a área de estudo. \square

Exercício 2.

(2,5 pontos). A tabela a seguir apresenta dados de uma amostra de 1000 ingressantes de uma universidade, com informações sobre a área de estudo e a classe sócio-econômica. Com base nessas observações, um investigador deseja concluir sobre a existência de associação entre essas duas variáveis.

Área	Classe sócio-econômica		
	Alta	Média	Baixa
Exatas	120	136	88
Humanas	91	95	80
Biologias	150	155	85

- (a) (0,5 pontos) Estipule as hipóteses estatísticas H e A adequadas para o problema.

Solução:

Considere a hipótese **H: A classe sócio-econômica e a área de estudo são independentes.**

Considere a hipótese **A: A classe sócio-econômica e a área de estudo não são independentes.**

As seguintes hipóteses também são válidas: Considere a hipótese **H: Não existe associação entre a classe sócio-econômica e a área de estudo.**

Considere a hipótese **A: Existe associação entre a classe sócio-econômica e a área de estudo.** \square

- (b) (0,5 pontos) No contexto do problema, dê uma interpretação para o erro do tipo I.

Solução:

O **Erro Tipo I** é Rejeitar **H** quando ela é verdadeira. No contexto do problema, o **Erro Tipo I** pode ser interpretado como concluir que há uma associação entre a classe sócio-econômica e a área de estudo, mas na verdade tal associação não existe. \square

- (c) (0,5 pontos) Supondo que a escolha da área seja independente da classe sócio-econômica, qual é o número esperado de alunos na área de humanas e na classe alta? E quantos foram observados?

Solução:

Se há independência entre as variáveis, então o número esperado de estudantes de ciências

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

humanas e que são de classe alta é $\frac{(120 + 91 + 150) \times (91 + 95 + 80)}{1000} = 96,026$. Foram observados 91 alunos \square

(d) (1 ponto) Conclua sobre suas hipóteses utilizando um nível de significância de 1%.

Solução:

A estatística de teste é da forma

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

onde r é o número de linhas e s é o número de colunas, portanto $(r - 1) \times (s - 1) = (3 - 1) \times (3 - 1) = 4$. Na tabela a baixo está apresentado os dados observados e os dados esperados.

Área do curso	Classe sócio-econômica						Total
	Alta		Média		Baixa		
	Observado	Esperado	Observado	Esperado	Observado	Esperado	
Exatas	120	124,184	136	132,784	88	87,032	344
Humanas	91	96,026	95	102,676	80	67,298	266
Biológicas	150	140,790	155	150,540	85	98,670	390
Total	361		386		253		1000

Assim,

$$\begin{aligned}\chi^2 &= \frac{(120 - 124,184)^2}{124,184} + \frac{(91 - 96,026)^2}{96,026} + \frac{(150 - 140,790)^2}{140,790} + \\ &+ \frac{(136 - 132,784)^2}{132,784} + \frac{(95 - 102,676)^2}{102,676} + \frac{(155 - 150,540)^2}{150,540} + \\ &+ \frac{(88 - 87,032)^2}{87,032} + \frac{(80 - 67,298)^2}{67,298} + \frac{(85 - 98,670)^2}{98,670} = \\ &= 0,141 + 0,263 + 0,602 + 0,078 + 0,574 + 0,132 + 0,011 + 2,397 + 1,894 \\ &= 6,092.\end{aligned}$$

Então, $\chi^2 \sim \chi_4^2$ e o nível descritivo é $P(\chi_4^2 \geq 6,092) = 0,1924$. Com 1% de significância e um nível descritivo 19,24% decide-se por não rejeitar a hipótese **H**. Não há evidências que sustentem a associação da classe sócio-econômica e a área de estudo. \square

Exercício 3.

(2 pontos). Numa pesquisa realizada com 500 famílias, coletaram-se informações sobre a variável número de filhos. Os resultados são apresentados na tabela a seguir.

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Nº de filhos	1	2	3	4	5	6	7	8	9	Total
Frequência (Nº de famílias)	115	130	105	45	35	30	25	5	10	500

Estudos anteriores assumem que o número de filhos igual a 1, 2, 3,..., 8 e 9 ocorre com probabilidade 0,21, 0,24, 0,21, 0,12, 0,09, 0,06, 0,05, 0,01 e 0,01, respectivamente. Um estatístico foi consultado para verificar se houve mudança desse padrão, com base nessa amostra de 100 famílias.

- (a) (1 ponto). O que representam os valores dessa tabela?

Solução:

Na tabela são apresentados os resultados da pesquisa onde foram observados o número de filhos e o número de famílias que apresentaram esses determinado número de filhos. Em outras palavras, os números na tabela representam a frequência absoluta da distribuição do número de filhos. \square

- (b) (1 ponto). Pode-se afirmar que o número de filhos segue a distribuição especificada em estudos anteriores? Estabeleça as hipóteses estatísticas H e A apropriadas e utilize um teste estatístico adequado para testá-las. Conclua com base no nível descritivo, considerando nível de significância de 2%.

Solução:

Considere a hipótese **H: O número de filhos segue a distribuição especificada em estudos anteriores.**

Considere a hipótese **A: O número de filhos não segue a distribuição especificada em estudos anteriores.**

A estatística de teste é da forma $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$, onde K é o número de categorias, no problema $k = 9$.

A tabela abaixo apresenta as frequências observadas e esperadas, assim como a distribuição de probabilidades sob a hipótese nula. As frequências esperadas foram obtidas multiplicando-se, para cada número de filhos, o número total de famílias (500) pela probabilidade associada ao número de filhos, especificada pela hipótese nula (estudos anteriores).

Nº de filhos	1	2	3	4	5	6	7	8	9	Total
Frequência Observada	115	130	105	45	35	30	25	5	10	500
Frequência Esperada	105	120	105	60	45	30	25	5	5	500
Distribuição de probabilidade	0,21	0,24	0,21	0,12	0,09	0,06	0,05	0,01	0,01	1

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Assim,

$$\begin{aligned}\chi^2 &= \frac{(115 - 105)^2}{105} + \frac{(130 - 120)^2}{120} + \frac{(105 - 105)^2}{105} + \\ &+ \frac{(45 - 60)^2}{60} + \frac{(35 - 45)^2}{45} + \frac{(30 - 30)^2}{30} + \\ &+ \frac{(25 - 25)^2}{25} + \frac{(5 - 5)^2}{5} + \frac{(10 - 5)^2}{5} \\ &= 0,9523 + 0,8333 + 0 + 3,7500 + 2,2222 + 0 + 0 + 0 + 5 \\ &= 12,7579\end{aligned}$$

Então, $\chi^2 \sim \chi_8^2$ e o nível descritivo é $P(\chi_8^2 \geq 12,7579) = 0,1205$. Com 2% de significância e um nível descritivo 12,05% decide-se por não rejeitar a hipótese **H**. A distribuição de probabilidade do número de filhos de estudos antigos ainda é válida. \square

Exercício 4.

(3 pontos). Os dados da tabela a seguir foram extraídos de uma pesquisa realizada com crianças de até 12 anos de idade em Florianópolis. Tal pesquisa tem por objetivo analisar se a severidade das cáries dentárias observadas tem ou não relação com as condições familiares da criança.

	Categorias	Baixa Severidade(N)	Alta Severidade(N)
Tempo de residencia (em anos)	Até 2	50	40
	mais de 2	20	30
Número médio de residentes por cômodo	0,25 - 0,83	20	20
	0,84 - 1,25	30	20
	1,26 ou mais	30	40
Responsável pelo domicilio	Pai/Mãe	50	30
	Avós	20	20
	Outros	10	30
Responsável pelo sustento da família	Pai/Mãe	30	30
	Pai	30	20
	Mãe	20	20
	Outros	10	20

- (a) (0,5 pontos). Com base na tabela, deseja-se fazer quatro testes de hipóteses para verificar se as variáveis tempo de residência, número médio de residentes por cômodo, responsável no domicílio e responsável pelo sustento da família estão associadas com a severidade das cáries. Construa as hipóteses H e A adequadas para as quatro situações do problema.

Solução:

Considere as seguintes hipóteses:

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

- **H₁**: *A severidade das cáries é independente do tempo de residencia* (Não existe associação entre as variáveis)
- **A₁**: *A severidade das cáries não é independente do tempo de residencia* (Existe associação entre as variáveis)
- **H₂**: *A severidade das cáries é independente do número de residentes* (Não existe associação entre as variáveis)
- **A₂**: *A severidade das cáries não é independente do número de residentes* (Existe associação entre as variáveis)
- **H₃**: *A severidade das cáries é independente do responsável pelo domicílio* (Não existe associação entre as variáveis)
- **A₃**: *A severidade das cáries não é independente do responsável pelo domicílio* (Existe associação entre as variáveis)
- **H₄**: *A severidade das cáries é independente do responsável pelo sustento da família* (Não existe associação entre as variáveis)
- **A₄**: *A severidade das cáries não é independente do responsável pelo sustento da família* (Existe associação entre as variáveis)

□

(b) (2 pontos). Utilizando o R, calcule os valores observados da estatística χ^2 e os correspondentes níveis descritivos P para os testes de hipóteses em questão.

Solução:

Nas tabelas abaixo observam-se os valores observados e os valores esperados.

H₁ contra A₁

Tabela 1: Tempo de residencia (em anos)

Categorias	Baixa Severidade(N)		Alta Severidade(N)		Total
	Observado	Esperado	Observado	Esperado	
Até 2	50	45	40	45	90
Mais de 2	20	25	30	25	50
Total	70		70		140

Assim, $\chi^2 = 3,1111$ com 1 grau de liberdade, então $P(\chi_1^2 \geq 3,1111) = 0,0778$.

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

H₂ contra A₂

Tabela 2: Número médio de residentes por cômodo

Categorias	Baixa Severidade(N)		Alta Severidade(N)		Total
	Observado	Esperado	Observado	Esperado	
0,25 - 0,83	20	20	20	20	40
0,84 - 1,25	30	25	20	25	50
1,26 ou mais	30	35	40	35	70
Total	80		80		160

Assim, $\chi^2 = 3,4286$ com 2 graus de liberdade, então $P(\chi_2^2 \geq 3,4286) = 0,1801$.

H₃ contra A₃

Tabela 3: Responsável pelo domicílio

Categorias	Baixa Severidade(N)		Alta Severidade(N)		Total
	Observado	Esperado	Observado	Esperado	
Pai/Mãe	50	40	30	40	80
Avós	20	20	20	20	40
Outros	10	20	30	20	40
Total	80		80		160

Assim, $\chi^2 = 15$ com 2 grau de liberdade, então $P(\chi_2^2 \geq 15) = 0,0006$.

H₄ contra A₄

Tabela 4: Responsável pelo sustento da família

Categorias	Baixa Severidade(N)		Alta Severidade(N)		Total
	Observado	Esperado	Observado	Esperado	
Pai	30	30	30	30	60
Mãe	30	25	20	25	50
Avós	20	20	20	20	40
Outros	10	15	20	15	30
Total	90		90		180

Assim, $\chi^2 = 5,3333$ com 3 grau de liberdade, então $P(\chi_3^2 \geq 5,3333) = 0,1490$.

□

- (c) (0,5 pontos). Conclua sobre as hipóteses formuladas por meio do valor do nível descritivo obtido, adotando o nível de significância igual a 5% e comente.

Solução:

MAE116 – Noções de Estatística

Grupo A - 1 semestre de 2012

Lista de exercícios 11 - Qui-quadrado – C A S A (gabarito)

Com base nos níveis descritivos do item anterior, conclui-se que:

- Não rejeita-se H_1
- Não rejeita-se H_2
- Rejeita-se H_3
- Não rejeita-se H_4

A severidade das caries Não depende nem do tempo de residência, nem do número de residentes por cômodo e nem do responsável pelo sustento da família. Depende somente do responsável pelo domicílio. □