Trabalho 1 - Dados 4

Controle Estatístico de Qualidade

Matheus de Moraes Neves | 12321EST201

Introdução

Neste relatório serão resolvidos as questões do Trabalho 1 utilizando os comandos em aula em linguagem R, escrito no editor de texto Quarto no Rstudio. Importanto a biblioteca que será utilizada.

```r
library(qcc)
```

Peças de Plástico - Dados 4

As peças de plástico são fabricadas por um processo de moldagem por injeção e são submetidas a um teste de resistência à compressão.

Figura 1: Teste de Resistência de Compressão

a) Descrição e apresentação dos dados

Para um estudo preliminar com 20 amostras de 5 peças, as resistências à compressão (em psi) são apresentadas na tabela a seguir:

Tabela 1: Amostras Submetidas ao teste.

X1	**X2**	**X3**	**X4**	**X5**
83	81.2	78.7	75.7	77
88.6	78.3	78.8	71	67
85.7	75.8	84.3	74.2	81
80.8	74.4	82.5	74.1	75.7
83.4	78.4	82.6	78.2	78.9
75.3	79.9	87.3	89.7	81.8
74.5	78	80.8	73.4	79.7
79.2	84.4	81.5	86	74.5
80.5	86.2	76.2	84.1	80.2
75.7	75.2	71.1	82.1	74.3
80	81.5	78.4	73.8	78.1
80.6	81.8	79.3	73.8	81.7
82.7	81.3	79.1	82	79.5
79.2	74.9	78.6	77.7	75.3
85.5	82.1	82.8	73.4	71.7
78.8	79.6	80.2	79.1	80.8
82.1	78.2	75.5	78.2	82.1
84.5	76.9	83.5	81.2	79.2
79	77.8	81.2	84.4	81.6
84.5	73.1	78.6	78.7	80.6

Armazenando em *data frame* no R:

```r
# Importando os dados dentro do diretório do Rstudio
dados = read.table("dados4.txt", h = T)

# Mostrando os 5 primeiros registros
head(dados, 5)
```

```
    X1   X2   X3   X4   X5
1 83.0 81.2 78.7 75.7 77.0
2 88.6 78.3 78.8 71.0 67.0
3 85.7 75.8 84.3 74.2 81.0
4 80.8 74.4 82.5 74.1 75.7
5 83.4 78.4 82.6 78.2 78.9
```

Como as amostras são pequenas, i.e., $n < 10$, então será utilizado o gráfico de R. É possível adicionar as colunas de média e amplitude de cada amostra a este *data frame*:

```
# Criando as colunas de média e amplitude
dados$X_barra = 0
dados$R = 0

for (j in 1:20) {
  dados$X_barra[j] = sum(dados[j,-(6:7)])/5
  dados$R[j] = max(dados[j,-(6:7)]) - min(dados[j,-(6:7)])
}

head(dados)
```

```
    X1   X2   X3   X4   X5 X_barra    R
1 83.0 81.2 78.7 75.7 77.0   79.12  7.3
2 88.6 78.3 78.8 71.0 67.0   76.74 21.6
3 85.7 75.8 84.3 74.2 81.0   80.20 11.5
4 80.8 74.4 82.5 74.1 75.7   77.50  8.4
5 83.4 78.4 82.6 78.2 78.9   80.30  5.2
6 75.3 79.9 87.3 89.7 81.8   82.80 14.4
```

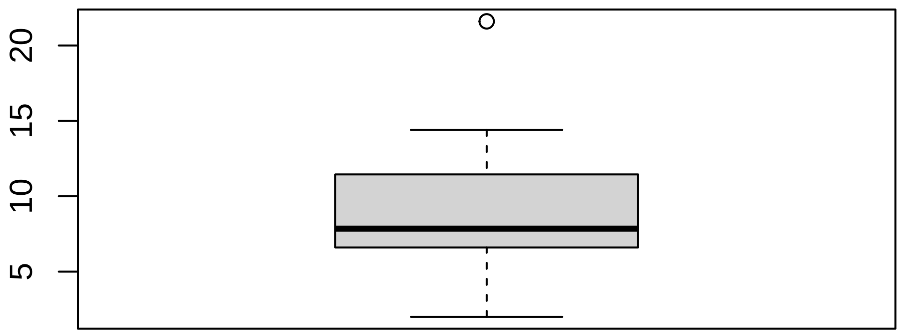
E para utilizar os dados sem as colunas de amplitude e média de cada amostra será feito da seguinte forma:

```
head(dados[,-(6:7)]) # retirando as duas ultimas colunas
```

```
    X1   X2   X3   X4   X5
1 83.0 81.2 78.7 75.7 77.0
2 88.6 78.3 78.8 71.0 67.0
3 85.7 75.8 84.3 74.2 81.0
4 80.8 74.4 82.5 74.1 75.7
5 83.4 78.4 82.6 78.2 78.9
6 75.3 79.9 87.3 89.7 81.8
```

Utilizando-se o gráfico de boxplot para visualizar a coluna de amplitudes por exemplo, é possível verificar a existência de um *outlier*, com seu valor acima do 20, ou seja, a segunda amostra com amplitude de $21,6$. Provavelmente esta amostra será retirada na etapa de criação do gráfico de R.

```r
# Boxplot da coluna de Amplitude
boxplot(dados$R)
```



b) Estimativas e Gráficos de Controle

Uma estimativa para média e desvio padrão quando o processo está em controle podem ser calculadas da seguinte forma:

$$\hat{\sigma}_0 = S_D = \frac{\bar{R}}{d_2} \qquad\qquad \hat{\mu}_0 = \bar{\bar{X}} = \frac{\sum_{i=1}^{m} \bar{X}_i}{m} \tag{1}$$

Como o tamanho das amostras são $n = 5$, então $d_2 = 2,326$. Calculando no R:

```
# Estimativa do desvio padrão em controle
mean(dados$R)/2.326
```

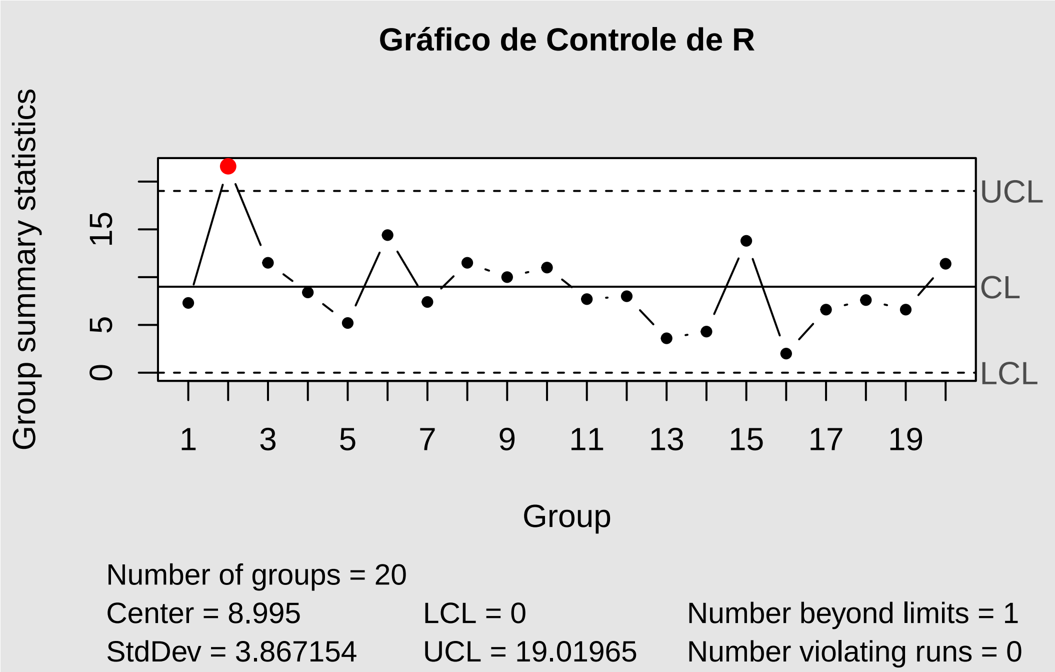
```
[1] 3.867154
```

```
# Estimativa da média em controle
mean(dados$X_barra)
```

```
[1] 79.351
```

Estimativas iniciais antes de qualquer exclusão de alguma amostra, então elas poderão ser alteradas. Criando os gráficos de controle de $\bar{X}$ e de R, com $k = 3$:

```
gráfico.R = qcc(dados[,-(6:7)], type = "R",
                title = 'Gráfico de Controle de R') # m = 20 amostras
```



Como suspeitado anteriormente, a segunda amostra, que era um outlier das amplitudes, ultrapassou um dos limites do gráfico de controle de R (limite superior), sendo necessária sua **exclusão** para construir os gráficos e **investigar** o que provocou esta alteração. Além disso, a estimativa do desvio padrão fornecida pelo gráfico coincide com a estimativa anterior.

Excluindo a segunda amostra:

```
# Excluindo manualmente
dados1 = dados[-2,]

head(dados1) # vendo os 5 primeiros registros
```

```
    X1   X2   X3   X4   X5 X_barra    R
1 83.0 81.2 78.7 75.7 77.0   79.12  7.3
3 85.7 75.8 84.3 74.2 81.0   80.20 11.5
4 80.8 74.4 82.5 74.1 75.7   77.50  8.4
5 83.4 78.4 82.6 78.2 78.9   80.30  5.2
6 75.3 79.9 87.3 89.7 81.8   82.80 14.4
7 74.5 78.0 80.8 73.4 79.7   77.28  7.4
```

```
# Caso não seja possível excluir facilmente, automatizando o processo:
```

```
m = 20 #amostras

LSC = (2.326+(3*0.864))*mean(dados$R)/2.326 # (d2+3*d3)*std_0

for (j in 1:m) {
  if(dados$R[j] > LSC){ # Se na linha j a Amplitude > LSC
    dados1 = dados[-j,] # Então exclui a linha j
  }
}

head(dados1)
```

```
    X1   X2   X3   X4   X5 X_barra    R
1 83.0 81.2 78.7 75.7 77.0   79.12  7.3
3 85.7 75.8 84.3 74.2 81.0   80.20 11.5
4 80.8 74.4 82.5 74.1 75.7   77.50  8.4
5 83.4 78.4 82.6 78.2 78.9   80.30  5.2
6 75.3 79.9 87.3 89.7 81.8   82.80 14.4
7 74.5 78.0 80.8 73.4 79.7   77.28  7.4
```

Vericando as estimativas deste novo *data frame*.

```
# Estimativa do desvio padrão em controle
mean(dados1$R)/2.326
```

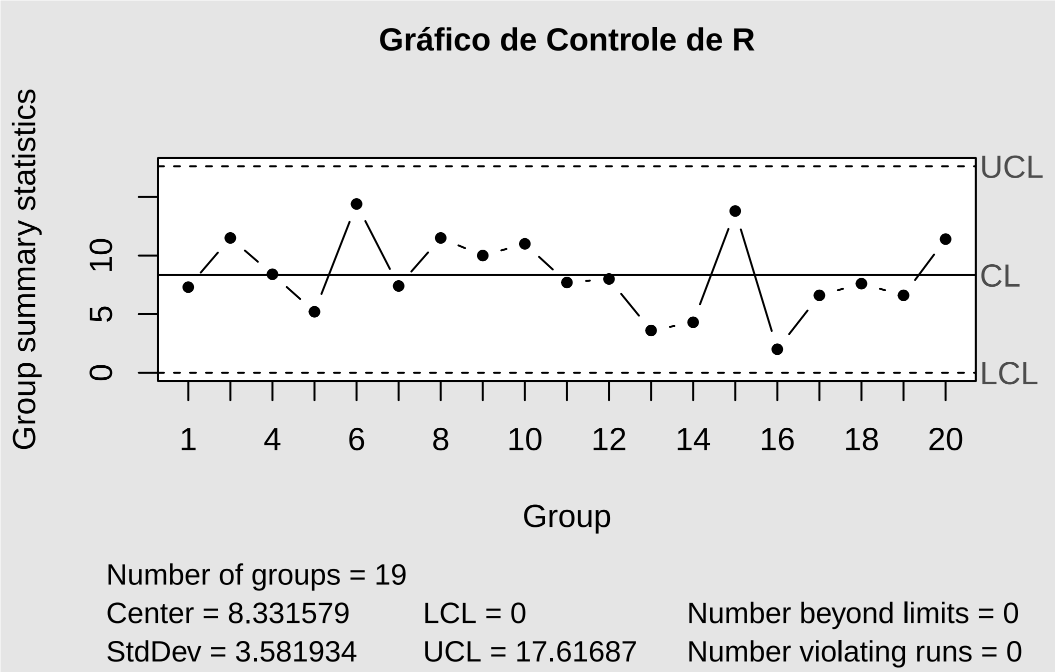
```
[1] 3.581934
```

```
# Estimativa da média em controle
mean(dados1$X_barra)
```

```
[1] 79.48842
```

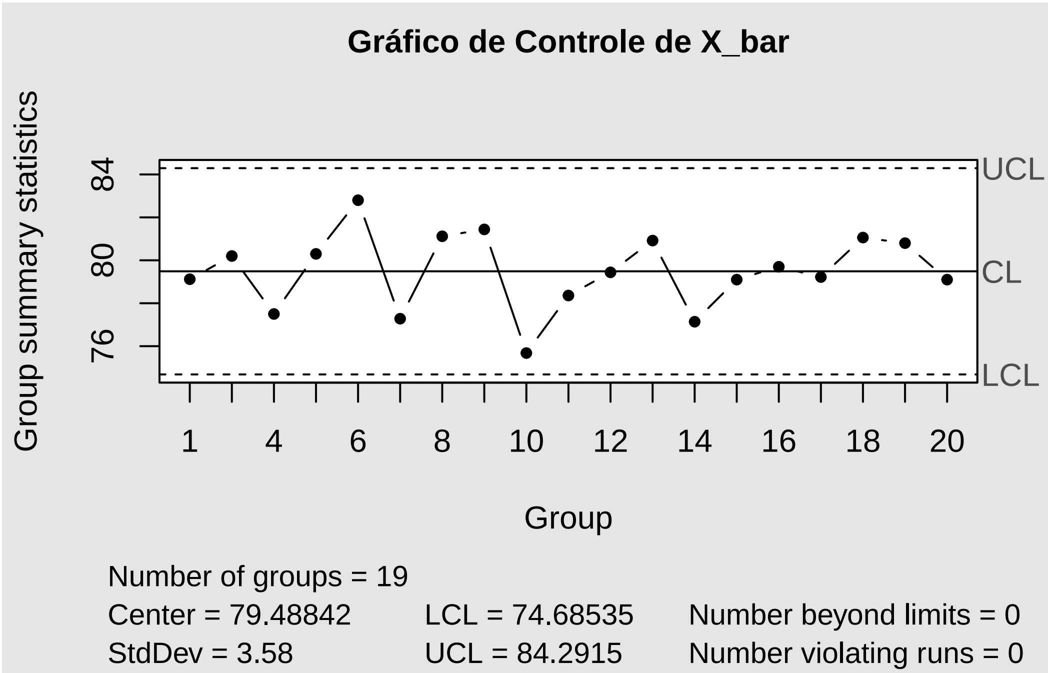
Com a amostra 2 retirada, será então feito o gráfico de R com 19 amostras:

```
gráfico.R1 = qcc(dados1[,-(6:7)], type = "R",
                title = 'Gráfico de Controle de R') # m = 19 amostras
```



Diminuindo o Limite Superior de R de 19 para 17,62. E a nova estimativa do desvio padrão em controle é $\hat{\sigma}_0 = 3,58$. Agora fazendo o gráfico de de $\bar{X}$ com desvio padrão definido:

```
gráfico.xbar1 = qcc(dados1[,-(6:7)], type = "xbar", std.dev = 3.58,
                    title = 'Gráfico de Controle de X_bar')# m = 19 amostras
```



Como não tem nenhuma amostra com média fora dos limites de controle, então este serão os gráficos finais para a fase 1. Portanto, a estimativa final para média e desvio padrão, fornecidas tanto pelos gráficos quanto pelo código autoral, para quando o processo está em controle será de:

$$\hat{\sigma}_0 = 3.58 \qquad \hat{\mu}_0 = 79.49 \tag{2}$$

c) Limites de Controle de R e X_barra

Como o limite superior de R é um valor menor que 0, então adota-se LIC = 0:

```
(2.326-(3*0.864))*mean(dados$R)/2.326 # (d2-3*d3)*std_0
```

```
[1] -1.028663
```

Organizando todos os limites de controles fornecidos pelos gráficos de controle ou pelo código R autoral, em tabela:

Tabela 2: Limites de Controle Simulatâneos

	R	**Xbar**
LSC	17.62	84.30
LM	8.33	79.49
LIC	0	74.69

d) NMA_0 do Gráfico da Média

Na fase 2, a probabilidade de se detectar algum ponto fora dos limites com o processo em controle, i.e, a probabilidade de se ter alarme falso é dada pela Equação (3).

$$\alpha = 2 \cdot \Pr(z < -k) \tag{3}$$

Como foi utilizado $k = 3$, então calculando a probabilidade acumulada abaixo do quartil -3 de uma Normal Padrão, e multiplcando por 2:

```r
# Cálculo de alfa
2*pnorm(-3)
```

```
[1] 0.002699796
```

Além disso, como a ocorrência de alarme falso segue uma distribuição geométrica, então o número médio de amostras até a ocorrência de um alarme falso, também chamado como NMA_0, será definido pela Equação (4):

$$NMA_0 = E(X) = \frac{1}{\alpha} \tag{4}$$

Em código R:

```r
1/(2*pnorm(-3))
```

```
[1] 370.3983
```

Portanto, em média terão 370 amostras até a ocorrência de um alarme falso, i.e, $NMA_0 = 370$.

e) Poder do gráfico da Média com Diferentes Parâmetros

Agora será calculado a probabilidade de se ter um alarme verdadeiro no gráfico da média, i.e, o Poder do Gráfico de $\bar{X}$ para diferentes tipos de alterações na média e no desvio padrão (testar com diferentes valores de λ e δ). Criando um *data frame* vazio para armazenar todas as informações:

```
# criando uma tabela vazia com 4 linhas
tabela = data.frame(delta = numeric(4),
                    lambda = numeric(4),
                    m1 = numeric(4),
                    std1 = numeric(4),
                    Pd = numeric(4),
                    NMA1 = numeric(4))

# preenchendo com os diferentes parâmetros de cada caso
tabela$delta = c(1.5,2,3,4)
tabela$lambda = c(1.5,2,2.5,3)

# calculando a alteração da média em cada caso (m1 = m0 + delta * std_0)
for (j in 1:4) {
  tabela$m1[j] = 79.49 + tabela$delta[j]*3.58
  tabela$std1[j] = tabela$lambda[j]*3.58
}
# calculando a alteração do desvio padrão em cada caso (std1 = lambda * std0)

tabela
```

```
  delta lambda    m1  std1 Pd NMA1
1   1.5    1.5 84.86  5.37  0    0
2   2.0    2.0 86.65  7.16  0    0
3   3.0    2.5 90.23  8.95  0    0
4   4.0    3.0 93.81 10.74  0    0
```

A probabilidade de detectar que o processo está fora de controle (Poder do Gráfico) é dada pela Equação (5).

$$Pd_{\bar{X}} = \Pr\left(z < \frac{-k - \delta\sqrt{n}}{\lambda}\right) + \Pr\left(z < \frac{-k + \delta\sqrt{n}}{\lambda}\right) \tag{5}$$

E analogamente ao que foi discutido na Equação (4), o NMA_1, i.e, o número médio de amostras até ocorrer um alarme verdadeiro:

$$NMA_1 = E(X) = \frac{1}{Pd} \tag{6}$$

Preenchendo as duas últimas colunas do *data frame*:

```
valor1 = 0
valor2 = 0

for (j in 1:4) {
  valor1 = (-3-tabela$delta[j]*sqrt(5))/tabela$lambda[j]
  valor2 = (-3+tabela$delta[j]*sqrt(5))/tabela$lambda[j]

  tabela$Pd[j] = pnorm(valor1) + pnorm(valor2)

  tabela$NMA1[j] = 1/tabela$Pd[j]
}

tabela
```

```
  delta lambda    m1  std1        Pd     NMA1
1   1.5    1.5 84.86  5.37 0.5933214 1.685427
2   2.0    2.0 86.65  7.16 0.7692488 1.299970
3   3.0    2.5 90.23  8.95 0.9310517 1.074054
4   4.0    3.0 93.81 10.74 0.9762624 1.024315
```

Organizando em tabela no LaTeX para melhor visualização:

Caso	δ	λ	μ_1	σ_1	**Poder**	**NMA$_1$**
1	1.5	1.5	84.86	5.37	0.5933	1.69
2	2	2	86.65	7.16	0.7692	1.30
3	3	2.5	90.23	8.95	0.9311	1.07
4	4	3	93.81	10.74	0.9763	1.02

f) Conlusão sobre o desempenho do Gráfico de $\bar{X}$

Portanto, o gráfico de controle da média são mais rápidos na detecção de grandes deslocamentos da média ($\delta = 4$), porém mais lentos em deslocamentos menores ($\delta = 1.5$), ou seja, o gráfico precisará de menos amostras defeituosas para detectar uma alteração na média quanto maior for esta alteração. Apesar do gráfico de controle da média também levar em conta a alteração no desvio padrão (λ), esta alteração não aparente fazer uma diferença considerável no valor de NMA_1, comparado à alteração da média. Em ambos os casos de alteração, seja na média ou no desvio padrão, quanto maior é mais fácil o gráfico consegue detectar.

Além disso, em todos os casos o gráfico de controle de $\bar{X}$ são adequados, i.e, são **bons** pelo critério de avaliação estabelicido em aula ($NMA_1 < 2$). Apesar de que o quarto caso é o melhor de todos ($NMA_1 = 1,02$), e o primeiro caso, o pior de todos, ter NMA_1 bem próximo de 2 amostras até o alarme verdadeiro ($NMA_1 = 1,69$).

g) Simulando Novos Dados

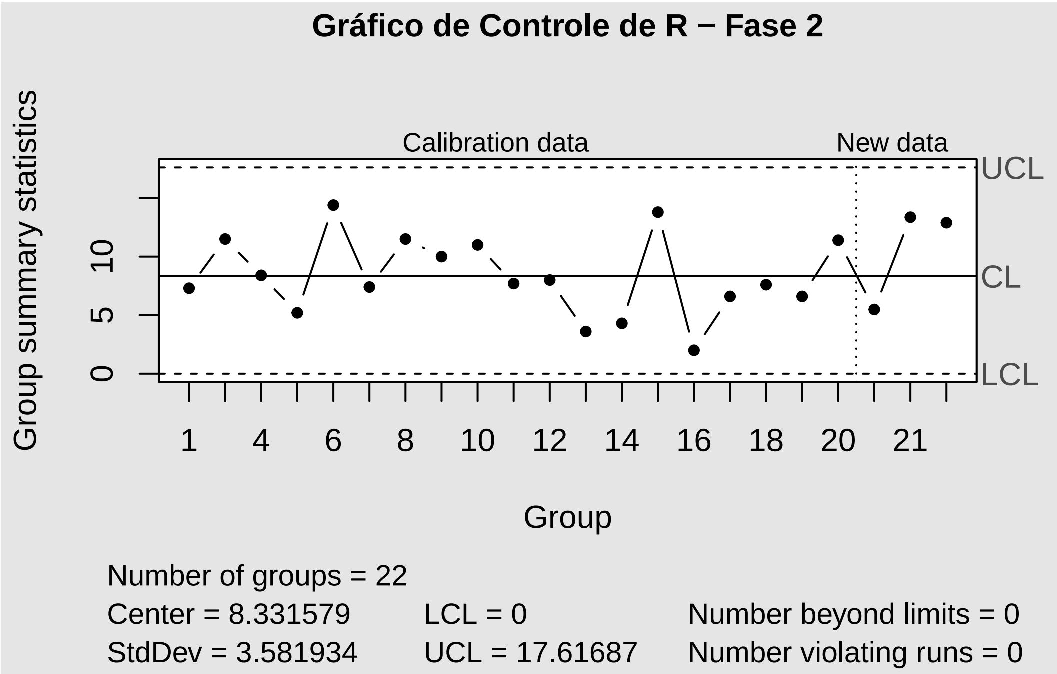
Agora simulando dados alaetórios para a fase 2, como se fosse monitoramente em tempo real.

```
# Retirando as colunas que não serão mais usadas
dados1 = dados1[,-(6:7)]

# Gerando 3 amostras aleatórias simulando a fase 2

set.seed(123) # setando a seed para replicabilidade do experimento

# Média e desvio padrão maiores (deslocamento do caso 1 da letra e) )
fase2 = matrix(rnorm(3*5, mean = 84.86, sd = 5.37), ncol = 5)

fase2
```

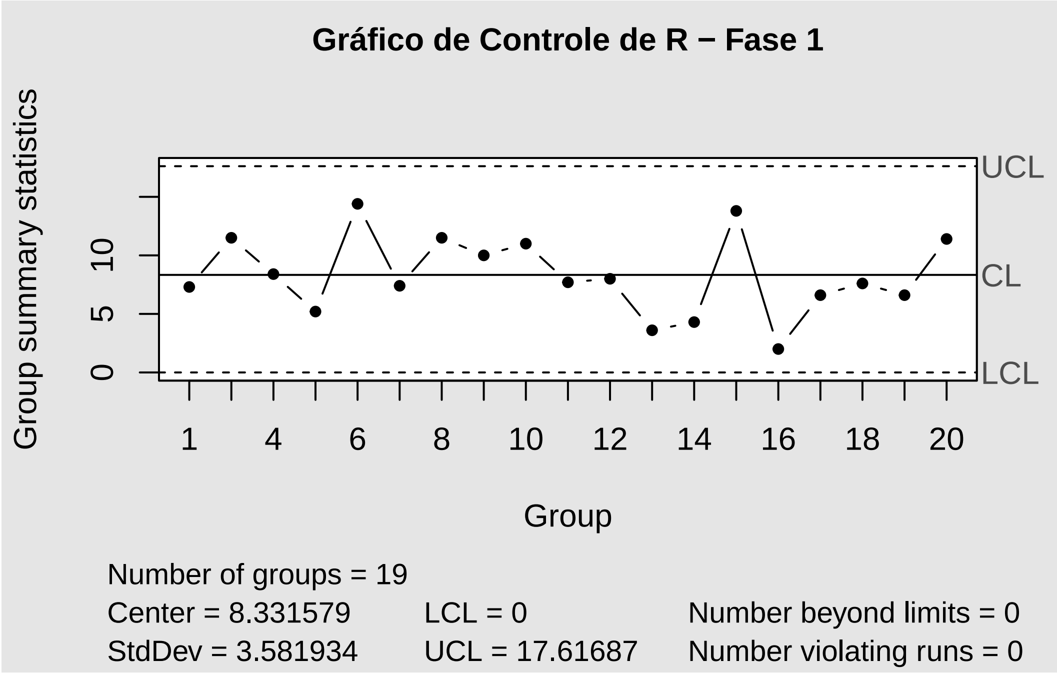
```
         [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 81.85025 85.23863 87.33512 82.46680 87.01214
[2,] 83.62395 85.55428 78.06662 91.43332 85.45437
[3,] 93.23026 94.06990 81.17160 86.79220 81.87513
```

Aplicando para o gráfico de Média e Amplitude, e comparando-os com o seus respectivos gráficos da fase 1.

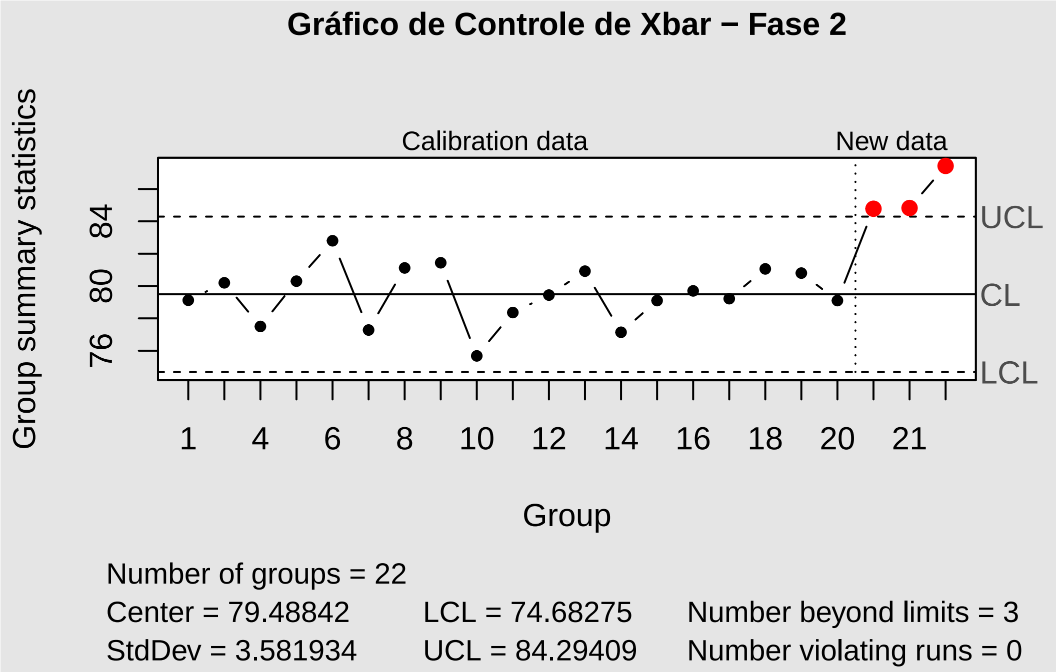
```
# Comparando os gráficos nas 2 fases
gráfico.R2 = qcc(dados1, type = 'R',
                 newdata = fase2,   # Inserindo os novos dados da fase 2
                 title = 'Gráfico de Controle de R - Fase 2')
```



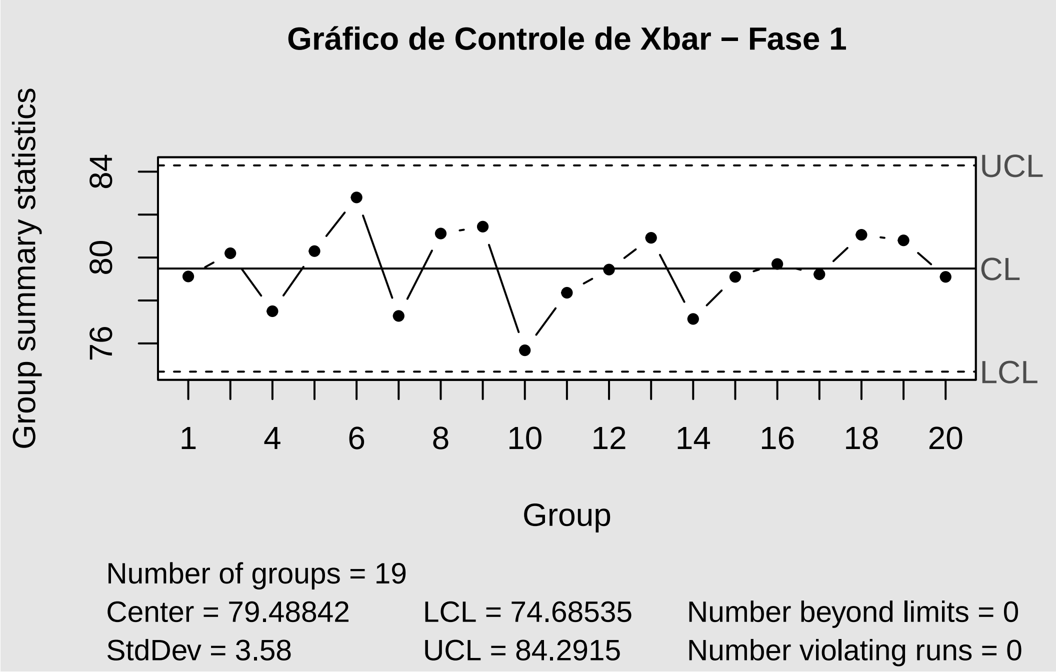

```
plot(gráfico.R1, title = 'Gráfico de Controle de R - Fase 1')
```



```r
gráfico.xbar2 = qcc(dados1, type = 'xbar', newdata = fase2, title = 'Gráfico de Controle de X
```




```r
plot(gráfico.xbar1, title = 'Gráfico de Controle de Xbar - Fase 1')
```



Entretando, o gráfico de $\bar{X}$ foi o único a detectar a alteração no processo de teste de resistência à compressão ao todo. A Tabela 3 mostra os dados das amostras nas 2 fases.

Tabela 3: Dados das Fases 1 e 2.

		X1	**X2**	**X3**	**X4**	**X5**
	1	83	81.2	78.7	75.7	77
	2	85.7	75.8	84.3	74.2	81
	3	80.8	74.4	82.5	74.1	75.7
	4	83.4	78.4	82.6	78.2	78.9
	5	75.3	79.9	87.3	89.7	81.8
	6	74.5	78	80.8	73.4	79.7
	7	79.2	84.4	81.5	86	74.5
	8	80.5	86.2	76.2	84.1	80.2
	9	75.7	75.2	71.1	82.1	74.3
Fase 1	10	80	81.5	78.4	73.8	78.1
	11	80.6	81.8	79.3	73.8	81.7
	12	82.7	81.3	79.1	82	79.5
	13	79.2	74.9	78.6	77.7	75.3
	14	85.5	82.1	82.8	73.4	71.7
	15	78.8	79.6	80.2	79.1	80.8
	16	82.1	78.2	75.5	78.2	82.1
	17	84.5	76.9	83.5	81.2	79.2
	18	79	77.8	81.2	84.4	81.6
	19	84.5	73.1	78.6	78.7	80.6
	20	81.85	85.24	87.34	82.47	87.01
Fase 2	21	83.62	85.55	78.07	91.43	85.45
	22	93.23	94.07	81.17	86.79	81.88