

DESM1 - Desafio do Módulo 1

Entrega 31 jul em 19:00

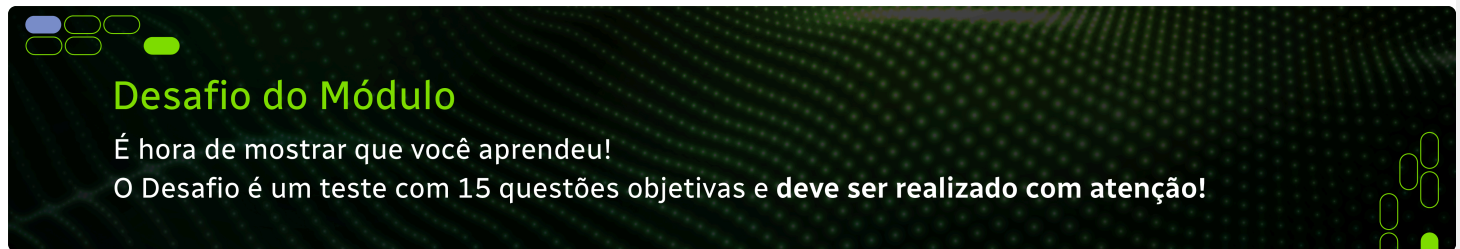
Pontos 45

Perguntas 15

Disponível até 31 jul em 19:00

Limite de tempo Nenhum

Instruções



Desafio do Módulo

É hora de mostrar que você aprendeu!
O Desafio é um teste com 15 questões objetivas e **deve ser realizado com atenção!**

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de Dúvidas sobre o Conteúdo do Módulo 1".

Para iniciá-lo clique em "Fazer teste". Você tem somente **uma** tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você concluí-lo. Antes de enviar confira todas as questões.

Caso o teste seja iniciado e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independente do progresso no teste. Fique atento ao seu teste e ao prazo final, pois novas tentativas só serão concedidas em casos de questões médicas.

O gabarito será disponibilizado a partir de sexta-feira, **02/08/2024**, às 23h59.

- O arquivo abaixo contém o enunciado do trabalho prático. Confira agora:

Enunciado do Desafio - Módulo 1 - Bootcamp Cientista de

dados.pdf (<https://online.igti.com.br/courses/7642/files/582823?wrap=1>)_ ↓

(https://online.igti.com.br/courses/7642/files/582823/download?download_frd=1)

Codigo_R_Desafio (1)-1.txt (<https://online.igti.com.br/courses/7642/files/582809?wrap=1>)_ ↓

(https://online.igti.com.br/courses/7642/files/582809/download?download_frd=1)

Bons estudos!

Atenciosamente,

Equipe XP Educação

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	<u>Tentativa 1</u>	6.952 minutos	45 de 45

⚠ As respostas corretas estarão disponíveis em 2 ago em 23:59.

Pontuação deste teste: **45** de 45
Enviado 30 jul em 6:36
Esta tentativa levou 6.952 minutos.

Pergunta 13 / 3 pts

Explore a variável resposta, que é o Consumo, e responda:

Pelo histograma, você diria que a variável segue uma distribuição normal?

☒

Não, pois os dados não são distribuídos simetricamente em torno de um valor central.

☐

Sim, pois os dados se concentram entre o primeiro quartil e a mediana.

☐

Sim, pois os dados se concentram entre a mediana e o terceiro quartil.

☐

Sim, pois os dados são distribuídos simetricamente em torno do desvio padrão.

Pergunta 23 / 3 pts

Explore a variável resposta, que é o Consumo, e responda:

Pelo boxplot, você consegue visualizar algum outlier?

☐

Não, pois apesar de ter pontos acima do limite superior, também deveria ter abaixo do limite inferior para se considerar outlier.

☐

Não, pois há somente 4 pontos acima do limite superior.

☐

Sim, há pontos outliers tanto acima do limite superior quanto do limite inferior.

☒

Sim, há pontos outliers acima do limite superior.

Pergunta 3**3 / 3 pts****Explore a variável resposta, que é o Consumo, e responda:**

Qual é o valor do primeiro quartil e terceiro quartil?

☒

Os valores do primeiro e terceiro quartil são 3903 e 13444, respectivamente, significa que 25% dos clientes consomem até 3903 e 75% dos clientes consomem até 13444.

☐

Os valores do primeiro e terceiro quartil são 1019 e 7430, respectivamente, significa que 25% dos clientes consomem até 1019 e 75% dos clientes consomem até 7430.



Os valores do primeiro e terceiro quartil são 7430 e 36936, respectivamente, significa que 25% dos clientes consomem até 7430 e 75% dos clientes consomem até 36936



Pelo fato de a variável não seguir uma distribuição normal, não é possível calcular o primeiro e o terceiro quartil.

Pergunta 4

3 / 3 pts

Você está trabalhando como analista de dados em uma empresa de marketing digital e precisa prever a probabilidade de um evento binário, no caso, o usuário clicar em um anúncio online. Qual a técnica recomendada para modelar tal problema?



Regressão linear simples



Regressão linear múltipla



Regressão não linear múltipla



Regressão logística

Pergunta 5

3 / 3 pts

Explore a relação entre as variáveis Consumo e Possui Imóvel Próprio, e responda:

Através de um teste t de Student para amostras independentes, existe diferença significativa entre o consumo médio do público que possui imóvel próprio quando comparado com o consumo médio do público não

possui imóvel próprio? Adote 95% de confiança ao realizar suas interpretações.



É estatisticamente diferente, pois o consumo médio de quem não tem imóvel próprio é de 12391,3 e quem tem imóvel próprio é de 10862,78.



Com 17.47 graus de liberdade, temos evidências para rejeitar a hipótese nula de igualdade de médias, ou seja, o consumo de quem tem imóvel próprio é estatisticamente diferente do consumo de quem não tem imóvel próprio.



Ao p valor de 72,26%, há evidências para rejeitar a hipótese nula de igualdade de médias, ou seja, a média do consumo de quem tem imóvel próprio é estatisticamente diferente da média de consumo de quem não tem imóvel próprio.



Ao p valor de 72,26%, não há evidências para rejeitar a hipótese nula de igualdade de médias, ou seja, a média do consumo de quem tem imóvel próprio não é estatisticamente diferente da média de consumo de quem não tem imóvel próprio.

Pergunta 6

3 / 3 pts

Explore a relação entre as variáveis Consumo e Idade, e responda:

Pelo gráfico de dispersão, você identifica que existe relação linear entre o Consumo e a Idade? Se sim, a relação é positiva ou negativa?



Não existe correlação.



Sim, existe uma correlação linear positiva, pois à medida que a idade aumenta o consumo também aumenta.

Pergunta 7

3 / 3 pts

Explore a relação entre as variáveis Consumo e Idade, e responda:

Obtenha o valor do coeficiente de correlação linear de Pearson entre o Consumo e a Idade e interprete. Marque a alternativa CORRETA.



O coeficiente de correlação linear de Pearson entre o Consumo e a Idade é de 0,93, que é uma correlação negativa fraca.



O coeficiente de correlação linear de Pearson entre o Consumo e a Idade é de 0,93, que é uma correlação positiva fraca.



O coeficiente de correlação linear de Pearson entre o Consumo e a Idade é de 0,93, que é uma correlação negativa forte.



O coeficiente de correlação linear de Pearson entre o Consumo e a Idade é de 0,93, que é uma correlação positiva forte.

Pergunta 8

3 / 3 pts

Explore a relação entre as variáveis Consumo e Idade, e responda:

Se tentarmos utilizar somente a Idade para prever o Consumo, o quanto da variação do Consumo a variável Idade consegue explicar? Em outras palavras, interpretar o R^2 da regressão linear do Consumo em função da Idade.

☐

O R^2 é de 88,34%, ou seja, para cada aumento unitário na idade, o Consumo aumenta em média 11,66% ($1-0,8834$).

☐

O R^2 é de 88,34%, ou seja, para cada aumento unitário na idade, o consumo diminui em média 88,34%.

☐

O R^2 é de 88,34%, ou seja, para cada aumento unitário na idade, o Consumo aumenta em média 88,34%.

☒

O R^2 é de 88,34%, ou seja, a Idade consegue explicar 88,34% da variação do Consumo.

Pergunta 9

3 / 3 pts

Explore a relação entre as variáveis Consumo e Idade, e responda:

Execute um teste de normalidade para regressão linear entre Consumo e Idade e, de acordo com os resultados, informe se os resíduos seguem ou não seguem uma distribuição normal. Adote 95% de confiança.

☐

Ao p valor de 24,15%, só rejeitaríamos a hipótese de normalidade caso o nível de confiança fosse de 99%. Portanto o teste é inconclusivo.

☐

Ao p valor de 24,15%, há evidências para rejeitar a hipótese de normalidade, ou seja, os resíduos não seguem uma distribuição normal



Ao p valor de 24,15%, não há evidências para rejeitar a hipótese de normalidade, ou seja, os resíduos seguem uma distribuição normal.



Ao p valor de 24,15%, só rejeitaríamos a hipótese de normalidade caso o nível de confiança fosse de 90%. Portanto o teste é inconclusivo.

Pergunta 10

3 / 3 pts

Explore a variável Renda Mensal e responda:

Qual o valor do primeiro quartil e qual a sua interpretação CORRETA?



O primeiro quartil é 2292, isso nos diz que 75% dos clientes têm renda até 2292 e 25% têm renda acima desse valor.



O primeiro quartil é 3576, isso nos diz que 25% dos clientes têm renda mensal de até 3576.



O primeiro quartil é 2292, isso nos diz que 25% dos clientes têm renda mensal de até 2292.



O primeiro quartil é 2292, isso nos diz que 50% dos clientes têm renda mensal até 2292 e os outros 50% têm renda mensal acima desse valor

Pergunta 11

3 / 3 pts

Explore a variável Renda Mensal e responda:

Qual o valor da mediana e qual a sua interpretação CORRETA?

☐

A mediana é 3576, isso nos diz que 25% dos clientes têm renda mensal de até 3576.

☒

O valor da mediana é 3576, isso nos diz que 50% dos clientes têm renda até 3576 e os outros 50% têm renda maior que 3576.

☐

O valor da mediana é 3576, isso nos diz que o cliente que tem a maior renda possui renda de 3576.

☐

A mediana é 3576, isso nos diz que 75% dos clientes têm renda até 3576 e 25% têm renda acima desse valor.

Pergunta 12**3 / 3 pts**

Ajuste uma Regressão linear do Consumo em função da Idade e Renda Mensal, e responda:

O modelo de regressão linear ajustado é válido?

☐

Não, pois p valor do teste F é aproximadamente zero, então não há evidências para rejeitar a hipótese de que o modelo ajustado não é válido.



Não, pois a estatística F obtida foi de 235,6 com apenas dois graus de liberdade no numerador e 26 no denominador.



Sim, pois p valor do teste F é aproximadamente zero, então há evidências para rejeitar a hipótese de que o modelo ajustado não é válido.



Não, pois não se pode ajustar uma regressão linear com apenas dois preditores.

Pergunta 13

3 / 3 pts

Ajuste uma Regressão linear do Consumo em função da Idade e Renda Mensal, e responda:

Marque a alternativa com a interpretação CORRETA do coeficiente Beta da variável Idade.



O coeficiente Beta da variável idade é de $3,48e+02$, ou seja, mantendo as demais variáveis constantes, a cada aumento unitário na idade, o consumo aumenta em média 348 unidades



O coeficiente Beta da variável idade é de $3,48e+02$, isso nos diz que teremos um coeficiente de correlação de Pearson de 0,48.



O coeficiente Beta da variável idade é de $3,48e+02$, isso nos diz que a idade explica 3,48% da variação do consumo.



O coeficiente Beta da variável idade é de $3,48e+02$, isso nos diz que a idade não é uma variável significativa e deve ser removida do modelo.

Pergunta 14

3 / 3 pts

Ajuste uma Regressão linear do Consumo em função da Idade e Renda Mensal, e responda:

Marque a alternativa com a interpretação CORRETA do coeficiente Beta da variável Renda Mensal.



O coeficiente Beta da variável Renda Mensal é de 2,62, ou seja, mantendo as demais variáveis constantes, a cada aumento unitário na Renda Mensal, o consumo aumenta, em média, 2,62 unidades.



O coeficiente Beta da variável idade é de 2,62, isso nos diz que a Renda Mensal não é uma variável significativa e deve ser removida do modelo.



O coeficiente Beta da variável idade é de 2,62, isso nos diz que a Renda Mensal explica 2,62% da variação do consumo.



O coeficiente Beta da variável Renda Mensal é de 2,62, isso nos diz que teremos um coeficiente de correlação de Pearson de 0,62.

Pergunta 15

3 / 3 pts

Ajuste uma Regressão linear do Consumo em função da Idade e Renda Mensal, e responda:

Como os valores t são obtidos? (coluna t value)

- ☐ Elevando ao quadrado os valores da coluna Estimate.
- ☒ Dividindo a coluna Estimate pela coluna Std. Error.
- ☐ Dividindo a coluna Estimate pela coluna $\Pr(>|t|)$.
- ☐ Tomando a raiz quadrado dos valores da coluna Estimate.

Pontuação do teste: **45** de 45