

title: “TESTE DE HIPÓTESE E INTERVALO DE CONFIANÇA”

author: “Leonardo Gomes Lara”

date: Sexta Feira, 18 de Abril 2015"

output: html_document

ATIVIDADE AVALIATIVA DE ANÁLISE DE DADOS

1. A resolução dos exercícios deve ser elaborada em um arquivo RMarkdown.
2. Formule os problemas como um teste de hipótese e como o cálculo de um Intervalo de Confiança (IC).
3. Faça uma análise exploratória (breve).
4. As conclusões devem ser claramente estabelecidas.

Lista de Exercícios

Tema: Intervalos de Confiança e Testes de Hipóteses

1. A autorização de despejo para uma indústria requer que a concentração média mensal de Concentração de Oxigênio Dissolvido (COD) seja inferior a 50 mg/L. A indústria quer que isso seja interpretado como “50 mg/L está dentro do intervalo de confiança da média, que vai ser estimada a partir de 20 observações por mês.” Para as 20 observações seguintes, estaria a indústria em conformidade com esta interpretação do padrão?

```
# Inserindo os dados
cod <- c(57, 60, 49, 50, 51, 60, 49, 53, 49, 56, 64, 60, 49, 52, 69, 40, 44, 38, 53, 66)

# Examinando dos dados

summary(cod)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	38.00	49.00	52.50	53.45	60.00	69.00

```
# Ordenando o vetor de menor valor para o maior valor

sort(cod)
```

```
## [1] 38 40 44 49 49 49 49 50 51 52 53 53 56 57 60 60 60 64 66 69
```

```
# Dados gráfico de ramos e folhas
```

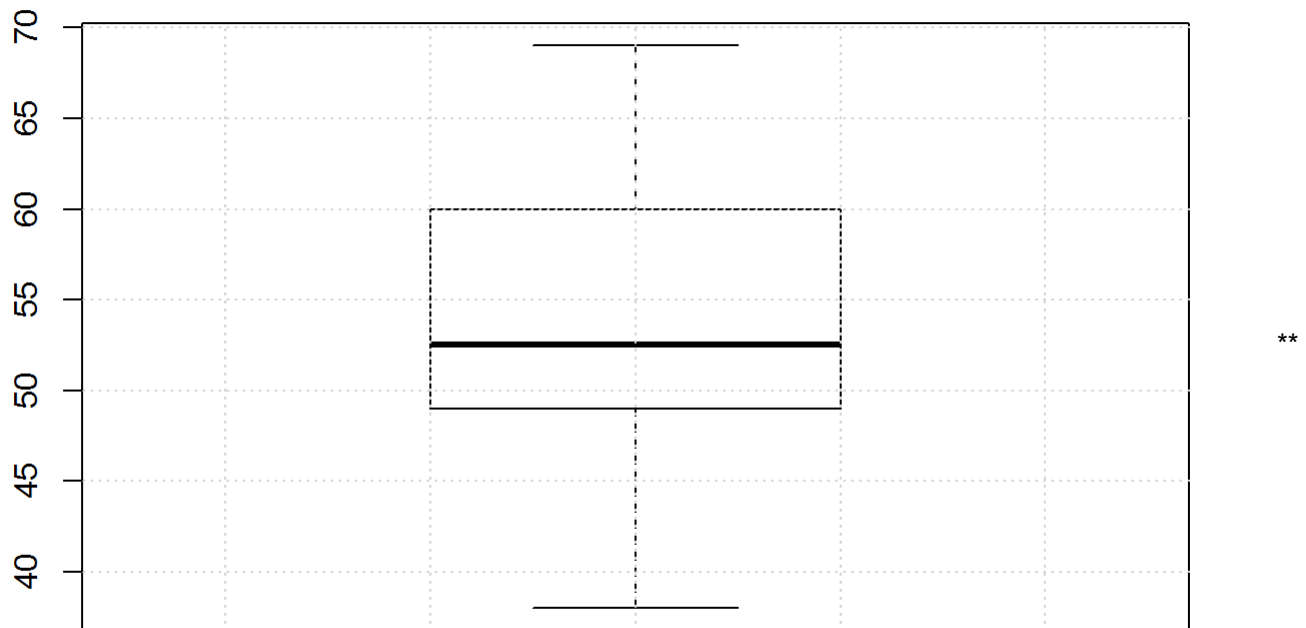
```
stem(cod)
```

```
##  
## The decimal point is 1 digit(s) to the right of the |  
##  
## 3 | 8  
## 4 | 049999  
## 5 | 0123367  
## 6 | 000469
```

```
# Observando os dados do gráfico boxplot
```

```
boxplot(cod,main="Gráfico Boxplot")  
grid()
```

Gráfico Boxplot



Considerações **

Ao analisar os dados apresentados pelo gráfico acima, percebe-se que há uma assimetria que não permite que as análises sejam conclusivas. Portanto, percebe-se importante verificar o intervalo de confiança.

```
t.test(cod, alternative = c ("two.sided") , mu = 50 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data:  cod
## t = 1.8894, df = 19, p-value = 0.07419
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  49.62825 57.27175
## sample estimates:
## mean of x
##      53.45
```

```
wilcox.test(cod,mu=50,conf.int=TRUE)
```

```
## Warning in wilcox.test.default(cod, mu = 50, conf.int = TRUE): cannot
## compute exact p-value with ties
```

```
## Warning in wilcox.test.default(cod, mu = 50, conf.int = TRUE): cannot
## compute exact confidence interval with ties
```

```
## Warning in wilcox.test.default(cod, mu = 50, conf.int = TRUE): cannot
## compute exact p-value with zeroes
```

```
## Warning in wilcox.test.default(cod, mu = 50, conf.int = TRUE): cannot
## compute exact confidence interval with zeroes
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  cod
## V = 139, p-value = 0.07905
## alternative hypothesis: true location is not equal to 50
## 95 percent confidence interval:
##  49.00001 58.00001
## sample estimates:
## (pseudo)median
##      53.66572
```

```
# Desvio Padrão
```

```
sd(cod)
```

```
## [1] 8.165879
```

**** Conclusão ****

Conclui-se que, por meio dos testes apresentados das amostras acima, os padrões de concentração média mensal de COD do período analisado, estão acima do estipulado pela autorização de despejo.

2. Um protocolo de garantia de qualidade laboratorial exige que as soluções padrão tenham 50 mg/l de COT - Concentração de Oxigênio Total, tais amostras são inseridas de forma aleatória no fluxo de trabalho. Os analistas são cegos a estas normas. Estime o viés e a precisão das 16 observações mais recentes deste padrão. A processo de medição de COT está em conformidade com o padrão? c(50.3, 51.2, 50.5, 50.2, 49.9, 50.2, 50.3, 50.5, 49.3, 50.0, 50.4, 50.1, 51.0, 49.8, 50.7, 50.2, 50.3, 50.5, 49.3, 50.0, 50.4, 50.1, 51.0, 49.8, 50.7, 50.6)

```
cot <- c(50.3, 51.2, 50.5, 50.2, 49.9, 50.2, 50.3, 50.5, 49.3, 50.0, 50.4, 50.1, 51.0, 49.8, 50.7, 50.2, 50.3, 50.5, 49.3, 50.0, 50.4, 50.1, 51.0, 49.8, 50.7, 50.6)
```

```
summary(cot)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  49.30   50.08   50.30   50.31   50.52   51.20
```

```
sort(cot)
```

```
## [1] 49.3 49.8 49.9 50.0 50.1 50.2 50.2 50.3 50.3 50.4 50.5 50.5 50.6 50.7
## [15] 51.0 51.2
```

```
mean(cot)
```

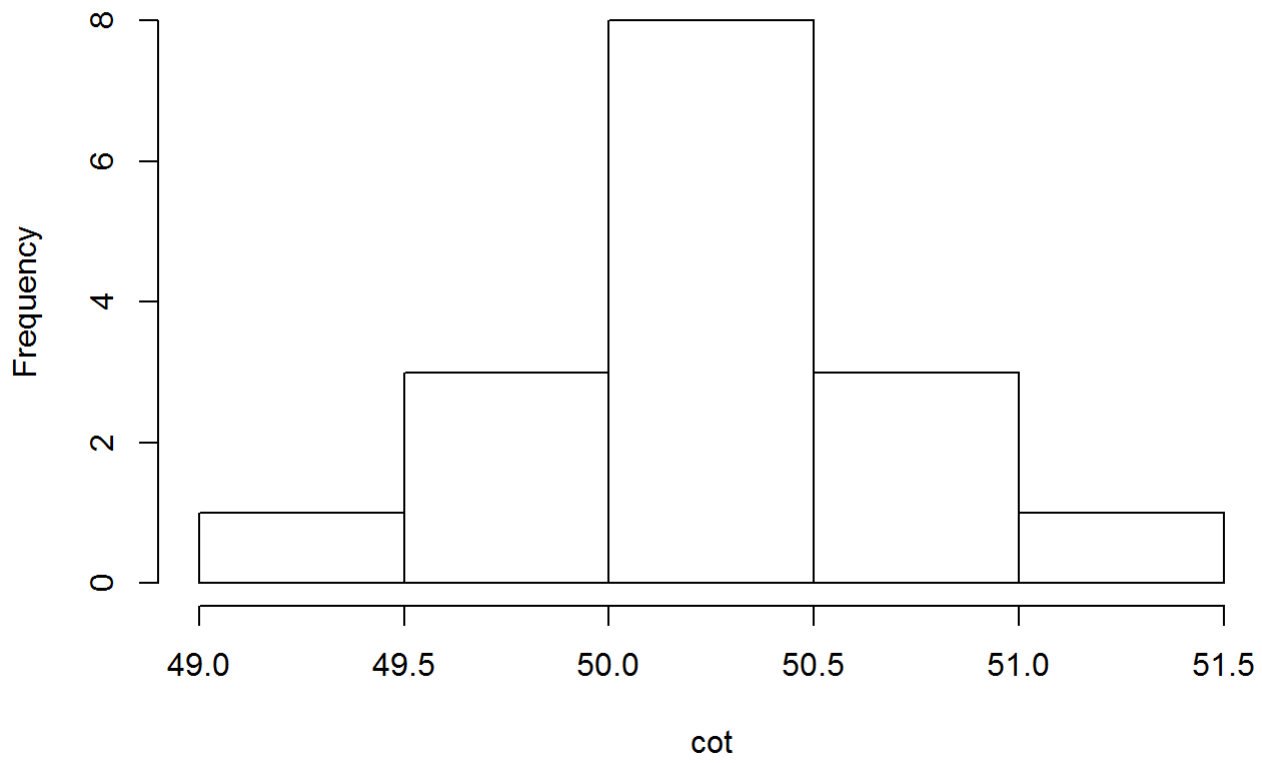
```
## [1] 50.3125
```

```
stem(cot)
```

```
##
## The decimal point is at the |
##
## 49 | 3
## 49 | 89
## 50 | 0122334
## 50 | 5567
## 51 | 02
```

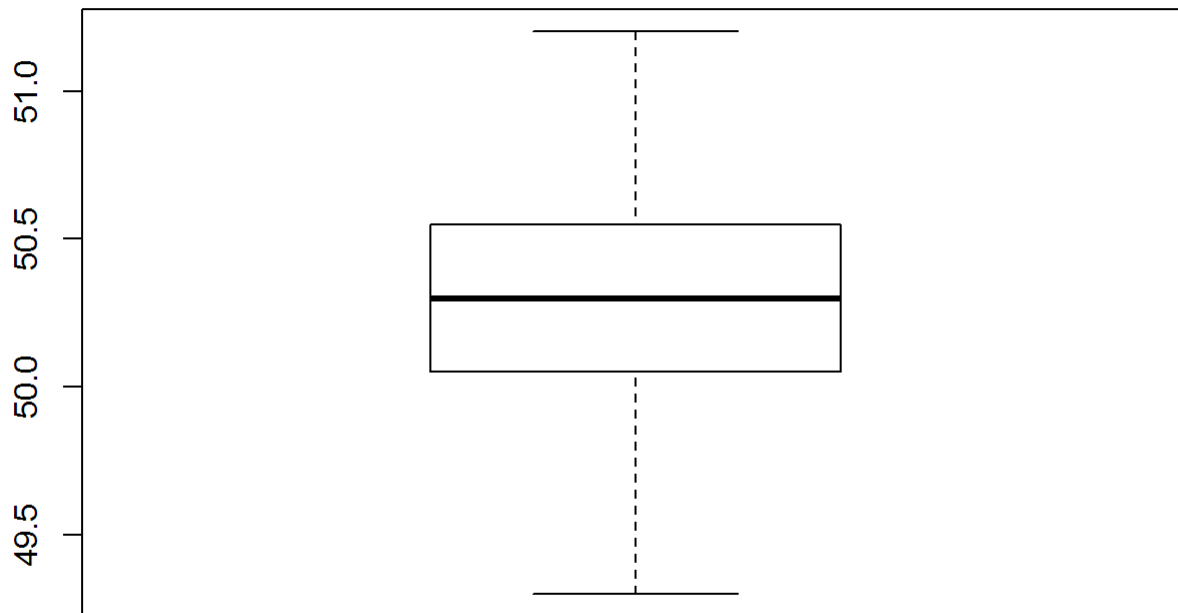
```
hist(cot)
```

Histogram of cot



```
boxplot(cot, main="Gráfico Boxplot")
```

Gráfico Boxplot



Percebe-se que há uma certa simetria dos dados amostrados no fluxo de trabalho. Portanto, esses dados não são conclusivos, sendo necessário efetuar o intervalo de confiança e teste de hipótese.

```
t.test(cot, alternative = c ("two.sided") , mu = 50 , conf.level = 0.95)
```

```
##  
## One Sample t-test  
##  
## data: cot  
## t = 2.7074, df = 15, p-value = 0.01622  
## alternative hypothesis: true mean is not equal to 50  
## 95 percent confidence interval:  
## 50.06648 50.55852  
## sample estimates:  
## mean of x  
## 50.3125
```

```
t.test(cot, alternative = c ("less") , mu = 50 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: cot
## t = 2.7074, df = 15, p-value = 0.9919
## alternative hypothesis: true mean is less than 50
## 95 percent confidence interval:
##      -Inf 50.51485
## sample estimates:
## mean of x
## 50.3125
```

```
t.test(cot, alternative = c ("great") , mu = 50 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: cot
## t = 2.7074, df = 15, p-value = 0.008108
## alternative hypothesis: true mean is greater than 50
## 95 percent confidence interval:
## 50.11015      Inf
## sample estimates:
## mean of x
## 50.3125
```

```
wilcox.test(cot,mu=50,conf.int=TRUE)
```

```
## Warning in wilcox.test.default(cot, mu = 50, conf.int = TRUE): cannot
## compute exact p-value with ties
```

```
## Warning in wilcox.test.default(cot, mu = 50, conf.int = TRUE): cannot
## compute exact confidence interval with ties
```

```
## Warning in wilcox.test.default(cot, mu = 50, conf.int = TRUE): cannot
## compute exact p-value with zeroes
```

```
## Warning in wilcox.test.default(cot, mu = 50, conf.int = TRUE): cannot
## compute exact confidence interval with zeroes
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: cot
## V = 102, p-value = 0.01823
## alternative hypothesis: true location is not equal to 50
## 95 percent confidence interval:
## 50.09998 50.59994
## sample estimates:
## (pseudo)median
## 50.34994
```

Conforme análise realizada pelos métodos t.teste = 50.3125, verificou-se que o viés é 0.3125, sendo confirmado o resultado pelo wilcox.test = 50.34994. Portanto, o processo de medição de COT está próximo da conformidade do padrão.

3. O gás produzido a partir da fermentação biológica é oferecido para a venda com a garantia de que o teor médio de metano de 72%. Uma amostra aleatória de n = 7 amostras de gás forneceu conteúdo de metano (em %) de 64, 65, 75, 67, 65, 74 e 75.

-efetue testes de hipóteses com níveis de significância de 0,10, 0,05, e 0,01 para determinar se é justo reivindicar uma média de 72.

-Calcular intervalos de confiança de 90%, 95% e 99% para avaliar a alegação de uma média de 72%.

```
met <- c(64, 65, 75, 67, 65, 74, 75)

summary(met)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  64.00   65.00   67.00   69.29   74.50   75.00
```

```
sort(met)
```

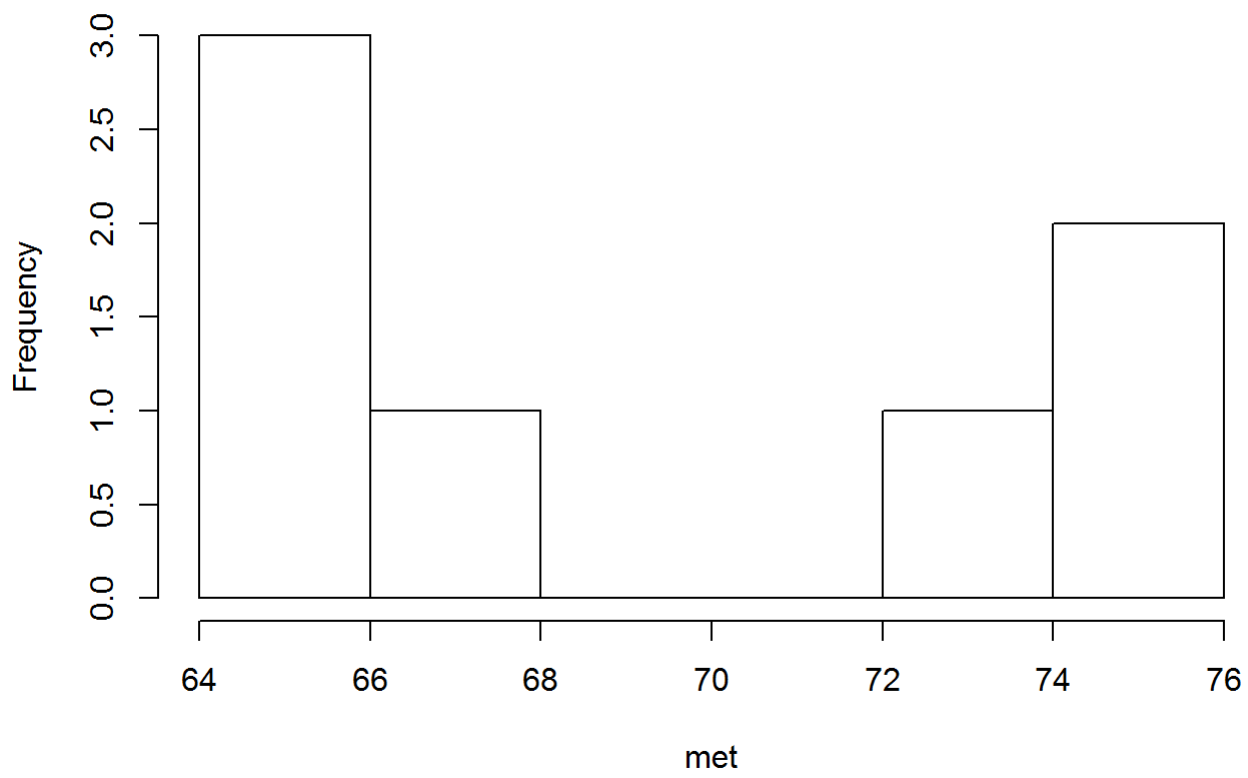
```
## [1] 64 65 65 67 74 75 75
```

```
stem(met)
```

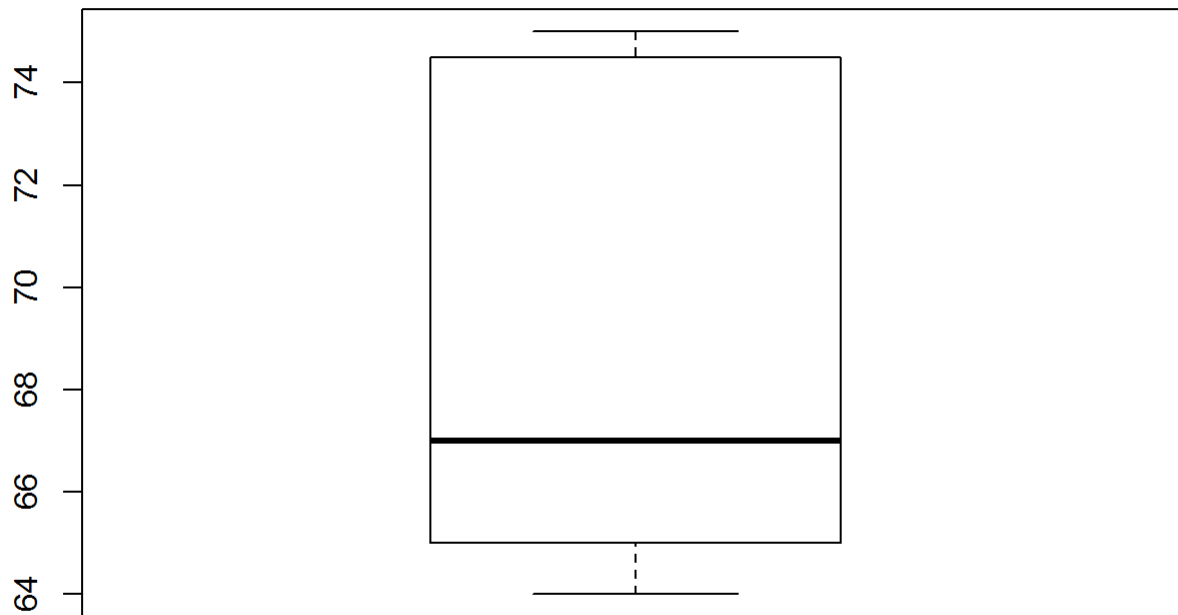
```
##
## The decimal point is 1 digit(s) to the right of the |
##
##  6 | 4
##  6 | 557
##  7 | 4
##  7 | 55
```

```
hist(met)
```


Histogram of met



```
boxplot(met)
```



*** Pré-análise dos dados ***

A distribuição dos dados apresentados estão aparentemente distribuídas assimetricamente, onde os dados não apresentam uma distribuição normal. É necessário a aplicação do teste de hipótese e intervalo de confiança.

```
t.test(met, alternative = c ("two.sided") , mu = 72 , conf.level = 0.90)
```

```
##
## One Sample t-test
##
## data: met
## t = -1.402, df = 6, p-value = 0.2105
## alternative hypothesis: true mean is not equal to 72
## 90 percent confidence interval:
##  65.52362 73.04781
## sample estimates:
## mean of x
##  69.28571
```

```
t.test(met, alternative = c ("two.sided") , mu = 72 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: met
## t = -1.402, df = 6, p-value = 0.2105
## alternative hypothesis: true mean is not equal to 72
## 95 percent confidence interval:
## 64.54836 74.02306
## sample estimates:
## mean of x
## 69.28571
```

```
t.test(met, alternative = c ("two.sided") , mu = 72 , conf.level = 0.99)
```

```
##
## One Sample t-test
##
## data: met
## t = -1.402, df = 6, p-value = 0.2105
## alternative hypothesis: true mean is not equal to 72
## 99 percent confidence interval:
## 62.10794 76.46349
## sample estimates:
## mean of x
## 69.28571
```

```
wilcox.test(met,mu=72.0,conf.int=TRUE)
```

```
## Warning in wilcox.test.default(met, mu = 72, conf.int = TRUE): requested
## conf.level not achievable
```

```
## Warning in wilcox.test.default(met, mu = 72, conf.int = TRUE): cannot
## compute exact p-value with ties
```

```
## Warning in wilcox.test.default(met, mu = 72, conf.int = TRUE): cannot
## compute exact confidence interval with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: met
## V = 6, p-value = 0.2033
## alternative hypothesis: true location is not equal to 72
## 90 percent confidence interval:
## 64.99999 74.50002
## sample estimates:
## (pseudo)median
## 69.50002
```

**** conclusões ****

Os testes demonstraram que quanto maior o valor de alfa, menor é o intervalo de confiança (amplitude dos dados), O valor da média amostral foi de 69,28, abaixo da valor solicitado, conforme demonstrado no gráfico de boxplot. O teste de Wilcox confirma o descrito acima com 69,5.

4. Os seguintes dados foram obtidos a partir de medidas pareadas de nitrito em água e em águas residuais por eletrodo direto de íon-seletivo e um método colorimétrico. Os dois métodos forneceram resultados consistentes?

Método: ISE

Medidas de Nitrito: 0.32 0.36 0.24 0.11 0.11 0.44 2.79 2.99 3.47

Método: Colorimetric

**** Medidas de Nitrico: 0.36 0.37 0.21 0.09 0.11 0.42 2.77 2.91 3.52****

**** Método ISE ****

```
ISE <- c(0.32, 0.36, 0.24, 0.11, 0.11, 0.44, 2.79, 2.99, 3.47)

summary(ISE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110   0.240   0.360   1.203   2.790   3.470
```

```
sort(ISE)
```

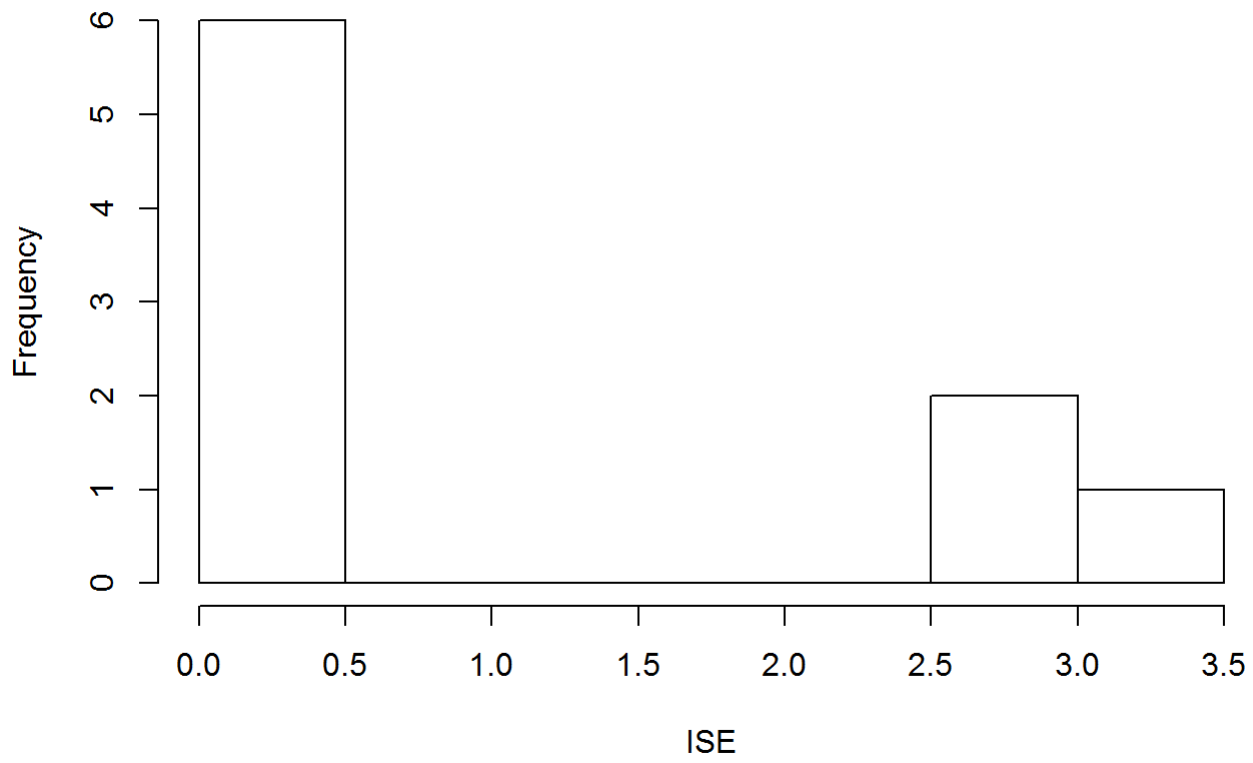
```
## [1] 0.11 0.11 0.24 0.32 0.36 0.44 2.79 2.99 3.47
```

```
stem(ISE)
```

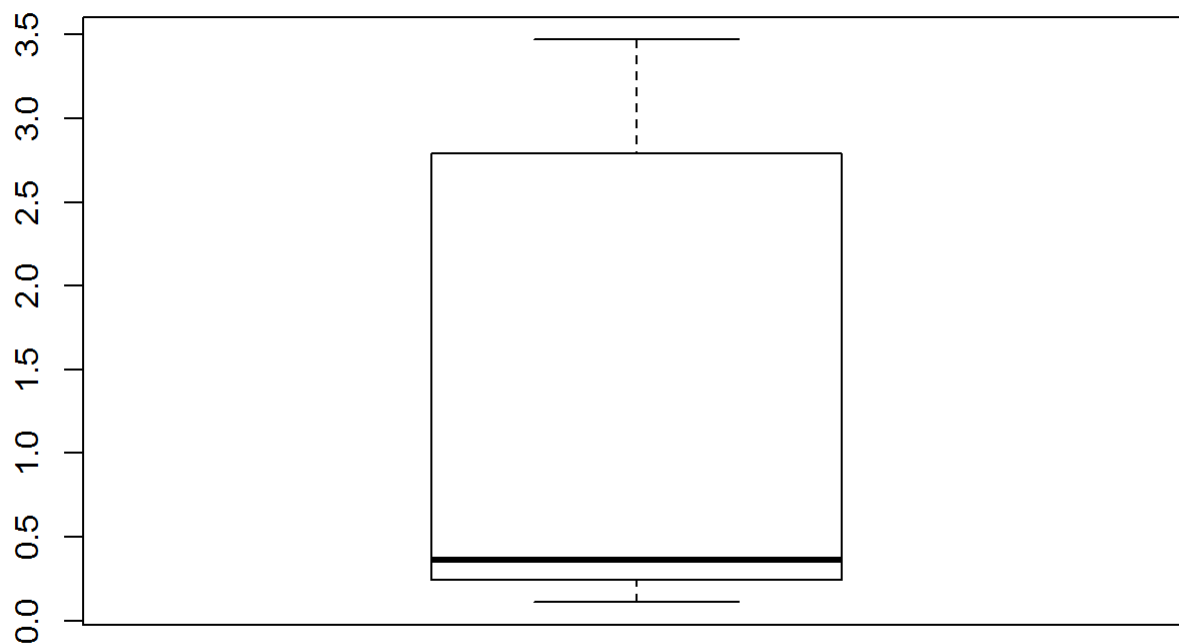
```
##
##  The decimal point is at the |
##
##  0 | 112344
##  1 |
##  2 | 8
##  3 | 05
```

```
hist(ISE)
```

Histogram of ISE



```
boxplot(ISE)
```



```
t.test(ISE, alternative = c ("two.sided") , mu = 0.0 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: ISE
## t = 2.5337, df = 8, p-value = 0.03505
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1081365 2.2985302
## sample estimates:
## mean of x
##  1.203333
```

Método: Colorimetric

```
color <- c(0.36, 0.37, 0.21, 0.09, 0.11, 0.42, 2.77, 2.91, 3.52)

summary(color)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.090   0.210   0.370   1.196   2.770   3.520
```

```
sort(color)
```

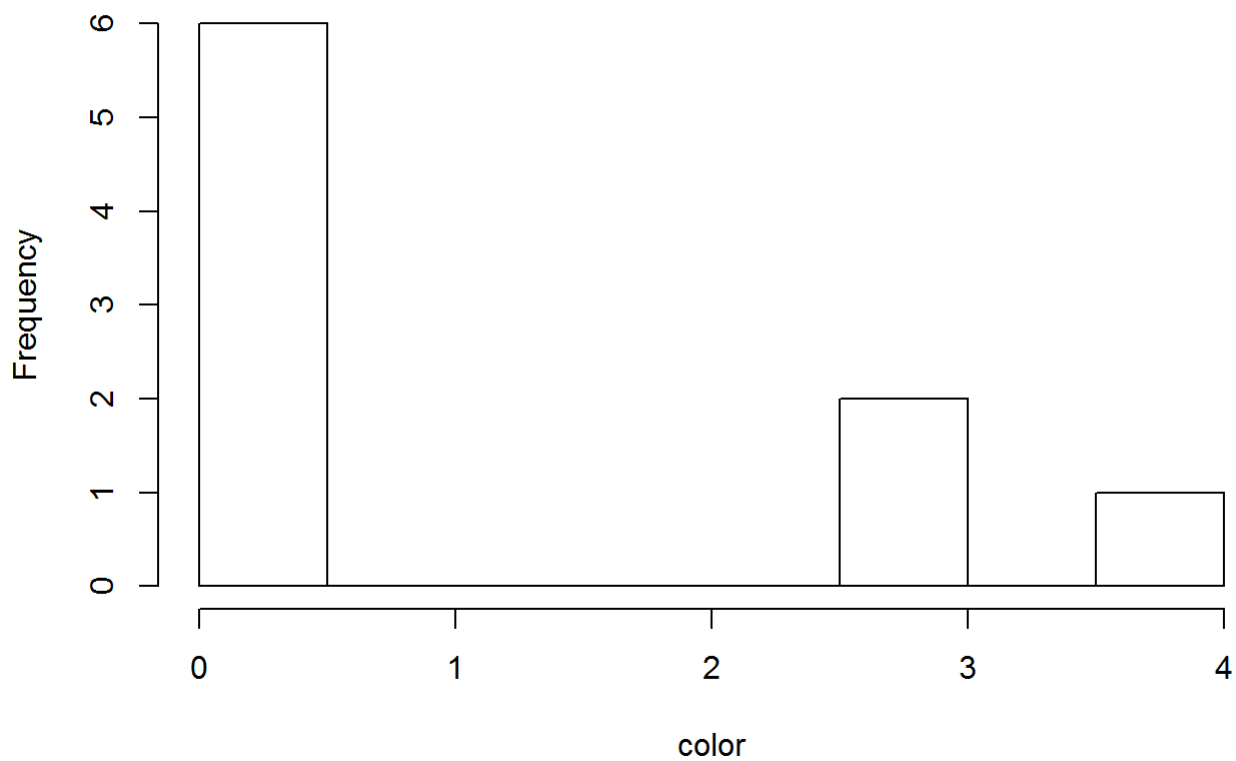
```
## [1] 0.09 0.11 0.21 0.36 0.37 0.42 2.77 2.91 3.52
```

```
stem(color)
```

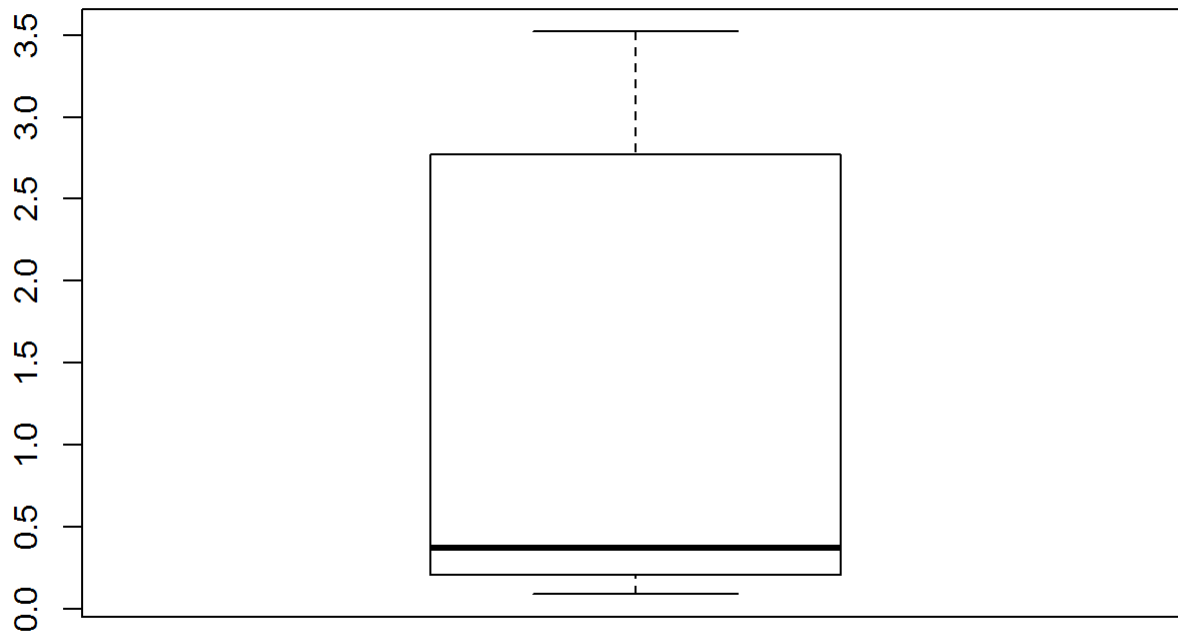
```
##  
## The decimal point is at the |  
##  
## 0 | 112444  
## 1 |  
## 2 | 89  
## 3 | 5
```

```
hist(color)
```

Histogram of color



```
boxplot(color)
```



```
t.test(color, alternative = c ("two.sided") , mu = 0.0 , conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data:  color
## t = 2.5224, df = 8, p-value = 0.03567
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1025856 2.2885255
## sample estimates:
## mean of x
##  1.195556
```

**** Conclusões ****

A analisar os dois métodos, percebe-se que eles são pereados, pois contêm o mesmo número de amostras. Observa-se que os gráficos apresentam dados com assimetria, porém, esses dados não podem ser conclusivos. Ao realizar o t.test nos dois métodos encontramos dados consistentes, pois os resultados estão próximos às médias.

5. Limites extremamente baixos existem agora para metais pesados em limites de águas residuais de efluentes. Pensa-se frequentemente que, sempre que a concentração de metais pesados é demasiado elevada, o problema pode ser corrigido forçando indústrias a interromper a descarga da substância nociva.

E possível, no entanto, que a concentração alvo do efluente ser tão baixa que elas podem ser ultrapassadas pela concentração em esgoto doméstico. Amostras de água potável foram coletadas de dois bairros residenciais, um servido pelo abastecimento de água da cidade e outro servido por poços particulares. As concentrações de mercúrio observados estão listados na tabela abaixo.

**** Para estudos futuros sobre concentrações de mercúrio em áreas residenciais, seria conveniente ser capaz de coletar dados em qualquer bairro, sem ter que se preocupar que o abastecimento de água pode afetar o resultado. Existe alguma diferença no teor de mercúrio das duas áreas residenciais?****

Origem: Cidade (n = 13)

Medidas de mercúrio: 0.34 0.18 0.13 0.09 0.16 0.09 0.16 0.10 0.14 0.26 0.06 0.26 0.07

Origem: Part. (n=10)

**** Medidas de mercúrio: 0.26 0.06 0.16 0.19 0.32 0.16 0.08 0.05 0.10 0.13****

```
cidade <- c(0.34, 0.18, 0.13, 0.09, 0.16, 0.09, 0.16, 0.10, 0.14, 0.26, 0.06, 0.26, 0.07)
```

```
Part <- c(0.26, 0.06, 0.16, 0.19, 0.32, 0.16, 0.08, 0.05, 0.10, 0.13)
```

```
summary(cidade)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06000 0.09000 0.14000 0.1569 0.18000 0.34000
```

```
summary(Part)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.05000 0.08500 0.14500 0.1510 0.1825 0.32000
```

```
sort(cidade)
```

```
## [1] 0.06 0.07 0.09 0.09 0.10 0.13 0.14 0.16 0.16 0.18 0.26 0.26 0.34
```

```
sort(Part)
```

```
## [1] 0.05 0.06 0.08 0.10 0.13 0.16 0.16 0.19 0.26 0.32
```

```
stem(cidade)
```

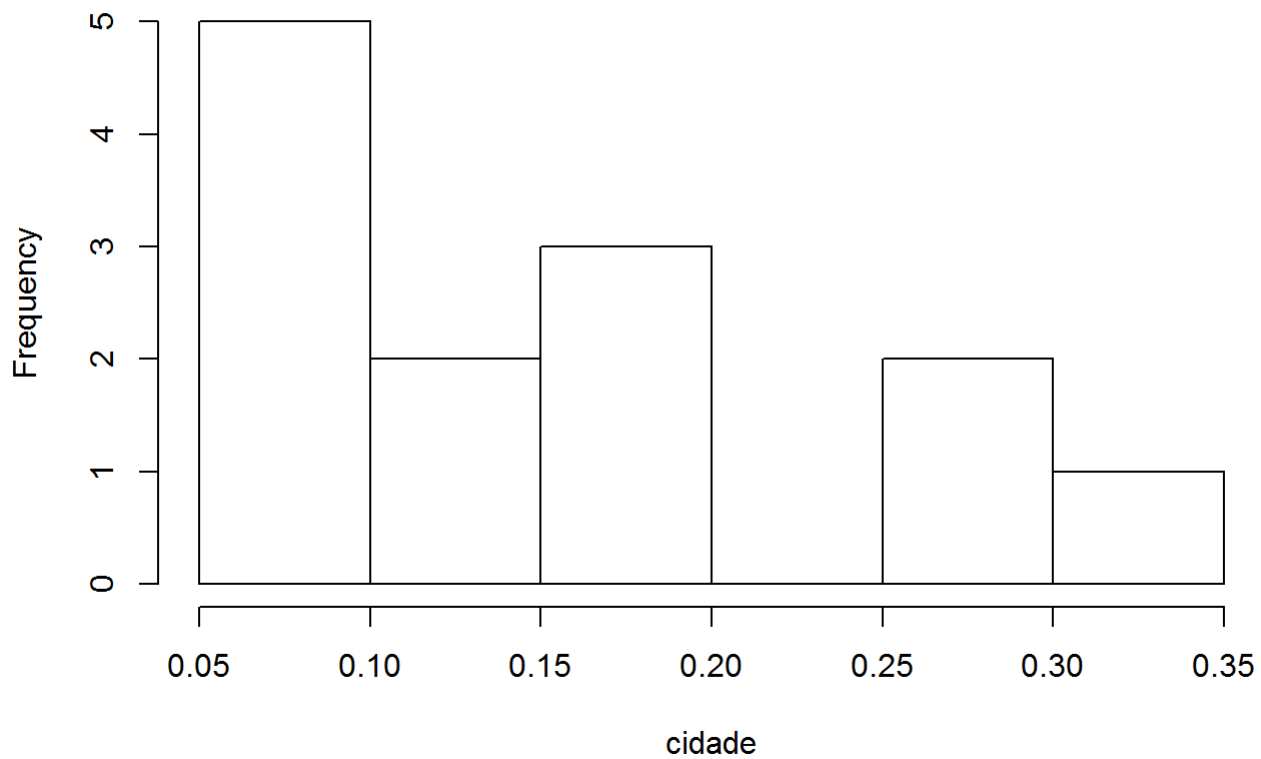
```
##
## The decimal point is 1 digit(s) to the left of the |
##
## 0 | 6799
## 1 | 034668
## 2 | 66
## 3 | 4
```

```
stem(Part)
```

```
##  
## The decimal point is 1 digit(s) to the left of the |  
##  
## 0 | 568  
## 1 | 03669  
## 2 | 6  
## 3 | 2
```

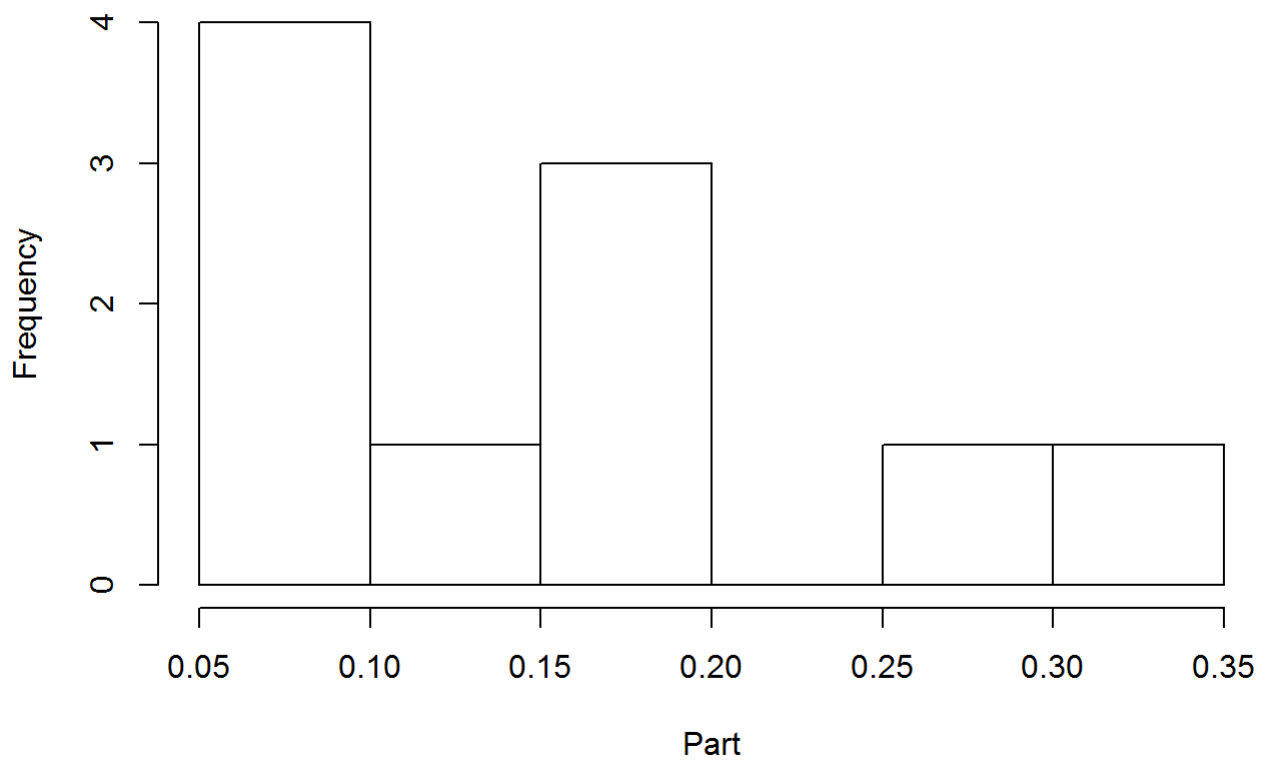
```
hist(cidade)
```

Histogram of cidade

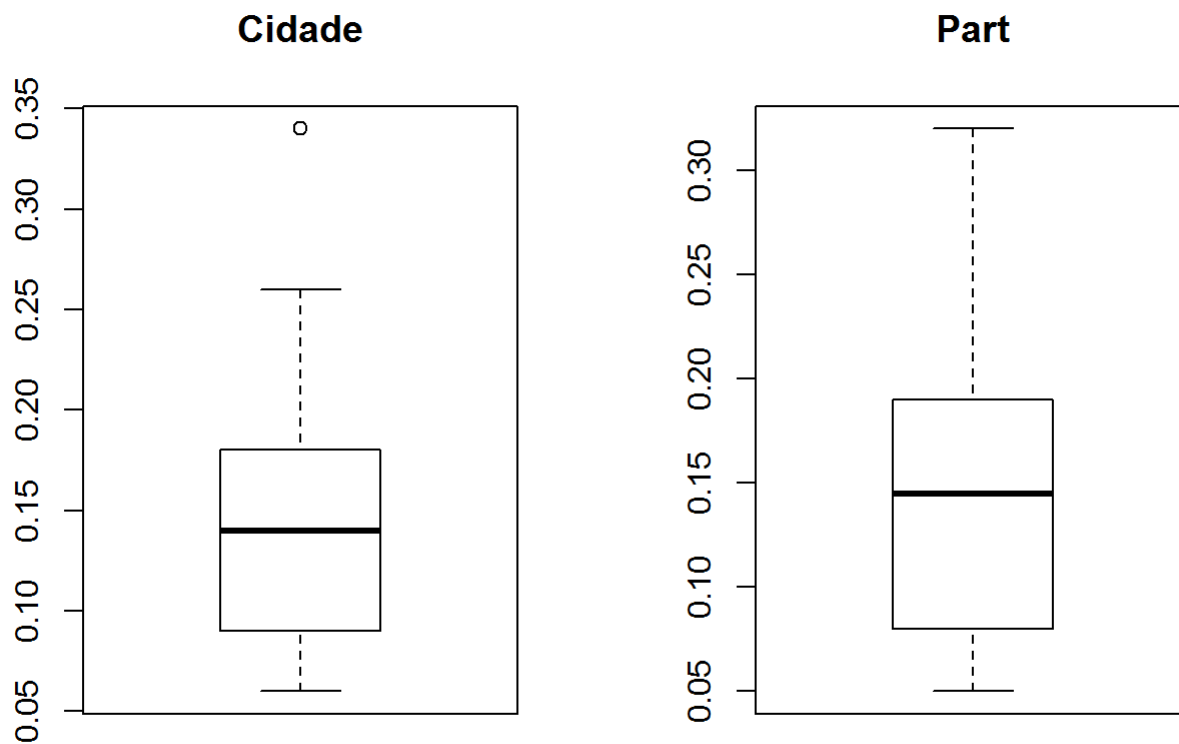


```
hist(Part)
```

Histogram of Part



```
par(mfrow=c(1,2))  
  
boxplot(cidade,main="Cidade")  
  
boxplot(Part,main="Part")
```



```
boxplot(cidade,Part, main="Cidade      Part", ylab="Valor",varwidth=TRUE,col="blue")  
  
sd(cidade)
```

```
## [1] 0.08439832
```

```
sd(Part)
```

```
## [1] 0.08736259
```

```
t.test(cidade,Part, paired=FALSE, var.equal=TRUE)
```

```
##
## Two Sample t-test
##
## data: cidade and Part
## t = 0.1643, df = 21, p-value = 0.871
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.06902507 0.08087122
## sample estimates:
## mean of x mean of y
## 0.1569231 0.1510000
```

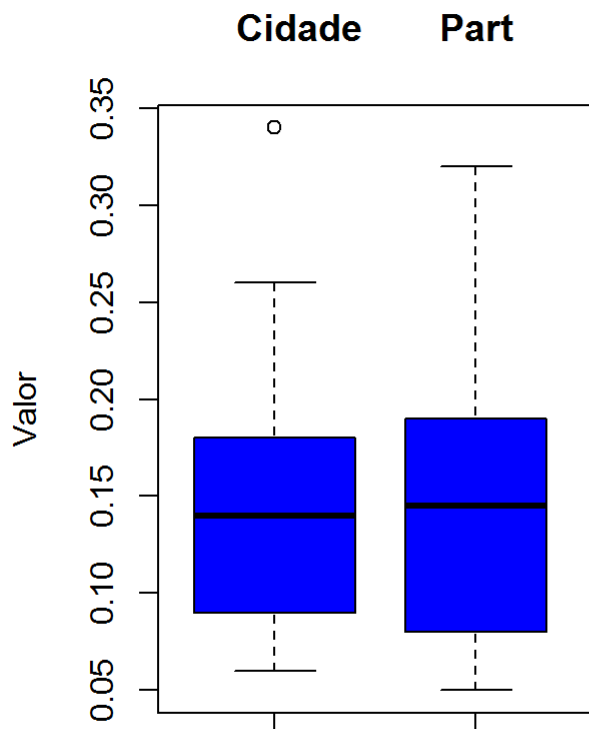
```
# Amplitude dos Dados
```

```
rcidade5<-t.test(cidade,alternative=c("two.sided"),mu=0,conf.level=0.95)
rcidade5$conf.int[2]-rcidade5$conf.int[1]
```

```
## [1] 0.1020028
```

```
rPart5<-t.test(Part,alternative=c("two.sided"),mu=0,conf.level=0.95)
rPart5$conf.int[2]-rPart5$conf.int[1]
```

```
## [1] 0.1249909
```



Conclusões:

Desvio Padrão - Cidades: 0.084 Part: 0.087 (amostras muito próximas)

Amplitude - Cidades: 0.102 Part: 0.124

Média - Cidade: 0.156 Part: 0.151

Os gráficos boxplot demonstraram uma similaridade nos dados das amostras. Os dados do gráfico de Part apresentaram uma maior amplitude. O tamanho da amostra de objetos da Cidade é maior, com amplitude menor. Desta forma, observa-se que não houve diferença significativa dos teores de mercúrio para as duas áreas residenciais.

THE END