

Estatística Geral I
Prof: Jony Arrais Pinto Junior
Lista 08

1. Utilizando os dados da Tabela 01 da Lista 01, faça o que se pede:
 - (a) Construa a distribuição de frequência conjunta para as variáveis grau de instrução e região de procedência.
 - (b) Qual a porcentagem de funcionários que têm o ensino médio?
 - (c) Qual a porcentagem daqueles que têm o ensino médio e são do interior?
 - (d) Dentre os funcionários do interior, quantos por cento têm ensino médio?
 - (e) Dentre os funcionários com ensino fundamental, quantos por cento são da capital?
 - (f) Qual a distribuição das proporções do grau de instrução segundo cada uma das regiões de procedência?
 - (g) Apresente um gráfico que ajude a visualizar a distribuição do item anterior.
 - (h) Baseado nos resultados dos dois itens anterior, você diria que existe dependência entre a região de procedência e o nível de educação do funcionário? (Faça uma análise das proporções e dos gráficos, não use nenhum coeficiente neste item).
 - (i) Apresente um coeficiente que calcule o grau de dependência destas variáveis e interprete o mesmo.
2. No estudo de uma certa comunidade foram entrevistados 1000 indivíduos e verificou-se que:
 - (i) A proporção de indivíduos solteiros é de 0,4.
 - (ii) A proporção de indivíduos que recebem até 10 salários mínimos é de 0,2.
 - (iii) A proporção de indivíduos que recebem até 20 salários mínimos é de 0,7.
 - (iv) A proporção de indivíduos casados entre os que recebem mais de 20 salários mínimos é de 0,7.
 - (v) A proporção de indivíduos que recebem até 10 salários mínimos entre os solteiros é de 0,3.
 - (a) Construa a tabela de dupla entrada das variáveis estado civil e faixa salarial e as respectivas distribuições de frequências marginais.
 - (b) Você diria que existe relação entre as duas variáveis?
3. Uma pesquisa para verificar a tendência dos alunos a prosseguir os estudos, segundo a classe social do respondente, mostrou o seguinte quadro:

Companhia	Classe Social		
	Alta	Média	Baixa
Sim	200	220	380
Não	200	280	720

- (a) Você diria que a distribuição de respostas afirmativas é igual a de respostas negativas?
- (b) Existe dependência entre os dois fatores? Dê uma medida quantificadora da dependência.
- (c) Se dos 400 alunos da classe alta, 160 escolhessem continuar e 240 não, você mudaria sua conclusão? Justifique.
4. A companhia de dedetização afirma que o processo por ela utilizado garante um efeito mais prolongado do que aquele obtido por seus concorrentes mais diretos. Uma amostra de vários ambientes dedetizados foi colhida e anotou-se a duração do efeito de dedetização. Os resultados estão na tabela abaixo. Você acha que existe alguma evidência a favor ou contra a afirmação feita pela Companhia A?

Companhia	Duração do efeito da dedetização		
	Menos de 4 meses	De 4 a 8 meses	Mais de 8 meses
A	64	120	16
B	104	175	21
C	27	48	5

5. Numa pesquisa sobre rotatividade de mão-de-obra, para uma amostra de 40 pessoas foram observados duas variáveis: número de empregos nos últimos dois anos (X) e salário mais recente, em número de salários mínimos (Y). Os resultados foram:
- (a) Usando a mediana, classifique os indivíduos em dois níveis, alto e baixo, para cada uma das variáveis, e construa a distribuição de frequências conjunta das duas classificações.
- (b) Qual a porcentagem de pessoas com baixa rotatividade e ganhando pouco?
- (c) Qual a porcentagem de pessoas que ganham pouco?
- (d) Entre as pessoas de baixa rotatividade, qual a porcentagem dos que ganham pouco?
- (e) A informação adicional dada em (d) mudou muito a porcentagem observada em (c)? O que isso significa?

Indivíduo	X	Y
1	1	1
2	3	1
3	2	2
4	3	1
5	3	3
6	2	3
7	2	4
8	3	5
9	1	6
10	2	1
11	3	2
12	4	2
13	4	3
14	4	2
15	2	3
16	1	4
17	2	5
18	3	5
19	4	5
20	2	2
21	3	1
22	1	1
23	3	1
24	4	1
25	2	4
26	4	5
27	1	2
28	1	1
29	1	5
30	3	6
31	2	6
32	2	6
33	3	1
34	1	2
35	4	1
36	1	2
37	2	3
38	2	4
39	1	4
40	3	5

6. O departamento de vendas de certa companhia foi formado há um ano com a admissão de 15 vendedores. Nessa época, foram observados para cada um dos vendedores os valores de três variáveis:

T : O resultado em um teste apropriado para vendedores;

E : anos de experiência de vendas;

G : conceito do gerente de vendas, quanto ao currículo do candidato.

O diretor da companhia resolveu agora ampliar o quadro de vendedores e pede sua colaboração para responder a algumas perguntas. Para isso, ele lhe dá informações

adicionais sobre duas variáveis:

V : volume médio mensal de vendas em s.m.

Z : zona da capital para a qual o vendedor foi designado.

Os resultados encontram-se abaixo:

Vendedor	T	E	G	V	Z
1	8	5	Bom	54	Norte
2	9	2	Bom	50	Sul
3	7	2	Mau	48	Sul
4	8	1	Mau	32	Oeste
5	6	4	Bom	30	Sul
6	8	4	Bom	30	Oeste
7	5	3	Bom	29	Norte
8	5	3	Bom	27	Norte
9	6	1	Mau	24	Oeste
10	7	3	Mau	24	Oeste
11	4	4	Bom	24	Sul
12	7	2	Mau	23	Norte
13	3	3	Mau	21	Sul
14	5	1	Mau	21	Oeste
15	3	2	Bom	16	Norte

- Faça um histograma da variável V em classes de tamanho 10, tendo por limite inferior da primeira classe o tamanho 15.
 - Encontre a média e a variância da variável V . Suponha que um vendedor seja considerado excepcional se seu volume de vendas é dois desvios padrões superior à média geral. Quantos vendedores excepcionais existem na amostra?
 - O diretor de vendas anunciou que transferirá para outra praça todos os vendedores cujo volume de vendas for inferior ao 1o quartil da distribuição. Qual o volume mínimo de vendas que um vendedor precisa ter para não ser transferido?
 - Os vendedores argumentam com o diretor que esse critério não é justo, pois há zonas de vendas privilegiadas. A quem você daria razão?
 - Qual das três variáveis observadas na admissão do pessoal é mais importante para julgar um futuro candidato ao emprego?
 - Qual o grau de associabilidade entre o conceito do gerente e a zona a que o vendedor foi designado? Você tem explicação para este resultado?
 - Qual o grau de associação entre o conceito do gerente e o resultado do teste? E entre zonas e vendas?
7. Para cada par de variáveis abaixo, esboce o diagrama de dispersão. Diga se você espera uma dependência linear e nos casos afirmativos avalie o coeficiente de correlação.
- Peso e altura dos alunos do primeiro ano de um curso de Administração.
 - Peso e altura dos funcionários de um escritório.
 - Quantidade de trigo produzida e quantidade de água recebida por canteiros numa estação experimental.
 - Notas de Cálculo e Estatística de uma classe onde as duas disciplinas são lecionadas.

- (e) Acuidade visual e idade de um grupo de pessoas.
 - (f) Renda familiar e porcentagem dela gasta com alimentação.
 - (g) Número de peças montadas e resultado de um teste de inglês por operário.
8. Abaixo estão os dados referentes à porcentagem da população economicamente ativa e empregada no setor primário e o respectivo índice de analfabetismo para algumas regiões metropolitanas brasileiras.

Regiões metropolitanas	Setor primário	Índice de analfabetismo
São Paulo	2,0	17,5
Rio de Janeiro	2,5	18,5
Belém	2,9	19,5
Belo Horizonte	3,3	22,2
Salvador	4,1	26,5
Porto Alegre	4,3	16,6
Recife	7,0	36,6
Fortaleza	13,0	38,4

- (a) Faça o diagrama de dispersão.
 - (b) Você acha que existe uma dependência linear entre as variáveis?
 - (c) Calcule o coeficiente de correlação e interprete.
 - (d) Existe alguma região com comportamento diferente das demais? Se existe, elimine o valor correspondente e recalcule o coeficiente de correlação. Interprete o mesmo e compare com o do item anterior.
9. Prove que

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)} \sqrt{(\sum y_i^2 - n\bar{y}^2)}}.$$

10. Usando os dados da Tabela 01 da Lista 01 responda:
- (a) Existe dependência entre a média geral do aluno e a idade entre os alunos do gênero masculino?
 - (b) E do sexo feminino?
11. Abaixo encontram-se os dados de acuidade visual e idade de 10 pacientes.

Indivíduo	Idade	Acuidade visual
1	20	90
2	20	100
3	20	80
4	20	90
5	25	100
6	25	90
7	25	80
8	25	90
9	30	70
10	30	90

- (a) Quem seria a variável dependente e independente?
- (b) Qual o modelo recomendado? Encontre a reta de mínimos quadrados e interprete o(s) coeficiente(s).
12. Um jornal quer verificar a eficácia de seus anúncios na venda de carros usados. A tabela abaixo mostra o número de anúncios publicados e o correspondente número de carros vendidos por seis companhias que usaram apenas esse jornal como veículo de propaganda.

Companhia	Anúncios	Carros vendidos
A	74	139
B	45	108
C	48	98
D	36	76
E	27	62
F	16	57

- (a) Ajuste a regra de mínimos quadrados adequada.
- (b) Interprete o(s) coeficiente(s) envolvido(s) na reta de regressão.
- (c) Como você convenceria a companhia F para que ela aumentasse o número de anúncios no seu jornal?
13. Os dados abaixo correspondem às variáveis renda familiar e gasto com alimentação numa amostra de dez famílias, representadas em salários mínimos.

Renda familiar	Gasto com alimentação
3	1,5
5	2,0
10	6,0
20	10,0
30	15,0
50	20,0
70	25,0
100	40,0
150	60,0
200	80,0

- (a) Ajuste a reta de mínimos quadrados, do gasto com alimentação como função da renda familiar.
- (b) Interprete o(s) coeficiente(s) envolvido(s) na reta de mínimos quadrados.
- (c) Qual a previsão de gasto com alimentação para uma família com renda de 1000 reais? Considere que um salário mínimo é de 500 reais.

Usando o R responda: Leia o *Banco Cancer Mama*. Foram avaliadas 30 pacientes com câncer de mama e 27 sem câncer de mama. Somente as pacientes com câncer de mama foram submetidas a um tratamento durante os últimos 6 meses com o objetivo de diminuir o número de células cancerígenas. No banco de dados encontram-se as seguintes variáveis:

- Câncer (0 - Não e 1 - Sim)

- Peso (Kg)
- Primeira menstruação antes dos 12 anos (0 - Não e 1 - Sim)
- Grau de parentesco com alguém que já teve câncer de mama na família (0 - 3º Grau, 1 - 2º Grau e 2 - 1º Grau)
- Número de células cancerígenas (antes do tratamento)
- Número de células cancerígenas (depois do tratamento)

Todas as variáveis para os pacientes sem câncer foram medidas em uma única avaliação. Para as pacientes com câncer foram medidas todas as variáveis em uma primeira avaliação e na segunda (depois dos 6 meses de tratamento) foi medida somente o número de células cancerígenas.

Com base no estudo, responda os seguintes questionamentos:

14. Você diria que existe associação entre ter a primeira menstruação antes dos 12 anos e câncer de mama? Apresente uma tabela e um gráfico que ajude a visualizar o resultado.
15. A literatura da área médica aponta a obesidade como um fator de risco para câncer de mama. Será que existe relação entre peso e câncer de mama nesta amostra? Mostre um gráfico que ajude a visualizar o resultado.
16. Pode-se dizer que existe associação entre grau de parentesco com pessoas portadoras de câncer e câncer de mama? Mostre um gráfico que ajude a visualizar o resultado.
17. Pode-se dizer que existe relação entre o peso e o número de células cancerígenas antes do tratamento para as pacientes com câncer de mama? E entre peso e o número de células cancerígenas depois do tratamento? Mostre um gráfico que ajude a visualizar o resultado.