MAE116 – Noções de Estatística

Grupo A - 1º semestre de 2012

Gabarito - Lista de Revisão 2

Exercício 1 (Estimação)

Um laboratório realizou um experimento que envolvia 100 lâmpadas de sua fabricação, medindo o tempo (em dias) até elas queimarem.

Dias	Frequência
0-30	49
30-45	32
45-60	13
60-80	6
	100

a) Calcule uma estimativa pontual da proporção de lâmpadas que duram pelo menos 45 dias.

Resposta: Seja p a proporção de lâmpadas que duram pelo menos 45 dias. Uma estimativa pontual de p é $\hat{p} = {}^{(13+6)}/_{100} = 0{,}19$.

b) Construa um intervalo de confiança para essa proporção com um coeficiente de confiança igual a 0,98.

Resposta: Neste caso, $\gamma = 0{,}98$ e portanto, $z = A(0{,}99) = 2{,}33$. Sendo assim,

$$IC(p; 98\%) = \left(\hat{p} - z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$= \left(0{,}19 - 2{,}33 \cdot \sqrt{\frac{0{,}19 \cdot (1 - 0{,}19)}{100}}; 0{,}19 + 2{,}33 \cdot \sqrt{\frac{0{,}19 \cdot (1 - 0{,}19)}{100}}\right)$$

$$= (0{,}099; 0{,}281)$$

c) Qual é o tamanho amostral necessário para estimar a proporção de lâmpadas que duram pelo menos 53 dias se o laboratório admite um erro de estimação de 0,05 com probabilidade de 0,95 e sabe-se que essa proporção é, no máximo, 25%?

Resposta: Neste caso, $\gamma = 0{,}95 \Rightarrow z = A(0{,}975) = 1{,}96$; $\varepsilon = 0{,}05$ e $p \leq 0{,}25$. Assim, temos que

$$n \leq \left(\frac{z}{\varepsilon}\right)^2 p(1-p) = \left(\frac{1{,}96}{0{,}05}\right)^2 0{,}25(1 - 0{,}25) = 288{,}12 \cong 289$$

Portanto, uma amostra de pelo menos 289 lâmpadas garante a estimação com as condições pedidas.

Exercício 2 (Teste de hipótese I)
Em pesquisa realizada em 2009, constatou-se que 25% dos alunos da USP não tomavam café da manhã. Uma nutricionista acredita que essa proporção tenha aumentado. Para verificar isso, fez uma pesquisa com 18 alunos selecionados aleatoriamente na USP; desses, a metade afirmou não tomar café da manhã.

a) Estabeleça as hipóteses do teste.

Resposta:
H: $p = 0{,}25$ (a proporção de alunos que não tomam café da manhã se manteve)
A: $p > 0{,}25$ (a proporção de alunos que não tomam café da manhã aumentou em relação à proporção de 2009)

b) Interprete os erros de tipo I e tipo II do teste no contexto do problema.

Resposta:
Erro de tipo I: Conclui-se que a proporção de alunos atual que não tomam café da manhã é superior a 25% sendo que, na realidade, essa proporção é igual a 25%.
Erro de tipo II: Conclui-se que a proporção de alunos atual que não tomam café da manhã é 25% sendo que, na realidade, essa proporção é superior a 25%.

c) Baseado na amostra selecionada, qual é a conclusão ao nível de significância de 5%?

Resposta: Seja X: número de alunos na amostra que não tomam café da manhã. Se H for verdadeira, então $X \sim b(18;\ 0{,}25)$. A seguir, é apresentada a tabela da distribuição de probabilidades de X.

k	Pr	≈
0	5.637710E-03	0.0056
1	3.382626E-02	0.0338
2	9.584107E-02	0.0958
3	1.703841E-01	0.1704
4	2.129802E-01	0.2130
5	1.987815E-01	0.1988
6	1.435644E-01	0.1436
7	8.203680E-02	0.0820
8	3.760020E-02	0.0376
9	1.392600E-02	0.0139

k	Pr	≈
10	4.177800E-03	0.0042
11	1.012800E-03	0.0010
12	1.969333E-04	0.0002
13	3.029744E-05	0.0000
14	3.606838E-06	0.0000
15	3.206078E-07	0.0000
16	2.003799E-08	0.0000
17	7.858034E-10	0.0000
18	1.455192E-11	0.0000

Precisamos k tal que $P(X \geq k\ |p = 0{,}25) \leq 0{,}05$.
Além disso, $P(X \geq 8|p = 0{,}25) = 0{,}0569$ e $P(X \geq 9|p = 0{,}25) = 0{,}0193$.

Assim, $k = 9$ e a região crítica do teste é RC=$\{X \geq 9\}$.

Como foi observado X=9, então rejeitamos a hipótese nula ao nível de significância de 5% e, portanto, há evidências de que a proporção de alunos da USP que não tomam café da manhã aumentou em relação ao ano 2009.

Exercício 3 (Teste de Hipótese II)

Sabe-se que 70% dos pacientes submetidos a uma certa cirurgia não apresentam um problema pós-operatório. Uma equipe médica desenvolveu uma nova técnica cirúrgica e afirma que a proporção de pacientes que não apresenta problema pós-operatório, quando submetidos à cirurgia com esta nova técnica, é superior a 70%. Para pôr à prova a afirmação da equipe médica, um hospital aplica a nova técnica a alguns pacientes.

a) Formule este problema como um problema de testes de hipótese (quem é p?)

Resposta:
Sendo p a proporção de pacientes que não apresentam problema pós-operatório quando submetidos à nova técnica, podemos formular o teste da seguinte forma:

$$H : p = 0{,}7$$
$$A : p > 0{,}7$$

b) Interprete os erros de tipo I e tipo II no contexto do problema

a) Erro tipo I: <u>Rejeitar H quando H é verdadeira</u>, ou seja, afirmar que a proporção de pacientes que não apresentam problemas pós-operatórios quando submetidos à nova técnica é maior que 70%, quando na verdade essa proporção é igual a 70%.
b) Erro tipo II: <u>Não rejeitar H quando H é falsa</u>, ou seja, afirmar que a proporção de pacientes que não apresentam problemas pós-operatórios quando submetidos à nova técnica é igual a 70%, quando na verdade essa proporção é maior que 70%.

c) Se entre 19 pacientes submetidos à nova técnica 17 não apresentaram o problema, qual o nível descritivo e qual a decisão a ser tomada, adotando α=3%?

Resposta:
Seja X: número de pacientes que não apresentam problema pós-operatório dentre 19 pacientes submetidos à nova técnica.

Assim, $X \sim b(19\ ;\ 0{,}7)$. Então, o nível descritivo do teste é dado por:

$$P = P(X \geq 17 \mid p = 0{,}7) = 0{,}0358 + 0{,}0093 + 0{,}0011 = 0{,}0462$$

Portanto, como o nível descritivo é maior que o nível de significância do teste (0,03) não rejeitamos a hipótese nula. Podemos afirmar que, ao nível de 3%, não há evidências de que esta nova técnica reduz a incidência de problema pós-operatório.

Tabela da distribuição binomial com n=19 e p=0,7

x	P(X = x)	≈
0	1.162e-10	0.0000
1	5.153e-09	0.0000
2	1.082e-07	0.0000
3	1.431e-06	0.0000
4	1.335e-05	0.0000
5	9.347e-05	0.0000
6	5.089e-04	0.0006
7	2.205e-03	0.0022
8	7.719e-03	0.0077
9	2.201e-02	0.0220
10	5.136e-02	0.0514
11	9.805e-02	0.0981
12	1.525e-01	0.1525
13	1.916e-01	0.1916
14	1.916e-01	0.1916
15	1.491e-01	0.1491
16	8.695e-02	0.0870
17	3.580e-02	**0.0358**
18	9.282e-03	**0.0093**
19	1.140e-03	**0.0011**

d) Se dentre 100 pacientes submetidos à nova técnica 21 apresentarem o problema, qual a decisão a ser tomada? Responda usando o nível descritivo do teste. Use um nível de significância α=3%.

Resposta:
Neste caso, iremos utilizar a aproximação da distribuição binomial pela normal, dado que o tamanho da amostra é grande. Temos que, se a hipótese nula for verdadeira,

$$E[X] = n \times p = 100 \times 0{,}7 = 70$$

$$Var[X] = n \times p \times (1-p) = 21$$

Assim, se a hipótese nula for verdadeira, X tem distribuição aproximadamente normal de média 70 e variância 21. Como 21 pacientes apresentam o problema pós-operatório, 79 não apresentaram. Assim, o nível descritivo é dado por:

$$P = P(X \geq 79 \mid p = 0{,}70) = P(Z \geq \frac{79-70}{\sqrt{21}}) = P(Z \geq 1{,}964) = 1 - A(1{,}964) =$$

$$1 - 0{,}9752 = 0{,}025$$

Portanto, como o nível descritivo é menor que o nível de significância do teste (0,03), há evidências de que a nova técnica reduz a incidência de problema pós-operatório ao nível de significância de 3%.

Exercício 4 (Qui-quadrado)

O exame da OAB (Ordem dos Advogados do Brasil) é composto de duas provas em que a aprovação é obrigatória para exercer a advocacia no Brasil. Com a finalidade de analisar a faixa etária (jovem, meia idade, adulto) dos bacharéis do curso de Direito que estão em fase de pré-exames da OAB e verificar se existe alguma associação entre essa variável e o resultado na primeira fase do exame, foram pesquisados 237 alunos, escolhidos ao acaso entre aqueles que fizeram algum curso preparatório para esse exame em 2009. Sabe-se que entre os 68 jovens, 24 foram aprovados na primeira fase, que entre os 171 reprovados na primeira fase, 79 eram adultos, e que 7 foram aprovados e tinham meia idade.

a) Escreva as informações da pesquisa em uma tabela de distribuição conjunta de frequências.

	aprovado	reprovado	total
J	24	44	68
MI	7	48	55
A	35	79	114
total	66	171	237

b) Se a faixa etária do bacharel não está associada com a aprovação na primeira fase, quantos alunos de meia idade e aprovados são esperados? Quantos foram observados nesse caso?

São esperados 55x66/237 = 15,3 (aproximadamente 15). Foram observados 7.

c) Escreva as hipóteses e informe o número de graus de liberdade da estatística do teste apropriado.

H: não há associação entre faixa etária e resultado no exame
A: há associação entre faixa etária e resultado no exame

número de graus de liberdade = (3 – 1) x (2 -1) = 2

d) A saída do R para o teste de hipótese apropriado forneceu χ^2= 8,597 e P = 0,014. Por meio do nível descritivo, conclua sobre suas hipóteses utilizando um nível de significância de 5%.

Como P<0,05, H é rejeitada ao nível de significância de 5%. Há, portanto, evidências de associação entre faixa etária e resultado no exame.

	aprovado	reprovado	Total
J	24 (35%)	44 (65%)	68 (100%)
MI	7 (13%)	48 (87%)	55 (100%)
A	35 (31%)	79 (69%)	114 (100%)
total	66 (28%)	171(72%)	237 (100%)

Nota-se menor proporção de aprovados entre os de “meia idade”.