

Aluno(a): _____

Curso: _____ Nº de matrícula: _____

Turma: _____ Professor(a): _____

CIÊNCIA DOS DADOS

Avaliação FINAL

12/06/2018

Prezado(a) Aluno(a),

Você terá 120 minutos a partir do início oficial da prova para concluir esta avaliação, administre bem o seu tempo. Leia atentamente as instruções a seguir e as questões da prova antes de começar a resolvê-la.

1. Identifique-se com letra legível em **todas** as folhas de prova.
2. Esta Avaliação Intermediária é composta por **2** partes. As questões da Parte 1 e as definições das variáveis da Parte 2 são apresentadas aqui a qual tem um total de **8** páginas. Verifique se a prova está completa e/ou se há problemas de impressão e comunique o aplicador **antes** de iniciar a prova. Comunicação posterior não será considerada.
3. Para a resolução das questões, utilize apenas os campos demarcados e não destaque as folhas de prova.
4. A resolução da prova poderá ser feita a lápis ou a caneta. Avaliações feitas a lápis, no entanto, **não serão revisadas pelo professor**.
5. Em caso de dúvida sobre alguma questão desta avaliação, redija um texto na folha de prova explicitando-a para que o professor avalie a pertinência durante a correção.
6. Portar qualquer material estranho à folha de questões – celular (ligado ou desligado), tablet, notebook, anotações e livro – constituirá violações ao Código de Ética e de Conduta e acarretará sanções nele previstas.
7. Consulta a colegas e a qualquer material estranho à folha de questões (celular, tablet, notebook, anotações e livro) constituirão violações ao Código de Ética e de Conduta e acarretarão sanções nele previstas. Faça o seu trabalho de maneira ética!
8. Você somente poderá sair da sala depois de entregar a prova. Caso necessite sair durante a realização da avaliação, peça autorização antecipadamente ao aplicador.

Boa Prova!

Parte	Valor	Nota
1	5,0	
2 (Jupyter)	5,0	

Para uso exclusivo do Professor

Prova passível de revisão?

() **SIM** () **NÃO**

Aluno(a): _____

Curso: _____ **Nº de matrícula:** _____

Turma: _____ **Professor(a):** _____

INSTRUÇÕES COMPLEMENTARES

1. Você receberá instruções complementares sobre como obter acesso à parte eletrônica da prova, a qual fará referência de Parte 2. Veja a folha a seguir sobre precauções de segurança para evitar ter sua prova anulada. **Essa Parte 2 deve ser entregue em HTML via Blackboard. Só poderá ser feita apenas UMA tentativa.**

2. Você pode utilizar qualquer material individual para consulta – anotações, livros, etc. Mas não pode compartilhar este material com seus colegas. Portar celular, mesmo que desligado, constituirá infração ao código de ética.

3. Lembre-se de entregar a prova, isto é de responsabilidade do aluno!!!

4. O **verso** das folhas pode ser usado como rascunho, porém, **não será levado em consideração durante a correção das questões.**

5. Leia atentamente cada questão e responda o que for pedido. Erros conceituais serão penalizados, mesmo que o conceito não se relacione com o que foi pedido na questão.

6. Caso, em algum item você necessite do resultado de um item anterior que você não conseguiu fazer, admita um valor razoável para esse resultado e faça o item normalmente. Indique na questão, caso isso aconteça.

7. **Todos os resultados devem ser justificados. Números que apareçam sem uma explicação de como foram encontrados serão ignorados na correção.**

8. Se utilizar o Python para resolver algum item da Parte 1, deixe anotado no próprio item qual foi o código utilizado.

Aluno(a): _____

Curso: _____ **Nº de matrícula:** _____

Turma: _____ **Professor(a):** _____

PRECAUÇÕES RELACIONADAS AO USO DE *LAPTOP* DURANTE A PROVA

1. Todas as precauções abaixo podem ser resumidas da seguinte forma: consulte o material que quiser desde que não sejam LIVROS; não seja nenhuma pessoa; e não compartilhe informações com ninguém.
2. Desligar o som do computador e não utilizar fones de ouvido.
3. Você só pode usar a rede Wifi ou Ethernet do Insper, é proibido usar redes 4G ou quaisquer outros meios. .
4. Confira se todas as abas e navegadores abertos não infringem nenhuma norma da prova.
5. Desligue todos os programas que sincronizam arquivos com a nuvem: Google Drive, Dropbox, OneDrive, etc. Você pode acessar suas pastas locais destes serviços normalmente.
6. Desligue todos os programas que acessam a Internet sem você pedir, por exemplo o Steam. Tome cuidado com páginas que realizam acessos automáticos.
7. Você não pode enviar nenhum tipo de mensagem durante a prova para qualquer pessoa dentro ou fora dela. Por exemplo: e-mail, Skype, mensagens do Facebook, mensagens do WhatsApp, mensagens via Twitter, etc
8. Você não pode compartilhar nenhum tipo de informação em nenhum site, incluindo Pastebin, Codeshare, wikis, etc
9. Você não pode perguntar questões da prova no StackOverflow, StackExchange, MathExchange, etc.
10. Você não pode trabalhar a partir de máquinas remotas com LogMeIn, SSH, VNC ou VPNs.
11. Você pode procurar a resposta de questões no Google.
12. Você pode consultar qualquer arquivo no seu computador.
13. Todos os logs dos acessos durante a avaliação dos alunos estarão disponíveis para consulta dos avaliadores, cabe ao aluno ter o cuidado de não transgredir nenhuma norma descrita nesse documento. Esses dados podem ser analisados após a realização da prova.
14. Qualquer transgressão das normas acima podem zerar a nota da prova e implicar em infração ao código de ética do Insper.

Aluno(a): _____

Curso: _____ Nº de matrícula: _____

Turma: _____ Professor(a): _____

Parte 1 – Questão 1 (1,0 ponto)

(Questão extraída do ENADE 2017)

Suponha que determinado programa de computador seja executado por meio de 13 etapas, com tempo médio de 50 segundos ao todo e dispersão relativa de 10% em torno da média.

Considere que uma equipe de engenharia propõe um novo algoritmo que reduz em 30% o tempo de execução de todas as 13 etapas desse programa.

Nesse contexto, avalie as afirmações a seguir, a respeito do tempo de execução do novo algoritmo.

- I. O tempo médio por etapa será 32,5 segundos.
- II. O desvio-padrão permanecerá inalterado.
- III. A dispersão relativa em torno da média permanecerá inalterada.

É correto o que se afirma em: **(ESTE EXERCÍCIO SERÁ ANULADO SE NÃO HOUVER DESENVOLVIMENTO DE RACIOCÍNIO PARA A ESCOLHA DA ALTERNATIVA, MESMO SE ASSINALADA A RESPOSTA CORRETA)**

- A) I, apenas
- B) III, apenas
- C) I e II, apenas
- D) II e III, apenas
- E) I, II e III

Deixe aqui seu RACIOCÍNIO feito para A ESCOLHA DA ALTERNATIVA ASSINALADA!! Se essa parte estiver EM BRANCO, seu exercício será anulado!

Aluno(a): _____

Curso: _____ Nº de matrícula: _____

Turma: _____ Professor(a): _____

Parte 1 – Questão 2 (4,0 pontos)



Você está em » Home » Semanal - Resumo 1 » Capitais »

Síntese dos Preços Praticados - Brasil - Capitais
Resumo 1 - Gasolina R\$/l
Período : de 13/05/2018 a 19/05/2018

https://www.anp.gov.br/preco/prc/Resumo_Semanal_Tipologia.asp

Segundo a Agência Nacional de Petróleo (ANP), o preço da gasolina na semana dos dias 13/05 a 19/05 desse ano era, em média, R\$ 4,048. Considere esse valor como populacional para a cidade de São Paulo.

No dia 24/05/2018, decorrente a greve dos caminhoneiros, o Brasil todo sofreu com a falta de combustível. Alguns especialistas relataram que, em todos os postos da cidade de São Paulo, o preço da gasolina aumentou em mais do que 50% em relação aos valores cobrados na semana dos dias 13/05 a 19/05.

Assuma que uma amostra de 20 postos foi coletada aleatoriamente na cidade de São Paulo e que o desvio padrão populacional após aumento era conhecido e igual a $\sigma = \text{R\$ } 0,297$.

- (0,75 ponto) Descreva a variável de interesse X para o problema acima descrito. Ainda, você conhece a distribuição do estimador \bar{X} ? Se não, é possível fazer alguma suposição?
- (0,75 ponto) Baseado nas informações descritas no enunciado, formule as hipóteses em termos do problema e em termos estatísticos.
- (0,75 ponto) Para um nível de significância de 10%, construa a RC.
- (0,75 ponto) Se a amostra de 20 postos forneceu média de R\$ 6,1699 no preço da gasolina, qual a conclusão deste teste?
- (1,0 ponto) Refaça a **conclusão** do teste de hipóteses, porém utilizando o valor-p?

Aluno(a): _____

Curso: _____ Nº de matrícula: _____

Turma: _____ Professor(a): _____

PARTE 2



KAGGLE:

O [Kaggle.com](https://www.kaggle.com/) é um site de competições internacionais de *Data Science* no qual empresas oferecem prêmios para os analistas que fizerem a melhor modelagem preditiva de diversos problemas.

Em uma das competições recente, o desafio era:

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

Na parte Python da avaliação, vocês receberão um conjunto de dados bem menor em termos de número de variáveis disponível no Kaggle.

SIGNIFICADO DAS VARIÁVEIS:

A base de dados que terão acesso traz informações de 1460 imóveis cujas variáveis selecionadas possuem as seguintes descrições traduzidas (mantendo os nomes das variáveis e das categorias do problema original - em inglês):

- ✓ **SalePrice:** preço de venda da propriedade (em dólares). Essa é a variável alvo que você irá tentar prever.
- ✓ **OverallQual:** avalia de maneira geral o material e o acabamento da casa, com as seguintes categorias:
 1. Very Poor Péssimo
 2. Poor Ruim
 3. Fair Justo
 4. Below Average Abaixo da média
 5. Average Médio
 6. Above Average Acima da média
 7. Good Bom
 8. Very Good Muito bom
 9. Excellent Excelente
 10. Very Excellent Muito excelente

Aluno(a): _____

Curso: _____ Nº de matrícula: _____

Turma: _____ Professor(a): _____

- ✓ **LotArea:** tamanho do lote (em pés quadrados).
- ✓ **X1stFlrSF:** tamanho do primeiro andar (em pés quadrados).
- ✓ **X2ndFlrSF:** tamanho do segundo andar (em pés quadrados).
- ✓ **GarageCars:** tamanho da garagem em capacidade de carros (número de carros).
- ✓ **Fireplaces:** número de lareiras.
- ✓ **CentralAir:** ar condicionado central, com as seguintes categorias:
 - 0. N Não
 - 1. Y Sim
- ✓ **YearBuilt:** ano da construção do imóvel.
- ✓ **YrSold:** ano de venda do imóvel.

O arquivo <houseAF.xlsx> traz todas essas 10 variáveis.

Responda as perguntas referentes à Parte 2 dessa avaliação descritas no .ipynb.
