Lista de exercícios

Esta lista foi ampliada em relação à versão de 2022, portanto teremos algumas questões sem resposta.

Introdução

1. Para cada uma das amostras abaixo, informar o tipo do processo de amostragem: P - Amostragem probabilística; NP - Amostragem não probabilística.

   Para uma pesquisa sobre os hábitos dos estudantes. Uma amostra foi construída com o seguinte procedimento:

   ( ) Todos os meus colegas da faculdade (tenho telefone e e-mail de todos eles).

   ( ) Fiquei na única porta de entrada da escola abordando todos os meus conhecidos.

   ( ) Fiquei na única porta de entrada da escola e a cada 12 pessoas que entravam, eu abordava.

   ( ) Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista.

   ( ) Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados.

2. Um estudo sobre o desempenho dos vendedores de uma grande cadeia de lojas de varejo está sendo planejado. Para tanto, deve ser colhida uma amostra probabilística dos vendedores. Classifique cada uma das amostras abaixo conforme a seguinte codificação:

   (A) Amostragem casual simples

   (B) Amostragem Sistemática

   (C) Amostragem estratificada

   (D) Amostragem por meio de conglomerados

   ( ) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei todos vendedores que ocupavam posições múltiplas de 15 (15ª posição, 30ª posição, 45ª posição, 60ª posição, 75ª posição, 90ª posição, 105ª posição, etc).

   ( ) Escolhi casualmente 3 lojas da rede. A amostra foi composta de todos os vendedores que atualmente em cada uma destas 3 lojas.

   ( ) Em cada uma das lojas, identifiquei todos os vendedores (lista de vendedores por loja). Selecionei aleatoriamente k vendedores da loja, onde k é um número inteiro proporcional à quantidade de vendedores da loja.

   ( ) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei aleatoriamente n vendedores.

3. Abaixo é apresentada uma pequena parte dos dados que uma empresa mantém a respeito de seus funcionários:

Nome	Idade	Sexo	Raça	Salário	Ocupação
Ana Costa	39	Feminino	Branca	2.210,00	Gerente
Jair Freitas	27	Masculino	Negra	1.750,00	Técnico
Ruoh-Lin	22	Masculino	Amarela	1.525,00	Técnico

Tab. 1: Dados dos funcionários

   a) Além dos nomes dos funcionários, há mais 5 variáveis descritas. Destas, diga quais são variáveis que representam categorias?

   b) Quais variáveis são quantitativas? Basendo-se nos dados da tabela, quais devem ser as unidades de medida de tais variáveis.

Estatística descritiva

4. Um paciente fez 6 exames de sangue em 6 meses consecutivos para medir seu nível de fosfato por decilitro de sangue. Os resultados obtidos foram: 5.6 5.2; 4.6; 4.9; 5.7; 6.4. Calcule a média, mediana e o desvio padrão do nível de fosfato por decilitro de sangue.

5. Os salários de 20 funcionários de um banco, ordenados em ordem crescente são: 1400, 1400, 1400, 1400, 1400, 1400, 1400, 1400, 1550, 1550, 1650, 1650, 1800, 1800, 1900, 2000, 2050, 2250, 2300, 2400.

   a) Calcule média, mediana e moda para estes dados.

   b) Resuma os dados empregando o esquema de 5 números. Interprete os resultados.

   c) Calcule os Q_1 e Q_3. Os valores fora do intervalo: $[Q_1 - 1.5 \times IQR; Q_3 + 1.5 \times IQR]$ com $IQR = Q_3 - Q_1$ podem ser visto como outliers. Empregando este critério há presença de outliers neste conjunto de dados?

   d) Esboce o box-plot deste conjunto de dados.

6. Tempos para montar uma peça foram coletados e organizados conforme a tabela abaixo. Calcule média e o desvio padrão do tempo:

Classe	Frequência	Classe	Frequência
150|–180	3	270|–300	33
180|–210	8	300|–330	40
210|–240	10	330|–360	35
240|–270	13	360|–390	8

7. Para o conjunto de números a seguir: 20; 25; 25; 27; 28; 31; 33; 34; 36; 37; 44; 50; 59; 85; 86

   a) Calcule a média e o desvio padrão;

   b) Encontre os dois principais "outliers" empregando o critério da questão anterior, deixe-os de fora e calcule novamente a média e o desvio padrão. Como os "outliers" afetam os valores da média e do desvio padrão?

   c) Calcule a mediana e os primeiro e terceiros quartis. Remova os outliers do item anterior e recalcule a mediana e os primeiro e terceiros quartis. Estes outliers impactaram na determinação da mediana e dos quartis?

8. A tabela abaixo apresenta 40 empréstimos pessoais de uma firma de crédito ao consumidor:

300	300	350	350	450	450	500	500	550	550
600	600	650	700	750	750	850	850	900	900
950	1000	1000	1000	1100	1200	1200	1250	1300	1400
1500	1500	1600	1650	1800	1900	2000	2000	2500	3000

   a) Calcule os Q_1 e Q_3. Os valores fora do intervalo: $[Q_1 - 1.5 \times IQR; Q_3 + 1.5 \times IQR]$ com $IQR = Q_3 - Q_1$ podem ser visto como outliers. Empregando este critério há presença de outliers neste conjunto de dados?

   b) Esboce o box-plot para este conjunto de dados.

   c) Pelo item b, a distribuição é simétrica?

   d) Construa uma tabela de distribuição de frequências para estes dados fazendo o limite inferior da 1ª classe igual a \$ 300 e o intervalo de classe igual a \$ 400.

   e) Utilizando a tabela de distribuição de frequências construa um histograma.

   f) Calcule média, mediana, moda e o desvio pdrão da tabela distribuição de frequências do item d

9. Os dados a seguir representam o tempo de sobrevivência (em dias) de ratos infectados por uma determinada bactéria em um experimento de laboratório.

Tab. 2: Tempos de sobrevivência.

43	45	53	56	56	57	58	66	67	73	74	79	80	80	81	81	81	82
83	83	84	88	89	91	91	92	92	97	99	99	100	100	101	102	102	102
103	104	107	108	109	113	114	118	121	123	126	128	137	138	139	144	145	147
156	162	174	178	179	184	191	198	211	214	243	249	329	380	403	511	522	598

   a)Construa o histograma e descreva suas principais características. Que formato ele apresenta?

   b) Forneça as medidas resumo que fazem parte do esquema de 5 números. Interprete seus resultados.

10. A seguir está o resultado de um estudo que avalia se existe uma associação entre anos de serviço (X) e o número de clientes (Y) de corretores de uma cia. de seguros.

Corretor	A	B	C	D	E	F	G	H	I	J
Anos	2	3	4	5	4	6	7	8	8	10
# de clientes	48	50	56	52	43	60	62	58	64	72

a) Faça um diagrama de dispersão entre X e Y.

b) Obter o coeficiente de correlação entre X e Y. Interprete seu resultado.

11. Os dados abaixo se referem aos resultados de operadores num um teste sobre conhecimento de língua estrangeira (X) e o tempo que cada um demorou para aprender a operar uma determinada máquina (Y).

Operador	A	B	C	D	E	F	G	H	I	J
Teste	42	53	64	75	74	76	77	88	80	90
# de clientes	348	350	356	342	343	360	362	358	364	372

a) Faça um diagrama de dispersão entre X e Y.

b) Obter o coeficiente de correlação entre X e Y. Interprete seu resultado.

12. Uma amostra de 200 habitantes de uma ilha foi selecionada para declarar sua opinião sobre um projeto ambiental. Os dados estão reproduzidos abaixo:

	Local da residência		
Opinião	Urbano	Suburbano	Rural
A favor	60	35	15
Contra	30	25	35

a) Calcule as proporções em relação ao total das colunas

b) Calcule as proporções em relações ao total das linhas

c) A opinião independe do local da residência?

Estimação

13. Seja $x_1; x_2; \ldots, x_5$ uma amostra retirada de uma população com média μ e desvio padrão σ. São sugeridos os seguintes esstimadores de μ :

$$\hat{\mu}_1 = x_1;\ \hat{\mu}_2 = \frac{x_1 + x_2}{2};\ \hat{\mu}_3 = \frac{x_1 + 2x_5}{3};\ \hat{\mu}_4 = \frac{x_1 + 2x_2 + \ldots + 4x_4}{10}; \hat{\mu}_5 = \frac{x_1 + x_2 + \ldots + x_5}{5}$$

a) Os estimadores são não viciados?

b)Qual destes estimadores propostos é o melhor? Justifique sua resposta.

14. Em um campeonato de jogos de tabuleiro, são empregados 4 dados, com 6, 8, 12 e 20 lados. Cada dado tem um parâmetro (θ) que está associado a uma cor específica: azul (AZ), verde(VD), vermelho(VM), branco (BR) respectivamente:para dados de 6, 8, 12, e 20 lados. A probabilidade de sair cada número da face de cada dado está na tabela a seguir:

x	P(X=x| θ = AZ)	P(X=x| θ = VD)	P(X=x| θ = VM)	P(X=x| θ = BR)
1	1/6	1/8	1/12	1/10
2	1/6	2/8	1/12	1/10
3	1/12	1/8	1/12	1/20
4	2/6	1/8	1/12	1/20
5	1/12	1/16	1/12	1/20
6	1/6	1/16	1/12	1/20
7		1/8	1/12	1/20
8		1/8	1/12	1/20
9			1/12	1/20
10			1/12	1/20
11			1/12	1/20
12			1/12	1/20
13				1/20
14				1/20
15				1/20
16				1/20
17				1/40
18				1/40
19				1/40
20				1/40

Determine o estimador de máxima verossimilhança de θ para cada valor de x.

15. Deseja-se estimar o número de trutas em um lago. Para tanto, recolheram-se 100 trutas, marcaram-se estas e depois foram devolvidas ao lago. Após uma semana, recolheram-se novamente 200 trutas.

    a) Se foram encontradas 20 trutas, forneça uma estimativa do número de trutas no lago?

    b) Se o número de trutas no lago (N) for 1000, qual a probabilidade de se encontrar 20 trutas marcadas?

    c) Qual a estimativa de máxima verossimilhança do número de trutas que maximiza a probabilidade de ocorrer o resultado encontrado (20 trutas)?

Intervalo de confiança - uma amostra

16. Para γ igual a 90%, 95% e 99%, determine uma constante a tal que $P(|\bar{x} - \mu| \leq a) = \gamma$, considerando que o desvio padrão σ é conhecido.

17. Considere uma população com média desconhecida μ, porém com variância conhecida $\sigma^2 = 25$. Em uma amostra de tamanho 36 encontramos $\bar{x} = 8$.

    a) Construa um intervalo de 95% de confiança para μ.

    b) Construa um intervalo de 90% de confiança para μ.

    c) Construa um intervalo de 99% de confiança para μ.

18. Para uma amostra aleatória com média amostral igual a 10, construa intervalos de confiança com os níveis de confiança γ, tamanhos de amostras n e desvios padrões populacionais σ dados abaixo:

Tab. 3: Dados.

	I	II	III	IV	V	VI
n	9	9	16	100	25	25
σ	4	12	5	5	8	8
γ	95	95	95	95	99	95

19. Da experiência passada sabe-se que o desvio padrão de altura de crianças da 5a série é 5 cm.

    a) Colhendo uma amostra de 36 destas crianças observou-se média amostral de 150 cm. Qual o intervalo de confiança de 95% para a média da população?

    b) Que tamanho deve ter uma amostra para que o intervalo de confinaça para média 150 $\pm$ 0,98 tenha 95% de confiança?

20. Qual o nível de confiança de cada um dos seguintes intervalos para média?

    a) $\overline{X} \pm \dfrac{1.96 \cdot \sigma}{\sqrt{n}}$

    b) $\overline{X} \pm \dfrac{1.64 \cdot \sigma}{\sqrt{n}}$

21. Em um teste de sensitividade, uma amostra de 18 válvulas resultou em um valor médio de 3,2 uv. Construa um intervalo de confiança para a sensitividade média com 98% de confiança, supondo variância de 4 uv.

22. O diretor do Banco Incerteza pretende analisar o endividamento médio dos clientes que fizeram empréstimo junto ao banco. Para tanto, uma amostra de 20 clientes apresentou média \$ 587,25. O setor bancário trabalha com um desvio padrão de \$ 93,76, considerado aceitável para este banco.

    a) Construa um intervalo de 95% de confiança assumindo que a distribuição dos empréstimos é normal.

    b) Quantos clientes deveriam ser consultados para que, com 90% de confiança, o erro amostral seja reduzido a \$25?

23. Para uma variável W com distribuição t de Student com ν graus de liberdade determine w_α tal que:

    a) $P(|W| \leq w_\alpha) = 95\%$ com $\nu = 10$

    b) $P(|W| \leq w_\alpha) = 90\%$ com $\nu = 15$

    c) $P(W \leq w_\alpha) = 90\%$ com $\nu = 10$

    d) $P(W \geq w_\alpha) = 95\%$ com $\nu = 14$

24. Para uma variável W com distribuição χ^2 com ν graus de liberdade determine w_α tal que:

    a) $P(W \geq w_\alpha) = 90\%$ com $\nu = 10$

    b) $P(W \leq w_\alpha) = 90\%$ com $\nu = 15$

    c) $P(W \geq w_\alpha) = 90\%$ com $\nu = 7$

25. A precipitação pluviométrica anual em uma certa região tem desvio padrão igual a 3.1. Nos últimos 9 anos, observaram-se 28.3 / 31.7 / 29.8 / 30.5 / 34.1 / 27.9 / 35 / 26.9 /30.2

    a) Construa um intervalo para a precipitação média com 98% de confiança.

    b) Utilizando ainda estes dados, construa um intervalo para a precipitação média com 98% de confiança caso o desvio padrão não fosse conhecido.

    c) Os técnicos têm interesse em estudar precipitações superiores a 30. Determine I.C. para a probabilidade deste evento com 98% de confiança.

26. Um engenheiro está avaliando a precisão de uma máquina que produz esferas para rolamentos. Ele colheu uma amostra e analisou o diâmetro das esferas cujos dados estão resumidos a seguir: $n = 20$; Média Amostral = 32 mm ; Desvio Padrão Amostral S= 1,2 mm.

    a) Construa um intervalo de confiança ao nível de 90% para variância populacional.

    b) Construa um intervalo de confiança ao nível de 95% para média populacional.

27. Empregando os dados do exercício 6

    a) Calcule um intervalo de confiança de 95% para a média populacional

    b) Calcule um intervalo de confiança de 95% para a variância

    c) A empresa se preocupa com tempos gastos para montar a superior a 300 s. Determine um intervalo de confiança de 98% para a probabilidade deste evento.

28. Uma amostra de 10 mil itens de um lote de produção foi inspecionada e o número de defeitos por item foi registrado:

    Tab. 4: Defeitos por item.

Defeitos	0	1	2	3	4
Quantidade	6000	3000	600	350	50

    a) Determine um intervalo de confiança para a proporção de itens defeituosos com 98% de confiança.

    b) determine um intervalo de confiança para o número médio de defeitos nos itens com 94% de confiança.

29. Um técnico precisa determinar o tempo médio gasto para perfurar 3 orifícios em uma peça de metal. Qual deve ser o tamanho da amostra para que tenhamos 95% de confiança em que sua média amostral esteja a mesmo de 15 segundos da verdadeira média. Sabe-se por experiência que o desvio padrão é de 40 segundos.

30. Uma amostra preliminar de 500 famílias verificou que 340 delas possuem forno de micro-ondas. Para estimar a proporção de famílias com este utensílio, qual o tamanho de amostra necessário para que tenhamos 95% de confiança em que o erro de nova estimativa não seja superior a 0.02.

31. Qual deve ser o tamanho da amostra para que a diferença entre a média amostral e a média da população em valor absoluto seja no máximo 1 com confiança de 95%. Sabe-se que o desvio padrão é igual a 10.

32. Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis ao seu candidato. Uma amostra piloto de 100 eleitores revelou que 60% deles eram favoráveis ao candidato. Qual o tamanho da amostra tal que o erro cometido seja no máximo 0.01 com 80% de confiança.

33. A resistência à tração de 20 corpos de prova é:

    Tab. 5: Resistência à tração.

131	132	134	135	135	138	139	139	140	142
143	144	144	145	146	147	148	149	150	138

    a) Construa um intervalo de confiança de 95% para a variância.

    b) Construa um intervalo de confiança de 95% para a média.

34. Uma empresa vai lançar um novo produto. O gerente de marketing encomenda uma pesquisa de mercado, entrevistando 500 pessoas na qual apenas 157 pessoas manifestam intenção de comprar o novo produto.

    a) Construa um I.C. com 92% de confiança.

    b) Se soubermos que a real proporção de compradores é inferior a 40%, qual seria o intervalo?

Intervalo de confiança - 2 amostras independentes

35. O tempo de execução de uma certa tarefa foi medida em duas equipes. A equipe A obteve um tempo médio de 10 min, com o um grupo de 50 funcionários. a equipe B obteve um tempo médio de 12 min, com um grupo de 70 funcionários. Admitindo-se que a variância do tempo seja igual a 100 min^2 para as 2 equipes pede-se:

    a) Determine um intervalo de confiança para o tempo médio da equipe A, com 92% de confiança.

    b) Idem para a equipe B.

    c) Determine um intervalo de confiança para a diferença entre os tempos médios das equipes A e B com 92% de confiança. Interprete o resultado.

36. Foi feita uma pesquisa eleitoral no bairro A com 500 eleitores sendo que 100 deles manifestam intenção de votar no candidato X. No bairro B, foram entrevistados 1000 eleitores, e 300 deles manifestaram interesse em votar em X.

    a)Construa I.C. com 95% de confiança para a probabilidade de votar no candidato X, no bairro A.

    b)Construa I.C. com 95% de confiança para a probabilidade de votar no candidato X, no bairro B.

    c) Determine um intervalo de confiança para a diferença entre as proporções de votantes ao candidato X dos bairro A e B com 95% de confiança. Interprete o resultado.

37. Dois processos de conservação de alimentos estão sendo utilizados e a variável de interesse é o tempo de duração. Duas amostras independentes A com 16 latas apresentaram um tempo médio de 50 dias e a amostra B com 25 latas apresentou 60 dias. Construa um intervalo para a diferença das médias com 95% de confiança e interprete os resultados. Use como o desvio padrão do processo igual a 10.

38. Em uma faculdade, duas amostras de 150 alunos (uma com os alunos do curso engenharia ambiental e outra com os alunos de curso de nutrição) foram selecionadas aleatoriamente para responder uma pesquisa sobre consumo de comida vegana. 60 alunos do curso de nutrição e 45 dos alunos do curso da engenharia afirmam que consomem este tipo de alimento. Construa um intervalo de confiança para a diferença das proporções com 95% de confiança.

Testes de hipótese - uma amostra

39. Um estudo pretende identificar se um grupo de indígenas pertence a uma determinada tribo. Sabe-se que os índios da tribo A possuem altura média de 120 cm, enquanto que os da tribo B possuem altura média 145 cm. O desvio padrão nos dois casos é 40. O critério de decisão é o seguinte: se para uma amostra de 100 pessoas for observada média amostral superior a 130 considera-se que o grupo é da tribo B, caso contrário é da tribo A. Você desconfia que os indivíduos são do grupo B

    a) Quais são as hipóteses nula e alternativa?

    b) Qual é o erro tipo I? Determine a probabilidade de erro tipo I (α).

    c) Qual é o erro tipo II? Determine a probabilidade de erro tipo II (β).

    d) Qual deve ser o critério de decisão para que $\alpha = 5\%$?

    e) Utilizando o critério do item anterior, determine a probabilidade do erro do tipo II (β).

40. A carga média de ruptura especificada de um parafuso é de 50 kg, sendo o desvio-padrão dessas cargas igual a 4 kg, suponha que o comprador especifique que:

    - Se o lote satisfaz à especificação, o comprador deseja limitar a 5% a probabilidade de concluir que o lote é insatisfatório;
    - Se o lote tiver uma resistência média ligeiramente menor que 50 kg, tal fato não causa preocupação, porém deseja-se que, se a verdadeira média for inferior a 48 kg, tal fato seja identificado com pelo menos 90% de probabilidade.

a) Quais são os limites da região crítica?

b) Nessas condições, qual o tamanho da amostra mínima necessária tal que atende os dois critérios especificados pelo comprador?

41. Fazendo o seguinte teste de hipótese: $\begin{cases} H_o : \mu = 1150; \sigma = 150 \\ H_1 : \mu = 1200; \sigma = 200 \end{cases}$

Com $n = 100$ obteve-se a seguinte região crítica RC = [1170, ∞] (se $\overline{X} \in RC \to$ H_0 é falsa)

a) Qual a probabilidade α de rejeitar H_0 quando verdadeira?

b) Qual a probabilidade β de aceitar H_0 quando H_1 é verdadeira?

c) Qual deve ser a região crítica para $\alpha = 0.05$?

d) E empregando a região anterior, qual a probabilidade β de aceitar H_0 quando H_1 é verdadeira?

42. Uma máquina enche sacos de meio quilo de pó de café. Sabe-se que o desvio padrão desta máquina é de 30 g. Desconfia-se que a máquina esteja mal calibrada, porém o custo de interromper o processo é por demais elevado. Uma amostra de 35 sacos foi retirada e obteve-se $\overline{x}$ = 506,24 g. Analise os resultados quanto à interrupção do processo ao nível de 5% de significância nos seguintes casos:

a) O produtor preocupa-se apenas com o custo de interrupção do processo.

b) O produtor preocupa-se com a calibração adequada da máquina.

c) O produtor preocupa-se apenas com as reclamações dos clientes.

43. Na ausência de treinamento, os escores de um exame de admissão em um MBA variam normalmente com média 475 e desvio padrão 100. Suponhamos, que o treinamento possa melhorar a média, mas não altere o desvio padrão. Uma equipe treina 100 estudantes. Suas notas acusam $\overline{x} = 478$.

a) Para um nível de significância de 5% é possível afirmar que as notas aumentaram? Determine o valor-p.

b) Qual é a conclusão se a amostra tivesse 1000 e não 100 alunos?

44. Um teste para aceitação de lotes de uma empresa automobilística consiste em avaliar o diâmetro de peças. As peças são consideradas dentro da especificação se sua média for $\mu = 61$u. Um lote de 16 peças é analisado. Admita $\sigma = 5$. Se o critério adotado no exercício anterior é: rejeite o lote quando $\overline{x} > 62.5$ ou $\overline{x} < 57.5$.

a) Qual a probabilidade de erro tipo I?

b) Qual a probabilidade de erro tipo II, caso a real média tenha sido 60u?

45. Uma amostra de 25 elementos extraída de um lote de 5000 peças forneceu:

$\sum x = 500$mm e $\sum x^2 = 10.051,67$mm^2.

Sabe-se ainda que dos 25 elementos, 16 são superiores a 19 mm. Com base nestas amostras, ao nível de 5% de significância, pode-se afirmar que:

a) A média do lote é superior a 19 mm? Forneça o valor-p deste teste.

b) O desvio padrão do lote é superior a 1,5 mm? Forneça o valor-p deste teste.

c) O número de elementos no lote inteiro com medidas superiores a 19 é superior a 3000 elementos?

d) Pode-se utilizar somente esta amostra para estimar a média do lote com desvio máximo da estimativa em relação ao verdadeiro valor de 0,01mm e 90% de confiança?

46. O número de furos em chapas de alumínio obedece a uma distribuição de Poisson com $\lambda = 0,3$ defeitos por cm. Queremos inspecionar uma chapa de comprimento L (cm) de forma a caracterizar o percentual do comprimento que não apresenta furos.

a)Encontre L tal que com 90% de certeza, haja, no máximo, 5 furos.

b) Você pretende comprar um lote de chapas com 18 cm de comprimento. O fabricante afirma que seu produto tem $\lambda = 0,25$ defeitos por cm. Você deve decidir se compra o lote ou não. Para isso, escolhe aleatoriamente 5 chapas do lote e, se duas ou mais chapas apresentarem mais do que 7 furos (cada uma), você rejeita o lote todo. Defina os erros do tipo I e II e calcule o α.

47. Duas empresas diferentes concentraram-se em oferecer serviços de televisão a cabo em uma determinada região. Seja p a proporção de todos os assinantes em potencial que favorecem a primeira empresa com relação à segunda. Considere o teste de H_0: p=0,5 versus H_a: p≠0,5 com base em uma amostra aleatória

de 25 indivíduos. Seja X o número na amostra que favorece a primeira empresa e x o valor observado de X.

a) Qual das regiões de rejeição a seguir é mais apropriada e por quê? $R_1 = \{x : x \leq 7 \text{ ou } x \geq 18\}, R_2 = \{x : x \leq 8\}, R_3 = \{x : x \geq 17\}$

b) No contexto da situação deste problema, descreva quais são os erros do tipo I e tipo II.

c) Qual é a distribuição de probabilidades da estatística do teste X quando H_0 for verdadeira? Use-a para calcular a probabilidade de um erro tipo I.

d) Calcule a probabilidade de um erro tipo II para a região selecionada quando p=0,3, novamente quando p=0,4, e também para p=0,6 e p=0,7.

e) Usando a região selecionada, a que conclusão você chegaria se 6 dos 25 questionados favorecessem a empresa 1?

48. Pretende-se testar a honestidade de uma moeda com base no número de caras obtidas em 15 lançamentos. (Atenção: neste caso não é possível utilizar a distribuição normal. Deve-se recair em uma distribuição binomial!)

a) Qual o teste de hipóteses mais adequado?

b) Qual deve ser a região crítica com $\alpha = 0.1$

49. Um certo fabricante de cartuchos de tinta afirma que estes têm duração segundo uma distribuição normal com média de 45.000 folhas. Uma empresa adquiriu um lote deste produto, retirou uma amostra de 16 cartuchos que forneceu $\bar{x} = 44.175$ folhas e $s = 300$.

a)Você rejeitaria o lote? Considere o nível de significância igual a 5%.

b) Obter o valor-p deste teste.

50. Desconfiando-se de que uma moeda fosse viciada, realizou-se um experimento que consiste em lançar essa moeda cem vezes. Obtivemos 59 caras.

a) Quais sãos as hipóteses nula e alternativa?

b)Pode-se afirmar a existência de vício? Considere o nível de significância igual a 5%.

c) Obter o valor-p do teste.

51. Uma amostra de 10 elementos extraída de uma população normal forneceu uma variância amostral igual a 12,4.

a) Este resultado é suficiente para se concluir, ao nível de significância 5%, que a variância desta população é inferior a 25?

b) Forneça o valor-p deste teste.

52. Um processo produz cabos para uma companhia de telefone. Quando o processo está operando corretamente, o diâmetro do cabo segue uma distribuição normal com média 1,6 cm e desvio padrão 0,5. Uma amostra aleatória de 16 pedaços de cabo apresentou o diâmetro médio de 1.615 cm e desvio padrão amostral de 0,86 cm.

a) Teste ao nível de 10% se a média populacional é 1,6 cm contra a alternativa de que é diferente de 1,6 cm.

b) Teste ao nível de 10% a hipótese nula de o desvio padrão populacional é 0,5 contra a alternativa de que é maior.

c) Forneça os valores-p dos itens a e b.

53. Uma empresa de fertilizantes afirma que em média seu produto aumenta a produção de milho em 8 toneladas por acre. Uma amostra aleatória de 16 observações mostra que em média a produção cresceu 6,8 toneladas por acre com desvio padrão amostral de 2,4.

a) Assumindo que o aumento da produção é normal, teste a afirmação da firma ao nível de 5% de significância.

b) Teste se a variância da produção é inferior a 5 ton^2/acre com 5% de significância.

c) Forneça valores-p dos dois testes.

54. Suponha que o salário inicial de recém formados seja uma variável aleatória com média desconhecida, porém com desvio padrão de \$200. Uma amostra de 36 recém formados forneceu $\overline{x} = \$2.050$.

    a) Faça o teste de hipóteses $\begin{cases} H_0 : \mu = 2100 \\ H_1 : \mu < 2100 \end{cases}$, ao nível de 5%.

    b) Admita que a variância seja desconhecida e a amostra retirada forneceu $s_x = 150$. Qual seria sua conclusão?

    c) Forneça os valores-p dos testes.

55. Uma companhia de cigarros anuncia que 100% dos cigarros que fabrica possuem índice de nicotina abaixo de 23 mg por cigarro. Sabe-se que o índice de nicotina se distribui normalmente nos cigarros. Um laboratório realiza 5 análises desse índice obtendo 24, 21, 25, 26, 22.

    a) Pode-se aceitar, ao nível de 10%, a afirmação do fabricante?

    b) Forneça valor-p deste teste.

56. Amostras de água são coletadas da água usada para resfriamento quando está sendo despejada de uma usina de energia em um rio. Determinou-se que, desde que a temperatura média da água despejada seja de no máximo 150°F, não haverá efeitos negativos sobre o ecossistema do rio. Para investigar se a usina está em conformidade com as regulamentações que proíbem uma temperatura média de água de descarga acima de 150°, 50 amostras de água serão tiradas em tempos selecionados aleatoriamente e a temperatura de cada amostra será registrada. Os dados resultantes serão usados para testar as hipóteses: H0: µ=150°F versus Ha: µ>150°F. No contexto dessa situação, descreva os erros tipo I e tipo II. Que tipo de erro você consideraria mais sério? Explique.

57. Uma amostra de dez empresas, cada uma com 100 funcionários, forneceu o número de empregados do sexo feminino conforme tabela abaixo. Aponte quais empresas não atendem à hipótese: $\begin{cases} H_0 : p = 0.45 \\ H_a : p > 0.45 \end{cases}$ ao nível de nível de 10%.

Tab. 6: Empregados do sexo feminino por empresa.

Empresa	A	B	C	D	E	F	G	H	I	J
Mulheres	51	52	54	51	47	49	51	46	45	58

58. Quando realizam-se tarefas extenuantes, a pulsação de 25 empregados aumenta em média 18,4 batimentos por minuto com desvio padrão de 4,9 batimentos por minuto. Teste a hipótese de que a variância $\sigma^2 = 30$, contra a hipótese alternativa $\sigma^2 < 30$, ao nível de 5%.

59. Você está convencido de que o salário médio de profissionais de uma dada área é alto, porém acredita que haja alta variabilidade entre eles, superior à média nacional $\sigma^2 = 256$.

    a) Qual a hipótese alternativa que você utilizaria para verificar esta afirmação?

    b) Se os salários possuem distribuição normal, qual a estatística adequada a este teste?

    c) Qual a região crítica apropriada se $\alpha = 0,025$ e $n = 31$?

    d) Você aceitaria ou rejeitaria H_0 se a amostra resultasse em $\sigma^2 = 350$?

60. Um pesquisador está analisando a variância do preço de uma ação. A hipótese alternativa é que $\sigma^2 < 900(\$^2)$. Uma amostra de 25 dias forneceu $s_x^2 = 870(\$^2)$. Assuma que o preço da ação possua distribuição normal. O que se pode concluir ao nível de significância de 1%?

61. A companhia telefônica está estudando a duração de chamadas telefônicas, bem como sua variabilidade. Admita que a variabilidade nacional seja $\sigma^2 = 4$ minutos^2. A companhia pretende verificar se em uma certa cidade a variabilidade do tempo das chamadas difere do padrão nacional. A duração das chamadas possui distribuição normal.

    a) Qual a hipótese nula e a hipótese alternativa?

    b) A partir de que valores da variância amostral você rejeitaria a hipótese H_0? ($n = 25$ e $\alpha = 0,05$)

    c) Qual sua decisão se uma amostra de 25 chamadas fornecesse $\sigma^2 = 2,5$?

62. Experiências passadas indicam que o tempo para que alunos veteranos do ensino médio completem um teste padronizado é uma variável aleatória normal, com média de 35 minutos. Se uma amostra aleatória de 20 alunos levou uma média de 33,1 minutos para completar o teste, com um desvio-padrão de 4,3 minutos, teste a hipótese, no nível de significância 0,05, de que $\mu = 35$ minutos contra a alternativa de que $\mu < 35$ minutos.

Testes de hipótese - duas amostras independentes

63. Duas máquinas, A e B, são usadas para empacotar pó de café. A experiência passada garante que o desvio padrão para ambas é de 10g. Porém, suspeita-se que elas têm médias diferentes. Para verificar, sortearam-se duas amostras: uma com 25 pacotes da máquina A e outra com 16 pacotes da máquina B. As médias foram, respectivamente, $\overline{x}_A = 502.74$g e $\overline{x}_B = 496.60$g. Com esses números, e com nível de 5%, qual seria a conclusão do teste $\mathrm{H}_0 : \mu_A = \mu_B$?

64. Os dados abaixo foram coletados em três empresas diferentes no dia 18/05 e referem-se ao tempo (minutos) que profissionais gastaram com pesquisas na internet naquele dia.

Tab. 7: Tempos gasto com pesquisas na internet.

Empresa	1	2	3	4	5	6	7	8	9	10
A	12,2	13	11,5	12,6	10,5	8,4	9,7	8,1	12,3	
B	11,2	23,1	12,4	10,4	12,1	19,3	17,5	11,1	12,4	16,6
C	9,1	9,9	8,7	8,4	8,6	8,3	9,2			

a) Suponha que o desvio padrão da empresa A seja $\sigma_A = 2,3$. Teste a hipótese de que o tempo médio na empresa A é inferior a 11. Forneça valor-p.

b) Refaça o item a admitindo variância desconhecida.

c) Teste a hipótese de que o tempo médio gasto na empresa C é inferior a 9,5 minutos.

d) Teste a hipótese de que a variância do tempo gasto na empresa A é inferior à variância da empresa B.

e) Suponha que o desvio padrão da empresa A seja $\sigma_A = 2,3$ e a da empresa B seja $\sigma_B = 3$. Teste a hipótese de que o tempo médio é maior na empresa B.

f) Refaça o item e assumindo que as variâncias são iguais, porém desconhecidas.

g) Teste a hipótese de que na empresa A, a proporção dos funcionários que gastam mais do que 10 minutos com internet é superior a 50%.

65. Quando trabalhamos com duas amostras aleatórias provenientes de populações normais cujas variâncias parecem ser distintas, utiliza-se a seguinte estatística de teste: $\tilde{t} = \dfrac{(\overline{X_1} - \overline{X_2})}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ cuja distribuição de probabilidade é aproximadamente t de Student com graus de liberdade v, estimados a partir dos valores observados das variâncias amostrais s_1^2 e s_2^2 como: $v = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$.

Utilize esta estatística para resolver o seguinte problema: Para comparar a resistência de pára-choques, seis de cada tipo foram montados sobre um certo tipo de carro compacto. Em seguida cada carro é projetado contra uma parede a 5 milhas por hora, e os seguintes custos de reparo por tipo de para choque são apresentados a seguir:

Tipo 1: 107 148 123 165 102 119

Tipo 2: 134 115 112 151 133 129

Use nível de significância $\alpha = 0,01$ para testar a existência de diferença entre as médias.

66. Para se estudar o desempenho de duas companhias corretoras de ações, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada, computou-se a porcentagem do lucro apresentada durante um período fixado de tempo. Os dados estão a seguir:

Corretora A: 45, 60, 54, 62, 55, 70, 38, 48, 64, 55, 56, 55, 54, 59, 48, 65, 55, 60

Corretora B: 57, 55, 58, 50, 52, 59, 59, 55, 56, 61, 52, 53, 57, 57, 50, 55, 58, 54, 59, 51, 56

a) Para verificar a homogeneidade das duas populações, um estatístico sugeriu que se usasse a razão quociente $F = \dfrac{Var(lucro|A)}{Var(lucro|B)}$ e adotou a seguinte regra de decisão: se $F < 2.12$ então elas possuem a mesma variância. Aplicando este procedimento o que se conclui?

Corretora A: $\sum X_i = 1004; \sum(X_i - \overline{x})^2 = 1001,24$

Corretora B: $\sum X_i = 1164; \sum(X_i - \overline{x})^2 = 201,14$

b) Para decidir se os desempenhos das duas corretoras são iguais ou não, adotou-se o seguinte teste: sejam caso $|t| < 2$ os desempenhos são semelhantes, caso contrário são diferentes. Qual a conclusão?

67. Pretende-se comparar a variância do preço de duas ações (com distribuição de preços normal). A hipótese alternativa de interesse é de que $\sigma_A^2 > \sigma_B^2$. Amostras de tamanho 21 e 25 forneceram variâncias amostrais igual a: $S_A^2 = (67,233)^2$ e $S_B^2 = (37,128)^2$. Teste com $\alpha = 0{,}01$ se as duas variâncias são iguais.

68. Uma companhia de seguros quer lançar um novo produto especificamente para a população de aposentados. Quer escolher entre a região do Nordeste e Sudeste para lançar o produto. Uma pesquisa envolvendo uma amostra de 1000 habitantes de cada região forneceu que as proporções de aposentados são respectivamente $\hat{p}_{NE} = 0,156$ e $\hat{p}_{SE} = 0,10$.

    a) Teste com nível de significância de 5% a hipótese que a proporção de aposentados no sudeste é igual a do nordeste.

    b) Forneça o valor-p deste teste.

69. Uma máquina automática enche latas com base no peso líquido, com variabilidade praticamente constante e independente dos ajustes na média, dada por um desvio padrão de 5g. Duas amostras retiradas em dois períodos consecutivos de 10 e 20 latas forneceram pesos médios líquidos, de respectivamente 184,6 g e 188,9 g. Desconfia-se que a regulagem da máquina quanto ao peso médio possa ter se alterado entre a coleta das 2 amostras. Qual a conclusão no nível de significância 5%? Forneça o valor-p deste teste.

70. Duas técnicas de venda são aplicadas por dois grupos de vendedores: a técnica A por 12 vendedores e a técnica B por 15 vendedores. Espera-se que a técnica B produza melhores resultados. No final do mês obteve-se:

    Tab. 8: Técnica A x Técnica B.

	A	B
n	12	15
$\overline{X}$	68	76
S^2	50	75

    Testar no nível de significância de 5%, se a expectativa quanto à técnica B se confirmou. Suponha vendas com distribuição normal, com variância comum e desconhecida.

71. Duas amostras independentes, respectivamente com 10 e 15 elementos, extraídas de populações normais forneceram variâncias amostrais S^2 igual a 6,34 e 18,7 . Com $\alpha = 5\%$ podemos afirmar que as variâncias são iguais? Forneceá o valor-p do teste.

72. Foram sorteados aleatoriamente dois grupos de nove estudantes de um curso de estatística. Duas técnicas de ensino foram adotadas, uma em cada grupo, e as notas finais são apresentadas a seguir:

    Tab. 9: Notas finais.

	1	2	3	4	5	6	7	8	9
A	76	92	99	87	65	87	86	85	88
B	98	57	85	89	60	53	52	93	95

    a) É possível concluir que as notas foram inferiores com a técnica B ao nível de 5% de significância? (Faça hipóteses apropriadas sobre as variâncias). Fornecer valor-p.

    b) Suponha que as variâncias fossem conhecidas nos dois casos iguais a 25 e 28 para o grupo A e B, respectivamente. Qual sua conclusão? Forneça o valor-p.

73. Queremos verificar se 2 máquinas produzem peças com a mesma homogeneidade e variabilidade quanto à resistência à tração. Os dados são:

A:	145	127	136	142	141	137
B:	143	128	132	138	142	132

a) Verifique se as duas máquinas produzem peças homogêneas quanto à tração

b) Verifique se as duas máquinas produzem peças com a mesma variabilidade quanto à tração (5% de significância para os dois casos).

74. Um determinado vício em redes sociais afeta dois estados brasileiros distintos. Uma amostra probabilística com n_1 entrevistados no primeiro forneceu o seguinte IC para a proporção de pessoas que se declaram serem viciadas na rede social testada: $2/3 \pm 1.96 \times 0.07857$, no segundo estado a pesquisa resultado no IC: $0.4898 \pm 1.96 \times 0.071414$.

a) Determine n_1, n_2 e quantas pessoas disseram ser viciadas na rede social testada em cada um dos estados.

b) Para $\alpha = 5\%$, teste a hipótese de que a proporção no primeiro estado é ao menos 15 pontos percentuais maior do que no segundo estado.

Teste t pareado

75. A abertura comercial e a valorização cambial aumentaram a concorrência dos produtos nacionais com os importados. Com isso muitas empresas nacionais encontram dificuldade para serem competitivas. Uma empresa de produção de autopeças pretende instalar um curso intensivo de treinamento para os seus funcionários com o intuito de melhorar a produtividade e recuperar o espaço perdido para os importados. Foi selecionada uma amostra de 25 funcionários e mediu-se o número de peças feitas por dia para cada um. Então esses mesmos funcionários receberam o treinamento e mediu-se o desempenho de cada um após o treinamento.

Dados: • Antes do treinamento: $\sum X_i = 512$; • Após o treinamento: $\sum X_i = 514,64$ e • Desvio-padrão da amostra das diferenças = 7,2.

a) Teste ao nível de 5% se o treinamento aumentou a produtividade dos empregados.

b) Forneça o valor-p do teste.

76. Comparamos a receita bruta de 10 empresas do setor comercial nos anos de 2015 e 2016 (em milhões). Teste a hipótese H_0: $\mu_{15} = \mu_{16}$ contra H_0: $\mu_{15} \neq \mu_{16}$ ($\alpha = 5\%$). Fornece valor-p do teste.

Empresa	1	2	3	4	5	6	7	8	9	10
Receita em 2015	21,2	57,4	18,4	29,6	12,7	28,4	40,4	30,1	23,6	31,8
Receita em 2016	18,9	54,1	19,3	32,7	faliu	27,6	45,8	29,8	24,9	33,4

77. No mês de dezembro, 9 pessoas foram para um spa. Elas são pesadas no dia de chegada e no dia de saída, obtendo-se o seguinte:

Tab. 10: Pesagens.

	1	2	3	4	5	6	7	8	9
Chegada	76	92	99	87	65	87	86	85	88
Saída	98	57	85	89	60	53	52	93	95

Você pode concluir (5%) que passar o mês no spa reduz o peso de pacientes?

78. Cinco operadores de um certo tipo de máquina são treinados em máquinas de 2 marcas diferentes, A e B. Mediu-se o tempo que cada um deles gastou na realização de uma mesma tarefa, obtendo-se:

Operador	Marca A	Marca B
01	80	75
02	72	70
03	65	60
04	78	72
05	85	78

a) No nível de 10% podemos afirmar que a tarefa realizada na máquina A demora mais que na máquina B?

b) Forneça valor-p deste teste.

Teste de aderência

79. Verifique se que os dados a seguir provém de uma distribuição normal. Use a 5% de significância, ? E a 1%? (utilize o teste de Kolmogorov-Smirnov).

25,7	26,3	25,5	23,9	20,7
25,3	24,7	20,1	22,7	28,7
18,6	24,9	28,4	28,8	22,4
26,1	16,6	20,2	22	21,5
25,3	21,9	23,4	23,1	22,6

80. Verifique se os dados a seguir provém de uma distribuição exponencial. Use 5% de significância. Refaça empregando o teste de Kolmogorov-Smirnov.

38,7	17,9	9,7	4,8	11,7	22,6
86,4	12,8	37,8	22,3	100,5	27,6
18,8	16,2	53,3	116,7	12,1	15,6
7,1	5,9	24,5	8,1	2,2	6,1
6,1	42,7	15,9	27,7	21,6	24,6

81. Um dado lançado 300 vezes tiveram seus resultados coletados e anotados conforme tabela abaixo:

Face	1	2	3	4	5	6	Total
Frequência Observada	43	49	56	45	66	41	300

Com as observações queremos saber se o dado é honesto, use 5% de significância.

82. Em genética dos pássaros uma certa população deve estar classificada em quatro categorias com probabilidades 0,656; 0,093; 0,093; 0,158. As frequências observadas são: 125; 18; 20; 34. Para um nível de significância de 5%, os dados observados aderem à distribuição teórica?

Análise de Variância

83. Um experimento foi conduzido para avaliar o desempenho de sete fertilizantes: T1, T2, T3, T4, T5, T6 e T7. Para isto comprimentos (cm) dos ramos de árvores após um certo tempo fixo foram medidos. Para cada fertilizante três réplicas foram empregadas.

	Réplicas		
Fertilizante	1	2	3
T1	10	12	8
T2	12	13	8
T3	12	11	7
T4	13	13	16
T5	13	17	15
T6	17	15	13
T7	18	16	14

Faça uma análise de variância e teste de hipótese de igualdade dos tratamento a 5% de significância e a 1% de significância.

84. Os salários médios (em mil unidades monetárias) de mecânicos de automóveis foram coletados por um sindicato. A finalidade era determinar a eventual existência igualdade entre 4 localidades. Obteve-se:

Loc. A	Loc. B	Loc. C	Loc. D
6	12	11	9
9	11	8	7
9	10	12	10
6	8	9	10
5	9	10	9

No nível de significância 5% podemos afirmar que existe igualdade de salário médio entre as 4 localidades?

85. Um consumidor pretende verificar se não existe diferença no preço de aspirinas em locais de cidades e em diferentes tipos de loja. Os resultados coletados foram os seguintes:

Tipo de lojas	Centro	Oeste	Leste	Sul
Drogaria	2,46	2,85	2,44	2,51
Farmácia	2,27	2,61	2,35	2,17
Supermercado	2,72	2,64	2,59	2,84

a) Os dois fatores não afetam o preço de aspirinas? (use 5%)

86. Em uma pesquisa de rendimento por hora, entre assalariados segundo o grau de instrução, obtiveram-se os seguintes resultados:

Escolaridade	$\sum x_i$	$\sum x^2$	n
1° Grau	111,50	259,93	50
2° Grau	71,00	258,89	20
3° Grau	84,30	717,94	10

Monte a tabela de análise de variância e verifique se os rendimentos por hora são iguais entre diferentes escolaridades. (use 5%)

87. Duas pessoas realizaram uma mesma tarefa, duas vezes, em dois dias diferentes na semana. Os tempos (segundos) obtidos estão apresentados a seguir:

Nome	2ª feira	3ª feira
Maria	471 413	385 434
João	637 612	770 705

a) Monte a tabela de análise de variância;

b) (use $\alpha = 5\%$), verifique se os tempos são iguais entre diferentes pessoas; E entre dias da semana? Existe interação entre dias da semana e pessoa?.

c) Faça análise dos resíduos.

88. Quatros amostras de três marcas de molho (A, B e C) estão sendo avaliadas por 4 laboratórios quanto ao seu teor de gordura. Os resultados obtidos foram:

Marca	Lab 1	Lab 2	Lab 3	Lab 4
A	3,7	2,8	3,1	3,4
B	3,1	2,6	2,7	3,0
C	3,5	3,4	3,0	3,3

a) Monte a tabela de análise de variância;

b) (use $\alpha = 5\%$), verifique se os teores médios de gordura são iguais entre marcas? E entre os laboratórios?

89. As vendas de 3 vendedoras que atendem uma certa região, durante 8 semanas, são fornecidas a seguir.

	Semana							
Vendedora	1	2	3	4	5	6	7	8
Rosa	186	222	198	216	210	194	203	219
Sueli	197	203	194	208	220	209	190	205
Tânia	174	213	190	197	233	206	199	221

a) Monte a tabela de análise de variância;

b) (use $\alpha = 5\%$), verifique se as vendas médias semanais são iguais entre semanas? E entre as vendedores?

90. Os números de peças defeituosas produzidas por quatro operadores trabalhando, em turnos, em três diferentes máquinas são dadas a seguir:

	Operador			
Máquina	A	B	C	D
1	35	38	41	32
2	31	40	38	31
3	36	35	43	25

Existem evidências de que os operadores ou máquinas não estejam produzindo com a mesma qualidade (a = 5%)?

91. Um supermercado deseja investigar se a posição onde o produto é colocado influencia a sua venda. Posições do corredor (frente ou fundo) e alturas da prateleira (alta ou baixa) foram avaliadas. Os resultados obtidos, em R$, foram:

	Prateleira	
Corredor	Alta	Baixa
Frente	86	70
	72	60
Fundo	60	28
	46	22

a) Monte a tabela de análise de variância;

b) (use $\alpha = 5\%$), verifique se a posição do corredor influencia na venda? E a altura da prateleira? Existe interação entre corredor e prateleria?

c) Efetue a análise dos resíduos.

Análise de Regressão

92. A tabela mostra o valor de aluguel e a idade dos automóveis

Idade (x)	10	13	5	7	20
Alguel (y)	4	3	6	5	2

a) Fazer um gráfico de dispersão entre x e y.

b) Ajuste um modelo de regressão linear. Interprete o significado prático dos coeficientes do modelo de regressão.

c) (use $\alpha = 5\%$). Teste a significância do coeficiente angular do modelo.

d) Efetua a análise dos resíduos.

93. Considere os dados abaixo onde y representa o número de produtos vendidos e x o valor (mil reais) aplicado em publicidade:

#	1	2	3	4	5	6	7	8
x	50	200	90	170	60	150	80	120
y	10	85	30	70	25	66	40	50

Dados: $\sum x_i = 920; \sum y_i = 376; \sum x_i y_i = 52700; \sum x_i^2 = 126400; \sum y_i^2 = 22206.$

a) Faça um diagrama de dispersão. Existe suspeita de uma relação linear?

b) Estime os coeficientes do modelo linear. Dê interpretações práticas sobre os coeficientes.

c) Com nível de significância 5% teste a significância da regressão? Podemos falar que as vendas aumentam devido à publicidade?

d) Forneça um IC (nível de confiança de 95%) para o coeficiente angular. Idem para o coeficiente linear.

e) Se temos 65 mil para investir, qual a previsão média de vendas? Forneça IC (nível de confiança de 95%) para a previsão média de vendas no caso de investimr 65 mil. E como fica IC (nível de confiança de 95%) para uma observação futura para este mesmo valor?

f) Faça análise dos resíduos e comente se as suposições feitas do modelo linear de regressão simples são válidas ou não.

94. Os modelos lineares são bastante utilizados para análise do comportamento de duas variáveis. Assim pode-se, sob hipóteses não muito restritivas, encontrar estimativas β_0 e β_1 , dos parâmetros da reta de regressão, $y = \beta_0 + \beta_1 x$. Podemos realizar testes de hipóteses sobre estes parâmetros. Para o coeficiente angular β_1, é importante testar se a variável Y permanece inalterada quando X varia, ou seja, se $H_0 : \beta_1 = 0$. Para tal emprega-se a seguinte estatística $t = \dfrac{\hat{\beta}_1}{\hat{\sigma}^2_{\beta_1}}$ que possui distribuição t-Student com $n-2$ graus de liberdade, n é o número de pares de pontos utilizado para construir a reta de regressão, onde $\hat{\sigma}^2_{\beta_1} = \dfrac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ e $\hat{\sigma}^2 = \dfrac{\text{SQResíduos}}{n-2}$.
Considere o seguinte conjunto de dados:

X	10	12	18	15	20	18	19	22	20	21	19	23
Y	62	65	72	70	81	77	72	77	75	90	82	95

a) Obtenha a reta de regressão

b) Construe a tabela de ANOVA referente ao modelo de regressão

c) (use $\alpha = 5\%$) para testar a $H_0 : \beta_1 = 0$ empregando a estatística t. Forneça o valor-p deste teste.
Dado: $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 168.917$

95. Considere um experimento em que se analisa a octanagem da gasolina (Y) em função da adição de um novo aditivo (X). Para isso, foram realizados ensaios com os percentuais de 1, 2, 3, 4, 5 e 6% de aditivo. Os resultados são mostrados na a seguir:

X	1	2	3	4	5	6
Y	80,5	81,6	82,1	83,7	83,9	85,0

a) Obtenha a reta de regressão

b) Construe a tabela de ANOVA referente ao modelo de regressão e (use $\alpha = 5\%$) para testar a $H_0 : \beta_1 = 0$ no modelo linear $y = \beta_0 + \beta_1 x$.

c) Faça a análise dos resíduos. As suposições do modelo de regressão estão satisfeitas?

96. Dados foram coletados sobre o número de imóveis construídos (Y) em função das taxas de juros (X) de financiamentos para o setor. Obteve-se a seguinte tabela:

Ano	00	01	02	03	04	05	06	07	08	09
Taxa de Juros (%)	8,95	9,68	11,15	13,95	16,52	15,79	13,43	13,80	12,28	10,07
Imóveis (1000)	1987	2020	1745	1292	1084	1062	1703	1750	1742	1805

a) Determine a reta de regressão (Y em função de X)

b) É possível concluir, ao nível de 5%, que o número de imóveis construídos decresce à medida que a taxa de juros aumenta?

c) Encontre um intervalo de confiança para a quantidade média de imóveis construídos quando a taxa de juros é de 10,0. Como você interpreta este intervalo?

d) Suponha que o modelo empregado fosse do tipo $y = \alpha \exp(\beta x)\phi$, onde ϕ é um erro aleatório, α e β, os parâmetros. Quais são as hipóteses que devem ser feitas sobre ϕ para que pudéssemos empregar o ferramental de regressão linear para construir um intervalo de confiança para β?

e) Estimate α e β considerando o modelo apresentado no item d.

97. Os dados abaixo se referem a medidas de raios cósmicos realizadas em diversas altitudes.

Altitude (X)	Dose (Y)
50	28; 27; 35
450	30; 36
780	32; 34
1.200	35; 36
4.440	51
4.800	55,4; 58; 56,2
5.300	69,3; 71

a) Faça um diagrama de dispersão.

b) Ajuste um modelo linear.

c) É possível concluir que os raios cósmicos são afetados pela altitude?

d) Faça a análise dos resíduos. Você acha que este modelo é o mais adequado para analisar o problema proposto? Justifique.

98. O faturamento anual de uma empresa de consultoria nos últimos anos é apresentado a seguir:

Ano	1	2	3	4	5	6
(R$)	112	149	238	354	580	867

a) Ajuste um modelo de regressão linear simples. Interprete os coeficientes.

b) Ajuste um modelo regressão polinomial de seguno grau.

c) Teste se a inclusão do segundo de grau no modelo de regressão é estatísticamente melhor. Use ($\alpha =$ 5%)?

d) Faça análise dos resíduos do modelo final.

99. Os dados a seguir se referem ao número de quartos (x_1) e ao número de banheiros (x_2) de oito casas e seu respectivo preço de venda recente (Y).

Casa	Quartos	Banheiros	Preço
1	3	2	788.000
2	2	1	743.000
3	4	3	838.000
4	2	1	742.000
5	3	2	797.000
6	2	2	749.000
7	5	3	884.000
8	4	2	829.000

a) Ajuste inicialmente um modelo de regressão simples escolhendo a variável auxiliar que mais se correlaciona com a variável resposta.

b) Ajuste um modelo de regressão com a inclusão da segunda variável. Interprete o significado prático dos coeficientes de cada variável auxiliar deste modelo.

c) Verifique se a inclusão da segunda variável é estatísticamente relevante.

d) Faça a análise dos resíduos do modelo final.

100. Uma empresa de vendas por catálogo está avaliando seu processo de distribuição. Deseja saber o que afeta o custo de distribuição. Dados dos últimos 24 meses foram obtidos do custo de distribuição, vendas e o número de pedidos processados, conforme a seguir:

Mês	Custo Distribuição	Vendas	Pedidos	Mês	Custo Distribuição	Vendas	Pedidos
1	52,95	386	4015	13	62,98	372	3977
2	71,66	446	3806	14	72,3	328	4428
3	85,58	512	5309	15	58,99	408	3964
4	63,69	401	4262	16	79,38	491	4582
5	72,81	457	4269	17	94,44	527	5582
6	68,44	458	4097	18	59,74	444	3450
7	52,46	301	3213	19	90,5	623	5079
8	70,77	484	4809	20	93,24	596	5735
9	82,03	517	5237	21	69,33	463	4269
10	74,39	503	4732	22	53,71	389	3708
11	70,84	535	4413	23	89,18	547	5387
12	54,08	353	2921	24	66,8	415	4161

a) Ajuste inicialmente um modelo de regressão simples escolhendo a variável auxiliar que mais se correlaciona com a variável resposta.

b) Ajuste um modelo de regressão com a inclusão da segunda variável. Interprete o significado prático dos coeficientes de cada variável auxiliar deste modelo.

c) Verifique se a inclusão da segunda variável é estatísticamente relevante.

d) Faça a análise dos resíduos do modelo final.

Respostas

1)

- NP - Amostrou apenas a população dos seus colegas o que não representa uma amostra probabilística da população.
- NP - Amostrou apenas os conhecidos o que não representa uma amostra probabilística da população.
- P - Adotou a estratégia de amostragem sistemática a qual é probabilística.
- P - Dado que houve a ordenação aleatória antes da seleção, ao selecionar os 20 primeiros gera um efeito igual a uma amostragem casual simples com 20 amostras.
- P - Semelhante ao item anterior, o efeito da amostragem é o mesmo da casual simples com 20 amostras.

---

2)

- B - A amostragem foi feito periodicamente a cada 15 posições da lista de vendedores.
- D - Dado que a população é divida em subdivisões (cada loja é uma subdivisão), ao sortear (casual simples) os conglomerados e selecionar seus elementos (vendedores) para a amostra é uma estratégia de amostragem por conglomerados.
- C - Diferentemente da estratégia por conglomerados onde os conglomerados são sorteados, na amostragem estratificada é retirado elementos de todos os estratos, a quantidade de quantos serão retirados pode variar dependendo do tipo de amostragem estratificada (uniforme, proporcional e ótima).
- A - Todos os vendedores tem a mesma probabilidade de pertencer a amostra.

---

3)
a) As categóricas são: sexo, raça e ocupação.
b) As variáveis quantitativas são idade, que deve ser medida em “anos”, e salário que deve ser medido em “R$/mês”.

---

4)
Média = 5,40
Mediana = 5,40
Desvio padrão = 0,64

---

5)
a)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 41,33$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = 20,607$$

b) os outliers são 85 e 86.
Após a remoção: grande impacto na média e desvio padrão

$$\bar{x} = 34,54$$

$$s = 10,96$$

c) pouco impacto após a remoção
antes da remoção:

$$Q_1 = 27; Q_2 = 34; Q_3 = 50$$

após a remoção:

$$Q_1 = 26; Q_2 = 33; Q_3 = 40.1$$

---

6)
d) Resumindo numa tabela de distribuição de frequências:

Classe	Frequência
300 - 700	13
700 - 1100	11
1100 - 1500	6
1500 - 1900	5
1900 - 2300	3
2300 - 2700	1
2700 - 3100	1

f)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = 1110$$

$$Md = Li_{Md} + \frac{h}{n_i} * [\frac{n}{2} - N_{i-1}] = 700 + \frac{400}{11} * [\frac{40}{2} - 13] = 954,54$$

$$Mo = Li_{Mo} + \frac{d_1 * h}{d_1 + d_2} = 300 + \frac{400 * 13}{15} = 646,66$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i - \frac{(\sum_{i=1}^{2} x_i f_i)^2}{\sum_{i=1}^{n} f_i}}{\sum_{i=1}^{n} f_i - 1}} = 627,5$$

---

7)

---

8)

---

9)

---

10)

---

11)

---

12)

---

13)
a) Todos os estimadores são não viciado, pois $E(\hat{\mu}_i) = \mu, = 1, 2, 3, 4, 5$
b) O melhor é $\hat{\mu}_5$ pois tem a menor variância.
$Var(\hat{\mu}_1) = \sigma^2; Var(\hat{\mu}_2) = \frac{\sigma^2}{2}; Var(\hat{\mu}_3) = \frac{5\sigma^2}{9}; Var(\hat{\mu}_4) = \frac{30\sigma^2}{100}; Var(\hat{\mu}_5) = \frac{\sigma^2}{5};$

---

14)
Se $x = \{1, 4, 6\}$; $\hat{\theta} =$ Azul;
Se $x = \{2, 3, 7, 8\}$; $\hat{\theta} =$ Verde;
Se $x = \{5\}$; $\hat{\theta} =$ Azul ou Vermelho;
Se $x = \{9, 10, 11, 12\}$; $\hat{\theta} =$ Vermelho;
Se $x = \{13, 14, 15, 16, 17, 18, 19, 20\}$; $\hat{\theta} =$ Branco.

---

15) a)

$$100 \times (\text{n°de trutas marcadas}) = N \times (\text{n°de trutas retiradas})$$

Assim:

$$100 \times 20 = N * 200$$

$$N = 1000$$

b)

$$P(X = 20) = \binom{200}{20} * 0.1^{20} * 0.9^{180} = 9,36$$

c) Em se tratando de n ensaios de Bernoulli, o estimador (p) de máxima verossimilhança para π é calculado como:

$$p = \frac{\sum_{i=1}^{n} x_i}{n}$$

Como π = 100/N; o valor calculado igualando o estimador ao valor do parâmetro populacional fica:

$$\frac{100}{N} = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$\frac{100}{N} = \frac{20}{200}$$
$$N = 1000$$

---

16)

- Para $\gamma = 90\%, z_{\alpha/2} = 1,645$ e $a = 1,645\frac{\sigma}{\sqrt{n}}$
- Para $\gamma = 95\%, z_{\alpha/2} = 1,96$ e $a = 1,96\frac{\sigma}{\sqrt{n}}$
- Para $\gamma = 99\%, z_{\alpha/2} = 2,576$ e $a = 2,576\frac{\sigma}{\sqrt{n}}$

---

17)

- Para $\gamma = 95\%, z_{\alpha/2} = 1,96$ e $IC = [6,37; 9,63]$
- Para $\gamma = 90\%, z_{\alpha/2} = 1,645$ e $IC = [6,63; 9,37]$
- Para $\gamma = 99\%, z_{\alpha/2} = 2,576$ e $IC = [5,85; 10,15]$

---

18)

- I. IC: $10 \pm 2,61$
- II. IC: $10 \pm 7,84$
- III. IC: $10 \pm 2,45$
- IV. IC: $10 \pm 0,98$
- V. IC: $10 \pm 4,12$
- VI. IC: $10 \pm 3,14$

---

19)
a)

- $n = 36$; $\bar{x} = 150$; $\sigma = 5$ $\gamma = 95\%(\alpha = 5\%)$; $z_{\alpha/2} = 1,96$
- IC: [148,37;151,63]

b) $e_o = 0,98, \gamma = 95\%, \sigma = 5$:

$$n = \left(\frac{z_{\alpha/2} * \sigma}{e_o}\right)^2 = \left(\frac{1,96 * 5}{0,98}\right)^2 = 100$$

---

20)
a) $z_{\alpha/2} = 1,96 \implies \alpha/2 = 2,5\% \implies \gamma = 95\%$
b) $z_{\alpha/2} = 1,64 \implies \alpha/2 = 5\% \implies \gamma = 90\%$

---

21)

- n = 18; $\sigma = 2$; $\bar{x} = 3,2$; $\gamma = 98\%(\alpha = 2\%)$; $z_{\alpha/2} = 2,33$ —-> IC: [2,102;4,298]

22) a)

- $n = 20$; $\bar{x} = 587,25$; $\sigma = 93,76$; $\gamma = 95\%(\alpha = 5\%)$ $\Longrightarrow$ $z_{\alpha/2} = 1,96$ $\Longrightarrow$ IC: [587,25±41,09]

b)

$$n = \left(\frac{z_{\alpha/2} * \sigma}{e_o}\right)^2 = \left(\frac{1,64 * 93,76}{25}\right)^2 = 37,83 \Longrightarrow \textit{entrevistar 38 pessoas}$$

23) a) $t_{5\%,10} = 1,812 \Longrightarrow W_\alpha = 1,812 * s/\sqrt{(}11)$
b)$t_{5\%,15} = 1,753 \Longrightarrow W_\alpha = 1,753 * s/\sqrt{(}16)$
c)$t_{10\%,15} = 1,372 \Longrightarrow W_\alpha = 1,372 * s/\sqrt{(}11)$

24) conferir
$\chi^2_{5\%,10} = 18,307 \Longrightarrow W_\alpha = 18,307 * D.P.(\bar{x})$
$\chi^2_{5\%,15} = 24,996 \Longrightarrow W_\alpha = 24,996 * D.P.(\bar{x})$
$\chi^2_{10\%,10} = 15,987 \Longrightarrow W_\alpha = 15,987 * D.P.(\bar{x})$

25)
a) Considerando os dados fornecidos para σ conhecido:

- n = 9; $\sigma = 3,1$; $\bar{x} = 30,489$; $\gamma = 98\%(\alpha = 2\%)$; $z_{\alpha/2} = 2,33$ —-> IC: [28,081;32,897]

b) Considerando os dados fornecidos para σ desconhecido:

- n = 9; $s = 2,734$; $\bar{x} = 30,489$; $\gamma = 98\%(\alpha = 2\%)$; $t_{\alpha/2;8} = 2,896$ —–> IC: [27,850;33,128]

c) Para os dados em questão, temos p = 5/9, 1-p = 4/9. Supondo ser possível aproximar pela normal (mesmo que n(1-p) < 5):

- n = 9; $p = 5/9$; $1 - p = 4/9$; $\sqrt{\frac{p(1-p)}{n}} = 0,166$; $\gamma = 98\%(\alpha = 2\%)$; $z_{\alpha/2} = 2,33$ ——> IC: $[0,556 \pm 0,386]$

26) a)

- N = 20; ν = N-1 = 19; $\bar{x} = 32$; $s = 1,2 \Longrightarrow s^2 = 1,44$; $\gamma = 90\% \Longrightarrow \alpha = 10\%$; χ^2_{sup} = 30,144; χ^2_{inf} = 10,117;

Temos:

$$P(\frac{(n-1)s^2}{\chi^2_{sup}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{inf}}) = \gamma = 1 - \alpha = P(\frac{(19) * 1,44}{30,144} \leq \sigma^2 \leq \frac{(19) * 1,44}{10,117}) = \gamma = 90\%$$

$$P(0,908 \leq \sigma^2 \leq 2,704) = 90\%$$

27)

28) a)

- p = 0,4; 1-p = 0,6; $\sqrt{\frac{p(1-p)}{n}} = 0,0049$; $z_{\alpha/2} = 2,33$ —–> IC: $[0,4 \pm 0,0114]$

29) Para dimensionar a amostra, com $e_o = 15, \gamma = 95\%, \sigma = 40$:

$$n = \left(\frac{z_{\alpha/2} * \sigma}{e_o}\right)^2 = \left(\frac{1,96 * 40}{15}\right)^2 = 27,32 \Longrightarrow \textit{amostra de } 28.$$

30)

- $p_{est} = \frac{340}{500}$ = 0,68; 1-p_{est} = 0,32; $e_o = 0,02$; $z_{\alpha/2} = 1,96$.

Assim:

$$n = \left(\frac{z_{\alpha/2}}{e_o}\right)^2 \quad p_{est} * (1 - p_{est}) = \left(\frac{1,96}{0,02}\right)^2 \quad 0,68 * (0,32) = 2089,83 \implies Entrevistar\ 2090\ pessoas.$$

---

31) Para dimensionar a amostra, com $e_o = 1, \gamma = 95\%, \sigma = 10$:

$$n = \left(\frac{z_{\alpha/2} * \sigma}{e_o}\right)^2 = \left(\frac{1,96 * 10}{1}\right)^2 = 384,16 \implies amostra\ de\ 385.$$

---

32) Considerando a proporção estimada a partir dos dados:

- $p_{est} = \frac{60}{100}$ = 0,6; 1-p_{est} = 0,4; $e_o = 0,01$; $z_{\alpha/2} = 1,28$.

Assim:

$$n = \left(\frac{z_\alpha/2}{e_o}\right)^2 \quad p_{est} * (1 - p_{est}) = \left(\frac{1,28}{0,01}\right)^2 \quad 0,6 * (0,4) = 3932.16 \implies Entrevistar\ 3933\ pessoas.$$

---

33) a)

- $s^2 = 32,787$; n = 2; ν = n-1 = 19; α = 5%; χ^2_{sup} = 32,85; χ^2_{inf} = 8,907;

Temos:

$$\left(\frac{(n-1)s^2}{\chi^2_{sup}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{inf}}\right) = \left(\frac{(19) * 32,787}{32,852} \leq \sigma^2 \leq \frac{(19) * 32,787}{8,907}\right) \implies (18,962 \leq \sigma^2 \leq 69,940)$$

---

34) Considerando a proporção de erros a partir dos dados:
a)

- p = $\frac{157}{500}$ = 0,314; 1-p = 0,686; $\sqrt{\frac{p(1-p)}{n}} = 0,0207$; $z_{\alpha/2} = 1,75$ —> IC: $[0,314 \pm 0,036]$

.

b)Caso fosse conhecido o intervalo no qual π se encontra, utilizamos $\sqrt{\frac{\pi(1-\pi)}{n}}$ em seu pior caso possível (mais próximo, em módulo, de 50%), ou seja, 0,4:

- p = $\frac{157}{500}$ = 0,314; 1-p = 0,686; $\sqrt{\frac{\pi(1-\pi)}{n}} = 0,022$; $z_{\alpha/2} = 1,75$ —–> IC: $[0,314 \pm 0,0385]$

.

---

35)
a) Para a equipe A:

- n = 50; $\sigma = 10$; $\bar{x} = 10$; $\gamma = 92\%(\alpha = 8\%)$; $z_{\alpha/2} = 1,75$ —–> IC: $[10 \pm 2,475]$

b) Para a equipe B:

- n = 70; $\sigma = 10$; $\bar{x} = 12$; $\gamma = 92\%(\alpha = 8\%)$; $z_{\alpha/2} = 1,75$ ——> IC: $[12 \pm 2,092]$

c) Para a diferença $(\mu_1 - \mu_2)$ com variâncias conhecidas:

- $n_1 = 50$; $n_2 = 70$; $\sigma_1 = 10$; $\sigma_2 = 10$; $D.P.(\bar{x_1} - \bar{x_2}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ = 1,852; $\bar{x_1} = 10$; $\bar{x_2} = 12$; $\bar{x_1} - \bar{x_2} = -2$; $\gamma = 92\%(\alpha = 8\%)$; $z_{\alpha/2} = 1,75$ —-> IC: $[-2 \pm 3,241]$

---

36) Considerando a proporção de erros a partir dos dados:
a)

- p = $\frac{100}{500}$ = 0,2; 1-p = 0,8; $\sqrt{\frac{p(1-p)}{n}} = 0,0179$; $z_{\alpha/2} = 1,96$ ——> IC: $[0,2 \pm 0,035]$

.

37) Para a diferença $(\mu_1 - \mu_2)$ com variâncias conhecidas:

- $n_1 = 16$; $n_2 = 25$; $\sigma_1 = 100$; $\sigma_2 = 100$; $D.P.(\bar{x_1} - \bar{x_2}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 19{,}149$; $\bar{x_1} - \bar{x_2} = -10$; $\gamma = 95\%(\alpha = 5\%) \implies z_{\alpha/2} = 1,96$ —–> IC: $[-10 \pm 37,531]$

---

38)

- $p_1 = \frac{45}{150} = 0,3$; $p_2 = \frac{35}{150} = 0,23333$; $1 - p_1 = 0,7$; $1 - p_2 = 0,76666$; DP $= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0{,}3(0{,}7)}{150} + \frac{0{,}23333(0{,}766667)}{150}} = 0{,}0509$; $z_{\frac{\alpha}{2}} = 1,96$;

I.C.: $[0,06667 \pm 0,0998]$= $[-0.0331 \ ; \ 0.166]$

---

39) a) Como há a desconfiança que os indivíduos são do grupo B temos:

$$H_0 : \mu = 145; H_1 : \mu < 145$$

O erro do tipo 1 é quando rejeitamos a hipótese nula sendo essa verdadeira. Neste casos, temos que seria rejeitarmos os indivíduos serem do grupo B sendo esses do grupo B.

b)

$$\bar{x}_1 = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 130 = 145 - z_\alpha \frac{40}{\sqrt{100}} \implies z_\alpha = 3,75$$

Com isso chegamos em:

$$\alpha = 0,0088\%$$

c)

$$\bar{x}_1 = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 145 - 1,64 \frac{40}{\sqrt{100}} = 138,44$$

---

40)

Sobre a afirmação a) temos que a média do lote é igual à média esperada, logo:$H_0 : \mu_0 = 50$

Nesse caso, dado que a hipótese nula é verdadeira, temos que a probabilidade de rejeitá-la sendo ela verdadeira deve se limitar a 5%, ou seja, temos que o erro do tipo 1 é deve ser limitado a 5%.

Sobre a afirmação b) temos que a média do lote é ligeiramente inferior à média esperada, logo: $H_1 : \mu_0 < 50$

Nesse caso temos que a hipótese nula é falsa, pois a média do lote é menor da esperada. Dessa forma, tal afirmativa trata sobre o erro do tipo 2 (o qual é a probabilidade de aceitar a hipótese nula sendo essa falsa). Nesse sentido, se a média do lote for inferior a 48 kg a chance de rejeitar a hipótese nula deve ser de no mínimo 90%, ou seja, o erro do tipo 2 para essa situação é de no máximo 10%.

Com essas duas informações temos que:

$$\alpha < 5\%; \beta_{\mu_1 = 48kg} < 10\%,$$

a) uma regra decisão tal que o erro do tipo 1 deva ser limitado a 5% é dada por:

$$\bar{x} = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 50 - 1,64 \frac{4}{\sqrt{35}} = 48,89$$

ou seja rejeitar H_0 se $\bar{x} < 48.89$

b) para determinar o tamanho da amostra tal que o erro do tipo 1 deve ser limitado a 5% quanto o erro do tipo 2 a 10%, podemos escrever:

$$P(\bar{x} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} | \mu_0 = 50) = 0.05; P(\bar{x} > \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} | \mu_1 = 48) = 0.10;$$

Após manipulações algébricas, vale a igualdade:

$$\mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_1 + z_\beta \frac{\sigma}{\sqrt{n}} \implies n = \left(\frac{(z_\alpha + z_\beta)\sigma}{\mu_0 - \mu_1} \right)^2$$

com

$$z_{5\%} = 1,64; z_{10\%} = 1,28; \sigma = 4 \implies n = 34,1$$

Logo temos que n = 35.

41) a)

$$P\left(Z > \frac{1170 - 1150}{150/10}\right) = P(Z > 4/3)$$

b)

$$P\left(Z < \frac{1170 - 12000}{200/10}\right) = P(Z < -1.5)$$

c)

$$P\left(Z > \frac{a - 1150}{150/10}\right) = 0.05 \implies a = 1150 + 1.64 \times 1.5$$

d)

$$P\left(Z < \frac{1150 + 1.64 \times 1.5 - 1200}{200/10}\right)$$

42)
Dados: desvio padrão conhecido $\sigma = 30; n = 35; \alpha = 5\%$,
a) Neste caso, as hipóteses nula e alternativa de interesse são:

$$H_0 : \mu = 500; H_1 : \mu > 500$$

Critério: Se $\bar{X} > 500 + 1.64 \times \frac{30}{\sqrt{35}} = 508.32$ rejeita a hipótese nula. A média das amostrada de 506,24 está fora da regrião crítica, não rejeitamos a hipótese nula.
b) Neste caso, as hipóteses nula e alternativa de interesse são:

$$H_0 : \mu = 500; H_1 : \mu \neq 500$$

Como temos um teste bilateral, temos que calcular dois limites para a região crítica, sendo esses:

$$\bar{x}_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{x}_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

com

$$z_{2,5\%} = 1,96; \bar{x}_1 = 490,1; \bar{x}_2 = 509,88$$

Rejeitaríamos a hipótese nula caso:

$$\bar{x} < \bar{x}_1 = 490,1 \text{ ou } \bar{x} > \bar{x}_2 = 509,8$$

A média das amostrada de 506,24 está fora da regrião crítica, não rejeitamos a hipótese nula. .
c) Neste caso, as hipóteses nula e alternativa de interesse são:

$$H_0 : \mu = 500; H_1 : \mu < 500$$

Critério: Se $\bar{X} < 500 - 1.64 \times \frac{30}{\sqrt{35}} = 491.68$ rejeita a hipótese nula. A média das amostrada de 506,24 está fora da regrião crítica, não rejeitamos a hipótese nula.

43)
a)

$$H_0 : \mu = 475; H_1 : \mu > 475$$

$$z_\alpha = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{478 - 475}{100/\sqrt{100}} = 0,3 < z_{5\%} = 1,64$$

Dessa forma não podemos rejeitar a hipótese nula, logo não podemos afirmar que as notas subiram.
b)

$$z_\alpha = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{478 - 475}{100/\sqrt{1000}} = 0,948 < z_{5\%} = 1,64$$

De novo continuamos sem poder rejeitar a hipótese nula, logo não podemos afirmar que as médias subiram.

44) a) Perceba que os limites da região crítica não são simétricos em relação à média esperada. Logo precisamos calcular a probabilidade do erro do tipo 1 para cada um dos extremos:

$$z_{\alpha_1} = \frac{U - \mu_0}{\sigma/\sqrt{n}} = \frac{62,5 - 61}{5/\sqrt{16}} = 1,2 \implies \alpha_1 = 11,5\%$$

$$z_{\alpha_2} = \frac{L - \mu_0}{\sigma/\sqrt{n}} = \frac{57,5 - 61}{5/\sqrt{16}} = -2,8 \implies \alpha_2 = 0,25\%$$

Com isso, temos que o erro do tipo 1 será: $\alpha = \alpha_1 + \alpha_2 = 11,75\%$

b) Sendo o erro do tipo 2 a probabilidade de aceitar a hipótese nula sendo essa falsa ou seja, vamos calcular

$$\beta = P(57.5 < \bar{X} < 62.5|\mu_1 = 60) = P\left(\frac{57.5 - 60}{5/\sqrt{16}} < Z < \frac{62.5 - 60}{5/\sqrt{16}}\right) = P(-2 < z < 2) = 0.9545$$

---

45) Dados: $n = 25; \bar{x} = \frac{\sum x}{n} = 20; s_x^2 = \frac{\sum x^2 - (\sum x)^2/n}{n-1} = 2,152 \implies s_x = 1,46;$

a) Nesta primeira parte trata de um teste de hipótese de média com desvio padrão desconhecido, temos:

$$H_0 : \mu = 19; H_1 : \mu > 19$$

$$t_{obs} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{20 - 19}{1,46/\sqrt{25}} = 3,4;$$

Como $t_{obs} > t_{(24,5\%)} = 1,711$ podemos rejeitar a hipótese nula, ou seja, a média é maior que 19.

b) Para o teste de desvio padrão temos:

$$H_0 : \sigma = 1,5; H_1 : \sigma > 1,5$$

Podemos calcular a estatística:

$$\chi^2_{n-1} = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{24 \times 2,152}{1,5^2} = 22,96 < \chi^2_{(24,5\%)} = 36,415$$

$$\chi^2_{24-5\%} = 36,415$$

$$\chi^2_{24} < \chi^2_{24-5\%}$$

Com isso não podemos rejeitar a hipótese nula, ou seja, o desvio padrão é igual a 1,5.

c) Agora temos que aplicar um teste de hipótese sobre a proporção nas amostras, com isso temos:

$$H_0 : p = \frac{3000}{5000} = 0,6$$

$$H_1 : p > 0,6$$

$$p\,' = 16/25 = 0,560.64$$

Com isso podemos calcular a estatística:

$$z = \frac{p\,' - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,64 - 0,6}{\sqrt{0,6(1-0,6)/25}} = 0,408 < z_{5\%} = 1,64$$

Com isso, não podemos rejeitar a hipótese nula, ou seja, o número de elementos no lote inteiro com medidas superiores a 19 não é superior a 3000 elementos.

d) Se tratando agora de um problema de estimação, o que queremos é analisar se a média estimada tem desvio máximo de 0,01 em relação ao verdadeiro valor é de 90% de confiança. Nesse sentido sabemos que:

$$e_0 = t_{n-1;\alpha/2}\frac{s_x}{\sqrt{n}}$$

$$t_{n-1;\alpha/2} = t_{24,5\%} = 1,711$$

$$e_0 = 1,711\frac{1,46}{\sqrt{25}} = 0,499$$

Vimos, portanto, que o desvio é maior que 0,01.

---

46) a) Com uma barra de tamanho L, esperamos:

$$\lambda_{ef} = \lambda * L = 0,3L$$

Queremos L tal que, no limite:

$$P(X > 5) = 10\%$$

Da distribuição de Poisson, temos que:

$$P(X > a) = 1 - \sum_{x=0}^{a} \frac{\lambda_{ef}^{x} \quad e^{-\lambda_{ef}}}{x!}$$

Para a = 5, $\lambda = 0,3L$ e P = 10%:

$$P(X > 5) = 1 - \sum_{x=0}^{5} \frac{(0,3L)^{x} \quad e^{(-0,3L)}}{x!} = 10\%$$

Substituindo e fazendo de maneira iterativa (usando *Excel* ou algum outro *software*), chegamos a (com precisão de 0,1): L = 10,5 cm

b)

H_0: λ_{ef} = 4,5cm ; H_1: λ_{ef} > 4,5cm

Com o critério de que se 2, 3, 4 ou 5 chapas apresentarem 7 ou mais furos, rejeitamos H_0.

A probabilidade de apresentar 7 ou mais furos em cada chapa é dada por 1 - P(X $\leq$ 6) - 0,1689 $\Longrightarrow$ probabilidade de não apresentar = 0,8311

A probabilidade de 2, 3, 4 ou 5 apresentarem será:

1 - P(0 chapas apresentarem 7 ou mais furos) - P(exatamente 1 apresentar 7 ou mais furos):

$$\alpha = P(E) = 1 - 0,8311^5 - C_{5,4} * (0,1689) * (0,8311) = 20\%$$

Concluindo, o erro do tipo 1 é rejeitarmos a hipótese nula sendo essa verdadeira (rejeitar um lote bom) e o erro do tipo 2 é aceitarmos a hipótese nula sendo essa falsa (aceitar um lote ruim).

---

47)

a)

Se tratando de um teste de hipótese bilateral, a melhore região de rejeição é a R1.

b)

A hipótese nula: há uma equidade entre as proporções da pessoas que favorecem a empresa A e a empresa B. Desse modo, o erro do tipo 1 é rejeitar essa hipótese, ou seja, decidirmos pela não equidade mas sendo a hipótese verdadeira (a equidade). Já o erro do tipo 2 é o caso decidimos pela equidade entre as proporções mas de fato a equidade é falsa.

c)

Sendo a hipótese nula verdadeira temos que a distribuição exata de X tem uma distribuição binominal que pode ser aproximada pela normal sob algumas condições. Tomando o intervalo de decisão R1 temos que os seus limites, em termo de proporções temos que:

$$\bar{p}_1 = \frac{7}{25} = 0,28$$

$$\bar{p}_2 = \frac{18}{25} = 0,72$$

Como se trata de um teste bilateral simétrico, precisamos calcular apenas a probabilidade do erro tipo 1 de um dos lados da distribuição, desse modo temos:

$$z_{\alpha/2} = \frac{\bar{p}_2 - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,72 - 0,5}{\sqrt{0,5(1-0,5)/25}} = 2,2$$

$$\alpha/2 = 1,39\% \Longrightarrow \alpha = 2,78\%$$

d)

Para cada caso, temos de calcular a probabilidade de que o valor encontrado na amostra caia na região de aceitação de H_0, dado que a distribuição real é aquela da hipótese alternativa. Assim sendo, temos de calcular a chance de que, para cada distribuição de p dada, o número se encontre entre $\frac{7}{25}$ e $\frac{18}{25}$. D.P.=$\sqrt{\frac{p(1-p)}{n}}$

- Com p=0,3; é a chance de estar entre -0,21 e 4,58; o erro seria de 58,71%.
- Com p=0,4; é a chance de estar entre -1,22 e 3,27; o erro seria de 88,83%.
- Com p=0,6; é a chance de estar entre -3,27 e 1,22; o erro seria de 88,83%.
- Com p=0,7;é a chance de estar entre -4,58 e 0,22; o erro seria de 58,71%.

e) Como $6 < 7$, rejeitamos a hipótese nula.

---

48) a)
Para amostras pequenas não podemos aproximar a distribuição binomial para uma normal, dessa forma devemos aplicar testes estatísticos que seguem a distribuição binomial que é o caso do teste de proporção bilateral.

---

49)

$$H_0 : \mu = 45000$$
$$H_1 : \mu < 45000$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{44175 - 45000}{300/\sqrt{16}} = -11$$

Como os valores de $t_{crit,+} = 2,131; t_{crit,-} = -2,131$, temos argumentos para rejeitar o lote .

---

50)
a)Estamos tratando de um teste de proporção populacional. Com isso temos que:

$$H_0 : p = 0,5$$
$$H_1 : p \neq 0,5$$
$$p\,' = \frac{59}{100} = 0,59$$

b)Podemos calcular a estatística:

$$z = \frac{p\,' - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,59 - 0,5}{\sqrt{0,5(1-0,5)/100}} = 1,8$$

Se tratando de um teste bilateral tem-se:

$$z_{2,5\%} = 1,96$$

Chegamos em:

$$z < z_{2,5\%}$$

Com isso não podemos rejeitar a hipótese nula, ou seja, a moeda não pode ser considerada viciada. Se utilizarmos um teste unicaudal, aí $z_{5\%} = 1,48$ e rejeitaríamos a hipótese nula.

---

51)
Estamos tratando de um teste de variância populacional, com isso temos:

$$H_0 : \sigma^2 = 25$$

$$H_1 : \sigma^2 < 25$$

Se tratando de um teste que utilizaremos a distribuição qui-quadrado vamos poder rejeitar a hipótese nula caso:

$$\chi^2_{n-1} < \chi^2_{n-1;1-\alpha}$$

Começando por calcular a estatística:

$$\chi^2_{n-1} = \chi^2_9 = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{9 \times 12,4}{25} = 4,464$$

$$\chi_{9;95\%} = 3,325$$

Chegamos em:

$$\chi^2_9 > \chi^2_{9;95\%}$$

Com isso não podemos rejeitar a hipótese nula, ou seja, não concluímos que a variância dessa população seja inferior a 25.

---

52)

a) Trata-se de um teste de hipótese de média com desvio padrão desconhecido, dessa forma, devemos usar a distribuição t-Student.

$$H_0 : \mu = 1,6$$

$$H_1 : \mu \neq 1,6$$

$$\alpha = 10\%$$

$$\bar{x} = 1,615$$

$$s_x = 0,86$$

Podemos calcular a estatística:

$$t_{n-1} = t_{15} = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{1,615 - 1,6}{0,86/\sqrt{16}} = 0,0697$$

$$t_{n-1;\alpha/2} = t_{15;5\%} = 1,753$$

Temos, portanto:

$$t_{15} < t_{15;5\%}$$

Com isso, não podemos rejeitar a hipótese nula, ou seja, a média populacional pode ser considera 1,6.

b) Para o desvio padrão, o teste segue de forma similar que o teste para variância, desse modo temos:

$$H_0 : \sigma^2 = 0,25$$

$$H_1 : \sigma^2 > 0,25$$

$$\alpha = 10\%$$

$$s_x = 0,86$$

Dessa forma, podemos calcular a estatística:

$$\chi^2_{n-1} = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{(15)0,86^2}{0,5^2} = 44,376$$

Para esse teste, devemos comparar o valor da estatística calculada com o valor tabelado de:

$$\chi^2_{n-1;\alpha} = \chi^2_{15;10\%} = 22,307$$

Chegamos, portanto, em:

$$\chi^2_{15} > \chi^2_{15;10\%} = 22,307$$

Com isso, podemos rejeitar a hipótese nula, ou seja, o desvio padrão da amostra não pode ser considerado de 0,5.

---

53) a)

Trata-se de um teste de hipótese de média com desvio padrão desconhecido, dessa forma, devemos usar a distribuição t-Student.

$$H_0 : \mu = 8$$

$$H_1 : \mu < 8$$

$$\alpha = 5\%$$

$$\bar{x} = 6,8$$

$$s_x = 2,4$$

Podemos calcular a estatística:

$$t_{n-1} = t_{15} = \frac{\mu_0 - \bar{x}}{s_x/\sqrt{n}} = \frac{8 - 6,8}{2,4/\sqrt{16}} = 2$$

$$t_{n-1;\alpha} = t_{15;5\%} = 1,753$$

Temos, portanto:

$$t_{15} > t_{15;5\%}$$

Com isso, podemos rejeitar a hipótese nula. Ou seja, temos que o aumento do produto estipulado pela empresa não se verifica com significância de 5%.

---

54) a) Trata-se de um teste de hipótese de média com desvio padrão conhecido:

$$H_0 : \mu = 2100$$

$$H_1 : \mu < 2100$$

$$\alpha = 5\%$$

$$\bar{x} = 2050$$

$$\sigma = 200$$

Podemos calcular a estatística:

$$z_{calc} = \frac{\mu_0 - \bar{x}}{\sigma/\sqrt{n}} = \frac{2100 - 2050}{200/\sqrt{36}} = -1,5$$

$$z_\alpha = -1,645$$

Temos, portanto, em módulo:

$$z_{calc} < z_{crit}$$

Com isso, não podemos rejeitar a hipótese nula. Ou seja, temos que as amostras advém de uma população com média 2100.

b) Se trata do mesmo exercício anterior, porém com desvio padrão desconhecido, temos então:

$$H_0 : \mu = 2100$$

$$H_1 : \mu < 2100$$

$$\alpha = 5\%$$

$$\bar{x} = 2050$$

$$s_x = 150$$

Podemos calcular a estatística:

$$t_{calc} = \frac{\mu_0 - \bar{x}}{s_x/\sqrt{n}} = \frac{2100 - 2050}{150/\sqrt{36}} = -2$$

$$t_\alpha = -2.03$$

Temos, portanto:

$$t_{calc} > t_{crit}$$

Com isso, não podemos rejeitar a hipótese nula. Ou seja, temos que as amostras advém de uma população com média igual a 2100.

---

55)

Temos um teste de hipótese de média da população sem desvio padrão populacional conhecido:

$$H_0 : \mu \leq 23$$

$$H_1 : \mu > 23$$

$$\alpha = 10\%$$

Com as informações:

- $\bar{x} = 23,6$; $D.P. = \frac{s}{\sqrt{n}} = \frac{2,0736}{\sqrt{5}} = 0,9274$; $t_{calc} = \frac{x-\mu}{D.P.} = \frac{23,6-23}{0,9274} = 0,647$; $t_{crit} = t_{4,10\%} = 3,747$

Como $t_{calc} < t_{crit}$, não rejeito H_0. Perceba que a amostra é muito pequena, o que dificulta rejeitar H_0.

56)

As hipóteses nulas e alternativas são: $H_0 : \mu = 150; H_1 : \mu > 150$

O erro do tipo 1 é a probabilidade de rejeitar a hipótese nula sendo essa verdadeira, nesse caso, seria a probabilidade de decidir que a temperatura da água despejada estar superior a 150F mas na realidade a temperatura da água despejada estar a 150F.

Já o erro do tipo 2 é a probabilidade de aceitar a hipótese nula sendo essa falsa, nesse caso, seria a probabilidade de decidir que a temperatura da água despejada estar a 150F mas na realidade a temperatura da água despejada estar superior a 150F.

Nesse tipo de situação em que a temperatura mais alta do que a estipulada pode causar muitos danos, o erro do tipo 2 se torna mais importante. Nesses casos, é melhor ter um modelo para detectar sempre quando a temperatura for maior mesmo que cometa mais erros do tipo 1. O erro do tipo 1 não causa danos tão sérios quanto o do 2 nessa situação.

57) Temos um teste de hipótese da proporção da população com:

$$H_0 : p = 0,45$$
$$H_1 : p > 0,45$$
$$\alpha = 10\%$$

Precisamos achar um critério de decisão dado por:

$$p_c = 0.45 + 1.28 \times \frac{\sqrt{0.45 \times 0.55}}{10} = 0.51.$$

As empresas B, C e J não atendem ao quesito de ter $p_0 = 0.45$

58)

Trata-se de um teste de variância populacional, portanto iremos utilizar a distribuição chi-quadrado.

$$H_0 : \sigma^2 = 30$$
$$H_1 : \sigma^2 < 30$$
$$\alpha = 5\%$$
$$s_x^2 = 4,9^2 = 24,01$$

Podemos calcular a estatística:

$$\chi^2_{n-1} = \chi^2_{24} = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{(25-1)24,01}{30} = 19,208$$

Agora podemos comparar com o valor tabelado de:

$$\chi^2_{n-1;1-\alpha} = \chi_{24;95\%} = 13,848$$

Chegamos em:

$$\chi^2_{24} > \chi_{24;95\%}$$

Com isso, não podemos rejeitar a hipótese nula, ou seja, a variância da população é de 30.

59)

a) A hipótese alternativa seria que o desvio padrão fosse maior que a média nacional de 256.

b) Sabemos que quando fazemos teste de hipótese para variância nós utilizamos a distribuição chi-quadrado.

c) Temos a estatística tabelada:

$$\chi^2_{n-1;\alpha} = \chi_{30;2,5\%} = 46,979$$

Com isso, podemos calcular o limite de decisão com:

$$s_x^2 = \frac{\sigma_0^2 \times \chi_{30;2,5\%}}{n-1} = \frac{256 \times 46,979}{31-1} = 400,88$$

d) Como o valor de 350 está dentro do intervalo de decisão, não poderíamos rejeitar a hipótese nula com esse valor observado.

60)
Trate-se de um teste de hipótese para variância. Temos:

$$H_0 : \sigma^2 = 900$$
$$H_1 : \sigma^2 < 900$$
$$\alpha = 1\%$$
$$s_x^2 = 870$$

Podemos calcular a estatística:

$$\chi^2_{n-1} = \chi^2_{24} = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{(25-1)870}{900} = 23,2$$

Agora podemos comparar com o valor tabelado de:

$$\chi^2_{n-1;1-\alpha} = \chi_{24;99\%} = 10,856$$

Chegamos em:

$$\chi^2_{24} > \chi_{24;99\%}$$

Dessa forma, não podemos rejeitar a hipótese nula, ou seja, a variância da população é de 900.

---

61)
a) Trata-se de um teste de hipótese de variância populacional. Temos que:

$$H_0 : \sigma^2 = 4$$
$$H_1 : \sigma^2 \neq 4$$

b)
Rejeitaremos a hipótese nula se:

$$\chi^2_{n-1} < \chi_{n-1;1-\alpha/2}$$

ou

$$\chi^2_{n-1} > \chi_{n-1;\alpha/2}$$

Das tabelas temos que:

$$\chi_{n-1;1-\alpha/2} = \chi_{24;95\%} = 13,848$$
$$\chi_{n-1;\alpha/2} = \chi_{24;5\%} = 36,415$$

Além disso, sabemos que:

$$\chi^2_{n-1} = \chi^2_{24} = \frac{(n-1)s_x^2}{\sigma_0^2}$$

Dessa forma temos as seguintes inequações

$$\chi^2_{n-1} < \chi_{n-1;1-\alpha/2} \implies \frac{(n-1)s_x^2}{\sigma_0^2} < 13,848 \implies s_x^2 < \frac{\sigma_0^2 * 13,848}{n-1} = 9,232$$

$$\chi^2_{n-1} > \chi_{n-1;\alpha/2} \implies \frac{(n-1)s_x^2}{\sigma_0^2} > 36,415 \implies s_x^2 > \frac{\sigma_0^2 * 36,415}{n-1} = 24,276$$

Portanto temos que se o valor da variância amostral for menor que 9,232 ou maior que 24,276 rejeitamos a hipótese nula.
c) Como 2,5 é menor que 9,232 temos a hipótese nula rejeitada.

---

62)
Trata-se de um teste de hipótese de média com desvio padrão amostral. Temos:

$$H_0 : \mu = 35$$
$$H_1 : \mu < 35$$
$$\alpha = 5\%$$
$$\bar{x} = 33,1$$

$$s_x = 4,3$$

Calculamos a estatística:

$$t = \frac{\mu - \bar{x}}{s_x/\sqrt{n}} = \frac{35 - 33,1}{4,3/\sqrt{20}} = 1,98$$

Ao buscar na tabela temos:

$$t_{n-1;\alpha} = t_{19;5\%} = 1,729$$

Como o valor calculado é maior que o valor tabelado podemos rejeitar a hipótese nula.

---

63)

$$H_0 : \mu_A = \mu_B \implies \mu_A - \mu_B = 0 = \Delta$$

$$H_1 : \mu_A - \mu_B \neq 0$$

Podemos calcular então a seguinte estatística:

$$x = \frac{(\bar{x}_A - \bar{x}_B) - \Delta}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = \frac{(502.74 - 496.60) - 0}{\sqrt{10^2/25 + 10^2/16}} = 1,9178$$

Lembre-se que estamos tratando de um teste bilateral:

$$z_{\alpha/2} = z_{2,5\%} = 1,96$$

$$z < z_{2,5\%}$$

Com isso não podemos rejeitar a hipótese nula.

---

64) a) Para o primeiro item temos um teste de hipótese de uma média populacional. Temos:

$$H_0 : \mu = 11$$

$$H_1 : \mu < 11$$

Calculando a média amostral chegamos em:

$$\bar{x} = 10,92$$

Considerando o teste de hipótese com um nível de 5%. Podemos calcular o limite de decisão com:

$$\bar{x}_c = \mu - z_{5\%}\frac{\sigma}{\sqrt{n}} = 11 - 1,65\frac{2,3}{\sqrt{9}} = 9.735$$

Como a média amostral é maior que o limite decisão, temos que não rejeitamos a hipótese nula para 5% de significância.

b) Para variância desconhecida precisamos, primeiro calcular a variância amostral:

$$s_x^2 = 3,37$$

$$t_{n-1;\alpha} = t_{8;0,05} = 1,86$$

Com isso podemos calcular o limite de decisão:

$$\bar{x}_c = \mu - t_{n-1;\alpha}\frac{s_x}{\sqrt{n}} = 11 - 1,86\frac{3,37}{\sqrt{9}} = 8.92$$

Novamente temos que o valor da média amostral é superior ao limite de decisão calculado, dessa forma, temos que não rejeitamos a hipótese nula para 5% de significância.

c) Temos um teste de hipótese de variância de duas populações. Temos portanto:

$$H_0 : \sigma_B^2 = \sigma_A^2$$

$$H_1 : \sigma_B^2 > \sigma_A^2$$

Primeiro temos que calcular as variâncias amostrais:

$$s_A^2 = 3,37$$

$$s_B^2 = 18,23$$

Dessa forma, podemos calcular a estatística:

$$\frac{s_B^2}{s_A^2} = F_{n_1-1;n_2-1} = F_{9;8} = \frac{18,23}{3,37} = 5,41$$

Rejeitaremos a hipótese nula se:

$$F_{9;8} = 5,41 > F_{9;8;\alpha}$$

Se considerarmos o α como 0,05 temos:

$$F_{9;8;0,05} = 3,18$$

Desse modo, rejeitamos a hipótese nula para essa significância. Logo, existem evidências que a variância de B é maior que a de A.

d) Trata-se de um teste de hipótese de média para duas populações com desvio padrão conhecido. Temos:

$$H_0 : \mu_B - \mu_A = \Delta$$

$$H_1 : \mu_B - \mu_A > \Delta$$

Como queremos saber se uma é maior que a outra, temos que o Δ vai ser igual a 0.

Calculando as médias amostrais:

$$\bar{x}_A = 10,92$$

$$\bar{x}_B = 14,61$$

Com isso podemos calcular a seguinte estatística:

$$z = \frac{(\bar{x}_B - \bar{x}_A) - \Delta}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = \frac{(14,61 - 10,92) - 0}{\sqrt{2,3^2/9 + 3^2/10}} = 2,93$$

Rejeitamos a hipótese nula se:

$$z > z_\alpha$$

Com α igual a 0,05 temos:

$$z_{0,05} = 1,65$$

Assim podemos rejeitar a hipótese nula. Ou seja, existem evidências de que o tempo médio de B é maior do que a de A.

e) Temos o mesmo teste de hipótese do exercício anterior, porém assumindo variâncias iguais mas desconhecidas. Dessa forma calculamos a variância amostral comum às duas populaçõescomo:

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{(9-1)10,92 + (10-1)14,61}{9 + 10 - 2} = 12,87$$

Com isso calculamos a estatística:

$$t_{n_A+n_B-2} = \frac{(\bar{x}_B - \bar{x}_A) - \Delta}{s_p\sqrt{1/n_A + 1/n_B}} = \frac{(14,61 - 10,92) - 0}{\sqrt{12,87}\sqrt{1/9 + 1/10}} = 2,24$$

Para $\alpha = 0{,}05$

$$t_{n_A+n_B-2;\alpha} = t_{9+10-2;0,05} = 1,740$$

Com isso temos:

$$t_{17} > t_{17;0,05}$$

Desse modo, rejeitamos a hipótese nula.

f) Trata-se de um teste de hipótese de proporção.

$$H_0 : p = 0,5$$

$$H_1 : p > 0,5$$

Ao calcular o valor de p_{obs}:

$$p_{obs} = \frac{6}{9} = 0,67$$

Podemos calcular a estatística:

$$z = \frac{p_{obs} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0,67 - 0,5}{\sqrt{0,67(1-0,67)/9}} = 1,084$$

Para α igual a 0,05 temos:

$$z_{5\%} = 1,65$$

Dessa forma:

$$z_{5\%} > z$$

Com isso não podemos rejeitar a hipótese nula.
Atenção para este item: Em função do tamanho da amostra, a aproximação da distribuição binomial pela normal pode não ser adequada. A solução mais correta é considerar a distribuição exata de p_{obs}, que neste caso é uma distribuição binomial e recalcular o critério de decisão.
g) Trata-se de um teste de hipótese de média com desvio padrão desconhecido.

$$H_0 : \mu = 9,5$$

$$H_1 : \mu < 9,5$$

Calculando a média amostral e o desvio padrão amostral:

$$\bar{x} = 8,88$$

$$s_x = 0,55$$

Podemos calcular a estatística:

$$t_6 = \frac{(\mu - \bar{x})}{s_x/\sqrt{n}} = \frac{(9,5 - 8,88)}{0,55/\sqrt{7}} = 2,98$$

Considerando um *alpha* de 0,05, temos:

$$t_{6;0,05} = 1,943$$

Dessa forma:

$$t_{6;0,05} < t_6$$

Portanto, não podemos rejeitar a hipótese nula, ou seja, aceitamos que o tempo médio é igual a 9,5.

---

65) Trata-se de um teste de hipótese de duas populações que tenham desvios-padrão diferentes e desconhecidos. Começamos por calcular as médias e variâncias amostrais:

$$\bar{x}_1 = 127,3$$
$$\bar{x}_2 = 129$$
$$s_1^2 = 597,9$$
$$s_2^2 = 202$$

Temos que o teste de hipótese é dado por:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Com isso, calculamos a estatística:

$$t = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{(129 - 127,3)}{\sqrt{597,9/6 + 202/6}} = 0,147$$

E também, podemos calcular o número de graus de liberdade pela fórmula dada na lista:

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 8,032$$

Portanto podemos considerar 8 graus de liberdade. Buscamos na tabela o valor de:

$$t_{8;0,5\%} = 3,355$$

Como o valor tabelado é maior que o calculado, não rejeitamos a hipótese nula.

---

66)
a) Começamos por calcular as variâncias amostrais:

$$s_A^2 = 58,92$$
$$s_B^2 = 10,06$$

Temos:

$$F = \frac{s_A^2}{s_B^2} = 5,86$$

Portanto, temos que as corretores não possuem a mesma variância.
b) Começamos por calcular as médias amostrais:

$$\bar{x}_A = 55,72$$
$$\bar{x}_B = 55,25$$

Pelo item anteriores, temos que as variâncias são desconhecidas e diferentes, podemos calcular a estatística empregada no exercício 65:

$$t = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{s_A^2/n_A + s_B^2/n_B}} = \frac{(55,72 - 55,25)}{\sqrt{58,92/18 + 10,06/21}} = 0,242$$

Pelo critério dado no enunciado, podemos dizer que os desempenhos são semelhantes.

---

67)
Trata-se de um teste de hipótese de variância de duas populações com amostras de tamanhos distintos. Temos:

$$H_0 : \sigma_A^2 = \sigma_B^2$$
$$H_1 : \sigma_A^2 > \sigma_B^2$$
$$\alpha = 0,01 = 1\%$$
$$n_A = 21$$
$$n_B = 25$$

Começamos por calcular a seguinte estatística:

$$F_{n_A-1;n_B-1} = F_{20;24} = \frac{S_A^2}{S_B^2} = \frac{67,233^2}{37,128^2} = 3,279$$

Para o valor tabelado:

$$F_{20;24;1\%} = 2,74$$

Rejeitaremos a hipótese nula se:

$$F_{20;24} > F_{20;24;1\%}$$

O que de fato acontece, logo podemos rejeitar a hipótese nula.

---

68)
Trata-se de um teste de hipótese para comparar duas proporções, temos:

$$H_0 : p_1 = p_2 \implies p_1 - p_2 = \Delta_p = 0$$
$$H_1 : p_1 \neq p_2 \implies p_1 - p_2 \neq 0$$
$$\alpha = 0,05 = 5\%$$
$$n_1 = n_2 = 1000$$
$$\hat{p}_1 = 0,156$$
$$\hat{p}_2 = 0,1$$

Podemos calcular a estatística sob hipótese nula:

$$D.P.(\hat{p}_1 - \hat{p}_1) = \sqrt{p'(1-p')\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},\ p' = \frac{x_1 + x_2}{n_1 + n_2} = \frac{156 + 100}{2000} = 0,128$$

$$\Longrightarrow \sqrt{p'(1-p')\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{0,128 \times 0.872 \times \left(\frac{1}{500}\right)} = 0,015$$

$$z_{obs} = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta_p}{D.P.(\hat{p}_1 - \hat{p}_2)} = \frac{0,056}{0,015} = 3.733$$

Temos que o valor tabelado é de:

$$z_{\alpha/2} = z_{2.5\%} = 1,96$$

Podemos rejeitar a hipótese nula se:

$$z > z_{2.5\%}$$

No caso, podemos rejeitar a hipótese nula.

---

69)
Trata-se de um teste de hipótese de comparação de duas médias. Temos:

$$H_0 : \mu_1 - \mu_2 = \Delta = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$
$$\alpha = 5\%$$
$$\bar{x}_1 = 184,6$$
$$\bar{x}_2 = 188,9$$
$$\sigma_1^2 = \sigma_2^2 = \sigma^2 = 25$$
$$n_1 = 10$$
$$n_2 = 20$$

Podemos calcular a estatística:

$$z = \frac{(\bar{x}_2 - \bar{x}_1) - \Delta}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}} = 2,22$$

Temos o valor tabelado de:

$$z_{\alpha/2} = z_{2,5\%} = 1,64$$

Podemos rejeitar a hipótese nula se:

$$z > z_{2.5\%}$$

Tal fato ocorre neste caso, dessa forma, podemos rejeitar a hipótese nula de que as médias das populações são iguais.

---

70)
Trata-se de um teste de hipótese de duas populações, tem-se:

$$H_0 : \mu_B - \mu_A = \Delta = 0$$
$$H_1 : \mu_B - \mu_A > 0$$
$$\alpha = 5\%$$
$$\bar{x}_A = 68$$
$$\bar{x}_B = 76$$
$$S_A^2 = 50$$
$$S_B^2 = 75$$
$$n_A = 12$$
$$n_B = 15$$

Assumindo variâncias iguais, porém desconhecidas, temos que calcular uma estimativa de variância comum às duas populações:

$$s_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} = 64$$

Com isso podemos calcular a estatística:

$$t = \frac{(\bar{x}_B - \bar{x}_A) - \Delta}{\sqrt{s_p^2(1/n_A + 1/n_B)}} = 2,582$$

Podemos comparar com o valor tabelado:

$$t_{n_A+n_B-2;\alpha} = t_{25;5\%} = 1,708$$

Como o valor calculado é maior que o tabelado podemos rejeitar a hipótese nula, ou seja, a técnica B aumentou os resultados.

---

71)

Trata-se de um teste de hipótese para comparar variância, temos:

$$H_0 : \sigma_A^2 = \sigma_B^2$$
$$H_1 : \sigma_A^2 \neq \sigma_B^2$$
$$\alpha = 5\%$$
$$n_A = 10$$
$$n_B = 15$$

Começamos por calcular a seguinte estatística:

$$F_{n_B-1;n_A-1} = F_{14;9} = \frac{S_B^2}{S_A^2} = \frac{18,7}{6,34} = 2,949$$

Para o valor tabelado:

$$F_{14;9;5\%} = 3.03$$

Rejeitaremos a hipótese nula se:

$$F_{14;9} > F_{14;9;5\%}$$

Não podemos rejeitar a hipótese nula.

---

72)

a) Trata-se de um teste de hipótese de duas populações que devemos determinar sobre as variâncias. Temos inicialmente:

$$S_a = 93 \ ; \ S_b = 388,19$$

$$F_{calc} = 4,174 \text{ (Maior/Menor)}$$

Como $F_{crit} = 3,44$, rejeitamos que as variâncias são iguais. Assim:

$$H_0 : \mu_A - \mu_B = \Delta = 0$$
$$H_1 : \mu_A - \mu_B > 0$$
$$\alpha = 5\%$$
$$\bar{x}_A = 85$$
$$\bar{x}_B = 75,8$$
$$s_A^2 = 93$$
$$s_B^2 = 388,2$$
$$n_A = 9$$
$$n_B = 9$$

Como a variância é diferente e desconhecida, temos que calcular:

$$D.P. = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = 7,312$$

$$T_{calc} = \frac{(x_a - x_b) - \Delta}{D.P.} = 1,26$$

O número de graus de liberdade é dado por uma aproximação que resulta em:

$$\frac{((\frac{s_1^2}{n_1}) + (\frac{s_2^2}{n_2}))^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}; \text{ arredondar pra baixo!}$$

Resultando em 11:
Podemos comparar com o valor tabelado:

$$t_{\nu;\alpha} = t_{11;5\%} = 1,796$$

Como o valor tabelado é maior que o calculado temos que não podemos rejeitar a hipótese nula. Portanto, não há diferença estatisticamente significativas entre as notas.
b) Com as variâncias conhecidas temos que calcular a estatística:
Podemos calcular a estatística:

$$z = \frac{(\bar{x}_A - \bar{x}_B) - \Delta}{\sqrt{\sigma^A/n_A + \sigma^B/n_B}} = 3,79$$

Comparando com o valor tabelado:

$$z_{5\%} = 1,64$$

Como o valor calculado é maior que o tabelado, podemos rejeitar a hipótese nula.

---

73) b) As hipóteses:

- H_0: $\sigma_a^2 = \sigma_b^2$; H_1: $\sigma_a^2 \neq \sigma_b^2$

Temos como fazer os cálculos:

- $s_1^2 = 40$; $s_2^2 = 36,97$; $f_{calc} = \frac{s_1^2}{s_2^2} = 1,082$; $f_{crit} = F_{5,5,5\%} = 5,05$

Assim, como $f_{calc} < f_{crit}$, não se rejeita a hipótese nula.

---

74)
a)
Dados os intervalos de confiança, torna-se fácil descobrir os valores de n_1 e n_2:
$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} = 0,07857 = \sqrt{\frac{0,66667(0,33333)}{n_1}} \implies n_1 = 36.$
Ademais, $\frac{2\times 36}{3} = 24$, assim, 24 pessoas responderam ser viciadas e 12 responderam não ser.
Similarmente $\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.071414 = \sqrt{\frac{0,4898(0,5102)}{n_2}} \implies n_2 = 49.$
Ademais, $0,4898 \times 49 = 24$, assim, 24 pessoas responderam ser viciadas e 25 responderam não ser.
b)
Temos um teste de hipóteses para a diferença de duas proporções, com valor de $\pi_1 - \pi_2$ diferente de 0:

$$H_0 : p_1 - p_2 = 0,15; H_1 : p_1 - p_2 > 0,15$$

A estatística do teste:

$$Z_{calc} = \frac{(\hat{p}_1 - \hat{p}_2) - 0.15}{\sqrt{\hat{p}_1 - \hat{p}_2}}$$

$$\sqrt{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0,66667(0,33333)}{36} + \frac{0,4898(0,5102)}{49}} = 0,1062$$

Assim:

$$Z_{calc} == \frac{0,02687}{0,1062} = 0,25$$

O valor de $z_{crit,5\%}$, para um teste unicaudal, será 1,645, então não rejeitamos a hipótese nula.

---

75) Trata-se de um teste de hipótese de dados emparelhados. Temos:

$$H_0 : \mu_\delta = 0$$

$$H_1 : \mu_\delta > 0$$
$$\alpha = 5\%$$
$$\bar{x} = \bar{x}_2 - \bar{x}_1 = 2,64$$
$$s_x = 7,2$$

Calculamos a estatística:

$$t = \frac{\bar{x} - \mu}{s_x/\sqrt{n}} = \frac{2,64}{7,2/\sqrt{25}} = 1,83$$

Ao buscar na tabela temos:

$$t_{n-1;\alpha} = t_{24;5\%} = 1,711$$

Como o valor calculado é maior que o valor tabelado podemos rejeitar a hipótese nula.

76)

77)

Como se trata de dados emparelhados, podemos calcular a diferença entre as duas amostragens e aplicar um teste de hipótese de uma população na diferença entre as amostragens. Isto é:

$$d_i = x_{A_i} - x_{B_i}$$

Com isso calculamos a média e o desvio padrão amostral sobre as amostras d_i e aplicamos um teste de hipótese comum de uma população com variância desconhecida.

Temos:

$$\bar{d} = 9,2$$
$$s_d = 21,2$$

Temos o teste de hipótese:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d > 0$$
$$\alpha = 5\%$$
$$n = 9$$

Podemos calcular a estatística:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = 1,301$$

e comparar com o valor tabelado:

$$t_{8;5\%} = 1,86$$

Como o valor tabelado é maior que o calculado, não podemos rejeitar a hipótese nula. Logo, não pode-se concluir que passar um mês no spa reduz o peso dos pacientes.

78)

Como se tratam de dados emparelhados, podemos calcular a diferença entre as duas amostragens e aplicar um teste de hipótese de uma população na diferença entre as amostragens. Isto é:

$$d_i = x_{A_i} - x_{B_i}$$

Com isso calculamos a média e o desvio padrão amostral sobre as amostras di e aplicamos um teste de hipótese comum de uma população com variância desconhecida.

Temos:

$$\bar{d} = 5$$
$$s_d = 1,87$$

Temos o teste de hipótese:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d > 0$$
$$\alpha = 10\%$$
$$n = 5$$

Podemos calcular a estatística:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = 5,97$$

e comparar com o valor tabelado:

$$t_{5;10\%} = 1,476$$

Como o valor tabelado é menor que o calculado podemos rejeitar a hipótese nula. Com isso, temos que a tarefa realizada na máquina A demora mais que na máquina B.

---

79)

As hipóteses H_0 : os dados seguem uma distribuição normal; H_1 : os dados não seguem uma distribuição normal;
Dados: A média amostral é 23,576 e o desvio padrão 3,113293.
A seguir está descrito um passo a passo para aplicar o teste de aderência Kolmogorov-Smirnov:

1. Colunas 1 e 2: Construir uma tabela com os valores dados em ordem crescente com o seu número de ocorrências;
2. Colunas 3 a 5:Calcular o incremento na frequência acumulada e preencher as três colunas correspondentes (a primeira com o incremento, a segunda com o valor da anterior, começando em 0 e a terceira com a soma das duas outras);
3. Coluna 6: Encontrar a probabilidade acumulada para cada observação, a partir da distribuição normal;
4. Colunas 7 e 8: Calcular as diferenças correspondentes à esquerda e à direita;
5. Pegar o maior valor das diferenças absolutas e comparar com o valor crítico.

Dado	#	$\frac{n_i}{n}$	F_{ant}	F_{obs}	$F_{\text{teórica}}$	$\|F_t - F_{ant}\|$	$\|F_t - F_{obs}\|$
16.6	1	0.04	0	0.04	0.012522	0.012522	0.027478
18.6	1	0.04	0.04	0.08	0.054987	0.014987	0.025013
20.1	1	0.04	0.08	0.12	0.132103	0.052103	0.012103
20.2	1	0.04	0.12	0.16	0.139098	0.019098	0.020902
20.7	1	0.04	0.16	0.2	0.1778	0.0178	0.0222
21.5	1	0.04	0.2	0.24	0.252444	0.052444	0.012444
21.9	1	0.04	0.24	0.28	0.295172	0.055172	0.015172
22	1	0.04	0.28	0.32	0.306352	0.026352	0.013648
22.4	1	0.04	0.32	0.36	0.352814	0.032814	0.007186
22.6	1	0.04	0.36	0.4	0.376953	0.016953	0.023047
22.7	1	0.04	0.4	0.44	0.389212	0.010788	0.050788
23.1	1	0.04	0.44	0.48	0.439241	0.000759	0.040759
23.4	1	0.04	0.48	0.52	0.477459	0.002541	0.042541
23.9	1	0.04	0.52	0.56	0.541443	0.021443	0.018557
24.7	1	0.04	0.56	0.6	0.640962	**0.080962**	0.040962
24.9	1	0.04	0.6	0.64	0.664681	0.064681	0.024681
25.3	2	0.08	0.64	0.72	0.710127	0.070127	0.009873
25.5	1	0.04	0.72	0.76	0.731711	0.011711	0.028289
25.7	1	0.04	0.76	0.8	0.752455	0.007545	0.047545
26.1	1	0.04	0.8	0.84	0.791236	0.008764	0.048764
26.3	1	0.04	0.84	0.88	0.809202	0.030798	0.070798
28.4	1	0.04	0.88	0.92	0.939367	0.059367	0.019367
28.7	1	0.04	0.92	0.96	0.950102	0.030102	0.009898
28.8	1	0.04	0.96	1	0.953323	0.006677	0.046677

Com base na tabela, a maior discrepância observada (entre a curva teórica e observada) foi **0.080962**.

Da tabela de K-S, com 25 dados, temos um valor crítico de 0,264 a 5% e 0,317 a 1%. Assim, em ambos os casos, não rejeitamos a hipótese nula (a distribuição dos dados adere a uma distribuição normal).

---

80)

As hipóteses H_0 : os dados seguem uma distribuição exponencial; H_1 : os dados não seguem uma distribuição exponencial
Seguir o passo a passo descrito no exercício 79.

Observação: O estimador de máxima verossimilhança para o parâmetro de uma exponencial é o inverso da média amostral, ou seja, $\bar{x}^{-1} = 0,036674817$.

A função de probabilidade acumulada será para calcularmos a coluna F_{terica}. $F = 1 - e^{-\lambda x}$. Assim sendo, a tabela fica:

Dado	#	$\frac{n_i}{n}$	$F_{anterior}$	F_{obs}	$F_{\text{teórica}}$	$\lvert F_t - F_{ant}\rvert$	$\lvert F_t - F_{obs}\rvert$
2.2	1	0.03	0.000	0.033	0.078	0.077515	0.04418
4.8	1	0.03	0.033	0.067	0.161	0.128081	0.09475
5.9	1	0.03	0.067	0.100	0.195	0.127905	0.09457
6.1	2	0.07	0.100	0.167	0.200	0.100458	0.03379
7.1	1	0.03	0.167	0.200	0.229	0.062583	0.02925
8.1	1	0.03	0.200	0.233	0.257	0.057005	0.02367
9.7	1	0.03	0.233	0.267	0.299	0.066016	0.03268
11.7	1	0.03	0.267	0.300	0.349	0.082235	0.0489
12.1	1	0.03	0.300	0.333	0.358	0.058384	0.02505
12.8	1	0.03	0.333	0.367	0.375	0.041313	0.00798
15.6	1	0.03	0.367	0.400	0.436	0.06901	0.03568
15.9	1	0.03	0.400	0.433	0.442	0.041851	0.00852
16.2	1	0.03	0.433	0.467	0.448	0.014625	0.01871
17.9	1	0.03	0.467	0.500	0.481	0.014659	0.01867
18.8	1	0.03	0.500	0.533	0.498	0.001834	0.03517
21.6	1	0.03	0.533	0.567	0.547	0.013808	0.01952
22.3	1	0.03	0.567	0.600	0.559	0.008047	0.04138
22.6	1	0.03	0.600	0.633	0.563	0.036551	0.06988
24.5	1	0.03	0.633	0.667	0.593	0.0405	0.07383
24.6	1	0.03	0.667	0.700	0.594	0.072343	0.10568
27.6	1	0.03	0.700	0.733	0.637	0.06341	0.09674
27.7	1	0.03	0.733	0.767	0.638	0.095412	**0.12875**
37.8	1	0.03	0.767	0.800	0.750	0.016663	0.05
38.7	1	0.03	0.800	0.833	0.758	0.04188	0.07521
42.7	1	0.03	0.833	0.867	0.791	0.042209	0.07554
53.3	1	0.03	0.867	0.900	0.858	0.008264	0.0416
86.4	1	0.03	0.900	0.933	0.958	0.057942	0.02461
100.5	1	0.03	0.933	0.967	0.975	0.04159	0.00826
116.7	1	0.03	0.967	1.000	0.986	0.01949	0.01384

Com base na tabela, temos a maior diferença de **0.12875**.
Da tabela de K-S, com 30 dados, temos um valor crítico de 0,242 a 5% e 0,290 a 1%. Assim, em ambos os casos, não rejeitamos a hipótese nula (a distribuição adere a uma exponencial).

---

81) A resolução desta questão será feita através do teste qui-quadrado de Pearson.
As hipóteses: H_0 : o dado é honesto; H_1 : o dado não é honesto.

Face	1	2	3	4	5	6	Total
Frequência Observada o_i	43	49	56	45	66	41	300
Frequência Esperada e_i	50	50	50	50	50	50	
$\frac{(o_i - e_i)^2}{e_i}$	0.98	0.02	0.72	0.5	5.12	1.62	8.96

O valor de $\chi^2_{calc} = 8,96$, contra um valor tabelado de $\chi^2_{crit} = 11,07$. Assim, não rejeitamos a hipótese nula (de que o dado é não-viciado).

---

82) A resolução desta questão será feita através do teste qui-quadrado de Pearson.
As hipóteses: H_0 : os dados seguem a distribuição teórica; H_1 : os dados não seguem a distribuição teórica.

Obs	125	18	20	34	197
Esp	129.23	18.32	18.32	31.13	
$\frac{(o_i - e_i)^2}{e_i}$	0.14	0.01	0.15	0.27	0.56

O valor de $\chi^2_{calc} = 0,56$, contra um valor tabelado de $\chi^2_{crit} = 7,81$. Assim, não rejeitamos a hipótese nula (de que a população é de fato como a teoria dispôs).

---

83)
Hipótese nula: Todas as médias são iguais
Tabela da ANOVA

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit} 5%	F_{crtico} 1%
Entre grupos	120	6	20	4,242424	2,847726	4,45582
Dentro dos grupos	66	14	4,714286	-	-	-
Total	186	20	-	-	-	-

Conclusão: a 5%, ao menos um dos tratamentos não tem média igual aos demais; a 1%, não se pode rejeitar H_o.

84)
Hipótese nula: Todas as médias são iguais
Tabela da ANOVA

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit} 5%
Entre grupos	30	3	10	4	3,238871517
Dentro dos grupos	40	16	2,5	-	-
Total	70	19	-	-	-

Assim, rejeitamos a hipótese de que todas as médias sejam iguais.

85)
Temos duas hipóteses a serem testadas:
1) Hipótese nula: Todas as médias são iguais em relação ao tipo de estabelecimento;
2) Hipótese nula: Todas as médias são iguais em relação às diferentes regiões;
Tabela da ANOVA:

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit}
Entre linhas - tipo de lojas	0,24605	2	0,123025	5,123077	5,143253
Entre colunas - locais	0,108892	3	0,036297	1,51151	4,757063
Resíduo/Erro	0,144083	6	0,024014	-	-
Total (SQT)	0,499025	11	-	-	-

86)
Hipótese nula: O rendimento médio é igual para diferentes escolaridades
Tabela Anova:

SQ	GL	QM	Fcalc	Fcrit	
SQEntre	321,566	2	160,783	487,1062	3,115
SQDentro	25,416	77	0,330078	-	-
SQT	346,982	79	-	-	-

De fato, há diferença de médias (a $\alpha = 5\%$). Para verificar quais são as diferenças, pode-se utilizar um teste de Tukey/Scheffé ou comparação 2 a 2.

87) São 3 testes de hipótese neste caso:
1) Hipótese nula: Os tempos são iguais entre dias de semana;
2) Hipótese nula: Os tempos são iguais entre as diferentes pessoas;
3) Hipótese nula: Não existe interação entre dias de semana e pessoas.

Fonte da variação	SQ	GL	QM	Fcalc	F crítico
Entre pessoas	130305,13	1,00	130305,13	98,20	7,71
Entre dias de semana	3240,13	1,00	3240,13	2,44	7,71
Interações	10585,13	1,00	10585,13	7,98	7,71
Resíduo/Erro	5307,50	4,00	1326,88		
Total (SQT)	149437,88	7,00			

Existem diferença dos tempos médios entre pessoas e há interação entre pessoas e dias da semana. Não existe diferença dos tempos médios entre dias da semana.

88)

São 2 testes de hipótese neste caso:
1) Hipótese nula: Os teores de gordura são iguais entre marcas;
2) Hipótese nula: os teores de gordura são iguais entre diferentes laboratórios;

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit}
Marcas	0,49	2,00	0,24	6,64	10,92
Laboratório	0,54	3,00	0,18	4,91	9,78
Resíduo/Erro	0,22	6	0,04		
Total (SQT)	1,25	11			

Assim, a 1%, não se rejeita nenhuma das H_0 .

89)
São 2 testes de hipótese neste caso:
1) Hipótese nula: As vendas médias semanais são iguais entre vendedoras;
2) Hipótese nula: As vendas médias semanais são iguais entre diferentes semanas;

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit}
Vendedoras	31,58333	2	15,79167	0,175405	2,726468
Semanas	2921,958	7	417,4226	4,636496	2,193134
Resíduo/erro	1260,417	14	90,02976	-	-
Total (SQT)	4213,958	23	-	-	-

Assim, afirma-se que, a 10%, rejeita-se que as vendas semanais tenham mesma média. Não se rejeita que as vendedoras tenham a mesma média.

90)
São 2 testes de hipótese neste caso:
1) Hipótese nula: O número médio de peças defeituosas entre máquina é igual;
2) Hipótese nula: O número médio de peças defeituosas entre operadores é igual;

Fonte da variação	SQ	GL	QM	F_{calc}	F_{crit}
máquinas	7,166667	2	3,583333	0,353425	5,143253
operador	214,9167	3	71,63889	7,065753	4,757063
resíduo/erro	60,83333	6	10,13889		
Total (SQT)	282,9167	11			

Assim, rejeita-se, a 5%, que os operadores tenham médias equivalentes. Não se rejeita que as máquinas tenham médias iguais.

91)
São 3 testes de hipótese neste caso:
1) Hipótese nula: As vendas médias são iguais entre diferentes posições no corredor;
2) Hipótese nula: As vendas médias são iguais entre diferentes alturas da prateleira;
3) Hipótese nula: Não existe interação entre posição do corredor e altura das prateleiras .

Fonte da variação	SQ	GL	QM	Fcalc	F crítico
Corredor	1984,5	1	1984,5	32,27	7,71
Prateleira	760,5	1	760,5	12,37	7,71
Interação	60,5	1	60,5	0,98	7,71
Resíduo/erro	246	4	61,5	-	-
Total (SQT)	3051,5	7	-	-	

Assim, conclui-se que não há interação entre os fatores posição no corredor e altura da prateleira. As vendas médias não são iguais entre as posições do corredor. Também não são iguais as vendas médias entre diferentes alturas da prateleira empregando $\alpha = 5\%$; e apenas rejeição da hipótese de igualdade de média para as posições do corredor empregando $\alpha = 1\%$.

92)
Modelo de regressão: y = -0,2609x + 6,8696. Observa-se que quanto mais novo é um automóvel, maior é seu valor de aluguel.

93)
a) Para avaliar a suspeita de correlação linear podemos calcular o coeficiente de correlação de Pearson.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 9460$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 20600$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 4534$$

$$r = 0,978$$

Como é bem próximo de 1 pode-se observar que há suspeita de uma correlação linear.
b) Para estimar os parâmetros da reta começamos pelo coeficiente angular b.

$$b = \frac{S_{xy}}{S_{xx}} = 0,459$$

Com isso pode-se calcular o coeficiente linear a.

$$a = \bar{y} - b\bar{x} = -5.81$$

Portanto a equação do modelo de regressão linear é:

$$\hat{y} = -5,81 + 0,459x$$

c) Para efetuar o teste de hipótese temos:

$$H_0 : \beta = \beta_0 = 0$$

$$H_1 : \beta \neq 0$$

Podemos calcular a variância residual:

$$s_R^2 = \frac{S_{yy} - bS_{xy}}{n-2} = 31,62$$

Com isso, calculamos a estatística:

$$t_{n-2} = t_6 = \frac{b - \beta_0}{s_R\sqrt{S_{xx}}} = 11,72$$

Dado que o valor crítico é:

$$t_{6,5\%} = 1,943$$

Como o valor calculado é maior que o crítico, pode-se rejeitar a hipótese nula. Ou seja, o coeficiente angular não é igual a zero empregando 5% de significância.

---

94)
Podemos calcular as seguintes somas:

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 346,5$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 168,9$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 1003$$

Com isso calculamos o coeficiente angular b :

$$b = \frac{S_{xy}}{S_{xx}} = 2,051$$

e o coeficiente linear a.

$$a = \bar{y} - b\bar{x} = 39,4$$

Com esses resultados podemos aplicar o teste de hipótese sobre o coeficiente angular:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Vamos calcular a variância residual:

$$s_R^2 = \frac{S_{yy} - bS_{xy}}{n-2} = 29,22$$

Com isso, calculamos a estatística:

$$t_{calc} = t_{calc} = \frac{b-0}{\frac{s_R}{\sqrt{S_{xx}}}} = 4,932$$

Dado que o valor crítico é:

$$t_{10,5\%} = 1,812$$

Como o valor calculado é maior que o crítico, podemos rejeitar a hipótese nula.
d) Testando se o coeficiente linear é positivo (isto é, $H_0 : \alpha = 0; H_1 : \alpha > 0$:

- $S_a = S_r\sqrt{\frac{\sum x_i^2}{n*S_{xx}}} = 7,6815$
- $T_{calc} = \frac{a}{S_a} = 5.13$

Comparando com o valor crítico de 1,812; rejeitamos a hipótese nula.

---

95)
Vamos calcular os seguintes somas:

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 15,5$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 17,5$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 14,08$$

Com isso calculamos uma estimativa do coeficiente angular b:

$$b = \frac{S_{xy}}{S_{xx}} = 0,885$$

e o coeficiente linear a.

$$a = \bar{y} - b\bar{x} = 79,7$$

Segue que temos que a regressão linear é dada por:

$$\hat{y} = 79,7 + 0,885x$$

---

96)
a) Vamos obter a regressão linear da função Imóveis(Y)xJuros(X).

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = -7295$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 59,95$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 1089446$$

Com isso calculamos o coeficiente angular b :

$$b = \frac{S_{xy}}{S_{xx}} = -121,7$$

e o coeficiente linear a .

$$a = \bar{y} - b\bar{x} = 3148$$

e temos a reta de regressão dada por:

$$\hat{y} = 3148 - 121,7x$$

b) Para isso pode-se efetuar o teste de hipótese:

$$H0 : \beta = \beta_0 = 0$$

$$H0 : \beta < 0$$

Começamos por calcular a estatística:

$$t_{n-2} = \frac{b - \beta_0}{s_R/\sqrt{S_{xx}}} = -5,94$$

Comparando com o valor tabelado de:

$$t_{8;5\%} = -1,86$$

Como, o valor calculado é maior que o tabelado, pode-se rejeitar a hipótese nula, ou seja, o coeficiente b é negativo. Desse modo, afirmamos com 95% de confiança que número de imóveis construídos decresce à medida que a taxa de juros aumenta.

---

97)

S_{yy} = 3165,409; S_{xx}= 70402733; S_{xy} = 453199,7;
b =0,006437; a = 28,80913; r = 0,960019; R^2 = 0,921637

Assim sendo, a reta que melhor aproxima os dados é $\hat{y} = 28,81 + 0,0064x$

---

98)
temos as equações:
Linear: $\hat{y} = -135,07 + 148,11x$
Quadrático:
$\hat{y} = 164,6 - 76,6x + 32,11x^2$
Anova - Regressão linear simples :

	SQ	GL	QM	Fcalc	Fcrit
SQReg(linear)	383912,2	1	383912,2	39,15963	7,71
SQRes(linear)	39215,1	4	9803,776		
SQT	423127,3	5			

Anova - Regressão polinomial quadrático:

	SQ	GL	QM	Fcalc	Fcrit
SQReg(parábola)	422398	2	211199	868,7231	9,55
SQRes(parábola)	729,3429	3	243,1143		
SQT	423127,3	5			

Anova melhoria reta x parábola:

	SQ	GL	QM	Fcalc	Fcrit
SQMelhoria	38485,76	1	38485,76	158,3032	10,13
SQRes(parábola)	729,3429	3	243,1143		
SQT	423127,3	5			

Assim, a parábola apresenta melhor explicação, conforme suspeitava-se.

---

99) Preço = R$651916,67 + R$41333,33*Quartos + R$7583,33*Banheiros.

---

100)Custo = -2,688 + 0,047 vendas + 0,0119 Pedidos.