

Introdução à Probabilidade e à Estatística II

Solução da lista 4

1. Usando a distribuição de poisson($\lambda = 1, 5$), temos que

$$P(x = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \text{ pela questão } x \text{ varia em } x = 0, 1, 2, 3, 4, 5$$

x	0	1	2	3	4	5
P(X=x)	0,2231	0,3347	0,251	0,1255	0,0471	0,0141

Freq_esperada	0	1	2	3	4	5
P(X=x)*200	44,62	66,94	50,2	25,1	9,42	2,82

A frequência esperada para 5 ou mais é menor que 5 e, portanto muito pequena. A recomendação é juntar na classe 4 ou mais, assim

Freq_esperada	0	1	2	3	4 ou mais
P(X=x)*200	44,62	66,94	50,2	25,1	$(1 - (0,2231 + 0,3347 + 0,251 + 0,1255)) * 200 = 13,14$

Hipóteses:

H_0 : a população amostrada segue uma distribuição de poisson com média 1, 5.

H_1 : a população amostrada não segue uma distribuição de poisson com média 1, 5.

Temos $r = 5$ categorias

Estatística de teste

$$\chi^2 = \sum_{i=1}^s \frac{(O_i - E_i)^2}{E_i}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{r-1}^2$$

com O_i = valor observado, E_i = valor esperado.

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x \in \mathbb{R} | x > \chi_{4,95\%}^2\} = \{x \in \mathbb{R} | x > 9,49\}$$

Valor observado:

$$\chi_o^2 = \frac{(70 - 44,6)^2}{44,6} + \frac{(57 - 66,9)^2}{66,9} + \frac{(46 - 50,2)^2}{50,2} + \frac{(20 - 25,1)^2}{25,1} + \frac{(7 - 13,1)^2}{13,1} = 20,2.$$

como $\chi_o^2 > \chi_{4,95\%}^2$ então pertence a RC, então rejeitamos H_0 , ou seja, os dados não segue a distribuição poisson 1, 5.

2. vamos classificar esses dados em quatro intervalos, então de acordo com a tabela abaixo o valor esperado é dado pelo $(\text{total}/n^\circ \text{ de intervalos}) = (280/4) = 70$, , delimitados pelos quantis teóricos $Q(0,25)$, $Q(0,5)$ e $Q(0,75)$ da $N(\bar{x}, S^2)$ com $\bar{x} = 132,3$ e $S^2 = 124,8$ (para mais detalhes veja o cálculo da questão 9 da lista 2).

Chamando de $Z(p)$ os quantis da $N(0,1)$, temos

$$Q(0,25) = 132,3 + 11 * Z(0,25) = 132,3 + 11 * (-0,67) \cong 125 \text{ com } S = \sqrt{124,8} \cong 11$$

$$Q(0,5) = 132,3 + 11 * Z(0,5) = 132,3 + 11 * (0) \cong 132$$

$$Q(0,75) = 132,3 + 11 * Z(0,75) = 132,3 + 11 * (0,67) \cong 140$$

PAM	N. de Pacientes	Proporção	Esperado
<125	$10+30+20+20/2=70$	0,25	70
125 –132	$20/2+40 \cdot 2/3 = 37$	0,25	70
132 –140	$40/3+80=93$	0,25	70
>140	$70+10=80$	0,25	70

Hipóteses:

H_0 : a população amostrada segue uma distribuição normal(132, 2; 124, 8).

H_1 : a população amostrada não segue uma distribuição normal(132, 2; 124, 8).

Temos $r = 4$ categorias

Estatística de teste

$$\chi^2 = \sum_{i=1}^s \frac{(O_i - E_i)^2}{E_i}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{r-1}^2$$

com O_i = valor observado, E_i = valor esperado.

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x > \chi_{3;95\%}^2\} = \{x \in \Re | x > 7,81\}$$

Valor observado:

$\chi_o^2 = \frac{(70-70)^2}{70} + \frac{(37-70)^2}{70} + \frac{(93-70)^2}{70} + \frac{(80-70)^2}{70} = 24,55$ como $\chi_o^2 > \chi_{3;95\%}^2$ então pertence a RC, então rejeitamos H_0 , ou seja, os dados não seguem a distribuição normal.

3. Considere valor esperado ($E_i = \text{total}/n^\circ \text{ de casos}$) = $(216/7) = 30,9$, então

Dia	Dom	Seg	Ter	Qua	Qui	Sex	Sab	Total
O_i	40	24	25	28	29	32	38	216
E_i	30,9	30,9	30,9	30,9	30,9	30,9	30,9	216
$(O_i - E_i)^2/E_i$	2,71	1,52	1,11	0,26	0,11	0,04	1,65	7,42

Hipóteses:

H_0 : os acidentes fatais ocorrem com igual frequência .

H_1 :os acidentes fatais ocorrem mais na sexta ou sábado

Temos $r = 7$ categorias

Estatística de teste

$$\chi^2 = \sum_{i=1}^s \frac{(O_i - E_i)^2}{E_i}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{r-1}^2$$

com O_i = valor observado, E_i = valor esperado.

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x \in \Re | x > \chi_{6,95\%}^2\} = \{x \in \Re | x > 12,6\}$$

Valor observado:

$$\chi_o^2 = 2,71 + 1,52 + 1,11 + 0,26 + 0,11 + 0,04 + 1,65 = 7,42.$$

como $\chi_o^2 = 7,42 < \chi_{6,95\%}^2 = 12,6$ então não pertence a RC, então não rejeitamos H_0 , ou seja, os acidentes fatais ocorrem com igual frequencia.

4. Considere (Tabela observada)

Observado(O_{ij})	Polícia	Caixa	Taxista	Segurança	Total
Homicídios	82	107	70	59	318
Não Homicídio	92	9	29	42	172
Total	174	116	99	101	490

Tabela esperada (obtida pelo produto marginal da i-ésima linha com a j-ésima coluna e dividido pela soma total), por exemplo $113 \cong (174 \cdot 318)/490$ e o mesmo ocorre para os demais casos.

Esperado ($\hat{E}_{ij}$)	Polícia	Caixa	Taxista	Segurança	Total
Homicídios	113	75	64	66	318
Não Homicídio	61	41	35	35	172
Total	174	116	99	101	490

Hipóteses:

H_0 : A ocupação é independente da causa de morte ser homicídio .

H_1 :A ocupação é dependente da causa de morte ser homicídio

Estatística de teste

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{(r-1)(s-1)}^2$$

com O_{ij} = frequência observada, $\hat{E}_{ij}$ = estimativa da frequência esperada supondo a independência, r=numero de linhas, s=numero de colunas

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x \in \mathbb{R} | x > \chi_{3,95\%}^2\} = \{x \in \mathbb{R} | x > 7,81\}$$

Valor observado:

$\chi_o^2 = \frac{(82-113)^2}{113} + \frac{(107-75)^2}{75} + \frac{(70-64)^2}{64} + \frac{(59-66)^2}{66} + \frac{(92-61)^2}{61} + \frac{(9-41)^2}{41} + \frac{(29-35)^2}{35} + \frac{(42-35)^2}{35} = 65,53$ como $\chi_o^2 = 65,53 > \chi_{3,95\%}^2 = 7,81$ então pertence a RC, então rejeitamos H_0 , ou seja, existe relação entre a ocupação e a causa da morte.

5. Considere (Tabela observada)

Observado(O_{ij})	Itens c/ preços regulares	Itens em Oferta	Total
Preço abaixo	20	7	27
Preço acima	15	29	44
Preço correto	384	364	748
Total	419	400	819

Tabela esperada (obtida pelo produto marginal da i-ésima linha com a j-ésima coluna e dividido pela soma total), por exemplo $14 \cong (419 \cdot 27)/819$ e o mesmo ocorre para os demais casos.

Esperado($\hat{E}_{ij}$)	Itens c/ preços regulares	Itens em Oferta	Total
Preço abaixo	14	13	27
Preço acima	23	21	44
Preço correto	382	366	748
Total	419	400	819

Hipóteses:

H_0 : As taxas de erro são as mesmas para itens com os preços regulares e para itens anunciados em ofertas .

H_1 :as taxas de erro não são as mesmas para itens com os preços regulares e para itens anunciados em ofertas

Estatística de teste

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{(r-1)(s-1)}^2$$

com O_{ij} = frequência observada, $\hat{E}_{ij}$ = estimativa da frequência esperada supondo a independência, r=numero de linhas, s=numero de colunas

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x \in \mathbb{R} | x > \chi_{2,95\%}^2\} = \{x \in \mathbb{R} | x > 5,99\}$$

Valor observado:

$\chi_o^2 = \frac{(20-14)^2}{14} + \frac{(7-13)^2}{13} + \frac{(15-23)^2}{23} + \frac{(29-21)^2}{21} + \frac{(384-382)^2}{382} + \frac{(364-366)^2}{366} = 10,81$ como $\chi_o^2 = 10,81 > \chi_{2,95\%}^2 = 5,99$ então pertence a RC, então rejeitamos H_0 , ou seja, as taxas de erro são diferentes para itens com os preços regulares e para itens anunciados em ofertas.

6. Considere (Tabela observada)

Observado(O_{ij})	Homicídio	Roubo	Assalto	Total
Criminoso estranho à vítima	12	379	727	1118
Criminoso conhecido da vítima	39	106	642	787
Total	51	485	1369	1905

Tabela esperada (obtida pelo produto marginal da i-ésima linha com a j-ésima coluna e dividido pela soma total), por exemplo $30 \cong (51 \cdot 1118)/1905$ e o mesmo ocorre para os demais casos.

Esperado($\hat{E}_{ij}$)	Homicídio	Roubo	Assalto	Total
Criminoso estranho à vítima	30	285	803	1118
Criminoso conhecido da vítima	21	200	566	787
Total	51	485	1369	1905

Hipóteses:

H_0 : tipo de crime é independente do fato de o criminoso ser estranho.

H_1 : tipo de crime depende do fato de o criminoso ser estranho

Estatística de teste

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}. \text{ Sob } H_0 \quad \chi^2 \sim \chi_{(r-1)(s-1)}^2$$

com O_{ij} = frequência observada, $\hat{E}_{ij}$ = estimativa da frequência esperada supondo a independência, r=numero de linhas, s=numero de colunas

Região Crítica: Fixando $\alpha = 5\%$ (Consultando a Tabela Qui-quadrado) então:

$$RC = \{x \in \mathbb{R} | x > \chi_{2,95\%}^2\} = \{x \in \mathbb{R} | x > 5,99\}$$

Valor observado:

$\chi_o^2 = \frac{(12-30)^2}{30} + \frac{(379-285)^2}{285} + \frac{(727-803)^2}{803} + \frac{(39-21)^2}{21} + \frac{(106-200)^2}{200} + \frac{(642-566)^2}{566} = 119,33$ como $\chi_o^2 = 119,33 > \chi_{2,95\%}^2 = 5,99$ então pertence a RC, então rejeitamos H_0 , ou seja, o tipo de crime depende do fato de o criminoso ser estranho.

7.

a) $\bar{X}_1 = 49; \bar{X}_2 = 48; \bar{X}_3 = 50$, com $\bar{X}_1$ = média do grupo A, $\bar{X}_2$ = média do grupo B, $\bar{X}_3$ = média do grupo C.

b) $\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} = 49$, com $\bar{\bar{X}}$ = grande média.

c) $STQ = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2 = (48-49)^2 + (49-49)^2 + (50-49)^2 + (49-49)^2 + (47-49)^2 + (49-49)^2 + (48-49)^2 + (48-49)^2 + (49-49)^2 + (51-49)^2 + (50-49)^2 + (50-49)^2 = 14$, com c = número de fertilizantes, n_j = número de repetições em cada grupo.

d) $SQE = \sum_{j=1}^c n_j (\bar{X}_j - \bar{\bar{X}})^2 = 4(49-49)^2 + 4(48-49)^2 + 4(50-49)^2 = 8$.

e) $SQD = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2 = (48-49)^2 + (49-49)^2 + (50-49)^2 + (49-49)^2 + (47-48)^2 + (49-48)^2 + (48-48)^2 + (48-48)^2 + (49-50)^2 + (51-50)^2 + (50-50)^2 + (50-50)^2 = 6$.

f) A distribuição do F calculado é dado por:

$$F_o = \frac{MQE}{MQD}, \text{ com } MQE = \frac{SQE}{c-1}; MQD = \frac{SQD}{n-c}$$

Então $MQE = 8/(3-1) = 4$; $MQD = 6/(12-3) = 0,67$, assim $F_o = 4/0,67 = 6$, o valor crítico do F tabelado para $\alpha = 5\%$ é dado por $F_{(c-1, n-c)} = F_{(2,9,95\%)} = 4,26$.

Como $F_o = 6 > F_{(2,9,95\%)} = 4,26$, então ao nível de 5% existe diferença significativa entre os fertilizantes.

8. $\bar{X}_1 = 35,5; \bar{X}_2 = 36; \bar{X}_3 = 33,3; \bar{X}_4 = 31,3$, com $\bar{X}_1$ = média do pneu tipo A, $\bar{X}_2$ = média do pneu tipo B, $\bar{X}_3$ = média do pneu tipo C, $\bar{X}_4$ = média do pneu tipo D.

A grande média é dada por: $\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4}{4} = 34$

$STQ = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2 = (33-34)^2 + (38-34)^2 + \dots + (33-34)^2 + (31-34)^2 = 304,96$, com c = tipos de pneus, n_j = número de repetições em automóveis.

$SQE = \sum_{j=1}^c n_j (\bar{X}_j - \bar{\bar{X}})^2 = 6(35,5-34)^2 + 6(36-34)^2 + 6(33,3-34)^2 + 6(31,3-34)^2 = 82,83$.

$$SQD = \sum_{j=1}^c \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 = (33 - 35,6)^2 + \dots + (32 - 36)^2 + \dots + (31 - 33,3)^2 + \dots + (28 - 31,3)^2 = 222,17.$$

A distribuição do F calculado é dado por:

$$F_o = \frac{MQE}{MQD}, \text{ com } MQE = \frac{SQE}{c-1}; MQD = \frac{SQD}{n-c}$$

Então MQE=82,83/(4-1)=27,61; MQD=222,17/(24-4)=11,1, assim $F_o = 27,61/11,1 = 2,49$, o valor crítico do F tabelado para $\alpha = 5\%$ é dado por $F_{(c-1, n-c)} = F_{(3, 20, 95\%)} = 3,10$.

Como $F_o = 2,49 < F_{(3, 20, 95\%)} = 3,1$, então ao nível de 5% não há diferença entre os tipos de pneus.