Resolução de Atividades | **Avaliação 5**
Curso: Estatística para Ciência de Dados | Perído: Segundo trimestre 2020
Tutores: Caio, Guilherme, Isis, Marina, Matheus e Tobias | Criado em: 09 de Maio de 2020

Questão 1

Assista o vídeo abaixo, especialmente atento no minuto 1:00.

A descrição da Wikipedia sobre o teste A/B é a seguinte:

(Fonte: `https://pt.wikipedia.org/wiki/Teste_A/B`)

> "**Teste** A/B é um método de teste de *design* através do qual comparam-se elementos aleatórios com duas variantes, A e B, em que estes são o controle e o tratamento de uma experiência controlada, com o objetivo de melhorar a percentagem de aprovação. Estas experiências são muito utilizadas em desenvolvimento web e de marketing, e até mesmo em formas tradicionais de publicidade. Teste A/B também se designa por experiência aleatória controlada, experiência online controlada e teste de divisão. Em web design, o teste A/B é utilizado para identificar alterações nas páginas web que podem provocar mudanças positivas ou negativas no interesse dos utilizadores. Como o nome já diz, duas versões são comparadas, as quais são idênticas exceto por uma variante que pode impactar o comportamento do utilizador. A versão A pode ser a versão utilizada atualmente (controle), enquanto a Versão B é a modificada (tratamento)."

Suponha que um teste A/B será realizado para comparar duas versões diferentes de páginas *web* (aa e bb). Contou-se o número de visitantes em cada página e se clicaram ou não em algum dos *links* desejados, com os seguintes resultados:

	não clicou	**clicou**
aa	4514	498 (9,7%)
bb	4473	527 (10,5%)

Para a comparação dos dois tipos de *design* em relação à resposta dos visitantes (clicks nos *links*), pode-se realizar um teste de hipóteses. O teste de hipóteses que compara se a proporção de visitantes que clica no *link* é a mesma entre aqueles que acessam a página com *design* aa e bb resultou num p-valor de $0,185$.

Assinale a alternativa **incorreta**:

Alternativas:

(a) O teste de hipóteses para a comparação das proporções de clientes com resposta positiva (cliques nos *links*) entre os *designs aa* e *bb* deve ser adequado para comparar proporções de sucesso de variáveis com distribuição de Bernoulli, com parâmetro $p1$, para o *design aa*, e $p2$, para o *design bb*, e pode utilizar o Teorema Central do Limite, para aproximar a distribuição da proporção amostral pela Normal.

(b) Com um nível de significância de $18,5\%$, o *design bb* parece ter maior resposta dos visitantes (cliques nos *links*) e deve ser adotado para a página *web* em questão.

(c) O teste t de Student é um tipo de teste A/B, adequado para a comparação de dois grupos com relação à média de variáveis contínuas com distribuição Normal.

(d) O teste de hipóteses para a comparação das proporções de clientes com resposta positiva (cliques nos *links*) entre os *designs aa* e *bb* deve ser adequado para comparar proporções de sucesso de variáveis com distribuição de Bernoulli, com parâmetro $p1$, para o *design aa*, e $p2$, para o *design bb*.

(e) As proporções de resposta positiva (clicar nos *links*) é equivalente entre as duas versões de *design*, a um nível de significância de 5%.

Solução: **Alternativa b.**

Note que o p-valor é a probabilidade de obter um efeito pelo menos tão extremo quanto aquele em seus dados amostrais, assumindo-se que a hipótese nula seja verdadeira. Quando um p-valor é menor ou igual ao nível de significância, deve-se rejeitar a hipótese nula.

Questão 2

Em um teste de hipóteses bicaudal para a comparação de dois grupos com relação às médias de variáveis aleatórias com distribuição normal, considere α o nível de significância, T a estatística de teste, p-valor o nível descritivo do teste e RC a região crítica do teste. Com base nos dados apresentados, é correto afirmar que:

Alternativas:

(a) Aumentar o nível de significância α do teste, irá diminuir o valor do p-valor.

(b) A hipótese H_0 determina que as médias dos dois grupos são iguais e H_1 determina que as médias são diferentes.

(c) O erro tipo II ocorre quando se rejeita a hipótese nula, sendo ela verdadeira.

(d) O nível de significância é o valor fixado no teste que corresponde à probabilidade de decidir por H_0, quando na verdade H_1 é a hipótese verdadeira.

(e) O nível de significância α do teste é fixado de tal forma que some 100% com o poder do teste.

Solução: **Alternativa b.**

Tem-se que:

- Se a hipótese nula e alternativa de um teste de hipóteses são, respectivamente

$$H_0 : \mu_1 = \mu_2 \quad \text{e} \quad H_0 : \mu_1 \neq \mu_2$$

o teste é chamado de bilateral (ou bicaudal);

- Se a hipótese nula e alternativa de um teste de hipóteses são, respectivamente

$$H_0 : \mu_1 = \mu_2 \quad \text{e} \quad H_0 : \mu_1 < \mu_2$$

  o teste é chamado de unilateral (ou unicaudal à esquerda);

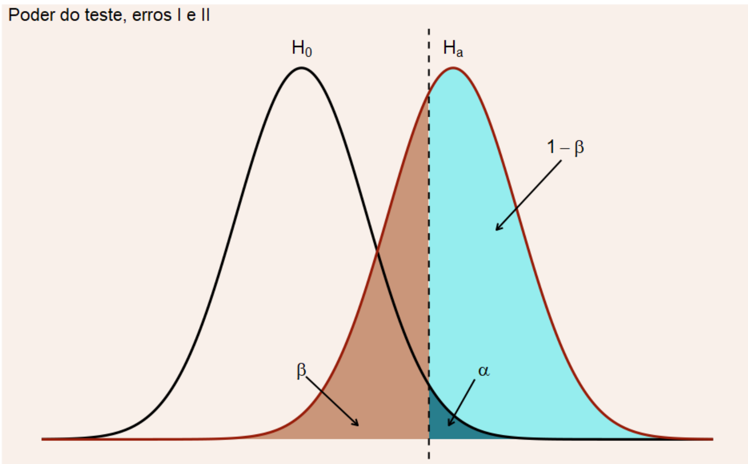
- Se a hipótese nula e alternativa de um teste de hipóteses são, respectivamente

$$H_0 : \mu_1 = \mu_2 \quad \text{e} \quad H_0 : \mu_1 > \mu_2$$

  o teste é chamado de unilateral (ou unicaudal à direita);

- O nível de significância, denotado por α, é a probabilidade de rejeição da hipótese nula quando ela é verdadeira ($P(\text{rejeitar} H_0 | H_0 \text{verdadeira})$ — probabilidade do erro tipo I);
- $\beta = P(\text{não rejeitar} H_0 | H_0 \text{falsa})$ — probabilidade do erro tipo II;
- O poder de um teste, denotado por $1 - \beta$, é a probabilidade do teste de se rejeitar corretamente a hipótese nula ($P(\text{rejeitar} H_0 | H_0 \text{falsa})$).

$\boldsymbol{H_0}$	**Rejeitar**	**Não Rejeitar**
verdadeira	erro tipo I	sem erro
falsa	sem erro	erro tipo II



Fonte: https://www.inf.ufsc.br/ andre.zibetti/probabilidade/teste-de-hipoteses.html

---

Questão 3

Importe as funções shapiro e normaltest do *scipy.stats*
`https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html`
`https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html`
em Python, para fazer 2 diferentes testes de hipóteses de distribuição Normal para as médias do ENEM, com os dados separados entre escolas públicas e privadas. Faça os testes, para cada tipo de escola, com a amostra total das 2901 escolas e, posteriormente, faça uma seleção aleatória de 10% delas (em cada tipo de administração) e repita as análises.

Considerando os resultados apresentados na videoaula e os que você obteve em Python, conforme descrito acima, pode-se concluir que:

Alternativas:

(a) Os testes de normalidade são muito sensíveis com relação ao tamanho amostral, ou seja, a variação da ordem de grandeza do n interfere no resultado do teste.

(b) Para a amostra total, a um nível de significância de 5%, podemos supor distribuição normal para as escolas privadas, mas não para as escolas públicas.

(c) A média amostral da variável média do ENEM tem, para grandes tamanhos de amostra, distribuição normal para as escolas privadas, mas não para as escolas públicas.

(d) As conclusões dos dois testes de normalidade serão sempre as mesmas, ou seja, ambos irão rejeitar ou aceitar que os dados seguem uma distribuição Normal, fixado o nível de significância do teste.

(e) A suposição de distribuição Normal é totalmente inadequada para tal variável para ambos os tipos de escola.

Solução: **Alternativa a.**

Note que variar o tamanho amostral n é um meio para equilibrar as probabilidades dos erros tipo I e II (α e β, respectivamente); mas com n muito grande, α e β podem ficar muito pequenos, detectando diferenças mínimas, quase irrisórias.

Questão 4

Com o conjunto de dados das Cotas dos deputados, usados na videoaula, gere amostras *bootstrap* para obter a distribuição amostral do total máximo reembolsado aos parlamentares nos primeiros 2 meses do ano. Ou seja, primeiro obtenha o valor total reembolsado a cada deputado (lista de exercícios 1 e videoaula 5), depois obtenha amostras *bootstrap* desses valores e, em cada amostra, calcule o máximo.

Alternativas:

(a) Não é possível obter a distribuição amostral desse estimador por *bootstrap*, pois é um método que serve apenas para a média amostral.

(b) A distribuição do estimador do máximo é igual à distribuição da variável amostrada.

(c) A seleção das amostras *bootstrap* é uma amostra aleatória da amostra original.

(d) A maioria das amostras *bootstrap* contém o valor máximo da amostra original.

(e) A distribuição amostral do máximo do total reembolsado aos deputados nos primeiros dois meses do ano é Normal.

Solução: **Alternativa c. Alternativa d.**

Note que a ideia básica do *bootstrap* é reamostrar de um conjunto de dados, diretamente ou via um modelo ajustado (versões não paramétrica ou paramétrica, respectivamente), a fim de criar réplicas dos dados, a partir das quais podemos avaliar a variabilidade de quantidades de interesse (os parâmetros, em geral).

Note também que:

1. A probabilidade de uma observação em particular ser a primeira selecionada para a primeira amostra *bootstrap* é $(1/n)$, sendo n, no nosso caso, 62.
2. Portanto, a probabilidade de essa observação **NÃO** ser a primeira selecionada para a amostra *bootstrap* 1 é $(1 - 1/n)$.
3. A probabilidade dessa observação **NÃO** ser a segunda selecionada para a amostra *bootstrap* é $(1-1/n)$ novamente, pois as seleções são feitas com reposição.

4. Assim sendo, a probabilidade desse observação **NÃO** estar na amostra *bootstrap* é $(1 - 1/n)^n$. Com nosso $n = 62$, isso é igual a $0,36$. Para um n bem maior, tipo 10.000, isso dá aproximadamente $36,8\%$.

Questão 5
No teste de hipóteses da eficácia do Sildenafil dado na videoaula, suponha que a proporção de indivíduos na população que conseguem ao menos 60% de sucessos nas tentativas de relação sexual seja inferior a $0,5$. Se essa proporção for de $0,4$, qual seria a probabilidade de cometer o erro tipo II do teste considerado no slide 9?

Alternativas:

(a) Não é possível calcular.

(b) 0,25

(c) 0,20

(d) 0,056

(e) Seria igual à probabilidade de erro tipo I

Solução: **Alternativa d.**

Tem-se que a estatística do teste é tal que: $\hat{p} \approx N\left(p; \frac{p(1-p)}{n}\right)$.

- Se H_0 é verdadeira, então: $\hat{p} \approx N\left(0,5; \frac{0,5(1-0,5)}{379}\right)$;
- Se H_a é verdadeira, então: $\hat{p} \approx N\left(0,4; \frac{0,4(1-0,4)}{379}\right)$.

A probabilidade do erro tipo II ($P(\text{decidir por} H_0 | H_a \text{ é verdadeira})$) nesse caso é dada por:

$$P(\hat{p} > x_{corte} = 0,44 | p = 0,4) = 1 - P(\hat{p} \leq x_{corte} = 0,44 | p = 0,4) = 1 - 0,944 = 0,056.$$
