

ESTATÍSTICA DESCRITIVA

QUESTÃO ED1

Para fins de vistoria ambiental, a concentração de monóxido de carbono (CO, mg/m³) foi medida em quatro chaminés. Os dados obtidos são apresentados a seguir,

Cham1	Cham2	Cham3	Cham4
40.5	41.64	35.0	44.5
41.5	58.36	37.0	45.0
42.5	42.29	42.0	45.5
43.5	57.71	53.9	46.0
44.5	42.93	53.0	46.5
45.5	57.07	50.6	47.0
46.5	43.57	50.5	47.5
47.5	56.43	53.8	48.0
48.5	44.21	52.5	48.5
49.5	55.79	53.6	49.0
50.5	44.86	50.4	49.5
51.5	55.14	52.2	50.0
52.5	45.50	52.7	50.5
53.5	54.50	52.4	51.0
54.5	46.14	52.7	51.5
55.5	53.86	51.4	52.0
59.5	46.79	53.8	52.5

Variavel	Media	Desvio	Minimo	Q1	Mediana	Q3	Maximo
Cham1	48.68	5.39	40.50	44.00	50.00	53.00	59.50
Cham2	49.81	6.33	41.64	NA	50.00	56.11	58.36
Cham3	49.85	NA	35.00	50.45	52.45	53.30	53.90
Cham4	48.50	2.52	44.50	46.25	NA	51.38	72.71

- Complete o quadro das estatísticas descritivas destes dados.
- Compare as distribuições quanto à simetria com base nos gráficos acima.
- Que chaminé apresenta maior variação relativa? Analise criticamente sua resposta.
- As estatísticas calculadas podem ser utilizadas para descrever o comportamento dos dados?

Resposta letra a)

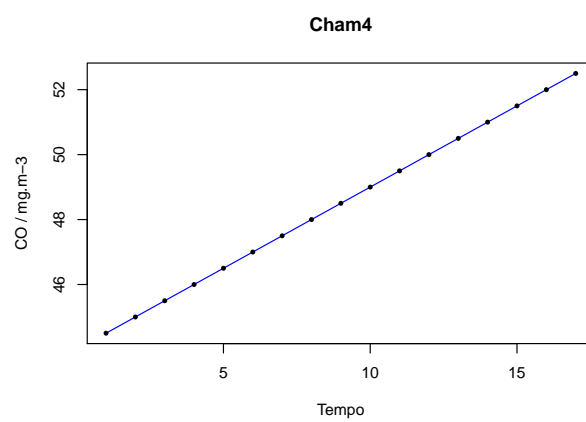
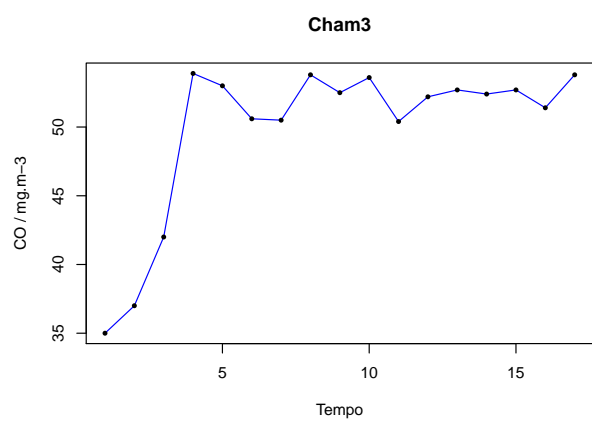
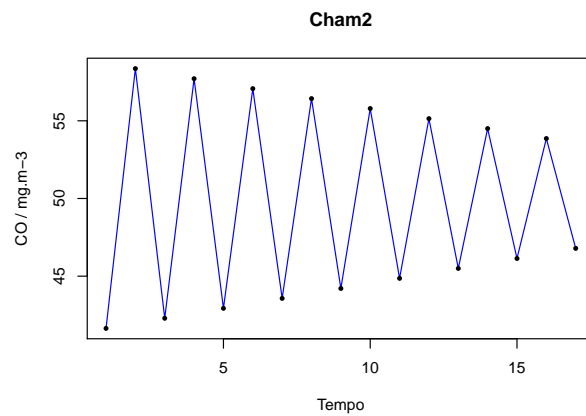
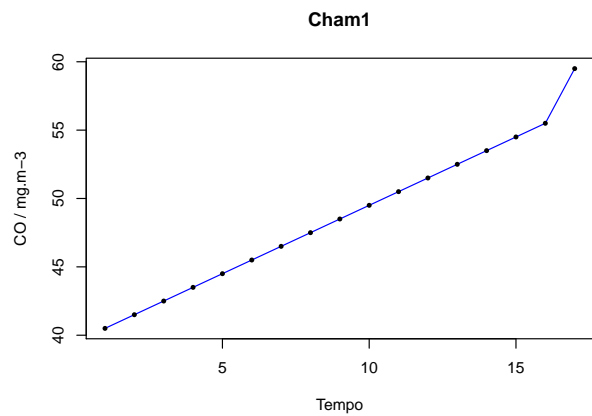
```
tb <- fdt(LQED1)
sumdf <- sapply(LQED1, summary)
sumdf <- round(sumdf,1)
sumdf <- data.frame(sumdf)
sddf <- sapply(LQED1, sd)
stddev <- round(sddf,1)
sd<- data.frame(rbind(stddev))
resumo <- rbind(sumdf, sd)
knitr::kable(resumo, align = 'c')
```

	Cham1	Cham2	Cham3	Cham4
Min.	40.5	41.6	35.0	44.5
1st Qu.	44.5	44.2	50.5	46.5
Median	48.5	46.8	52.4	48.5
Mean	48.7	49.8	49.9	48.5
3rd Qu.	52.5	55.8	53.0	50.5
Max.	59.5	58.4	53.9	52.5
stddev	5.4	6.3	5.9	2.5

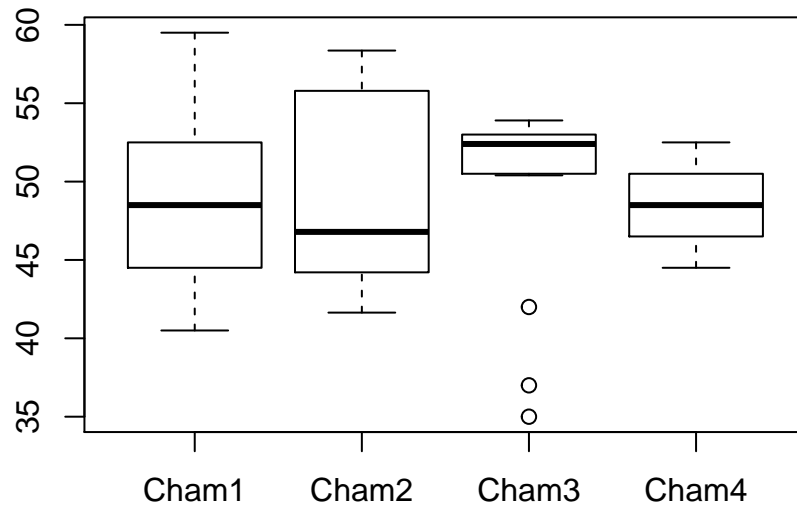
Usaremos a função **sapply()** para realizar o cálculo dos parâmetros estatísticos para cada coluna, arredondando com a função **round()** os valores para até quatro significativos, usaremos a função **sapply()** para o cálculo do desvio padrão para cada coluna, arredondando o valor do desvio padrão para até 4 significativos com a função **round()**. Transformando o objeto sddf para do tipo dataframe com a função **data.frame()**, mostrando o dataframe com todos os dados necessários.

Resposta letra b)

```
par(mfrow = c(2,2))
plot(LQED1$Cham1, type = "l", col = "blue",
     xlab = "Tempo", ylab = "CO / mg.m-3", main = "Cham1")
points(LQED1$Cham1, pch = 16, col = "black", cex = .7)
plot(LQED1$Cham2, type = "l", col = "blue",
     xlab = "Tempo", ylab = "CO / mg.m-3", main = "Cham2")
points(LQED1$Cham2, pch = 16, col = "black", cex = .7)
plot(LQED1$Cham3, type = "l", col = "blue",
     xlab = "Tempo", ylab = "CO / mg.m-3", main = "Cham3")
points(LQED1$Cham3, pch = 16, col = "black", cex = .7)
plot(LQED1$Cham4, type = "l", col = "blue",
     xlab = "Tempo", ylab = "CO / mg.m-3", main = "Cham4")
points(LQED1$Cham4, pch = 16, col = "black", cex = .7)
```

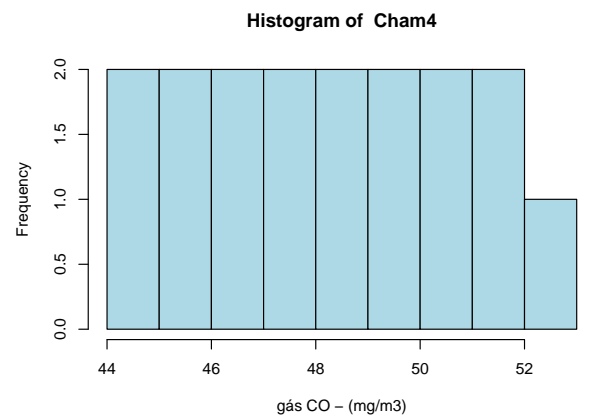
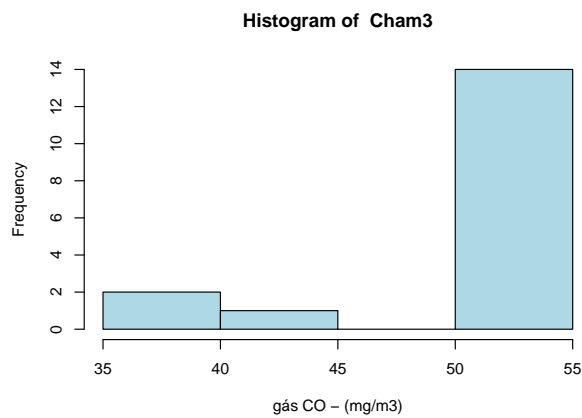
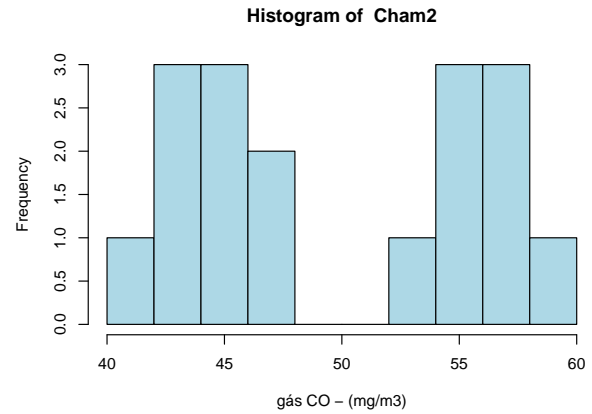
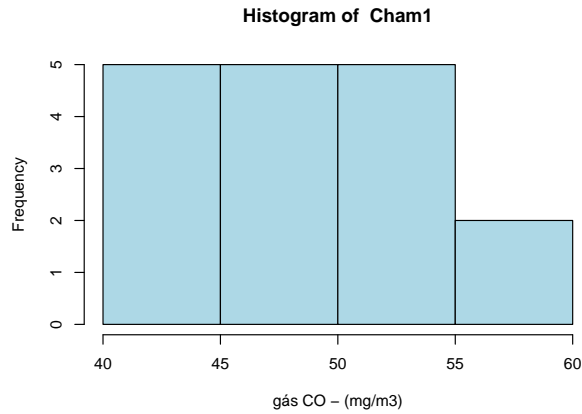


Plotando os boxplots



Plotando os histogramas

```
par(mfrow = c(2,2))
graph <- sapply(X = c("Cham1","Cham2","Cham3","Cham4"), FUN=function(s) {
  hist(LQED1[,s], xlab= c("gás CO - (mg/m3)"), main = paste("Histogram of ",s),
    col = "light blue")
})
```



```
Cvar <- sapply(LQED1, cv)
Coefvar <- round(Cvar,2)
kurt <- sapply(LQED1, kurtosis) # Curtose
kurt <- round(kurt,2)
skwns <- sapply(LQED1, skewness) # Assimetria
skwns <- round(skwns,2)
percentl <- sapply(LQED1, quantile)
percentl <- round(percentl,2)
otprm <- rbind(Coefvar, kurt, skwns, percentl)
resumo <- rbind(otprm, make.row.names = TRUE, stringsAsFactors=
  default.stringsAsFactors())
knitr::kable(resumo, align = 'c')
```

	Cham1	Cham2	Cham3	Cham4
Coefvar	11.07	12.71	11.84	5.21
kurt	2.17	1.26	4.43	1.79
skwns	0.24	0.08	-1.73	0.00
0%	40.50	41.64	35.00	44.50
25%	44.50	44.21	50.50	46.50
50%	48.50	46.79	52.40	48.50
75%	52.50	55.79	53.00	50.50
100%	59.50	58.36	53.90	52.50
make.row.names	1.00	1.00	1.00	1.00

	Cham1	Cham2	Cham3	Cham4
stringsAsFactors	1.00	1.00	1.00	1.00

Os gráficos e a tabela acima evidenciam que: para a Chaminé 1 e Chaminé 2 temos assimetria leve à direita como pode também ser notado pela variáveis *skwns*. Para a Chaminé 3 existem uma maior quantidade de leituras para valores mais baixos de concentração de monóxido de carbono, ou seja, existe uma assimetria à esquerda, com valor de *skwns* igual à -1,73, já a Chaminé 4 tem assimetria nula (*skwns* = 0), isso indica que a distribuição é normal ou simétrica.

Resposta letra c)

Ao Olhar a tabela acima pode-se notar que a variável *Coefvar* que corresponde ao Coeficiente de Variação *Cv*, está calculado para as quatro chaminés. Observando os valores calculados podemos notar que há uma ordem crescente onde o *Cv* pode ser ordenado dessa forma $5.21 > 11.07 > 11.84 > 12.71$, ou seja, o coeficiente de variação que corresponde à dispersão relativa dos dados, ordena em ordem crescente as chaminés em termos de dispersão relativa dessa forma:

```
knitr::kable(class, align = 'c')
```

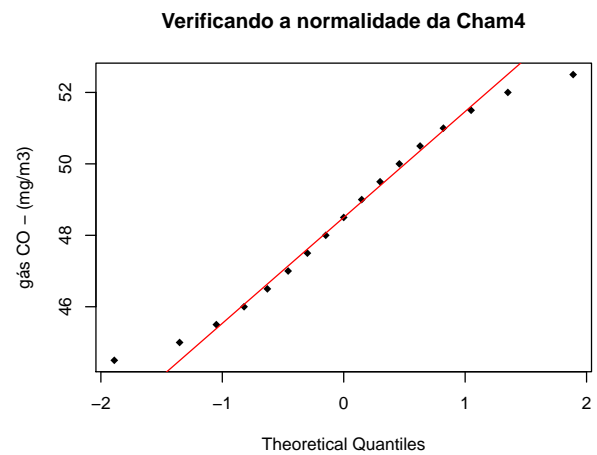
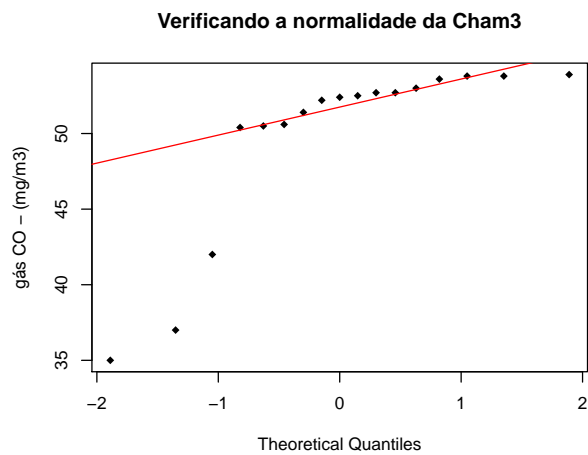
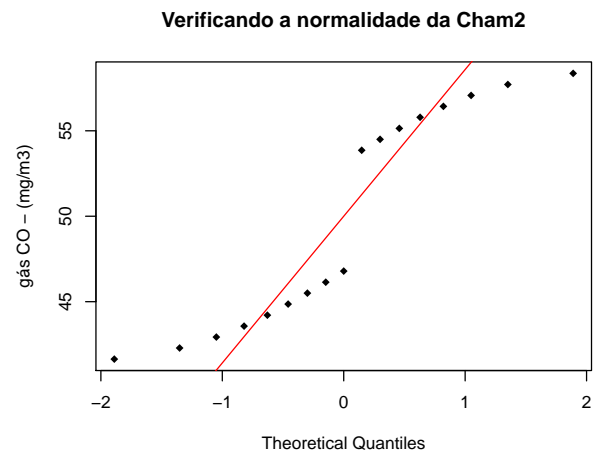
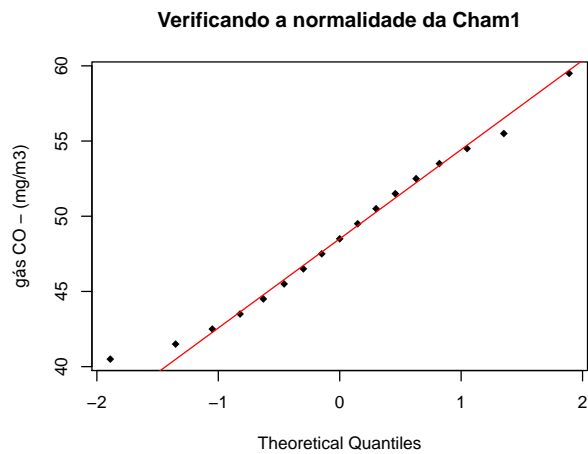
Cham4	Cham1	Cham3	Cham2
5.21	11.07	11.84	12.71

Ou seja a Chaminé 2 tem a maior dispersão relativa ou o maior coeficiente de variação dos dados.

Resposta letra d)

Não é possível usar as estatísticas para descrever os dados, pois a distribuição não é normal, e ainda na Chaminé 2, temos uma distribuição bimodal, ou seja, existem duas médias, e a média global dessas duas distribuições não representa a população. Os gráficos abaixo comparam a distribuição dos dados de cada chaminé, com a distribuição normal, podemos observar que a única chaminé que tem proximidade com a distribuição normal é a Cham4, por isso nesse caso não é possível usar as estatísticas para todas as chaminés.

```
par(mfrow = c(2,2))
sapply(X = c("Cham1","Cham2","Cham3","Cham4"), FUN=function(s) {
  qqnorm(LQED1[,s], ylab= c("gás CO - (mg/m3)"),
        main = paste("Verificando a normalidade da",s),
        pch = 18)
  qqline(LQED1[,s], col = "red")
})
```



```
## $Cham1
## NULL
##
## $Cham2
## NULL
##
## $Cham3
## NULL
##
## $Cham4
## NULL
```