

Aluno(a): \_\_\_\_\_

Curso: \_\_\_\_\_ Nº de matrícula: \_\_\_\_\_

Turma: \_\_\_\_\_ Professor(a): \_\_\_\_\_

**CIÊNCIA DOS DADOS**  
**Avaliação SUBSTITUTIVA**  
**14/06/2018**

**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
<b>1</b>	<b>5,0</b>	
<b>2 (Jupyter)</b>	<b>5,0</b>	

**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 (2,00 pontos)**

Considere que funcionários de uma empresa dividem seu tempo de trabalho utilizando dois tipos de aparelhos diferentes: em **notebook** ou pelo **celular**. Assuma que esses eventos sejam complementares, ou seja, que seja impossível que funcionários trabalhem em ambos os aparelhos simultaneamente. Ainda, considere que esses funcionários trabalhem em seus respectivos aparelhos navegando por uma das duas conexões possíveis: via **Wi-Fi** ou via **3G/4G**. Os resultados de uso do tipo de aparelho e tipo de conexão feita pelos funcionários dessa empresa são os seguintes:

- 50% das vezes esses funcionários utilizam **notebook** para trabalhar;
- Quando funcionários trabalham utilizam **notebook**, 92,5% das vezes a conexão feita é via **Wi-Fi**;
- Quando a conexão feita pelo funcionários é via **Wi-Fi**, 82,2% das vezes os funcionários estão trabalhando com **notebook**;
- Assuma que conexões “**Wi-Fi**” e “**3G/4G**” sejam eventos complementares (como de fato é na vida real).

Responda cada item abaixo, **definindo os eventos necessários e considerando, na sua resolução, as notações vistas em aula**. Utilize 3 casas decimais na resolução dos itens.

- a) Calcule a probabilidade dos funcionários dessa empresa trabalharem com conexão **Wi-Fi**. (1,0 ponto)
- b) Das vezes que a conexão feita é via **3G/4G**, calcule a probabilidade do tipo de aparelho utilizado pelos funcionários seja **celular**. (1,0 ponto)

Aluno(a): \_\_\_\_\_

Curso: \_\_\_\_\_ Nº de matrícula: \_\_\_\_\_

Turma: \_\_\_\_\_ Professor(a): \_\_\_\_\_

Parte 1 – Questão 2 (3,00 pontos)

**Greve de caminhoneiros provoca falta de alimentos e combustível em SP e no Rio**

A Associação Paulista de Supermercados (Apas) afirmou que há desabastecimento sobretudo em itens perecíveis que precisam de reposição diária nos mercados paulistas, como frutas e verduras. Os caminhões que transportavam o produto aderiram ao protesto e estão parados às margens da Via Dutra.

<https://economia.estadao.com.br/noticias/geral,greve-de-caminhoneiros-provoca-falta-de-alimentos-e-combustivel-em-sp-e-no-rio,70002320482>

Decorrente a greve dos caminhoneiros, os supermercados ficaram desabastecidos devido à ausência de entregas diárias de diversos produtos. Em um período normal de entregas, uma rede de supermercados recebe, em média, a quantidade de 350 caminhões diariamente. Alguns especialistas da área relataram que, na semana de greve dos caminhoneiros, a quantidade diária de caminhões fazendo entregas de produtos a essa rede de supermercado caiu em mais do que 80% em relação aos dias normais de entregas.

Assuma que foi observado a quantidade de caminhões fazendo entrega a essa rede de supermercado em 5 dias durante a greve e que o desvio padrão populacional dessa quantidade diária de caminhões durante a greve era conhecido e igual a  $\sigma=10$ .

- a) (0,75 ponto) Baseado nas informações descritas no enunciado, formule as hipóteses em termos do problema e em termos estatísticos.
- b) (0,75 ponto) Para um nível de significância de 10%, construa a RC.
- c) (0,75 ponto) Se a amostra de 5 dias durante a greve forneceu média de 60 caminhões abastecendo a tal rede de supermercados diariamente, qual a conclusão deste teste?
- d) (0,75 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):

- ✓ **Id:** número que identifica o imóvel na base de dados.
- ✓ **SalePrice:** preço de venda da propriedade (em dólares). Essa é a variável alvo que você irá tentar prever.
- ✓ **Foundation:** tipo de fundação, com as seguintes categorias:

BrkTil	Tijolo e telha
CBlock	Bloco de concreto
PConc	Concreto usinado
Slab	Laje
Stone	Pedra
Wood	Madeira

Aluno(a): \_\_\_\_\_

Curso: \_\_\_\_\_ Nº de matrícula: \_\_\_\_\_

Turma: \_\_\_\_\_ Professor(a): \_\_\_\_\_

- ✓ **OverallQual:** avalia de maneira geral o material e o acabamento da casa, com as seguintes categorias:

Very Poor	Péssimo
Poor	Ruim
Fair	Justo
Below Average	Abaixo da média
Average	Médio
Above Average	Acima da média
Good	Bom
Very Good	Muito bom
Excellent	Excelente
Very Excellent	Muito excelente
- ✓ **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.
- ✓ **Street:** tipo de acesso rodoviário à propriedade, com as seguintes categorias:

Grvl	Cascalho
Pave	Pavimentado
- ✓ **LandSlope:** inclinação da propriedade, com as seguintes categorias:

Gtl	inclinação suave
Mod	inclinação moderada
Sev	severo decline
- ✓ **RoofStyle:** tipo do telhado, com as seguintes categorias:

Flat	Telhado plano
Gable	Telhado de duas águas
Gambrel	Telhado estilo celeiro
Hip	Telhado de ripa
Mansard	Mansarda
Shed	Galpão
- ✓ **CentralAir:** ar condicionado central, com as seguintes categorias:

N	Não
Y	Sim
- ✓ **YearBuilt:** ano da construção do imóvel.
- ✓ **YrSold:** ano de venda do imóvel.

O arquivo <houseAS.xlsx> traz todas essas 15 variáveis.

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

---