01. Sabe-se que o nível de significância é a probabilidade de cometermos um determinado tipo de erro quando da realização de um teste de hipóteses. Então:

a) A escolha ideal seria um nível de significância igual a zero, o que eliminaria por completo qualquer possibilidade de erro no teste.

b) O nível de significância é uma questão típica de amostras pequenas - em amostras grandes ficam automaticamente eliminadas as possibilidades de erro.

c) Mantido o tamanho da amostra, quanto maior o nível de significância, menor a probabilidade de se cometer o erro Tipo II.

d) O nível de significância mede a probabilidade de rejeitarmos a hipótese nula (H_0) sendo a mesma verdadeira.

e) Se "α" é o nível de significância, então "$1 - \alpha$" é a probabilidade de não cometermos o erro Tipo I.

02. Um comerciante afirma que pelo menos 20% do público preferem seu produto. Uma amostra de 100 pessoas é escolhida para checar a afirmativa. Com um nível de significância de 5% podemos dizer que a porcentagem amostral, para que a afirmativa seja contestada, deve assumir o valor de:

a) 19% **b)** 20% **c)** Aproximadamente 13% **d)** 14% **e)** Qualquer número menor que 20%

03. (UFMG 2005) Considere um teste de hipóteses bilateral para média populacional com hipóteses H_0: $\mu = 100$ contra H_1: $\mu \neq 100$. Suponha que uma amostra aleatória de tamanho 100 foi utilizada ($n = 100$) obtendo-se os valores amostrais $\bar{x} = 99{,}608$ e $s = 2{,}00$. O valor da probabilidade de significância do teste é:

a) 0,025 **b)** 0,050 **c)** 0,100 **d)** 0,150 **e)** 0,010

04. (UFMG 2005) Dois lotes de laranjas de dois produtores diferentes devem ser testados quanto ao teor de carboidratos do suco. Duas amostras de 100 laranjas de cada lote foram analisadas, obtendo-se médias de 111,2g/l e 112,7 g/l, com desvios padrões amostrais 5,5 g/l e 5,1g/l respectivamente. Considere a hipótese nula de que o primeiro lote tem teor de carboidratos maior ou igual do que o segundo lote. Pelo experimento, a hipótese nula deve ser rejeitada a um nível de significância maior ou igual a:

a) 0,05 **b)** 0,04 **c)** 0,03 **d)** 0,02 **e)** 0,01

05. Com base em uma amostra de 100 unidades, quer-se testar, a um nível de significância de 5%, se a vida útil de determinado tipo de lâmpada é de 2000 horas (H_0: $\mu = 2000$) contra a hipótese de que ela seja inferior a 2000 horas (H_1: $\mu < 2000$). Pode-se, então, afirmar que:

a) Há uma probabilidade de 5% de rejeitarmos H_0 quando a vida média de tais lâmpadas for, de fato, 2000 h.

b) Rejeitaremos H_0 sempre que a média encontrada na amostra for 5% menor do que 2000.

c) O teste que iremos realizar é unilateral.

d) Há uma probabilidade de 95% de que a média da população seja igual a média da amostra.

e) Como a variável em causa é contínua, devemos usar no teste a distribuição de Bernoulli.

06. Para uma população normal com média μ e variância $\sigma^2 = 25$ deseja-se testar H_0: $\mu = 10$ contra H_1: $\mu = 5$. Então o tamanho da amostra, n, tal que $\alpha = \beta = 0{,}025$ será:

a) 1 **b)** 2 **c)** 4 **d)** 16 **e)** 25

07. Dadas as seguintes afirmativas sobre testes de hipóteses, é correto dizer que:

a) A probabilidade do erro tipo I é calculada utilizando-se a estatística de teste, para cujo cálculo presume-se que a hipótese nula é falsa.

b) Uma vez definida a região de confiança para um determinado parâmetro da população, várias hipóteses nulas podem ser testadas utilizando-se este intervalo de confiança.

c) Quanto maior o p-valor, maior a credibilidade da hipótese alternativa.

d) A aceitação de determinada hipótese nula implica que esta hipótese seja verdadeira.

e) O poder de um teste é a probabilidade de se rejeitar a hipótese nula quando esta for falsa.

08. Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes e normalmente distribuídas com média μ e variância σ^2. Em relação ao teste de hipótese da média $H_0: \mu = \mu_0$ contra $H_1: \mu < \mu_0$, são corretas as afirmativas:

a) Se o p-valor do teste for menor que o nível de significância, α, a hipótese H_0 deve ser rejeitada.

b) Se a variância σ^2 for conhecida, a estatística do teste segue a distribuição t-Student. Caso contrário, a distribuição do teste será a Normal Padrão.

c) Dados os parâmetros da população: $\mu_0 = 50$ e $\sigma^2 = 900$, suponha que a média de uma amostra aleatória de tamanho 36 retirada desta população seja $\overline{X} = 47$. Neste caso, o nível de significância do teste, α, será igual a 0,2743.

d) A função-poder para este teste de hipótese será uma função decrescente da média μ.

e) Se a hipótese alternativa fosse $H_a: \mu > \mu_0$, ainda assim a função-poder seria decrescente com a média μ.

09. É conhecido que num certo país, os fumantes usavam, na média, 10 cigarros por dia, com desvio padrão de 1,5 cigarros. O governo realizou recentemente uma campanha contra o uso de cigarros. Os produtores estão ansiosos para avaliar os efeitos desta campanha. Eles tiraram uma amostra aleatória simples de 144 fumantes e descobrem que a média amostral é de 8,5 cigarros por dia.

a) A hipótese nula dever ser formulada assim: H_0: μ = 10.

b) A hipótese alternativa deve ser formulada assim: H_1: $\mu \neq 10$.

c) Deve ser rejeitada a hipótese nula ao nível de significância de 5%.

d) Não pode ser feito qualquer teste sem o desvio padrão da amostra.

e) Para evitar cometer um erro, o teste da hipótese nula deve ser feito ao nível de significância de 1%.

10. Indique qual das seguintes considerações sobre os testes de hipótese é verdadeira.

a) O erro do tipo II é definido como a probabilidade de não se rejeitar uma hipótese nula quando esta for falsa e o erro do tipo I é definido como a probabilidade de se rejeitar a hipótese nula quando esta for verdadeira.

b) No teste de hipótese para proporções, se a variância da proporção populacional for desconhecida, a estatística t de Student com n-1 graus de liberdade (n é o tamanho da amostra) é a indicada para o teste.

c) Num teste de hipótese bi-caudal, o valor-p (ou valor de probabilidade) é igual a duas vezes a probabilidade da região extrema delimitada pelo valor calculado da estatística do teste.

d) Não se pode realizar um teste de hipótese para a variância populacional, pois a estatística do teste, que segue uma distribuição Qui-quadrado com n - 1 graus de liberdade (n é tamanho da amostra), não é simétrica.

e) No teste de hipótese para a média (H_0: μ = 0 contra H_1: $\mu \neq 0$), ao nível de significância α, se o intervalo de confiança com 1 - α de probabilidade não contiver μ = 0, não se poderá rejeitar H_0.

11. Uma amostra aleatória simples, de tamanho 4, de uma densidade normal com média μ apresentou os seguintes valores: 2,0 4,0 3,0 3,0. O problema é testar H_0: μ = 2,5 versus H_1: μ ≠> 2,5. O valor-p (significância) associado à estatística de teste usual é tal que:

a) $0,30 \leq p < 0,40$ **b)** $0,20 \leq p < 0,30$ **c)** $0,15 \leq p < 0,20$ **d)** $0,05 \leq p < 0,15$ **e)** $p < 0,05$

12. Para testar se as proporções populacionais referentes à classificação dos elementos populacionais em quatro categorias A, B, C e D são iguais a 20%, 30%, 30% e 20%, uma amostra aleatória simples de tamanho 400 foi obtida e as freqüências observadas foram: classe A: 80, classe B: 100, classe C: 120, classe D: 100. O valor da estatística qui-quadrado usual para esses dados é aproximadamente igual a:

a) 2,64 **b)** 4,06 **c)** 5,28 **d)** 6,78 **e)** 8,33

13. (Eletrobrás 2007) Uma amostra aleatória simples, de tamanho 4, de uma densidade normal com média μ apresentou os seguintes valores: 2,0 4,0 3,0 3,0. O problema é testar H_0: $\mu \leq 2,5$ versus H_1: $\mu > 2,5$. O valor-p (significância) associado à estatística de teste usual é tal que:

a) $p < 0,05$ **b)** $0,05 \leq p < 0,15$ **c)** $0,15 \leq p < 0,20$ **d)** $0,20 \leq p < 0,30$ **e)** $0,30 \leq p < 0,40$

14. ***(AGU 2006)*** Uma pesquisa para avaliar a dependência entre sexo e presença de certo fator resultou na seguinte tabela de contingências:

	Fator	
	Presente	Ausente
Masculino	40	60
Feminino	60	40

O coeficiente de contingência de Pearson associado a esses dados é aproximadamente igual a:

a) 0,2 **b)** 0,6 **c)** 0,9 **d)** 8,0 **e)** 16,0.

15. Para testar H_0: $p = 0,5$ versus H_1: $p = 0,8$, em que p representa uma proporção populacional de "sucessos", será usada uma amostra aleatória simples de tamanho 4 e o critério de decisão que rejeita H_0 se forem observados quatro "sucessos" na amostra. As probabilidades de erro tipo I e tipo II valem respectivamente:

a) 0,0625 e 0,4096

b) 0,1568 e 0,8432

c) 0,0625 e 0,5904

d) 0,1568 e 0,4634

e) 0,3880 e 0,6120

16. (AGU 2006) Para testar, ao nível de significância de 5%, H_0: $\mu \leq 20$ versus H_1: $\mu > 20$, onde μ representa a média de uma distribuição normal com variância 25, uma amostra aleatória de tamanho 100 será observada. A região crítica resultante será:

a) $\overline{x} \geq 20,40$; **b)** $\overline{x} \geq 20,43$; **c)** $\overline{x} \geq 20,52$; **d)** $\overline{x} \geq 20,64$; **e)** $\overline{x} \geq 20,82$.

17. (AGU 2006) Para testar H_0: $\mu \leq 5$ versus H_1: $\mu > 5$, em que μ representa a média de uma distribuição normal com parâmetros desconhecidos, foi usada uma amostra aleatória simples de tamanho 16, que forneceu as seguintes estatísticas: $\overline{x} = 6$ e $\sum_{i=1}^{16}(x_i - \overline{x})^2 = 60$. O p-valor associado à estatística de teste usual, que tem distribuição t-Student quando $\mu = 5$, é tal que:

a) $p < 0,001$ **b)** $0,001 < p < 0,025$ **c)** $0,025 < p < 0,05$ **d)** $0,05 < p < 0,10$ **d)** $p > 0,10$

18. Um fabricante de fibra têxtil está investigando um novo fio, que a companhia fornecedora afirma ter um alongamento médio de 12 kg. O pesquisador deseja testar a hipótese H_0: $\mu = 12$ contra H_1: $\mu < 12$, usando uma amostra aleatória de quatro fios. Considerando que a população apresenta distribuição normal e desvio padrão 0,5 kg, Calcule:

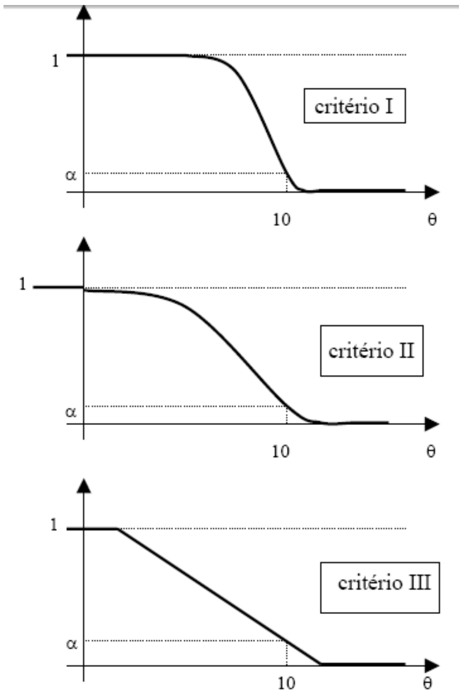
i) A probabilidade do erro tipo I (α) se a região de rejeição for definida como: $\overline{x} < 11,5$kg.

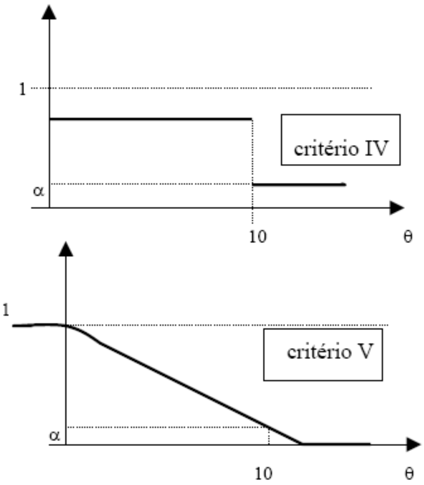
ii) A probabilidade do erro tipo II (β) para o caso em que o alongamento verdadeiro seja 11,25kg. [utilize a mesma região de rejeição do item i)]

Escolha a resposta correta que corresponda respectivamente aos itens i e ii.

a) 0,1587; 0,8413 **b)** 0,025; 0,8413 **c)** 0,05; 0,1587 **d)** 0,0228; 0,1587 e) 0,025; 0,1587

19. (AGU 2006) Para testar H_o: $\theta \geq 10$ versus H_1: $\theta < 10$, θ parâmetro unidimensional, ao nível de significância α, as funções poder de cinco critérios de decisão I, II, III, IV e V foram obtidas e estão apresentadas nos gráficos a seguir:





Entre os apresentados, o melhor critério é o:

(A) I; (B) II; (C) III;

(D) IV; (E) V.

20. (Eletrobrás 2007) Para testar se as proporções populacionais referentes à classificação dos elementos populacionais em quatro categorias A, B, C e D são iguais a 20%, 30%, 30% e 20%, uma amostra aleatória simples de tamanho 400 foi obtida e as frequências observadas foram: classe A: 90, classe B: 110, classe C: 110, classe D: 90. O valor da estatística qui-quadrado usual para esses dados é aproximadamente igual a:

a) 2,64 **b)** 4,17 **c)** 5,28 **d)** 6,78 **e)** 8,33

21. Para testar a aderência de conjunto de observações a uma densidade normal, os dados foram distribuídos em 10 classes e as freqüências observadas foram obtidas. As estimativas da média e da variância populacionais foram calculadas e seus valores foram usados para calcular as freqüências esperadas nas 10 classes. Em seguida, a estatística qui-quadrado usual foi calculada. Sob a hipótese nula de aderência, essa estatística tem distribuição qui-quadrado aproximada com o seguinte número de graus de liberdade:

a) 6 **b)** 7 **c)** 8 d) 9 **e)** 10

22. A tabela apresenta dados sobre faixas de pessoal ocupado por setor de atividade econômica de 100 empresas nacionais.

Faixas de pessoal ocupado	Setor de atividade			Total
	Alimentação	Esporte, saúde, beleza e decoração	Educação e treinamento	
Menos de 20 empregados	30	35	10	75
20 empregados ou mais	10	5	10	25
Total	40	40	20	100

Usando o teste qui-quadrado para testar as hipóteses:

H_0: a faixa de pessoal ocupado independe do setor de atividade

H_1: a faixa de pessoal ocupado depende do setor de atividade

a decisão sobre H_0, aos níveis de 1%, 5% e 10% de significância é:

	α = 1%	α = 5%	α = 10%
a)	Não rejeitar	Não rejeitar	Não rejeitar
b)	Não rejeitar	Não rejeitar	Rejeitar
c)	Não rejeitar	Rejeitar	Rejeitar
d)	Rejeitar	Rejeitar	Não rejeitar
e)	Rejeitar	Rejeitar	Rejeitar

23. A produção mensal de uma indústria se distribuía normalmente com variância 300. Foi introduzida uma nova técnica no processo de fabricação. Em 25 meses, verificou-se que a produção média mensal foi de 100 mil unidades com um desvio padrão de 20 unidades. Utilizando um nível de significância de 2% como o objetivo de testar se a viabilidade do processo de produção aumentou com a incorporação da nova técnica, elaborou-se um teste de hipóteses. Nestas condições, tem-se que:

	Valor da estatística teste	Região crítica	Decisão sob H_0
a)	32	$[40{,}270; +\infty)$	Não rejeitar
b)	32	$(-\infty; 10{,}856]\cup[42{,}980: +\infty)$	Não rejeitar
c)	1,6	$[40{,}270: +\infty)$	Rejeitar
d)	1,6	$(-\infty; 10{,}856]\cup[42{,}980: +\infty)$	Rejeitar
e)	0,07	$[2{,}06: +\infty)$	Não rejeitar

24. Em um teste hipótese, a probabilidade de não rejeitar a hipótese nula, quando ela e falsa, e a probabilidade de rejeitar a hipótese nula, quando ela e verdadeira, são denominados, respectivamente, como:

a) Erro tipo II e Erro tipo I.
b) Erro tipo I e Erro tipo II.
c) Erro tipo I e poder do teste.
d) Nível de significância e poder do teste.
e) Poder do teste e nível de significância.

25. Foi realizado um estudo para verificar se existe diferença na idade dos estudantes que ingressaram na universidade pública ao longo de 3 décadas. Uma amostra aleatória simples de 3000 alunos foi extraída da população dos alunos ingressantes nas décadas de 1990, 2000, 2010. Assumindo que a idade dos alunos segue uma distribuição Normal e que os grupos tem variâncias iguais, o teste de hipótese adequado seria:

a) Teste Kruskal-Walis. **b)** Teste Normal. **c)** Teste t de Student.
d) Analise de Variância. **e)** Teste qui-quadrado.

26. Uma amostra aleatória simples de tamanho n = 9 é selecionada de uma população normal com média μ e desvio padrão σ conhecido e igual a 3. Essa amostra é utilizada para testar H_0: $\mu = 0$ contra H_1: $\mu > 0$. Se a média amostral é $\overline{x} = 1{,}3$, o valor-p do teste é:

a) 0,0227 **b)** 0,0454 **c)** 0,0709 **d)** 0,0968 **e)** 0,1936

27. Considere o intervalo [5; 15] com 90% de confiança para a média μ de uma população com distribuição normal. No teste de hipótese H_0: $\mu = 0 \times H_1$: $\mu \neq 0$, tem-se que a hipótese nula de H_0:

a) não é rejeitada para qualquer nível.
b) é rejeitada para qualquer nível abaixo de 5%.
c) é rejeitada ao nível de 0,05%.
d) é rejeitada a qualquer nível maior que 5% e menor que 10%.
e) é rejeitada ao nível de 10%.

28. Um experimento foi conduzido para testar H_0: $\mu_A = \mu_B$ contra H_1: $\mu_A \neq \mu_B$, sendo μ_A e μ_B as médias de duas populações infinitas, independentes e normalmente distribuídas, isto é, X_A tem distribuição $N(\mu_A, \sigma_A)$ e X_B tem distribuição $N(\mu_B, \sigma_B)$ e $COV(X_A, X_B) = 0$. Amostras de tamanho $n_A = n_B = 5$ são extraídas das respectivas populações e as médias $\overline{X}_{.} = \frac{\sum X_{i.}}{n_{.}}$ e variâncias $S_{.}^2 = \frac{\sum_i (X_{i.} - \overline{X}_{.})^2}{n_{.} - 1}$, calculadas para permitir a realização do teste. Considerando que as variâncias das populações sejam desconhecidas e iguais, a estatística do teste e sua distribuição amostral são, respectivamente:

a) $\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$, com distribuição $N(0,1)$ $\left[S_p^2 = \dfrac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A + n_B - 2)}\right]$

b) $\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$, com distribuição $t_{n_A + n_B}$

c) $\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$, com distribuição $N(0,1)$

d) $\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$, com distribuição $t_{n_A + n_B - 2}$ $\left[S_p^2 = \dfrac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A + n_B - 2)}\right]$

e) $\dfrac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$, com distribuição T^2 de Hotelling, com 1 e $(n_A + n_B - 2)$ graus de liberdade

29. A tabela apresenta a distribuição de frequências dos funcionários promovidos e não promovidos de uma empresa, discriminados pela fluência no idioma inglês.

Fluência em Inglês	Promovidos	Não promovidos	Total
Sim	40	20	60
Não	44	36	80
Total	84	56	140

O valor mais próximo da estatística χ^2 de Pearson, calculada com base na tabela é:

a) 0,5 **b)** 1 **c)** 2 **d)** 3 **e)** 5

30. A função poder dos testes estatísticos é a probabilidade de:

a) Aceitar indevidamente uma hipótese nula falsa.
b) Aceitar indevidamente uma hipótese alternativa falsa.
c) Rejeitar corretamente uma hipótese alternativa falsa.
d) Rejeitar corretamente uma hipótese nula falsa.
e) Rejeitar toda e qualquer hipótese falsa.

31. Em um cálculo de regressão, se o coeficiente angular é zero, conclui-se que:

a) O modelo deve ser o múltiplo.
b) O tamanho da amostra é muito pequeno.
c) Não há relacionamento linear entre as variáveis.
d) As observações têm muita dispersão.
e) Não existe nenhum relacionamento entre as variáveis.

32. Para estimar-se, por regressão simples, o modelo Y = a + bX + U, coletou-se uma amostra de 25 elementos, chegando-se aos seguintes resultados: Y = 3,2 + 0,7X e R^2 = 0,81, então:

a) O coeficiente de correlação estimado é de 0,9;

b) Respeitadas as hipóteses básicas do modelo de regressão, o valor 3,2 é uma estimativa não tendenciosa do parâmetro â;
c) Uma vez que b = 0,7, a variável X explica 70% das variações de Y em torno de sua média;
d) Está faltando, na equação estimada, o termo referente à estimativa da variável "u";
e) Essa regressão não pode ser utilizada em previsões uma vez que o coeficiente de determinação estimado é menor que 90%.

33. Para os valores:

X	10	9	8	7	6
Y	1	2	3	4	5

o coeficiente de correlação será igual a:

a) 0,0 **b)** 1,0 **c)** -1,0 **d)** Um valor entre 0,0 e 0,5 **e)** Um valor entre 0,0 e -0,5

34. Dentre os itens abaixo, identifique as premissas básicas para o modelo de regressão.

I - Linearidade do fenômeno medido;

II - Variância não constante dos termos erro (hetrocedasticidade);

III - Normalidade dos erros;

IV - Erros correlacionados;

V - Presença de colinearidade.

São premissas APENAS os itens:

a) I e III **b)** II e III **c)** I, III e IV **d)** I, III e V **e)** I, II, III e V

35. Ajustou-se um modelo de regressão linear simples a dados provenientes de alguns experimentos executados por um fabricante de concreto, com o objetivo de determinar de que forma e em que medida a dureza de um lote de concreto depende da quantidade de cimento usada para fazê-lo. Quarenta lotes de concreto foram feitos com quantidades diferentes de cimento na mistura e a dureza de cada lote foi medida após sete dias. Sabendo-se que:

$SQR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = 5275,2$ $\quad SQE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 366,6$, o coeficiente de determinação é, aproximadamente:

a) 0 **b)** 0,064 **c)** 0,5 **d)** 0,94 **e)** 14,38

36. Em um estudo de observação, em uma indústria de semicondutores, foram coletadas 25 observações das variáveis, a resistência à tração (uma medida de força requerida para romper a cola), o comprimento do fia e a altura do molde. Suponha que um modelo de regressão linear múltipla foi definido para relacionar a resistência à tração ao comprimento do fio e à altura do molde. Logo $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. Onde Y = resistência à tração, X_1 = comprimento do fio e X_2 = altura do molde. Os resultados obtidos foram:

	Coeficientes estimados	Erro padrão	Estatística t	p-valor
Constante	2,263791	1,060066	2,136	0,044099
Comprimento do fio	2,744270	0,093524	29,343	< 2e-16
Altura do molde	0,012528	0,002798	4,477	0,000188

Com base nestes resultados pode-se concluir que:

a) a reta estimada é $\hat{Y} = 2{,}136 + 29{,}343X_1 + 4{,}477X_2$.
b) os coeficientes estimados são significativos ao nível de 1%.
c) se rejeita a hipótese H_0: $\beta_0 = 0$ ao nível de 1%.
d) se rejeita a hipótese H_0: $\beta_1 = 0$ ao nível de 5%.
e) não se rejeita a hipótese H_0: $\beta_2 = 0$ ao nível de 5%.

37. Considerando a regressão da questão anterior e a tabela da ANOVA (incompleta) a seguir:

Fontes de Variação	Graus de Liberdade	Soma dos Quadrados	Média Quadrática	Estatística F	p-valor
regressão (b0)			R	T	2,2e-16
regressão (b1 | b0)		104,9	S	U	0,0001
resíduo	P	115,2			
Total	Q	6.106,0			

tem-se que os valores de P, Q, R, S, T e U são, respectivamente:

	P	Q	R	S	T	U
a)	22	25	5885,9	104,9	1124,3	20,04
b)	23	25	5885,9	104,9	1175,14	20,94
c)	22	24	5885,9	104,9	1124,3	20,04
d)	21	24	5885,9	52,45	1072,95	9,56
e)	22	25	2942,95	104,9	562,02	20,03

38. Após ser ajustado um modelo de regressão linear entre X e Y, encontrou-se um modelo da forma Y=aX + b + E, onde **a** e **b** são os coeficientes da regressão e **E** o erro aleatório, e um coeficiente de determinação de 80%. O percentual de variação de Y considerado aleatório é de:

a) 80% **b)** 20% **c)** 0% **d)** 72% 36%.

39. Dados 5 pares de números (X_1, Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), (X_5, Y_5), considere os somatórios:

$$\sum_{i=1}^{5} x_i = 15,\ \sum_{i=1}^{5} y_i = 30,\ \sum_{i=1}^{5} x_i^2 = 55,\ \sum_{i=1}^{5} y_i^2 = 230,\ \sum_{i=1}^{5} x_i y_i = 110\ .$$

Ajustando um modelo de regressão linear simples aos pares (X^i, Y^i), a equação da reta de regressão obtida é:

a) y = 2 + x **b)** y = 3 + x **c)** y = 2x **d)** y = 4 + 3x **e)** y = 1 + 4x

40. Ajustando um modelo de regressão linear múltipla $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, obtemos o plano de regressão de equação $y = c + 5x_1 - 6x_2$. Se 30, 15 e 10 são, respectivamente, as médias dos valores de y, x_1 e x_2, o valor de **c** é:

a) -15 **b)** 15 **c)** 30 **d)** 50 **e)** 70

41. Um biólogo marinho realizou um experimento sobre a influência da temperatura da água T na expectativa do tempo de vida do peixe-boi V. Ele ajustou um modelo de regressão linear simples aos dados e obteve a equação da reta de regressão simples T = 1 + 5V/63, adotando V como variável independente e obteve a equação da reta de regressão simples V = 2 + 7T/5, adotando T como variável independente. O coeficiente de correlação linear de Pearson entre as variáveis T e V é:

a) 3/14 **b)** 5/16 **c)** 1/3 **d)** 3/16 **e)** 4/7

42. Em uma agência dos Correios de uma cidade, o gerente realizou um estudo para relacionar o peso total em kg de correspondências recebidas por dia ao número efetivo de correspondências. O levantamento foi realizado em 20 dias e ajustou-se um modelo de regressão linear simples. Logo $Y = \beta_0 + \beta_1 X + \varepsilon$, onde Y = peso total de correspondências e x = número de correspondências. Os resultados foram:

	Coeficientes estimados	Erro padrão	Estatística t	p-valor
Constante	−18,123	3,601	−5,032	8,65 e −05
Número de correspondências	7,777	0,627	12,403	2,93 e −10

Com base nos resultados acima, analise as seguintes afirmações:

I - A reta estimada é: $\hat{Y} = -18{,}123 + 7{,}777x$

II - Rejeita-se a hipótese H_0: $\beta_0 = 0$ ao nível de 5%

III - Admite-se a hipótese H_0: $\beta_1 = 0$ ao nível de 5%

IV - Os coeficientes estimados são significativos ao nível de 1%.

Estão corretas apenas as seguintes afirmações:

a) I **b)** I e II **c)** II e III **d)** I, II e III **e)** I, II e IV

43. Considerando os valores da questão anterior e sabendo que:

$SQR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = 1543{,}79$ $SQE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 180{,}41$. Se SQR é a soma dos quadrados da regressão, SQE é a soma dos quadrados dos erros e $\overline{Y}$ é a média, o coeficiente de determinação é, aproximadamente:

a) 0,10 **b)** 0,12 **c)** 0,89 **d)** 0,95 **e)** 8,55

44. Duas variáveis x e y apresentam covariância amostral Sxy = 100 e desvios padrões amostrais Sx = 10 e Sy = 20. Considere um modelo de regressão linear simples para explicar o comportamento de y a partir de x como sendo: $y = \beta_0 + \beta_1 x + \varepsilon$, sendo ε um ruído branco Gaussiano. Se estimarmos esse modelo, utilizando o método de mínimos quadrados ordinários, a estimativa do coeficiente de inclinação β_1 será:

a) 0,25 **b)** 0,5 **c)** 1 **d)** 5 **e)** 10

45. A figura mostra o resumo de uma regressão realizada no Excel por meio do suplemento Análise de Dados – Regressão. Deseja-se calcular na célula F3 o valor predito pelo modelo linear, quando a variável explicativa X_1 é igual a 10. A fórmula que deve ser escrita na célula F3 (em destaque), para realizar o cálculo desejado, é:

1	RESUMO DOS RESULTADOS						
2							
3	*Estatística de regressão*				10		
4	R múltiplo	0,96776282					
5	R-Quadrado	0,936564876					
6	R-quadrado ajustado	0,925992355					
7	Erro padrão	2,845062604					
8	Observações	8					
9							
10	ANOVA						
11		*gl*	*SQ*	*MQ*	*F*	*F de significação*	
12	Regressão	1	717,0393312	717,0393312	88,58482342	8,17431E-05	
13	Resíduo	6	48,56628733	8,094381221			
14	Total	7	765,6056185				
15							
16		*Coeficientes*	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*
17	Interseção	2,962621916	2,216854921	1,336407668	0,229862277	-2,461826653	8,387070474
18	Variável X 1	4,131871855	0,439002691	9,411951095	8,17431E-05	3,057670971	5,20607274

a) =B19*E3 + B17 **b)** =D17*10 + E17 **c)** =B17 + B18*E3 **a)** =D18*10 + D13 **a)** =E3*6 + 7

46. Considere um modelo de regressão linear simples, com intercepto, em que a variável dependente Y seja a demanda por petróleo em um país e a variável explicativa x seja o nível de industrialização do país. Este modelo foi ajustado a uma amostra de 122 países, fornecendo as seguintes estimativas de β_0 (intercepto) e β_1 (coeficiente de x), com os respectivos erros padrão estimados:

$\hat{\beta}_0 = 90$ (erro padrão estimado = 40)

$\hat{\beta}_1 = 66$ (erro padrão estimado = 30)

A partir desses resultados conclui-se que a regressão:

a) não é significativa aos 3 níveis usuais de 0,01, 0,05 e 0,1.

b) é significativa aos 3 níveis usuais de 0,01, 0,05 e 0,1.

c) é significativa aos níveis de 0,1 e 0,05, mas não ao nível de 0,01.

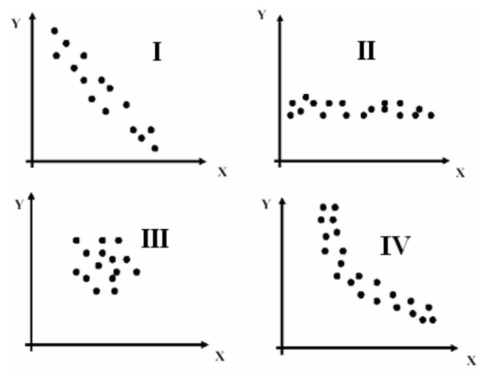
d) é significativa ao nível de 0,1, mas não aos níveis de 0,05 e 0,01.

e) é significativa aos níveis usuais de 0,01, 0,05, mas não ao nível de 0,1.

47. Um modelo de regressão linear com intercepto e 2 variáveis explicativas foi ajustado a uma amostra de tamanho 43, fornecendo um coeficiente de determinação R^2 = 0,8. O valor da estatística F que permite testar a significância do modelo é:

a) 40 **b)** 80 **c)** 86 **d)** 160 **e)** 164

48. Considere os seguintes diagramas de dispersão:



a) O gráfico I indica que existe forte correlação linear positiva entre as variáveis X e Y.

b) O gráfico II indica que o coeficiente de correlação de Pearson está próximo de 1.

c) O gráfico III indica que praticamente não há correlação entre as variáveis X e Y.

d) O gráfico IV indica que existe forte correlação entre X e Y.

e) O gráfico I e II indicam que o coeficiente de correlação de Pearson está próximo de zero.

49. Em um modelo de regressão linear do tipo $Y_i = \beta_0 + \beta_1 X_i + \varepsilon$,foram obtidos os seguintes resultados:

	Coeficientes	Erro padrão do coeficiente	Estatística t	valor p
Interseção $(\hat{\beta}_0)$	14,24430812	20,52146269	?	0,507260244
Área $(\hat{\beta}_1)$	1,030417061	0,110002636	?	0,0000138

Observando a tabela, podemos afirmar que, a um nível de significância de 5%:

a) O intercepto e o coeficiente angular foram estatisticamente significativos e o valor da estatística t para o coeficiente angular é 9,367.

b) O coeficiente angular foi estatisticamente significativo e o valor da estatística t para o coeficiente angular é 9,367.

c) O intercepto e o coeficiente angular não apresentaram significância estatística e o valor da estatística t para o coeficiente angular é 0,113.

d) Somente o intercepto foi estatisticamente significativo e o valor da estatística t para o intercepto é 1,441.

e) O coeficiente angular foi estatisticamente significativo e o valor da estatística t para o coeficiente angular é 0,113.

50. Uma empresa está estudando como varia a procura de certo produto em função do preço de venda e obteve as seguintes informações:

x	162	167	173	176	180
y	248	242	215	220	205

onde x = preço de venda (em reais) e y = vendas mensais (unidades). Neste caso, o coeficiente angular do modelo é:

a) -6,44 **b)** -2,45 **c)** 2,45 **d)** 3,50 **e)** 6,44