Universidade de São Paulo
MAE 219 – Introducão à Probabilidade e à Estatística I
Lista de Exercícios 2 - Semestre 1 de 2015 - FEA

Prof. Pedro Morettin e Prof. Chang Chiann

1. Problema 1

   a. Usando o pacote $R - project$, temos:

   - $X_{med} = 2$

   - $Y_{med} = 2.5$

   Então fazendo a classificação temos os seguintes resultados:

   - $X_{baixo} = \{x : x \ \leq X_{med}\} = \{1, 3, 5, 6, 8, 9, 11, 13, 14, 17, 18, 19, 20, 21, 24, 25, 28, 31, 32, 33, 34, 37, 39, 40\}$. 24 indivíduos

   - $X_{alto} = \{x : x > X_{med}\} = \{2, 4, 7, 10, 12, 15, 16, 22, 23, 26, 27, 29, 30, 35, 36, 38\}$.16 indivíduos

   - $Y_{baixo} = \{y : y \ \leq Y_{med}\} = \{2, 4, 6, 9, 10, 12, 15, 16, 19, 20, 22, 23, 26, 27, 31, 32, 33, 35, 36, 38\}$. 20 indivíduos

   - $Y_{alto} = \{y : y \ > Y_{med}\} = \{1, 3, 5, 7, 8, 11, 13, 14, 17, 18, 21, 24, 25, 28, 29, 30, 34, 37, 39, 40\}$. 20 indivíduos

Tabela 1: Distribuição Conjunta

	X_{baixo}	X_{alto}	Total
Y_{baixo}	7	13	20
Y_{alto}	17	3	20
Total	24	16	40

   b. A porcentagem das pessoas com baixa rotatividade e ganhando pouco é $\frac{7}{40} = 0.175 -> 17.5\%$.

   c. A porcentagem das pessoas que ganham pouco é $\frac{20}{40} = 0.5 -> 50\%$.

   d. O total das pessoas com baixa rotatividade é 24, das quais 7 ganham pouco é $\frac{7}{24} = 0.2916 -> 29.16\%$.

   e. Sim, a informação adicional em (d) mudou significativamente a porcentagem observada em (c), pois em as duas situações mudou-se o conjunto referente. No primeira são todas as pessoas que ganham pouco e na segunda são as pessoas que ganham pouco mas no conjunto das pessoas com baixa rotatividade. Os dados mostram que mais pessoas ganham pouco com alta rotatividade do que com baixa rotatividade.

2. Problema 2

   a. Definamos os seguintes variáveis:
   - $S :=$ defeitos na suspensão
   - $E :=$ defeitos no sistema elétrico
   - S^c :=não tem defeitos na suspensão

- E^c :=não tem defeitos no sistema elétrico

Temos a informação,

proporção de S é 0.25

proporção de E é 0.15

proporção de S e E é 0.10

proporção dos que não tem defeitos na suspensão e não tem defeitos no sistema elétrico é 0.70

proporção dos que tem defeitos na suspensão e não tem defeitos no sistema elétrico é 0.15

proporção dos que não tem defeitos na suspensão e tem defeitos no sistema elétrico é 0.05

Tabela 2: Distribuição Conjunta

	E	E^c	Total
S	0.10	0.15	0.25
S^c	0.05	0.70	0.75
Total	0.15	0.85	1

b. A proporção de carros com defeitos é $0.25 + 0.15 - 0.10 = 0.30$

c. Entre os carros que apresentam defeitos na suspensão, a proporção que apresenta defeito no sistema elétrico é $\frac{0.10}{0.25} = 0.4$

d. Entre os carros que não apresentam defeitos na suspensão, a proporção que apresenta defeito no sistema elétrico é $\frac{0.05}{0.75} = 0.066$

e. Calculemos o coeficiente de contingência

- $\chi^2 = 16.3181$
- Coeficiente de Contingência: $C = \sqrt{\frac{\chi^2}{\chi^2+100}} = \sqrt{\frac{16.3181}{116.3181}} = 0.3745.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/100}{(2-1)(2-1)}} = 0.4039$

Como C e T, são grandes, indica haver associação entre as variáveis.

3. Problema 3

Sejam H : homen, M : mulher, U : usa hospital e N : não usa hospital.

a. Usaram o hospital 100 homens e 150 mulheres, para um total de 250 indivíduos. Então a proporção homens entre os indivíduos que usaram o hospital é $\frac{100}{250} = 0.4$.

b. Não usaram o hospital 900 homens e 850 mulheres, para um total de 1750 indivíduos. Então a proporção homens entre os indivíduos que não usaram o hospital é $\frac{900}{1750} = 0.514$.

Tabela 3: Distribuição Conjunta

	H	M	Total
U	100(10%)	150(15%)	250(12.5%)
N	900(90%)	850(85%)	1750(87.5%)
Total	1000(100%)	1000(100%)	2000(100%)

c. Observando a tabela, vemos que as porcentagem dos homens (10% e 90%) e das mulheres S^c (15% e 85%) são próximas das marginais (12.5% e 87.5%). Esses resultados parecem indicar não haver dependência entre as duas variáveis. Para ter uma conclusão mais exata, calculamos as seguintes medidas.

d. Valores esperados

Tabela 4: Valores esperados assumindo independência

	H	M	Total
U	125(12.5%)	125(87.5%)	250(100%)
N	875(12.5%)	875(87.5%)	1750(100%)
Total	1000(50%)	1000(50%)	2000(100%)

- $\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} = \frac{(100-125)^2}{125} + \frac{(150-125)^2}{125} + \frac{(900-875)^2}{875} + \frac{(850-875)^2}{875} = 0.14$
- Coeficiente de Contingência: $C = \sqrt{\frac{\chi^2}{\chi^2+2000}} = \sqrt{\frac{0.14}{2000.14}} = 0.0036.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/2000}{(2-1)(2-1)}} = 0.008366$

Como C e T são pequenos, indica não haver associação entre as variáveis.

4. Problema 4

Tabela 5: Distribuição Conjunta

	Menos de 4 meses	De 4 a 8 meses	Mais de 8 meses	Total
X	64(32%)	120(60%)	16(8%)	200(100%)
Y	104(35%)	175(58%)	21(7%)	300(100%)
Z	27(34%)	48(60%)	5(6%)	80(100%)
Total	195(33%)	343(59%)	42(8%)	580(100%)

Calculemos χ^2 e os coeficientes C e T.

- $\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} = 0.751477$
- Coeficiente de Contingência: $C = \sqrt{\frac{\chi^2}{\chi^2+580}} = \sqrt{\frac{0.751477}{580.751477}} = 0.03597.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/580}{(3-1)(3-1)}} = 0.0039375$

Os coeficientes C e T são próximos a zero, portanto, podemos dizer que as variáveis são independentes. Isto é, a duração do efeito de dedetização é independente da companhia (X, Y ou Z) selecionada, assim não existe evidência a favor da afirmação feita pela companhia X.

Tabela 6: Valores esperados assumindo independência

	Menos de 4 meses	De 4 a 8 meses	Mais de 8 meses	Total
X	67(33%)	118(59%)	15(8%)	200(100%)
Y	101(33%)	177(59%)	21(8%)	300(100%)
Z	27(33%)	48(59%)	6(8%)	80(100%)
Total	195(33%)	343(59%)	42(8%)	580(100%)

5. Problema 5

   Definir os seguintes variáveis:

   - S : indivíduos solteiros
   - C : indivíduos casados
   - $A10$: recebem até 10 salários mínimos
   - $A20$: recebem até 20 salários mínimos
   - $M20$: recebem mais 20 salários mínimos

   a. Considere a seguinte tabela:

Tabela 7: Distribuição Conjunta

	até 10	de 10 até 20	mais de 20	Total
S	0.12	0.19	0.09	0.40
C	0.08	0.31	0.21	0.60
Total	0.2	0.5	0.3	1

Temos as seguintes informações

proporção de indivíduos solteiros: 0.4

proporção de indivíduos que recebem até 10 salários mínimos: 0.2

proporção de indivíduos que recebem até 20 salários mínimos: 0.7

proporção de indivíduos casados dos recebem mais de 20 salários mínimos: 0.7

proporção de indivíduos que recebem até 10 salários mínimos dos que são solteiros: 0.3

Completamos a tabela como segue:

- proporção de indivíduos casados e recebem até 10 salários mínimos: 0.08
- proporção de indivíduos que recebem mais de 20 salários mínimos: 0.3
- proporção de indivíduos casados: 0.6

- proporção de indivíduos que recebem até 10 salários mínimos e são solteiros: 0.12
- proporção de indivíduos que recebem mais de 20 salários mínimos e são casados: 0.21

Tabela 8: Porcentagem por linha

	até 10	de 10 até 20	mais de 20	Total
S	30%	47.5%	22.5%	100%
C	13.3%	51.67%	35%	100%
Total	20%	50%	30%	100%

b. Calculemos o coeficiente de contingência

- $\chi^2 = 4.666$
- Coeficiente de Contingência: $C = \sqrt{\frac{\chi^2}{\chi^2+100}} = \sqrt{\frac{4.666}{104.666}} = 0.2111.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/100}{(2-1)(3-1)}} = 0.1527$

Como C e T são pequenos, indica ter uma associação fraca. Dizemos então que existe evidência para afirmar que o estado civil é independente da faixa salarial.

6. Problema 6

a. Calcule as proporções em relação ao total das colunas

Tabela 9: Porcentagem por colunas

	urbano	suburbano	rural	Total
a favor	30 (33.33%)	35 (58.3%)	35(70%)	100(100%)
contra	60(66.66%)	25 (41.7%)	15 (30%)	100(100%)
Total	90(100%)	60(100%)	50(100%)	200(100%)

Tabela 10: Valores Esperados

	urbano	suburbano	rural	Total
a favor	45(50%)	30(50%)	25(50%)	100(100%)
contra	45(50%)	30(50%)	25(50%)	100(100%)
Total	90(45%)	60(30%)	50(25%)	200(100%)

b. A opinião é dependente do local de residência, pois as proporções por cada coluna não são semelhantes.

c. Calculando o estatístico χ^2 temos: $\chi^2 = 19.666$.

- Coeficiente de Contingência: $C = \sqrt{\frac{\chi^2}{\chi^2+200}} = \sqrt{\frac{19.666}{219.666}} = 0.2992.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/200}{(2-1)(3-1)}} = 0.2217$

Como C e T não são muito pequenos, indica haver uma associação. Dizemos então que existe evidência para afirmar que la opinião dependente do local de residência.

7. Problema 7

Participam	São Paulo	Campinas	R.Preto	Santos	Total
Sim	50(25%)	65(26%)	105(35%)	120(40%)	340(100%)
Não	150(75%)	185(74%)	195(65%)	180(60%)	710(100%)
Total	200(20%)	250(23%)	300(28%)	300(29%)	1050(100%)

Tabela 11: Valores Esperados

Participam	São Paulo	Campinas	R.Preto	Santos	Total
Sim	65(32%)	80(32%)	98(33%)	97(32%)	340(100%)
Não	135(68%)	170(68%)	202(67%)	203(68%)	710(100%)
Total	200(20%)	250(23%)	300(28%)	300(29%)	1050(100%)

Para estudar se a participação em atividades esportivas depende ou não da cidade, calculamos os estatísticos χ^2 e C. De fato:

- $\chi^2 = 18.5184$
- $C = \sqrt{\frac{\chi^2}{\chi^2+1050}} = \sqrt{\frac{18.5184}{1068.5184}} = 0.1316.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/1050}{(2-1)(4-1)}} = 0.0766$

Como os valores de C e T são pequenos, podemos afirmar que a participação é independente da cidade.

8. Problema 8

a. Considere a seguinte tabela, com as proporções feitas com respeito a cada fila:

	Alta	Média	Baixa	Total
Sim	200(25%)	220(27.5%)	380(47.5%)	800(100%)
Não	200(16.7%)	280(23.3%)	720(60%)	1200(100%)
Total	400(20%)	500(25%)	1100(55%)	2000(100%)

Notemos que não há uma diferença notável na classe média, enquanto nas classes alta e média a diferença é notável. Por tanto é possível dizer que as respostas dependem da classe social.

b. Da tabela original calculamos os estatísticos χ^2, C e T.

- $\chi^2 = 33.6353$

Tabela 12: Valores Esperados

	Alta	Média	Baixa	Total
Sim	160(20%)	200(25%)	440(55%)	800(100%)
Não	240(20%)	300(25%)	660(55%)	1200(100%)
Total	400(20%)	500(25%)	1100(55%)	2000(100%)

- $C = \sqrt{\frac{\chi^2}{\chi^2+1200}} = \sqrt{\frac{33.6353}{1233.6353}} = 0.16512.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/1200}{(2-1)(3-1)}} = 0.1183$

Como neste caso C e T são pequenos, indica uma associação fraca das variáveis.

c. Baixo estas condições a nova tabela e os novos valores dos estatísticos χ^2, C e T, são:

	Alta	Média	Baixa	Total
Sim	160(21%)	220(29%)	380(50%)	760(100%)
Não	240(19%)	280(23%)	720(58%)	1240(100%)
Total	400(20%)	500(25%)	1100(55%)	2000(100%)

Tabela 13: Valores Esperados

	Alta	Média	Baixa	Total
Sim	152(20%)	190(25%)	418(55%)	760(100%)
Não	248(20%)	310(25%)	682(55%)	1240(100%)
Total	400(20%)	500(25%)	1100(55%)	2000(100%)

- $\chi^2 = 16.9693$
- $C = \sqrt{\frac{\chi^2}{\chi^2+1200}} = \sqrt{\frac{16.9693}{1216.9693}} = 0.1180.$
- Estatístico T: $T = \sqrt{\frac{\chi^2/1200}{(2-1)(3-1)}} = 0.0840$

Com essa mudança, a associação fica mais fraca.

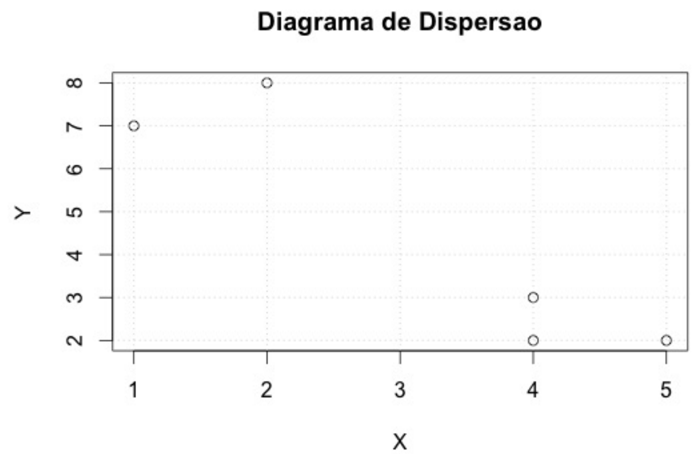
9. Problema 9

a. Calculemos o coeficiente de correlação usando a informação dada:

$$corr(X,Y) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = -0.9189002$$

b. Como $corr(X,Y) \approx -1$, isto é, existe uma correlação alta negativa entre as duas variáveis. Por tanto dizemos que X pode ser usada para explicar Y.

c. Usando R temos o seguinte diagrama de dispersão:

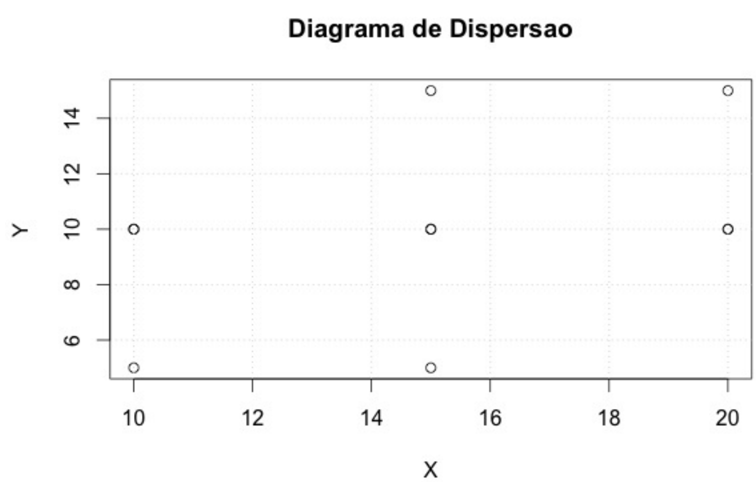


10. Problema 10

    a. $\bar{X} = 15$ e $dp(X) = 3.8729$

    b. $\bar{Y} = 10$ e $dp(Y) = 3.1622$

    c. Diagrama de dispersão



    d. $corr(X, Y) = 0.4082483$

    e. Salário médio familiar $SM = 25$

    Variância de salário familiar $V = 35$

    f. Salário médio familiar com desconto $SMF = 23.2$

    Variância de salário familiar com desconto $V = 30.18$