

Data: 29/04/2019

Duração: 1h40m

Prova 2

ICD / DCC-UFMG

Prof. Flavio Vinicius Diniz de Figueiredo

Nome:	Matrícula:
-------------	------------------

A prova é individual e sem consulta. O entendimento correto das questões faz parte da prova. Marcar as respostas a caneta, preenchendo toda a área do quadrado correspondente à alternativa selecionada, sem ultrapassar as bordas.

Aviso: Lembre-se de definir seu tema e equipe na planilha de grupos.

Compute intervalos com confiança de 95%. Ou seja, $z = 1.96$. Uma dica é usar $z = 2$ para deduzir a resposta as respostas.

1) Um cientista político está interessado em estudar um país com três estados. Cada estado contém 10 000, 20 000 e 100 000 cidadãos respectivamente. Para entender a população, uma amostragem de 25, 50 e 75 adultos votantes por estado foi feita. Qual foi o tipo de amostragem utilizado na pesquisa?

- ☐ A Bola de Fogo.
- ☐ B Uniforme com Reposição.
- ☒ C Estratificada.
- ☐ D Bola de Neve.
- ☐ E Nenhuma das alternativas.

2) Dois professores da UFMG estão lecionando a mesma disciplina. Neste semestre, os professores decidiram aplicar a mesma Prova 01. Por conta das atividades complementares noturnas, foi impossível realizar as duas provas no mesmo dia. Por consequência, uma prova foi aplicada na terça e outra na quinta seguinte. Curiosos, os professores decidiram levantar evidências se os alunos que realizaram a prova na terça-feira vazaram as perguntas para os alunos que realizaram a prova na quinta-feira. Assumindo que as duas turmas tem o mesmo número de alunos $n = 81$, foram computadas as estatísticas apresentadas na Tabela 1. Com base na tabela, qual é o valor do erro padrão (*standard error, SE*) de cada prova?

- ☐ A Terça = 0.060, Quinta = 0.074
- ☐ B Terça = 2.039, Quinta = 1.382
- ☐ C Terça = 2.520, Quinta = 3.060
- ☒ D Terça = 0.548, Quinta = 0.666
- ☐ E Terça = 0.226, Quinta = 0.153

Table 1: Dados para o grupo de perguntas PDS2.

	Média \bar{x}	Desvio Padrão s
Prova 1 (Terça)	13.09	2.52
Prova 2 (Quinta)	14.07	3.06

CORRECTED

3) Para realizar afirmativas sobre os dados acima, qual seria uma hipótese nula adequada?

- ☐ A) Nenhuma das alternativas.
- ☐ B) A média da quinta é menor que a média da terça.
- ☒ C) As duas médias são iguais.
- ☐ D) A média da terça é menor que a média da quinta.
- ☐ E) Os dois desvios padrão são diferentes.

4) Com base nas suas respostas acima, podemos afirmar que:

- ☒ A) Não temos evidências suficientes de que as médias são diferentes. Porém ainda é possível que estamos errado e houve sim vazamento.
- ☐ B) Não temos evidências suficientes de que as médias são diferentes. Portanto, temos certeza de que não houve vazamento de questões.
- ☐ C) Como temos evidências de que as médias são estatisticamente diferentes, houve um vazamento de questões pois a turma da terça tem média maior. Temos certeza deste vazamento.
- ☐ D) É impossível falar qualquer coisa usando apenas médias e desvios padrão.
- ☐ E) Como temos evidências de que as médias são estatisticamente diferentes, houve um vazamento de questões pois a turma da quinta tem média maior. De qualquer forma, ainda existe uma chance de erro.

5) Lembrando do Bootstrap, indique a alternativa falsa.

- ☒ A) O Bootstrap define uma hipótese nula.
- ☐ B) O Bootstrap é uma forma de estimar intervalos de confiança.
- ☐ C) Não podemos aplicar o Bootstrap para estimar o erro de valor extremos.
- ☐ D) Podemos testar hipóteses (não necessariamente no arcabouço de hipóteses nulas) com o Bootstrap.
- ☐ E) Para realizar o Bootstrap, precisamos fazer amostras com reposição.

6) Uma das vantagens do Bootstrap em relação as abordagens com base no erro padrão é que:

- ☐ A) Ao realizar amostras com reposição, o Bootstrap captura a média melhor.
- ☐ B) O Bootstrap serve para testar hipóteses.
- ☐ C) Nenhuma das alternativas.
- ☐ D) O Bootstrap é mais correto estimando intervalos com mais chances de capturar uma estatística.
- ☒ E) O Bootstrap pode ser usado para tendências além da média.

7) Assumindo um Bootstrap para a média, qual o motivo que precisamos de muitas amostras no Bootstrap?

- ☐ A) Para ter valores-p mais corretos.
- ☒ B) Para realizar estatísticas suficientes ao redor da média da amostra.
- ☐ C) Nenhuma das alternativas.
- ☐ D) Não sabemos nada da população, então com mais amostras pegamos mais dados da população.
- ☐ E) Para melhor estimar a média da amostra.

CORRECTED

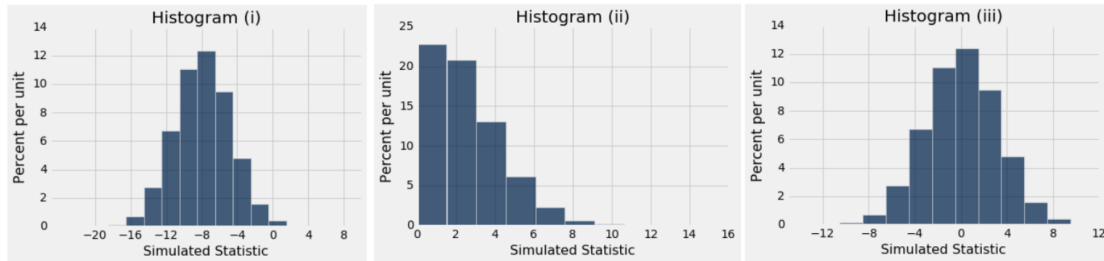


Figure 1: Use na pergunta 09.

8) Pesquisadores estão estudando a eficácia de uma vacina específica contra a gripe. Uma amostragem grande de pessoas foi feita em 2016. Dentre estas, 48% não ficaram doentes com a gripe. Em 2017, uma outra amostra foi feita onde o número de pessoas sem gripe foi de 40%.

Um pesquisador acredita que a vacina foi menos efetiva em 2017 quando comparada com 2016. Sabendo disto, qual é a hipótese nula que o pesquisador deve testar?

- ☐ A A vacina foi igualmente efetiva nas duas populações, a diferença observada nas duas populações foi devida ao acaso.
- ☐ B A vacina foi mais efetiva na amostra de 2016, a diferença da amostra de 2017 foi devida ao acaso.
- ☒ C A vacina foi igualmente efetiva nas duas amostras, a diferença observada nas duas amostras foi devida ao acaso.
- ☐ D Nenhuma das alternativas. Não podemos falar nada sobre vacinas.
- ☐ E A vacina foi menos efetiva na população de 2016, a diferença da população de 2017 foi devida ao acaso.

9) Qual dos histogramas na Figura 1 representa simulações da hipótese nula? Considere a alternativa mais provável. Teoricamente, existe uma chance qualquer uma pode ser verdade. Assuma um teste de permutação. O eixo-x da figura representa a estatística na simulação, o eixo-y é proporcional a quantidade de amostras em cada *bin*.

- ☐ A III
- ☐ B II
- ☒ C I
- ☐ D VI (sim não existe, mas preciso de 5 alternativas)
- ☐ E Nenhuma das alternativas.

10) Um valor-p caputro:

- ☐ A A probabilidade de aceitar a hipótese nula e rejeitar a hipótese alternativa.
- ☐ B A probabilidade de observar valores mais extremos do a estatística observada na hipótese alternativa.
- ☒ C A probabilidade de observar valores mais extremos do que a estatística observada na hipótese nula.
- ☐ D A probabilidade de rejeitar a hipótese nula e aceitar a hipótese alternativa.
- ☐ E A probabilidade de rejeitar a hipótese nula com base na probabilidade da estatística observada.

CORRECTED

11) Um intervalo de confiança com base no erro padrão captura:

- ☒ A média da população com 95% de chance.
- ☐ A mediana da amostra com 5% de chance.
- ☐ A distribuição cumulativa empírica da população com 95% de chance.
- ☐ A mediana da população com 95% de chance.
- ☐ A média da amostra com 95% de chance.

12) Uma turma de introdução à ciência de dados tem 100 alunos. Como parte de uma tarefa, cada aluno testa o viés de uma moeda. Então, usando uma mesma moeda, cada um dos 100 alunos joga a mesma para cima. Com base nos lançamentos, uma hipótese nula é levantada: **H0** a moeda é justa (sem viés). Usando uma significância de 5%, quantos alunos (em média) podem concluir que a moeda não é justa? Assuma que a moeda é sem viés.

- ☐ 15 alunos.
- ☐ 2 alunos.
- ☐ 95 alunos.
- ☐ 50 alunos.
- ☒ 5 alunos.

13) Para computar um erro tipo 2 (β) e o poder do teste ($1 - \beta$) precisamos formular uma hipótese alternativa. Caso esta alternativa seja **H1: a moeda tem um viés de 55% de para caras**. Qual das afirmativas é correta?

- ☒ A chance erro tipo 2 nos 100 lançamentos é alta.
- ☐ Sempre queremos mais poder, assim podemos identificar qualquer viés.
- ☐ Nenhuma das alternativas.
- ☐ Nosso poder é de 100% independente do número de amostras.
- ☐ Com 100 lançamentos já temos poder suficiente (acima de 80%).

14) Escolha uma alternativa abaixo. Esta pergunta vale 0.

- ☐ Correta.
- ☒ Correta.
- ☐ Correta.
- ☐ Correta.
- ☐ Correta.

CORRECTED

15) Para testar se alunos escolhem alternativas de forma uniformemente aleatória, qual seria a hipótese nula correta?

- ☐ A **total variation distance** da resposta acima é igual a um quando comparada com $< .95, .95, .95, .95, .95, .95 >$.
- ☒ A **total variation distance** da resposta acima é igual a zero quando comparada com $< .2, .2, .2, .2, .2, .2 >$.
- ☐ A **total variation distance** da resposta acima é igual a zero quando comparada com $< .5, .5, .5, .5, .5, .5 >$.
- ☐ A **total variation distance** da resposta acima é igual a zero quando comparada com $< .95, .95, .95, .95, .95, .95 >$.
- ☐ A **total variation distance** da resposta acima é igual a um quando comparada com $< .2, .2, .2, .2, .2, .2 >$.

16) Escreva um código para testar a hipótese acima. Pode ser em pseudocódigo, mas tem que ter ou uma permutação ou um bootstrap.

Marque (colorir) toda a área sem ultrapassar as bordas. Use uma caneta (preta ou azul). Caso não tenha, use um grafite marcando de forma bem escura. Por redundância, indique seu nome também!

Nome:

Matrícula (preencha coluna por coluna):

0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

Quadro de respostas

Preencha o quadrado **inteiro** – ■. Não rasure!

Q1:	A	B	■	D	E	Q9:	A	B	■	D	E
Q2:	A	B	C	■	E	Q10:	A	B	■	D	E
Q3:	A	B	■	D	E	Q11:	■	B	C	D	E
Q4:	■	B	C	D	E	Q12:	A	B	C	D	■
Q5:	■	B	C	D	E	Q13:	■	B	C	D	E
Q6:	A	B	C	D	■	Q14:	A	■	C	D	E
Q7:	A	■	C	D	E	Q15:	A	■	C	D	E
Q8:	A	B	■	D	E						