

LISTA DE EXERCÍCIOS 5

Disciplina: Estatística II

1. Quando que as amostras são consideradas grandes o suficiente, para que a distribuição de amostragem da diferença entre proporções de duas amostras seja (aproximadamente) normal?
2. Uma organização de pesquisas eleitorais, de âmbito nacional, realiza uma pesquisa telefônica junto a adultos a respeito do desempenho do presidente. Das pessoas pesquisadas, um total de 990 havia votado na eleição presidencial anterior. Das 630 que votaram no presidente, 72% afirmaram que seu desempenho era bom. Das 360 que não votaram nele, 60% se manifestaram favoravelmente. Teste a hipótese nula de que não há diferença nas proporções populacionais das pessoas que consideram bom o desempenho presidencial (use o nível de 0,05 de significância).
3. Um pesquisador entrevista, por telefone, uma amostra de âmbito estadual de pessoas com 18 anos ou mais acerca de sua atitude em relação a um controle mais rigoroso sobre a posse de armas. Com o conjunto de resultados a seguir, teste a significância da diferença entre as proporções de homens e de mulheres que são favoráveis a um controle mais rigoroso.

Tabela 1 - Proporção de homens e mulheres.

	Homens	Mulheres
A favor	92	120
Contra	74	85
<i>n</i>	166	205

4. Explique o significado do coeficiente de determinação.
5. Quando os pontos em um diagrama de dispersão se aglomeram em torno da reta de regressão, podemos dizer que a correlação é
 - a. fraca.
 - b. forte.
 - c. positiva.
 - d. negativa.
6. O que o coeficiente de correlação linear nos informa sobre a relação entre duas variáveis? Dentro de qual intervalo um coeficiente de correlação pode assumir um determinado valor?
7. Para um conjunto de dados amostrais, o coeficiente de correlação linear, r , apresenta um valor positivo. Qual dos seguintes itens é verdadeiro em relação à inclinação, b , da linha de regressão estimada para os mesmos dados amostrais?
 - a. O valor de b será positivo.
 - b. O valor de b será negativo.
 - c. O valor de b pode ser positivo ou negativo.
8. Para um conjunto de dados amostrais sobre duas variáveis, o valor do coeficiente de correlação linear é igual a zero. Isso significa que essas variáveis não estão relacionadas? Explique.

9. Um coeficiente de correlação expressa, em um único número,
- a intensidade de uma correlação.
 - a direção de uma correlação.
 - a intensidade e a direção de uma correlação.
10. $r = -0,17$ representa
- uma correlação forte e negativa.
 - uma correlação forte e positiva.
 - uma correlação fraca e negativa.
 - uma correlação fraca e positiva.
11. Qual das seguintes condições *não* é exigida pelo r de Pearson?
- Uma relação linear.
 - Dados nominais.
 - Amostragem aleatória.
 - Características distribuídas normalmente.
12. Explique os significados das palavras *simples* e *linear*, conforme utilizadas em *regressão linear simples*.
13. Explique o significado da variável dependente e o significado da variável independente, em relação a um modelo de regressão.
14. Explique a diferença entre relação linear e relação não-linear entre duas variáveis.
15. Explique a diferença entre y e \hat{y} .
16. Explique sucintamente as premissas do modelo de regressão da população.
17. Como você expressa os graus de liberdade para um modelo de regressão linear simples?
18. Na análise de regressão, predizemos o valor de Y a partir de X com base em:
- um coeficiente de correlação significativa.
 - uma equação matemática.
 - uma correlação positiva.
 - uma curva normal.
19. Qual a alternativa denota a inclinação ou o coeficiente de regressão para X ?
- O intercepto Y .
 - O termo a .
 - O termo do erro.
 - O termo b .
20. Uma medida da proporção da variância em Y explicada por X é conhecida como:
- coeficiente de não-determinação.
 - coeficiente de determinação.
 - regressão múltipla.
 - inclinação.

21. Com base nos dados da Tabela 2:

- Construa um diagrama de dispersão.
- Encontre a correlação entre a precipitação e a umidade relativa do ar e verifique a significância.
- Desenvolva uma equação de regressão estimada para esses dados e verifique a significância.

Tabela 2 - Variabilidade da precipitação e da umidade relativa do ar no ano de 2004 em uma área de pastagem a 15 km de Ouro Preto.

Mês	Precipitação (mm)	Umidade relativa do ar (%)
Jan	332,4	84,47
Fev	346,4	83,48
Mar	307,4	83,51
Abr	249,2	82,99
Mai	67,6	82,64
Jun	3,2	74,85
Jul	36	70,39
Ago	41,6	62,31
Set	10,6	65,70
Out	154,8	75,70
Nov	241	80,64
Dez	251,8	83,24

22. Os hábitos de ler e de assistir televisão concorrem quanto ao tempo de lazer? Para verificar, um especialista em comunicação entrevistou uma amostra de 10 crianças para saber o número de livros que elas leram no decorrer do último ano e o número de horas que passaram vendo televisão diariamente. Seus resultados são os seguintes:

Número de livros	Número de horas assistindo televisão
0	3
7	1
2	2
1	2
5	0
4	1
3	3
3	2
0	7
1	4

Quais são a intensidade e a direção da correlação entre número de livros lidos e de horas diárias assistindo televisão? A correlação é significativa?

23. Uma pesquisadora educacional está interessada no efeito do desempenho acadêmico no ensino médio sobre o desempenho acadêmico na faculdade. Para tanto, consultou os registros escolares de 12 bacharéis (todos frequentaram o mesmo colégio), a fim de determinar sua nota média no ensino médio (X) e sua nota média na faculdade (Y).

- Faça um diagrama de dispersão dos dados.
- Calcule a inclinação e o intercepto Y da reta de regressão.

c. Verifique a validação da regressão.

X	Y
3,3	2,7
2,9	2,5
2,5	1,9
4,0	3,3
2,8	2,7
2,5	2,2
3,7	3,1
3,8	4,0
3,5	2,9
2,7	2,0
2,6	3,1
4,0	3,2

24. Os dados fornecidos a seguir correspondem aos resultados de final de semestre em um curso para uma amostra de 10 alunos e aos resultados das avaliações dos alunos em relação ao professor. (Nos resultados da avaliação do professor, 1 representa o resultado mais baixo e 4 representa o resultado mais elevado.).

Avaliação do professor	3	2	3	1	2	4	3	4	4	2
Resultado teste semestral	90	75	97	64	47	99	75	88	93	81

- Encontre a regressão dos resultados da avaliação do professor em relação aos resultados dos testes semestrais.
- Teste, no nível de significância de 1%, se B é significativo.

25. Em um estudo com 1700 adolescentes entre 15 e 19 anos, metade respondeu a uma pesquisa por escrito e metade usando anonimamente um programa de computador. Dos que fizeram a pesquisa escrita, 7,9% afirmaram ter portado uma arma nos últimos 30 dias. Dos que fizeram a pesquisa por computador, 12,4% afirmaram ter portado uma arma nos últimos 30 dias (com base em dados do Instituto Urbano).

- As porcentagens amostrais de 7,9% e 12,4% são obviamente diferentes, mas a diferença é significativa? Explique.

26. Um artigo no *Journal of Environmental Engineering* (v. 115, n. 3, 1989, p. 608-619) reportou os resultados de um estudo a respeito da ocorrência de sódio e cloreto nas correntes superficiais de um rio na parte central de *Rhode Island*. Os dados a seguir se referem à concentração de cloreto (em mg/L), y ; e a área (em %) das encostas exploradas para análise, x .

y	4,4	6,6	9,7	10,6	10,8
x	0,19	0,15	0,57	0,70	0,67
y	10,9	11,8	12,1	14,3	14,7
x	0,63	0,47	0,70	0,60	0,78
y	15,0	17,3	19,2	23,1	27,4
x	0,81	0,78	0,69	1,30	1,05
y	27,7	31,8	39,5		
x	1,06	1,74	1,62		

- Desenhe um diagrama de dispersão dos dados. Um modelo de regressão linear simples parece apropriado aqui?
- Ajuste um modelo de regressão linear simples.

27. Refira-se ao exercício anterior que apresenta dados sobre concentração de cloreto y e áreas da encosta x .

- Que proporção da variabilidade total na concentração de cloreto é explicada pelo modelo de regressão?
- Plote os resíduos contra \hat{y} e contra x . interprete esses gráficos.
- Verifique a validação da regressão.

28. Uma empresa de energia elétrica está interessada em desenvolver um modelo relacionando o pico de demanda horária (y em kW) ao consumo mensal total de energia (x em kWh) durante um mês. Dados para 50 consumidores residenciais são mostrados na tabela abaixo.

- Desenhe um diagrama de dispersão de y versus x .
- Ajuste um modelo de regressão linear simples.
- Teste a significância de regressão, usando $\alpha = 0,05$.
- Plote os resíduos versus \hat{y} e comente as suposições sobre a regressão em foco. Especificamente, você acha que a suposição de igualdade de variância foi satisfeita?
- Encontre um modelo de regressão linear simples, usando \sqrt{y} como a resposta. Essa transformação em y estabiliza o problema da desigualdade de variância, notado no item (d)?

Consumidor	$x(\text{kWh})$	$y(\text{kW})$	Consumidor	$x(\text{kWh})$	$y(\text{kW})$	Consumidor	$x(\text{kWh})$	$y(\text{kW})$
1	679	0,79	18	1276	1,88	35	1130	4,79
2	292	0,44	19	745	0,77	36	463	0,51
3	1012	0,56	20	795	3,70	37	770	1,74
4	493	0,79	21	540	0,56	38	724	4,10
5	582	2,70	22	874	1,56	39	808	3,94
6	1156	3,64	23	1543	5,28	40	790	0,96
7	997	4,73	24	1029	0,64	41	783	3,29
8	2189	9,50	25	710	4,00	42	406	0,44
9	1097	5,34	26	1434	0,31	43	1242	3,24
10	2078	6,85	27	837	4,20	44	658	2,14
11	1818	5,84	28	1748	4,88	45	1746	5,71
12	1700	5,21	29	1381	3,48	46	895	4,12
13	747	3,25	30	1428	7,58	47	1114	1,90
14	2030	4,43	31	1255	2,63	48	413	0,51
15	1643	3,16	32	1777	4,99	49	1787	8,33
16	414	0,50	33	370	0,59	50	3560	14,94
17	354	0,17	34	2316	8,19			

*Viver é a coisa mais rara do mundo,
A maioria das pessoas apenas existe.*

Oscar Wilde (1856-1900), escritor irlandês.