Atividade Final Inferência

A atividade deverá ser feita em duplas e enviada para o email danilosouto@gmail.com com o assunto UP2020-PTI-INFERENCIA até o dia 23/11 com seu nome completo no corpo do email.

Aluno: Rodrigo Renie de Braga Pinto

Informe o método utilizado e justifique. A formulação da resposta faz parte da avaliação. Exibir o código utilizado no R.

1. A empresa fictícia TI Systems utiliza o tempo de ponto de função para estimar o custo de um sistema. É considerado 2 horas como o custo de mão de obra por medida. Sabe-se que o desvio padrão é de 0,5 hora. No mês passado os tempos por ponto de função coletado foram de:

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

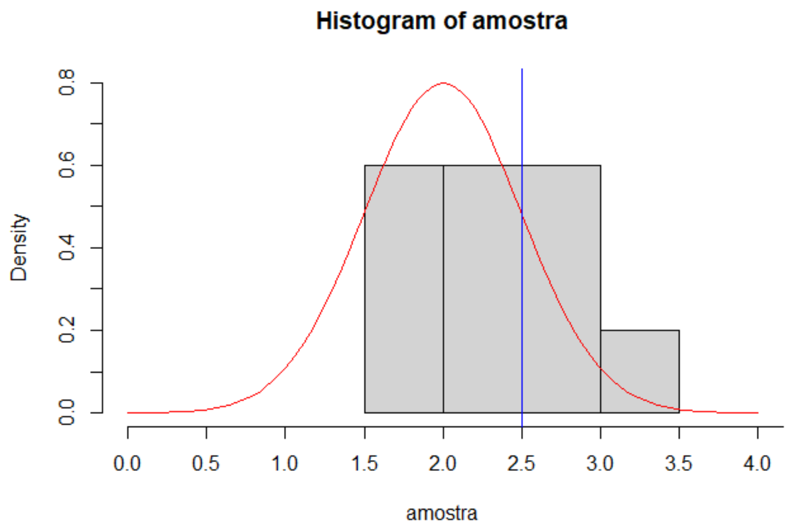
   Usando p=0,05, verifique se o custo excede 2 horas. Qual é sua conclusão e que recomendações você consideraria fazer aos gerentes?

```
library(TeachingDemos)

amostra <- c(1.9, 1.7, 2.8, 2.4, 2.6, 2.5, 2.8, 3.2, 1.6, 2.5)
mu <- 2
sigma <- 0.5

# Realizando os cálculos manualmente

p <- 0.05
k1 <- p/2
k2 <- 1 - p/2
media <- median(amostra)
n <- length(amostra)

Z <- (media - mu)/(sigma / sqrt(n))

Zk2 <- qnorm(k2)

if (Z > Zk2) {
  print('Rejeita-se a Hipósete Nula (H0)')
} else {
  print('Aceita-se a Hipósete Nula (H0)')
}

# Utilizando a função de teste-Z pronta

pvalue <- z.test(x=amostra ,mu=mu, stdev=sigma, alternative='greater')$p.value

if (pvalue >= p) {
  print('Aceita-se a Hipósete Nula (H0)')
} else {
  print('Rejeita-se a Hipósete Nula (H0)')
}

hist(amostra, probability=T, xlim=c(0,4), ylim=c(0,0.8), xaxt='n', yaxt='n')
axis(side=1, at=seq(0, 6, by=0.5), labels=T)
axis(side=2, at=seq(0, 0.8, by=0.1), labels=T)
curve(expr = dnorm(x, mean=mu , sd=sigma), col = "red", add=T)
abline(v=median(amostra), col='blue')
```

Neste exercício é utilizado o teste Z pois o desvio padrão da população é conhecido. As seguintes hipóteses foram formuladas:

- H_0: $\mu = 2$ (os tempos de ponto de função são iguais a 2)
- H_1: $\mu > 2$ (os tempos de ponto de função são maiores que 2)

Como o valor de Z (3.162278) e significativamente maior que o k2 (1.96), rejeitamos a hipótese nula, ou seja, os custos estão excedendo 2 horas, que pode ser confirmado com a análise do gráfico resultante, que indica que os custos estão alocados bem mais à direita do gráfico, indicando que estão maiores que o valor de 2 horas:



2. Uma indústria testa a aplicação de uma nova liga metálica na fabricação de seus produtos. Analisa-se a resistência de 7 linhas de produtos distintas**. Há aumento da resistência dos produtos ao adotar o metal 2?** Utilize 90% de confiança.

Marca	Metal 1	Metal 2
A	68	61
B	75	69
C	62	64
D	86	76
E	52	52
F	46	38
G	72	68

```
metal1 <- c(68, 75, 62, 86, 52, 46, 72)
metal2 <- c(61, 69, 64, 76, 52, 38, 68)
confianca <- 0.9
significancia <- (1 - confianca)

pvalue.v <- var.test(metal1, metal2)$p.value

if (pvalue.v > significancia) {
  var <- TRUE
} else {
  var <- FALSE
}

pvalue.t <- t.test(metal1, metal2, alternative = 'two.sided',
                   conf.level=confianca, var.equal=var)$p.value

# Se o valor do p-value é inferior ou igual ao nível de significância, então
# podemos rejeitar a Hipótese Nula (H0) e aceitar a Hipótese Alternativa (H1).
if (pvalue.t > significancia) {
  print('Aceita-se a Hipósete Nula (H0)')
} else {
  print('Rejeita-se a Hipósete Nula (H0)')
}
```

Neste exercício é utilizado o teste T pela amostra ser menor que 30 e desconhecermos o desvio padrão da população. Para verificar se houve um aumento na resistência do Metal 2 em relação ao Metal 1, foram consideradas as seguintes hipóteses:

- Hipótese Nula: A resistência do Metal 1 é igual à resistência do Metal 2
  - H_0: $\mu_{\text{resistência metal 1}} = \mu_{\text{resistência metal 2}}$
- Hipótese Alternativa: A resistência do Metal 1 é diferente da resistência do Metal 2
  - H_1: $\mu_{\text{resistência metal 1}} \neq \mu_{\text{resistência metal 2}}$

Se for identificado que a resistência dos dois materiais tem diferença significativa, na sequência será considerado as seguintes hipóteses:

- Hipótese Nula: A resistência do Metal 2 é maior que a do Metal 1
  - H_0: $\mu_{\text{resistência metal 2}} > \mu_{\text{resistência metal 1}}$
- Hipótese Alternativa: A resistência do Metal 2 é menor ou igual ao do Metal 1
  - H_1: $\mu_{\text{resistência metal 1}} \leq \mu_{\text{resistência metal 2}}$

Além disso, como o grau de confiança a ser considerado é de 90%, temos que o grau de significância é de 0.1.

Ao executar o teste para analisar a primeira hipótese, temos que:

- significancia = 0.1
- p-value = 0.516655

Portanto, como o valor de p-value é maior que o valor de significância, podemos concluir que a resistência dos dos metais são significativamente iguais e, portanto, **não houve um aumento na resistência ao adotar o Metal 2**.

3. Uma linha de produção tem os seguintes pesos em kg:

5.4, 4.5, 4.7, 4.0, 3.9, 5.3, 5.4, 5.1, 5.9, 7.1, 4.5, 2.7, 6.0, 4.3, 4.3, 6.0
4.7, 3.8, 5.2, 4.9, 5.0, 5.4, 4.6, 5.6, 4.8, 5.3, 6.1, 5.4, 4.7, 6.1, 6.0, 5.5
5.2, 4.4, 6.4, 4.4, 7.2, 6.5, 4.8, 4.0

```
lprod <- c(5.4, 4.5, 4.7, 4.0, 3.9, 5.3, 5.4, 5.1, 5.9, 7.1, 4.5, 2.7, 6.0, 4.3,
           4.3, 6.0, 4.7, 3.8, 5.2, 4.9, 5.0, 5.4, 4.6, 5.6, 4.8, 5.3, 6.1, 5.4,
           4.7, 6.1, 6.0, 5.5, 5.2, 4.4, 6.4, 4.4, 7.2, 6.5, 4.8, 4.0)
```

Calcule no R:

a) Mediana:

```
lprod.median <- median(lprod)

## 5.15
```

b) Média:

```
lprod.mean <- mean(lprod)

## 5.1275
```

c) O desvio padrão:

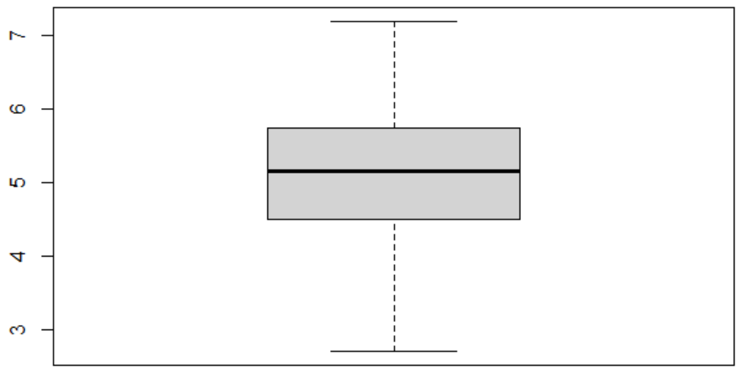
```
lprod.sd <- sd(lprod)


## 0.9229239
```

d) Calcule o 1o quartil, 2o quartil e 3o quartil:

```
lprod.1qt <- as.numeric(quantile(lprod)['25%'])
lprod.2qt <- as.numeric(quantile(lprod)['50%'])
lprod.3qt <- as.numeric(quantile(lprod)['75%'])

## lprod.1qt: 4.5
## lprod.2qt: 5.15
## lprod.3qt: 5.675
```

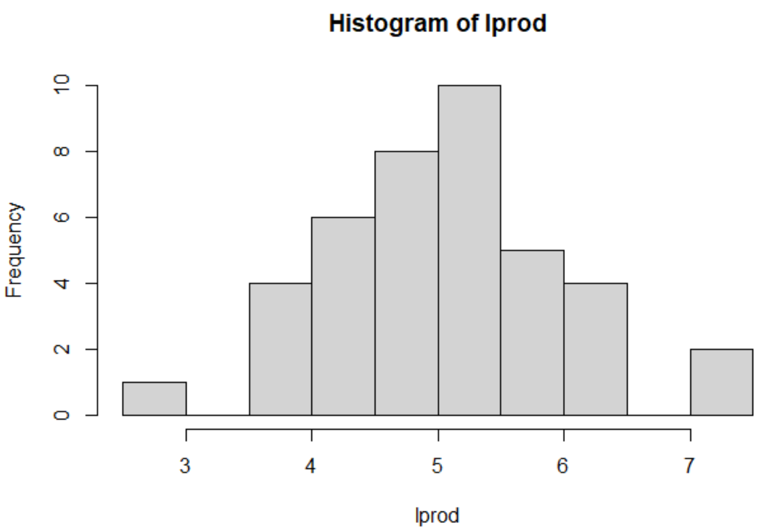
e) Plote o *boxplot* (diagrama de caixa) da amostra:

```
boxplot(lprod)
```



f) plote o histograma da amostra:

```
hist(lprod)
```



g) qual a probabilidade um produto ter um peso entre 4.2 e 5.2?

```
lprod <- c(5.4, 4.5, 4.7, 4.0, 3.9, 5.3, 5.4, 5.1, 5.9, 7.1, 4.5, 2.7, 6.0, 4.3,
           4.3, 6.0, 4.7, 3.8, 5.2, 4.9, 5.0, 5.4, 4.6, 5.6, 4.8, 5.3, 6.1, 5.4,
           4.7, 6.1, 6.0, 5.5, 5.2, 4.4, 6.4, 4.4, 7.2, 6.5, 4.8, 4.0)

lprod.median <- median(lprod)
lprod.mean <- mean(lprod)
lprod.sd <- sd(lprod)

## A probabilidade de X estar entre 4.2 e 5.2 é igual
## a probabilidade de X ser menor que 5.2 menos
## a probabilidade do X ser menor que 4.2, portanto:
## P(4.2 < X < 5.2) = P(X < 5.2) - P(X < 4.2)

getz <- function(x, mean=0, sd=0) {
  if (missing(mean)) {
    mean <- mean(x)
  }

  if (missing(sd)) {
    sd <- sd(x)
  }

  return ((x - mean)/sd)
}

PZ_5.2 <- pnorm(getz(5.2, lprod.mean, lprod.sd))
PZ_4.2 <- pnorm(getz(4.2, lprod.mean, lprod.sd))
P <- PZ_5.2 - PZ_4.2

curve(dnorm(x, mean=mean(lprod), sd=sd(lprod)),
      from = min(lprod), to = max(lprod),
      ylab="Densidade", xlab="Peso", xaxt='n')
axis(side=1, at=seq(min(lprod), max(lprod), by=0.5), labels=T)

seqx <- seq(4.2, 5.2, by=0.2)
xp <- c(4.2, seqx, 5.2)
yp <- c(0, dnorm(seqx, lprod.mean, lprod.sd), 0)
polygon(xp, yp, col="green")
abline(v=lprod.median, lty="dashed")
text(5, 0.05, sprintf("P = %.2f", P) , adj=c(1,0))

## A probabilidade de X estar entre 4.2 e 5.2 é de 0.3738481 (37.38%)
```