Terceira avaliação
Prof. Thiago Gentil Ramires - Probabilidade e Estatística

Nome: Filipe Augusto Parreira Almeida

RA: a2320622

Para todas as questões abaixo, **interprete os resultados e apresente os códigos**. As bases de dados estão em anexo do google Classroom, já salvas em CSV, com o separador decimal em Inglês, ou seja, as decimais estão separadas por ponto. **Cada questão vale 2 pontos.**

---

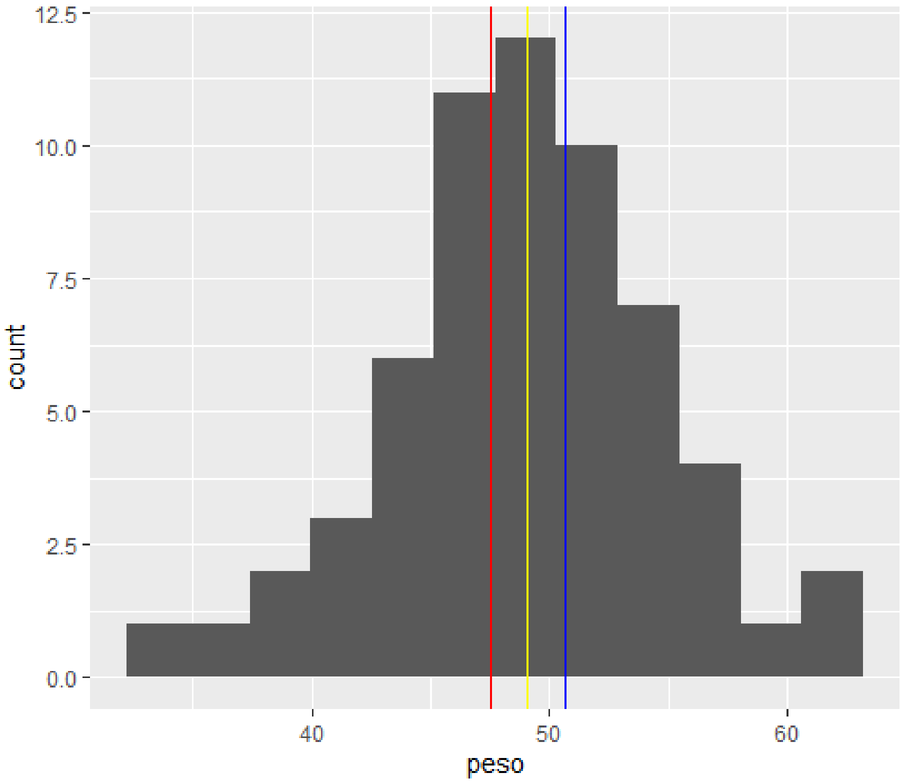
ex1) Considerando os dados p3ex1.csv, construa um intervalo de confiança para a variável peso, considerando 4% de significância. Apresente um histograma da variável, apresentando a média e os limites do intervalo encontrados. (dica: utilize geom_vline() no ggplot para adicionar as linhas)

Resposta: Pode-se concluir, com a análise do gráfico e analisando o intervalo de confiança, que para um nível de 4% de significância, ou seja, 96% de confiança, temos que a verdadeira média de peso está entre 47.51 (linha vermelha) e 50.63 (linha azul), e obtém-se um valor médio de 49.07 (linha amarela), tal que este valor não representa a verdadeira média de pesos.

Códigos:
```
head(dadosEx1)

t.test(dadosEx1$peso, conf.level = 0.96)

ggplot(dadosEx1, aes(peso)) + geom_histogram(bins = 12) +
  geom_vline(xintercept = 50.63513, col = "blue") + geom_vline(xintercept = 47.51532, col = "red") +
  geom_vline(xintercept = 49.07522, col = 'yellow')
```



ex2) Em uma amostra de 100 peças, observou-se quais estavam com o tamanho correto ou não (p3ex2.csv). Determine a estimativa de peças com o tamanho incorreto, bem como uma estimativa confiável para a verdadeira proporção de equipamentos com tamanho incorreto, considerando um nível de significância de 4%. (Dica: Table())

Resposta: Através dos dados, a estimativa de peças com tamanho incorreto é de 53 peças em 100, logo, 53%. Desta forma, aplicando o teste de proporção com um nível de confiança de 96%, conclui-se que a verdadeira estimativa de peças com tamanho incorreto está entre aproximadamente 42% e 63%, ou, 0.4235209 e 0.6339348.

Códigos:
```
head(dadosEx2)

table(dadosEx2$tamanho)

#De 100, 53 estão quebrados, nivel de confiança de 96%

prop.test(53, 100, conf.level = 0.96)
```

ex3) O tempo médio ideal **máximo** para a produção de determinado componente é de **50h ou menos**. Para verificar se uma nova máquina atende os requisitos mínimos, foi realizado o teste em 40 produtos (p3ex3.csv). Verifique, considerando um nível de significância de 5%, se a máquina nova atende o tempo médio ideal, ou seja, se ela será comprada ou não. (Dica: H0: Compra x H1: não compra)

Resposta: H0: média = 50; H1: média > 50.

Considerando um nível de confiança de 95% (0.95), ou seja, um nível de significância de 5% (0.05), pode-se rejeitar a hipótese nula (H0) pois o p-valor obtido foi de aproximadamente 0.0001, ou seja, menor que o α que é de 0.05. Conclui-se então, pelo método de teste de hipótese, que não é indicado comprar a máquina, pois seu tempo médio máximo foi maior que 50h.

Códigos:
```
##Questão 03

head(dadosEx3)

#nivel de confiança de 95% | H0 : media = 50 -> Compra | H1 : media > 50 -> Não Compra

#nivel de significancia 5% ou 0.05, p-valor 0.0001, logo H0 é rejeitada.

t.test(dadosEx3$tempo, alternative = "greater", mu = 50, conf.level = 0.95)
```

ex4) Em um estudo, 30 ligas de metal foram testadas para verificar a resistência **das mesmas, antes e após** a aplicação de um produto. As forças necessárias para o rompimento das ligas foram anotadas (p3ex4.csv). Verifique, considerando 6% de significância, se o produto utilizado teve efeito nas resistências das ligas.

Resposta: H0: mediaAntes – mediaDepois = 0 ou mediaAntes = mediaDepois

H1: mediaAntes – mediaDepois < 0 ou mediaAntes < mediaDepois

Com a análise do teste de hipótese do problema apresentando, considerando um nível de confiança de 94% (0.94), ou seja, um nível de significância (α) de 6% (0.06); obteve-se um p-valor de aproximadamente 0.0002, logo, menor que o valor de α. Sendo assim podemos excluir a hipótese nula (H0) e podemos afirmar que a média de resistência das ligas de metal antes da aplicação do produto é menor que a média de resistência depois da aplicação do produto. Portanto, aplicando o produto nas ligas, obtém-se uma melhor resistência.

```
Códigos: ##Questão 04

head(dadosEx4)

#antes e apos -> dependente

#confiança -> 94% ou 0.94 | alpha -> 6% ou 0.06

#H0 : antes = depois | H1 : antes < depois

t.test(dadosEx4$antes, dadosEx4$depois, alternative = "less", conf.level = 0.94, paired = T)

#p-valor = 0.0002991
```

ex5) Uma empresa pretende comprar uma nova máquina, porém está preocupada com a durabilidade dos itens produzidos pelas máquinas existentes no mercado. Para escolher entre as duas máquinas mais baratas, um teste de produção de 60 peças foi realizado em **cada uma das máquinas** (p3ex5.csv), ou seja, totalizando 120 testes. Determine, ao nível de 8% de significância, se a durabilidade média das máquinas testadas são as mesmas ou não. Caso não, qual a melhor?

Resposta: Teste 1 -> H0: mediaA – mediaB = 0 ou mediaA = mediaB

H1: mediaA – mediaB != 0 ou media A != mediaB

Teste 2 -> H0: mediaA – mediaB = 0 ou mediaA = mediaB

H1: mediaA – mediaB < 0 ou mediaA < mediaB

Para este problema foi realizado dois testes de hipóteses, o primeiro para verificar se a durabilidade média das máquinas testadas são iguais, e o segundo para verificar qual máquina é melhor, ambos os testes considerando um nível de significância (α) de 8% (0.08), ou seja, um nível de confiança de 92%. No primeiro teste, obteve-se um p-valor de 0.3487, logo, maior que o α, sendo assim, não é descartado a hipótese nula (H0), ou seja, não podemos afirmar que a durabilidade média das máquinas são diferentes. Já no segundo teste, obteve-se um p-valor de 0.1743, logo, maior que o α, portanto, não se pode, novamente, descartar a hipótese nula, ou seja, não podemos afirmar que a verdadeira durabilidade média da máquina A é menor que a verdadeira durabilidade média da máquina B, porém, ao obter computacionalmente as durabilidades médias é encontrado que a durabilidade média da máquina A é de 595.2724 e de B é de 612.3158, mas estas média não são as verdadeiras.

Códigos:

```
##Questão 05
head(dadosEx5)
#maquinas diferentes -> independente
#confiança -> 92% ou 0.92 | alpha -> 8% ou 0.08
#H0 : mediaA - mediaB = 0 | H1 : mediaA - mediaB != 0
t.test(dadosEx5$A, dadosEx5$B, alternative = 'two.side', conf.level = 0.92)
#p-valor = 0.3487 > alpha | não rejeita-se H0, não afirma-se que os dois são diferentes


#H0 : mediaA - mediaB = 0 | H1 : mediaA > mediaB
t.test(dadosEx5$A, dadosEx5$B, alternative = 'less', conf.level = 0.92)
#p-valor = 0.1743
```