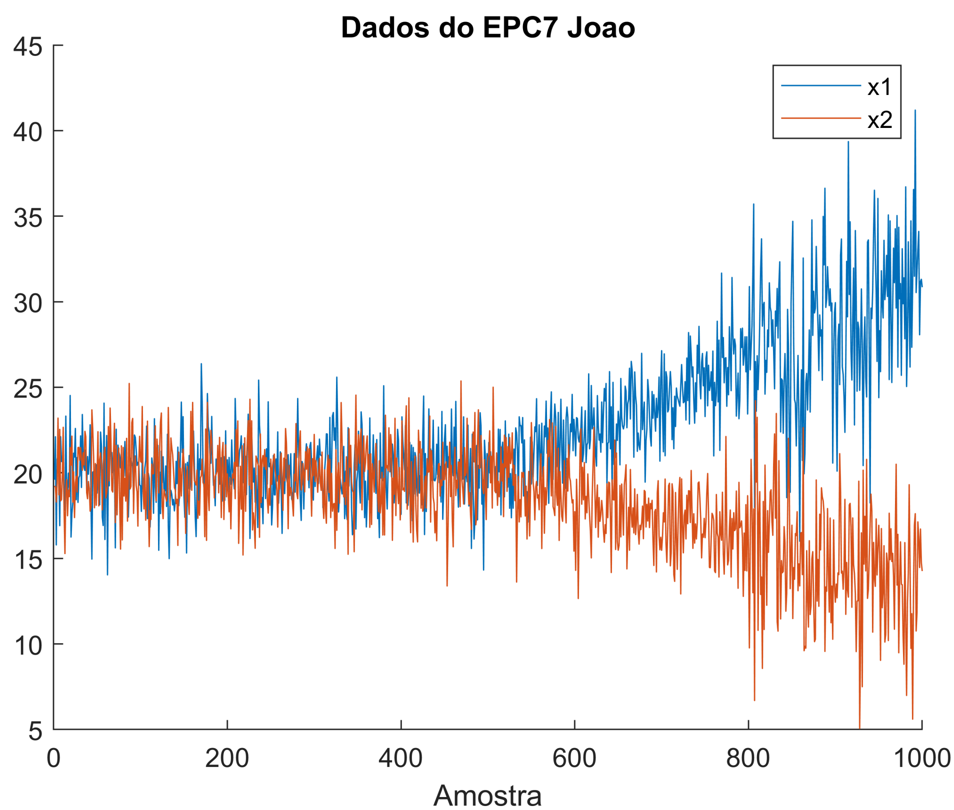
EPC 7 - REPOSIÇÃO

João Gabriel Santos Custodio

Loading de dados:

```
load("epc7_Joao.mat")

figure()
hold on
plot(1:1:1000,x1)
plot(1:1:1000,x2)
title("Dados do EPC7 Joao")
legend("x1","x2")
xlabel("Amostra")
```



Atividade 1: Fazer um gráfico de controle para média, desvio padrão e range, usando as 500 primeiras amostras de x_1 , para n=5 e para n=20.

• Mostrar e comentar os dados utilizados.

• Explicar como são calculados $\overline{\overline{x}}, \overline{s}, \overline{r}$ e os valores usados de n e m em cada caso.

• Mostrar os gráficos informando em que amostras rejeitou-se a hipótese nula de que o processo está em controle estatístico e comparar as diferenças para n=5 e n=20 .

Resposta:

Os cálculos de $\bar{\bar{x}}, \bar{s}, \bar{r}$ são feito utilizando as equações abaixo.

Temos uma função que reitera o cálculo feito anteriormente, em conjunto com a tabela de coeficientes presente no livro, para construir os gráficos de controle estatístico. Isso é realizado por meio das equações que estabelecem os valores de UCL, CL e LCL.

$\bar{\bar{x}}$

$$UCL = \bar{\bar{x}} + 3\frac{\bar{s}}{c_4\sqrt{n}} \qquad CL = \bar{\bar{x}} \qquad LCL = \bar{s} - 3\frac{\bar{s}}{c_4\sqrt{n}}$$

$\bar{s}$

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1 - c_4^2} \qquad CL = \bar{s} \qquad LCL = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1 - c_4^2}$$

$\bar{r}$

$$UCL = D_4\bar{r} \qquad CL = \bar{r} \qquad LCL = D_3\bar{r}$$

Também é utlizada a tabela de coeficientes:

Table X Factors for Constructing Variables Control Charts

	Factor for Control Limits						
	$\overline{X}$ Chart			R Chart		S Chart	
n^*	A_1	A_2	d_2	D_3	D_4	c_4	n
2	3.760	1.880	1.128	0	3.267	0.7979	2
3	2.394	1.023	1.693	0	2.575	0.8862	3
4	1.880	.729	2.059	0	2.282	0.9213	4
5	1.596	.577	2.326	0	2.115	0.9400	5
6	1.410	.483	2.534	0	2.004	0.9515	6
7	1.277	.419	2.704	.076	1.924	0.9594	7
8	1.175	.373	2.847	.136	1.864	0.9650	8
9	1.094	.337	2.970	.184	1.816	0.9693	9
10	1.028	.308	3.078	.223	1.777	0.9727	10
11	.973	.285	3.173	.256	1.744	0.9754	11
12	.925	.266	3.258	.284	1.716	0.9776	12
13	.884	.249	3.336	.308	1.692	0.9794	13
14	.848	.235	3.407	.329	1.671	0.9810	14
15	.816	.223	3.472	.348	1.652	0.9823	15
16	.788	.212	3.532	.364	1.636	0.9835	16
17	.762	.203	3.588	.379	1.621	0.9845	17
18	.738	.194	3.640	.392	1.608	0.9854	18
19	.717	.187	3.689	.404	1.596	0.9862	19
20	.697	.180	3.735	.414	1.586	0.9869	20
21	.679	.173	3.778	.425	1.575	0.9876	21
22	.662	.167	3.819	.434	1.566	0.9882	22
23	.647	.162	3.858	.443	1.557	0.9887	23
24	.632	.157	3.895	.452	1.548	0.9892	24
25	.619	.153	3.931	.459	1.541	0.9896	25

*$n > 25$: $A_1 = 3/\sqrt{n}$ where n = number of observations in sample.

Agora, para resolução da questão, primeiramente serão feitos os cálculos com n = 5.

```
n = 5;
m = 500/n;
% 500 primeiras amostras
x1_1_500 = x1(1:500);
% cria uma matriz de dimensão mxn
matrix_m_n = zeros(m,n);
% cria um vetor de dimensão mx1
vetor_x_barra = zeros(m,1);
% cria um vetor mx1
vetor_r = zeros(m,1);
```

```
% cria um vetor mx1
vetor_s = zeros(m,1);
% Loop responsável por iterar dentro do vetor contendo as primeiras 500
% amostras
for i = 1:1:m
    % Vetor n que percorre as amostras de n em n por loop
    % ex: i = 1, vet_n = (1:5); i=2 vet_n = (6:10)...
    vetor_n = x1_1_500(n*i-(n-1):n*i);
    x_barra = mean(vetor_n);
    r = max(vetor_n) - min(vetor_n);
    s = std(vetor_n);
    vetor_x_barra(i) = x_barra;
    vetor_r(i) = r;
    vetor_s(i) = s;
    matrix_m_n(i,:) = vetor_n;
end

% Calculo das "m" médias

x1_barrra_barra = mean(vetor_x_barra)
```

```
x1_barrra_barra = 20.0863
```

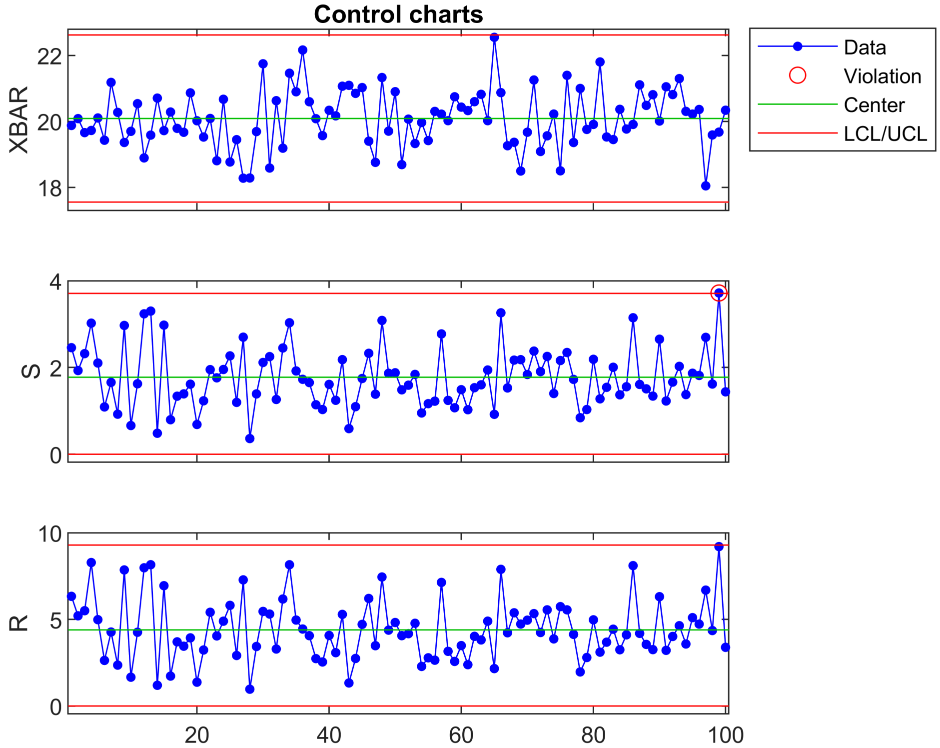
```
r_barra = max(vetor_r) - min(vetor_r)
```

```
r_barra = 8.2314
```

```
s_barra = std(vetor_s)
```

```
s_barra = 0.6926
```

```
controlchart(matrix_m_n,'charttype',{'xbar' 's' 'r'})
```



Agora, para n = 20:

```
n = 20;
m = 500/n;
% 500 primeiras amostras
x1_1_500 = x1(1:500);
% cria uma matriz de dimensão mxn
matrix_m_n = zeros(m,n);
% cria um vetor de dimensão mx1
vetor_x_barra = zeros(m,1);
% cria um vetor mx1
vetor_r = zeros(m,1);
% cria um vetor mx1
vetor_s = zeros(m,1);
% Loop responsável por iterar dentro do vetor contendo as primeiras 500
% amostras
for i = 1:1:m
    % Vetor n que percorre as amostras de n em n por loop
    % ex: i = 1, vet_n = (1:5); i=2 vet_n = (6:10)...
    vetor_n = x1_1_500(n*i-(n-1):n*i);
    x_barra = mean(vetor_n);
    r = max(vetor_n) - min(vetor_n);
    s = std(vetor_n);
    vetor_x_barra(i) = x_barra;
    vetor_r(i) = r;
    vetor_s(i) = s;
    matrix_m_n(i,:) = vetor_n;
```

```
end

x1_barrra_barra = mean(vetor_x_barra)
```

```
x1_barrra_barra = 20.0863
```

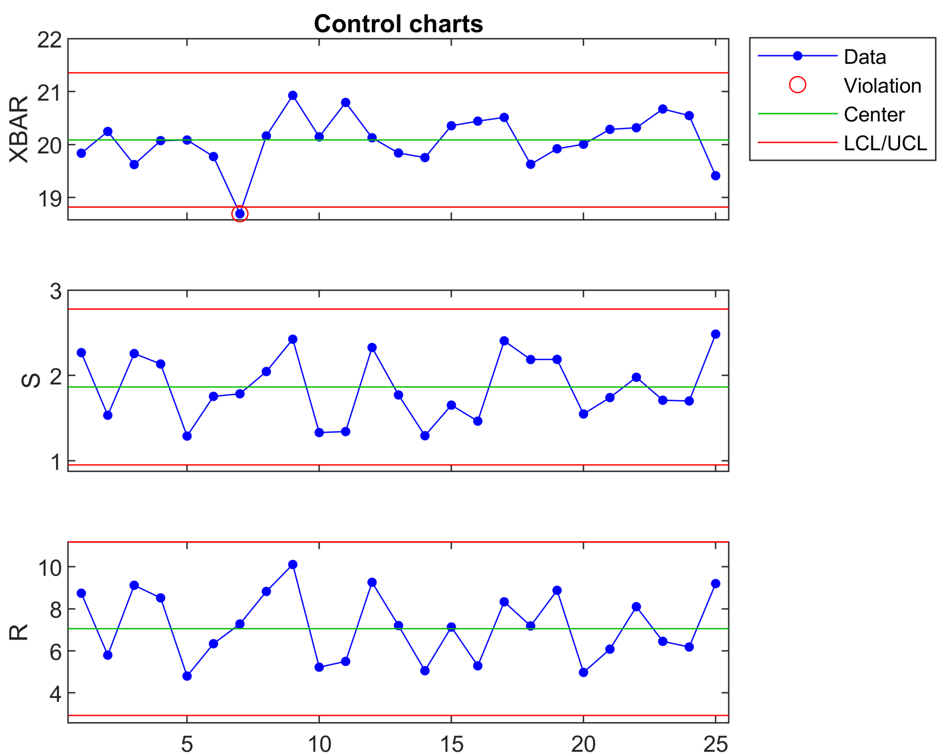
```
r_barra = max(vetor_r) - min(vetor_r)
```

```
r_barra = 5.3088
```

```
s_barra = std(vetor_s)
```

```
s_barra = 0.3856
```

```
controlchart(matrix_m_n,'charttype',{'xbar' 's' 'r'})
```



Observa-se que o aumento do números de amostras por subgrupo para n = 20 ocultou um subgrupo que estava fora de controle, o subgrupo 99 quando n = 5. Além disso, para n = 20 o subgrupo 7 ficou fora de controle estatístico. Dessa forma podemos ter a conclusão que para um processo de verificação de qualidade é melhor optar por uma menor quantidade de amostrar por subgrupo para que sejam avaliados mais subgrupos do que o inverso. Sendo assim, temos um aumento na sensibilidade à detecção de pequenas variações do control chart.

Atividade 2: Variar n para melhorar o desempenho das métricas recall e TD no gráfico de média.

• Use as amostras 1 a 500 de x_1 para obter UCL, LCL, CL para n=5, n=15, n=25, e então avalie as métricas recall e TD usando as amostras 501 a 1000 de x_1.

• Mostre cada um dos gráficos e informe o valor de recall e TD, comentando os resultados.

Primeiramente, temos que as métricas são definidas sendo elas:

$$\text{Recall} = \frac{\text{TP}}{\text{Tp} + \text{FN}}$$

Posteriormente, calculamos o Recall e a Taxa de Deteção (TD) para cada valor de n (5, 10 e 25). O Recall avalia a capacidade do gráfico de controle em identificar eventos fora dos limites de controle. Um aumento no recall indica que o gráfico está se tornando mais sensível à detecção de eventos fora de controle. Buscamos identificar o valor de n que resulta no maior recall sem comprometer significativamente a eficiência.

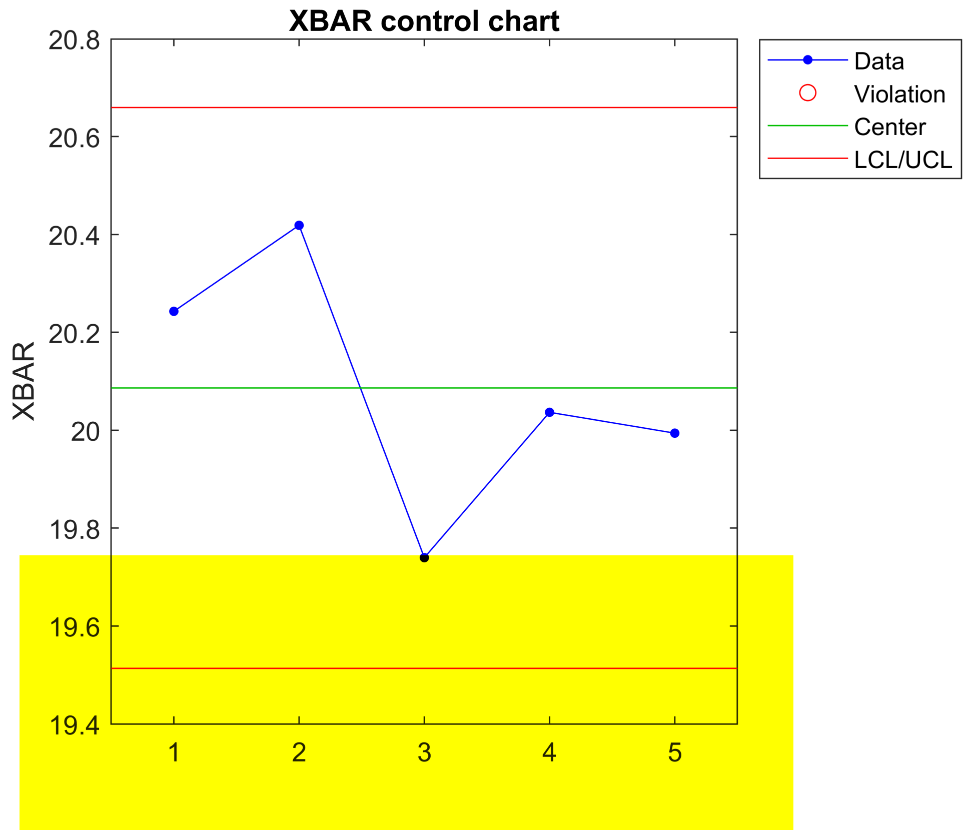
Por outro lado, a Taxa de Deteção (TD) avalia a capacidade do gráfico de controle em identificar eventos dentro dos limites de controle. Um aumento na TD indica que o gráfico está se tornando mais eficiente na identificação de eventos dentro de controle. Buscamos identificar o valor de n que resulta na maior TD sem sacrificar muito em termos de recall. Vale ressaltar que a TD também mede o tempo necessário para a detecção dos eventos, expresso em número de amostras.

Finalmente, os resultados são representados graficamente para uma compreensão visual mais clara.

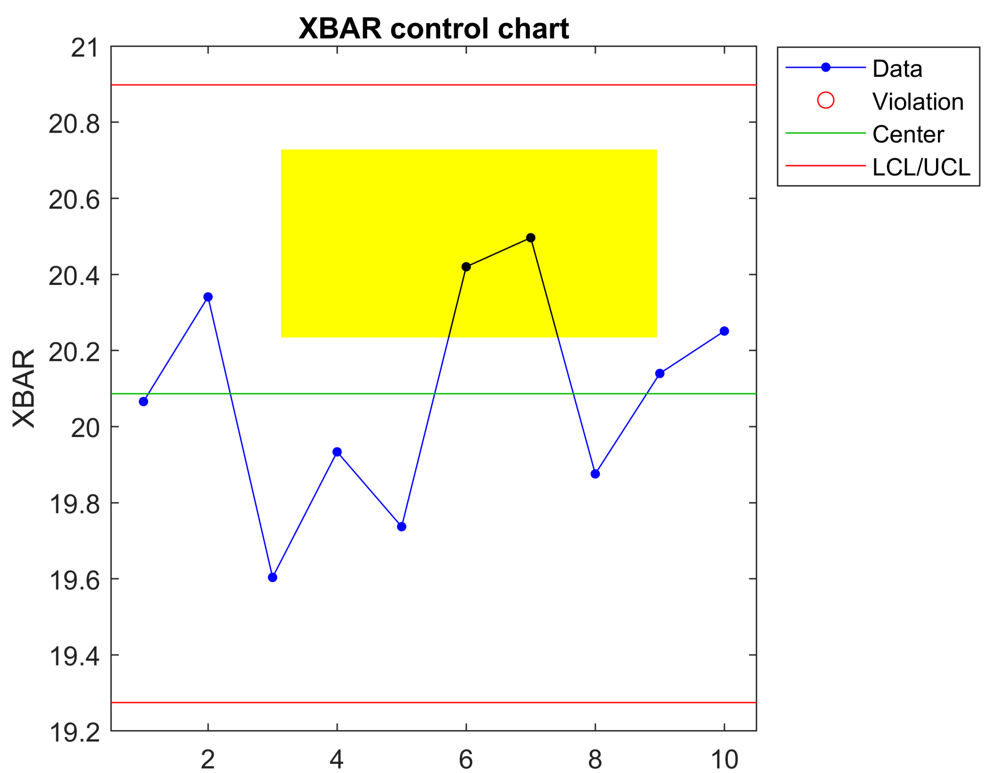
Agora para os cálculos, primeiramente, é necessário dividir as amostras a serem utilizadas. A amostra_1 consiste nos elementos de 1 a 500 de x1, enquanto a amostra_2 abrange os elementos de 501 a 1000 de x1. Em seguida, utilizamos essas amostras para obter os valores de UCL, LCL e CL para diferentes tamanhos de subgrupos (n), especificamente 5, 10 e 25.

Um ponto importante aqui a ser destacado é que o enunciado nos pede para n=15, entretanto 500 não é divisível por 15. Dessa forma, foi optado pela utilização de n = 10.

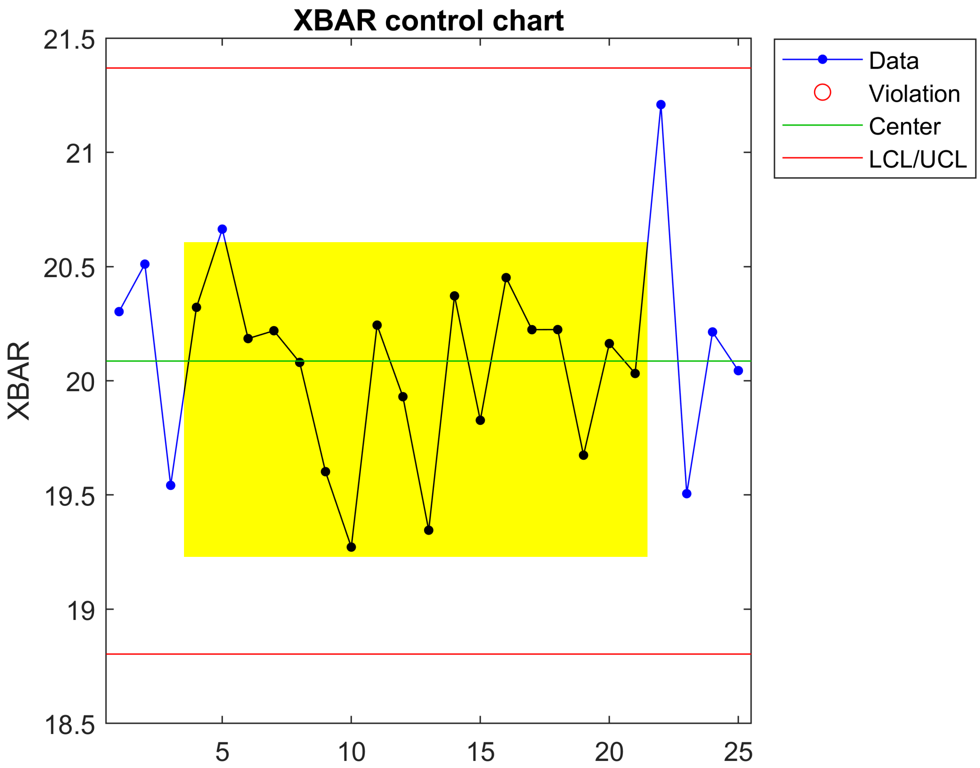
```
samples_1=x1(1:500,1);
samples_2=x1(501:1000,1);

subgrup_1_5=reshape(samples_1,5,[]);
subgrup_1_10=reshape(samples_1,10,[]);
subgrup_1_25=reshape(samples_1,25,[]);
[stats5,plotdata5] = controlchart(subgrup_1_5,'charttype',{'xbar'});
```




```
[stats10,plotdata10] = controlchart(subgrup_1_10,'charttype',{'xbar'});
```



```
[stats25,plotdata25] = controlchart(subgrup_1_25,'charttype',{'xbar'});
```




```
ucl_5=plotdata5.ucl(1);
lcl_5=plotdata5.lcl(1);
cl_5=plotdata5.cl(1);
ucl_10=plotdata10.ucl(1);
lcl_10=plotdata10.lcl(1);
cl_10=plotdata10.cl(1);
ucl_25=plotdata25.ucl(1);
lcl_25=plotdata25.lcl(1);
cl_25=plotdata25.cl(1);

subgrup_2_5=reshape(samples_2,5,[]);
subgrup_2_10=reshape(samples_2,10,[]);
subgrup_2_25=reshape(samples_2,25,[]);

med_5=mean(subgrup_2_5);
med_10=mean(subgrup_2_10);
med_25=mean(subgrup_2_25);

out_lim_5=find(med_5>ucl_5|med_5<lcl_5);
out_lim_10=find(med_10>ucl_10|med_10<lcl_10);
out_lim_25=find(med_25>ucl_25|med_25<lcl_25);

TD_5=5*(out_lim_5(1)-1)
```

```
TD_5 = 0
```

```
TD_10=10*(out_lim_10(1)-1)
```

TD_10 = 30

```
TD_25=25*(out_lim_25(1)-1)
```

TD_25 = 50

```
recall_5=numel(out_lim_5)/numel(med_5)
```

recall_5 = 0.9700

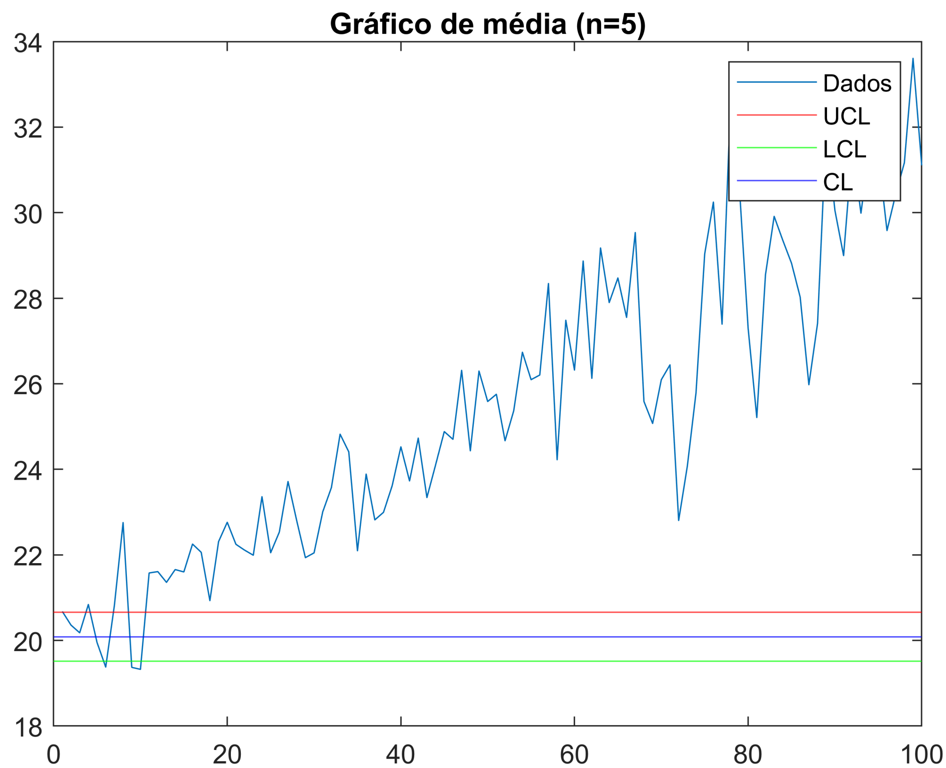
```
recall_10=numel(out_lim_10)/numel(med_10)
```

recall_10 = 0.9200

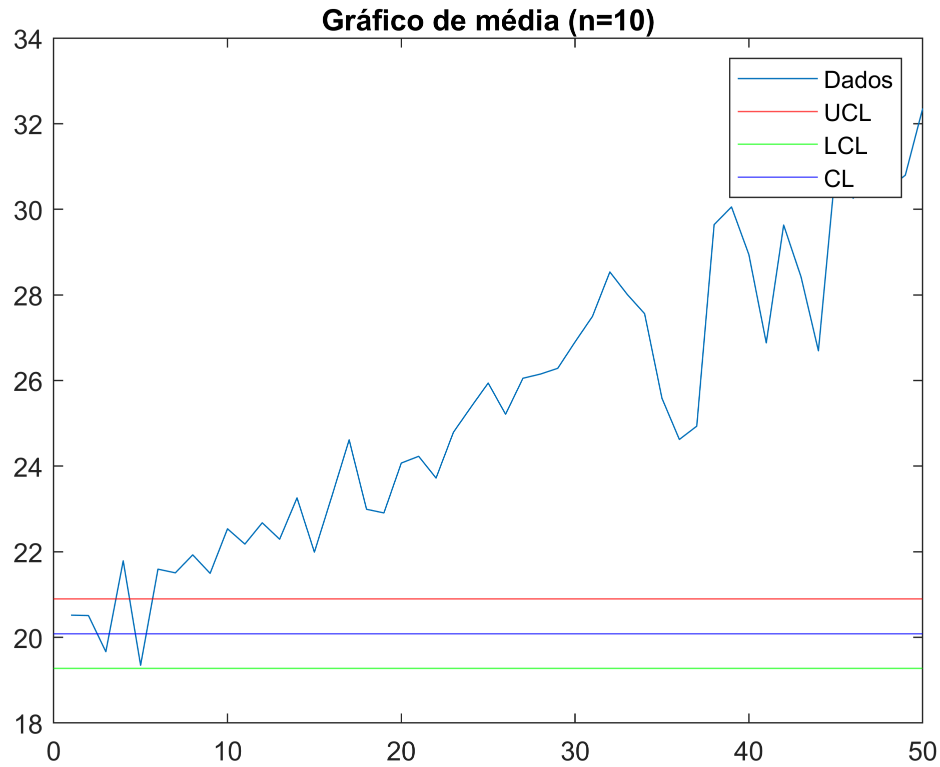
```
recall_25=numel(out_lim_25)/numel(med_25)
```

recall_25 = 0.9000

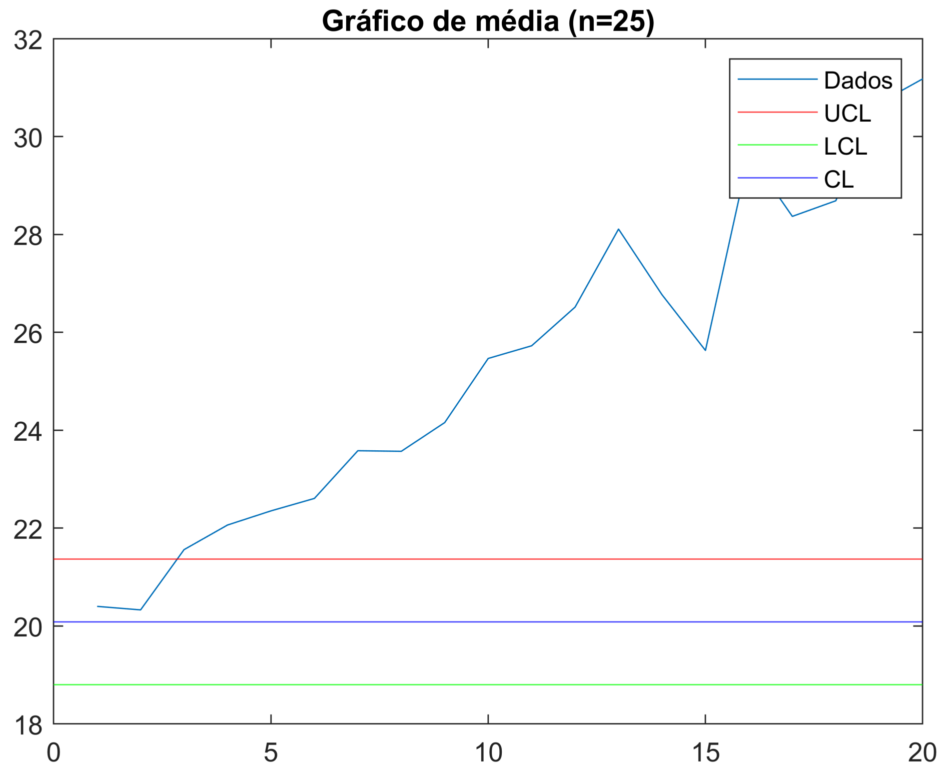
```
figure
plot(med_5)
yline(ucl_5,'color', "r")
yline(lcl_5,'color', "g")
yline(cl_5,'color', "b")
title('Gráfico de média (n=5)')
legend('Dados','UCL','LCL','CL')
```



```
figure
plot(med_10)
yline(ucl_10,'color', "r")
yline(lcl_10,'color', "g")
yline(cl_10,'color', "b")
title('Gráfico de média (n=10)')
legend('Dados','UCL','LCL','CL');
```




```
figure
plot(med_25)
yline(ucl_25,'color', "r")
yline(lcl_25,'color', "g")
yline(cl_25,'color', "b")
title('Gráfico de média (n=25)')
legend('Dados','UCL','LCL','CL')
```



Analisando os resultados podemos chegar nas seguintes conclusões:

- Se o objetivo principal é maximizar o Recall, ou seja, a capacidade de identificar eventos fora dos limites de controle, então o valor de "n" associado ao maior Recall é preferível. Neste caso, "n=5" possui um Recall mais alto (0.97), indicando que o gráfico é mais sensível à detecção de eventos fora de controle.
- No entanto, se o foco estiver na maximização da Taxa de Deteção (TD), ou seja, identificar eficientemente eventos dentro dos limites de controle, então o valor de "n" associado à maior TD seria mais adequado. Neste caso, "n=25" tem a maior TD (50), sugerindo uma detecção mais eficiente de eventos dentro dos limites de controle.

Em resumo, a escolha entre "n=5", "n=10" e "n=25" dependerá das prioridades do sistema, se é mais crítico detectar eventos fora dos limites de controle (maior Recall) ou identificar eficientemente eventos dentro dos limites de controle (maior TD). Em alguns casos, pode ser necessário fazer um compromisso entre Recall e TD, escolhendo um valor de "n" que atenda às necessidades específicas do seu contexto.

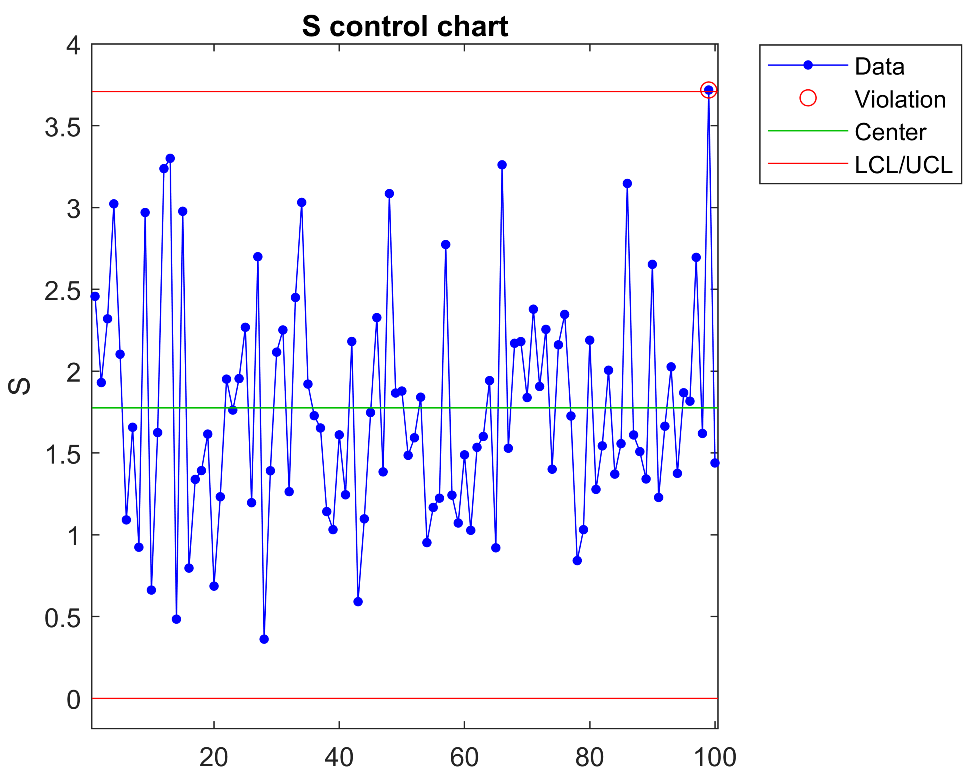
Atividade 3: Variar n para melhorar o desempenho das métricas recall e TD no gráfico de desvio padrão.

• Use as amostras 1 a 500 de x_1 para obter UCL, LCL, CL para n=5, n=15, n=25, e então avalie as métricas recall e TD usando as amostras 801 a 1000 devx_1 .

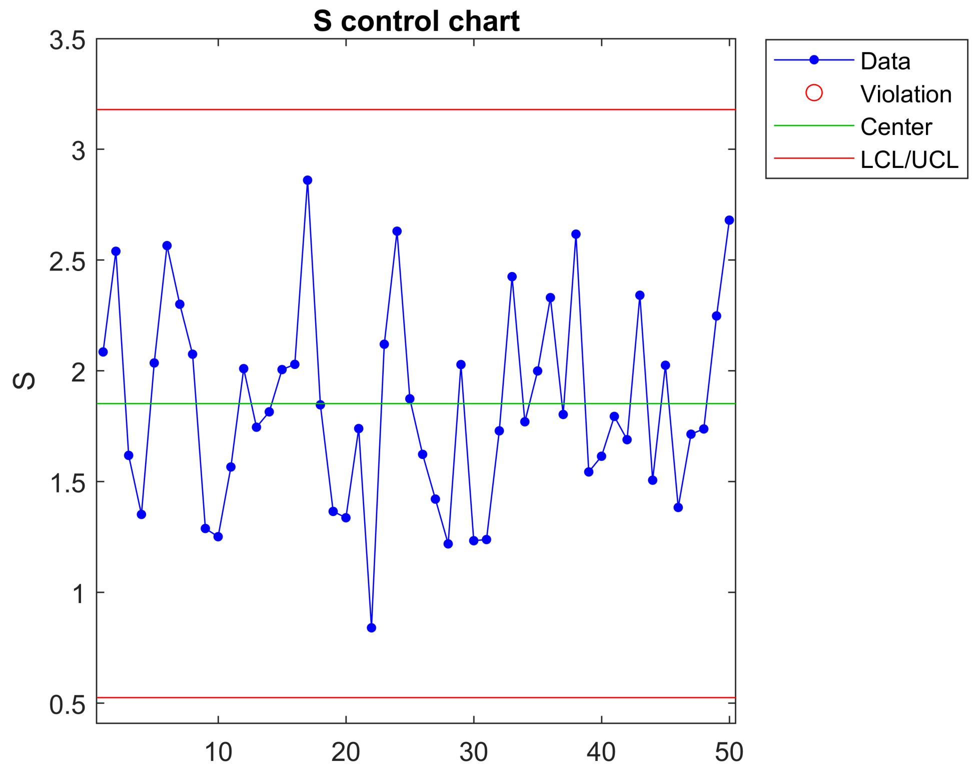
• Mostre cada um dos gráficos e informe o valor de recall e TD, comentando os resultados.

Da mesma forma que foi feita a atividade 2, faremos um processo consideravelmente parecido No entanto, em vez de examinar as médias, concentramo-nos na análise dos desvios padrões.

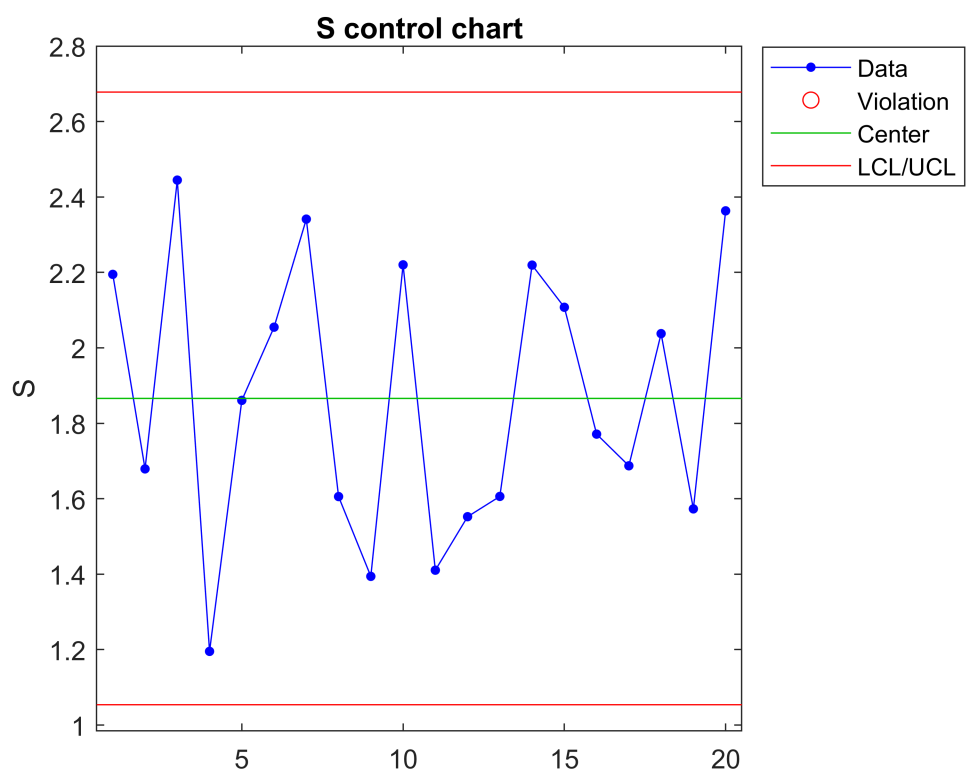
```
samples_1=x1(1:500,1);
samples_2=x1(801:1000,1);

subgrup_1_5=reshape(samples_1,5,[]);
subgrup_1_10=reshape(samples_1,10,[]);
subgrup_1_25=reshape(samples_1,25,[]);

[stats5,plotdata5] = controlchart(subgrup_1_5','charttype',{'s'});
```




```
[stats10,plotdata10] = controlchart(subgrup_1_10','charttype',{'s'});
```




```
[stats25,plotdata25] = controlchart(subgrup_1_25','charttype',{'s'});
```



```
ucl_5=plotdata5.ucl(1);
lcl_5=plotdata5.lcl(1);
cl_5=plotdata5.cl(1);
ucl_10=plotdata10.ucl(1);
lcl_10=plotdata10.lcl(1);
cl_10=plotdata10.cl(1);
ucl_25=plotdata25.ucl(1);
lcl_25=plotdata25.lcl(1);
cl_25=plotdata25.cl(1);

subgrup_2_5=reshape(samples_2,5,[]);
subgrup_2_10=reshape(samples_2,10,[]);
subgrup_2_25=reshape(samples_2,25,[]);

std_5=std(subgrup_2_5);
std_10=std(subgrup_2_10);
std_25=std(subgrup_2_25);

out_lim_5=find(std_5>ucl_5|std_5<lcl_5);
out_lim_10=find(std_10>ucl_10|std_10<lcl_10);
out_lim_25=find(std_25>ucl_25|std_25<lcl_25);

TD_5=5*(out_lim_5(1)-1)
```

```
TD_5 = 5
```

```
TD_10=10*(out_lim_10(1)-1)
```

```
TD_10 = 0
```

```
TD_25=25*(out_lim_25(1)-1)
```

```
TD_25 = 0
```

```
recall_5=numel(out_lim_5)/numel(med_5)
```

```
recall_5 = 0.1800
```

```
recall_10=numel(out_lim_10)/numel(med_10)
```

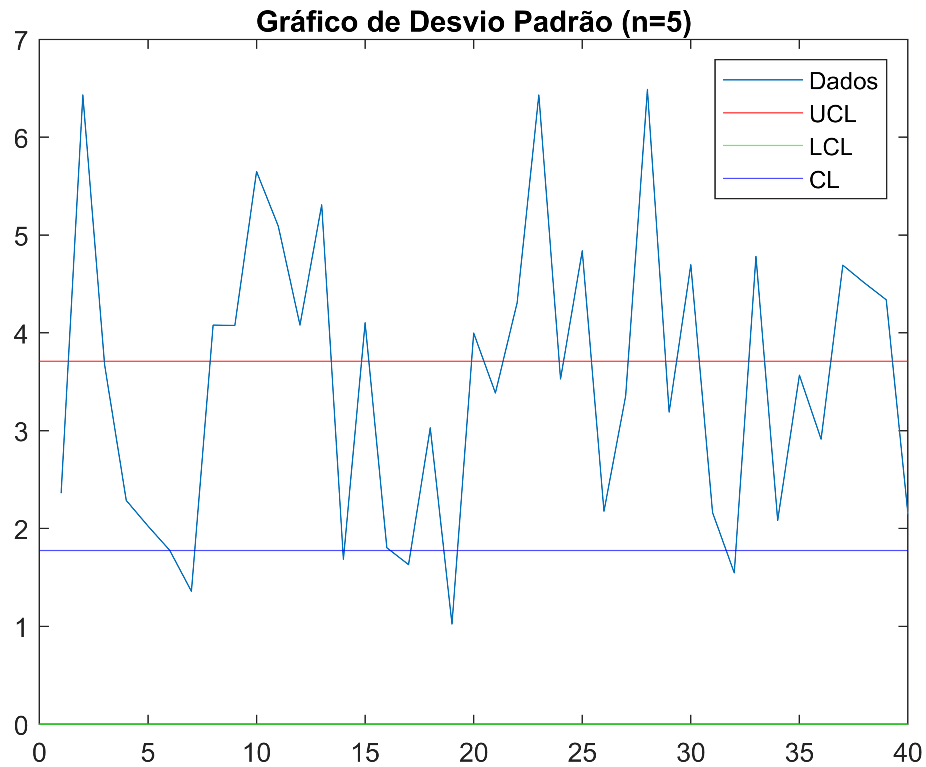
```
recall_10 = 0.3200
```

```
recall_25=numel(out_lim_25)/numel(med_25)
```

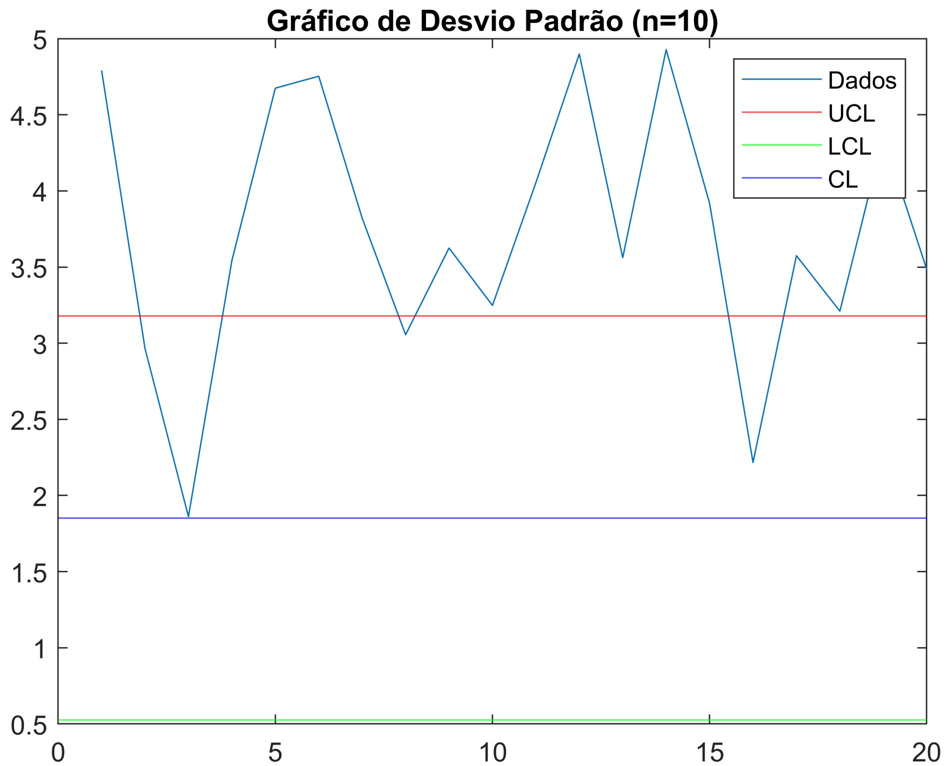
```
recall_25 = 0.4000
```

```
figure
plot(std_5)
yline(ucl_5,'color', "r")
```

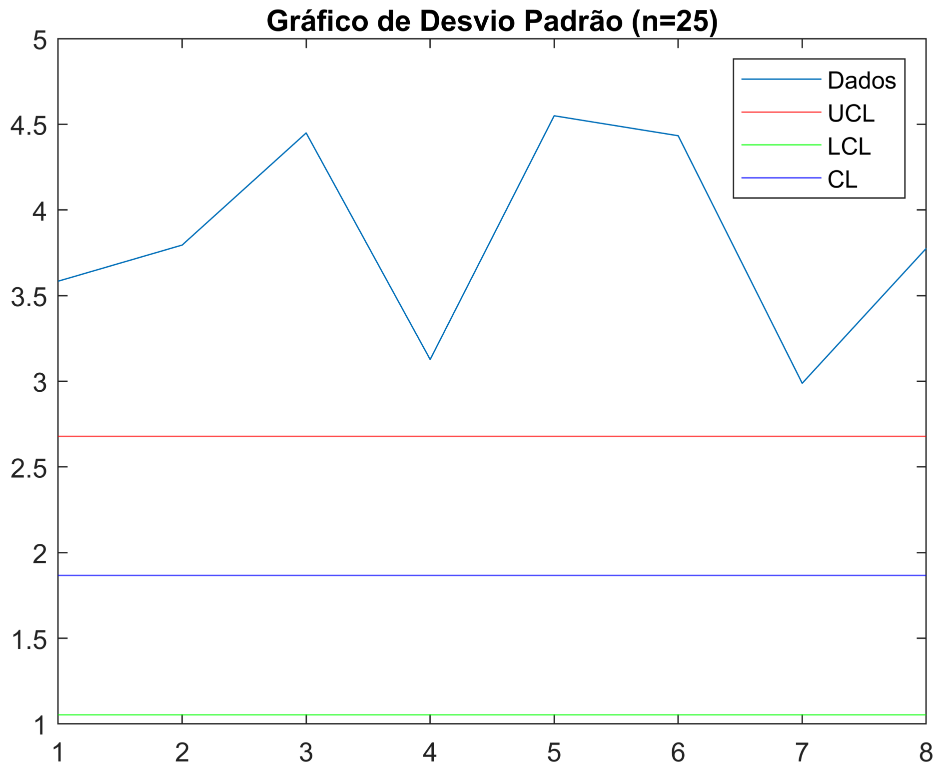
```
ylline(lcl_5,'color', "g")
yline(cl_5,'color', "b")
title('Gráfico de Desvio Padrão (n=5)')
legend('Dados','UCL','LCL','CL')
```




```
figure
plot(std_10)
yline(ucl_10,'color', "r")
yline(lcl_10,'color', "g")
yline(cl_10,'color', "b")
title('Gráfico de Desvio Padrão (n=10)')
legend('Dados','UCL','LCL','CL')
```




```
figure
plot(std_25)
yline(ucl_25,'color', "r")
yline(lcl_25,'color', "g")
yline(cl_25,'color', "b")
title('Gráfico de Desvio Padrão (n=25)')
legend('Dados','UCL','LCL','CL')
```



A escolha do "n" para otimizar Recall e Taxa de Deteção (TD) ao analisar o desvio padrão segue um princípio semelhante ao processo utilizado ao analisar a média. O objetivo ainda é encontrar um equilíbrio entre sensibilidade na detecção de eventos fora dos limites de controle (medida pelo Recall) e eficiência na identificação de eventos dentro dos limites de controle (medida pela TD).

Sendo assim, escolhe-se n = 5 pois é onde temos um mínimo controle no TD, com TD = 5 e Recall = 0.18 com o melhor resultado em relação a n = 10 e 25.

Para n = 10, temos TD = 0, Recall = 0.32. E para n = 25, temos TD = 0 e Recall = 0.40.

Atividade 4: Fazer os gráficos de controle para média e desvio padrão nos dados de teste .

• Use os valores de UCL, LCL e CL obtidos para o valor de n da atividade 2 usando x_1 que deram o melhor recall.

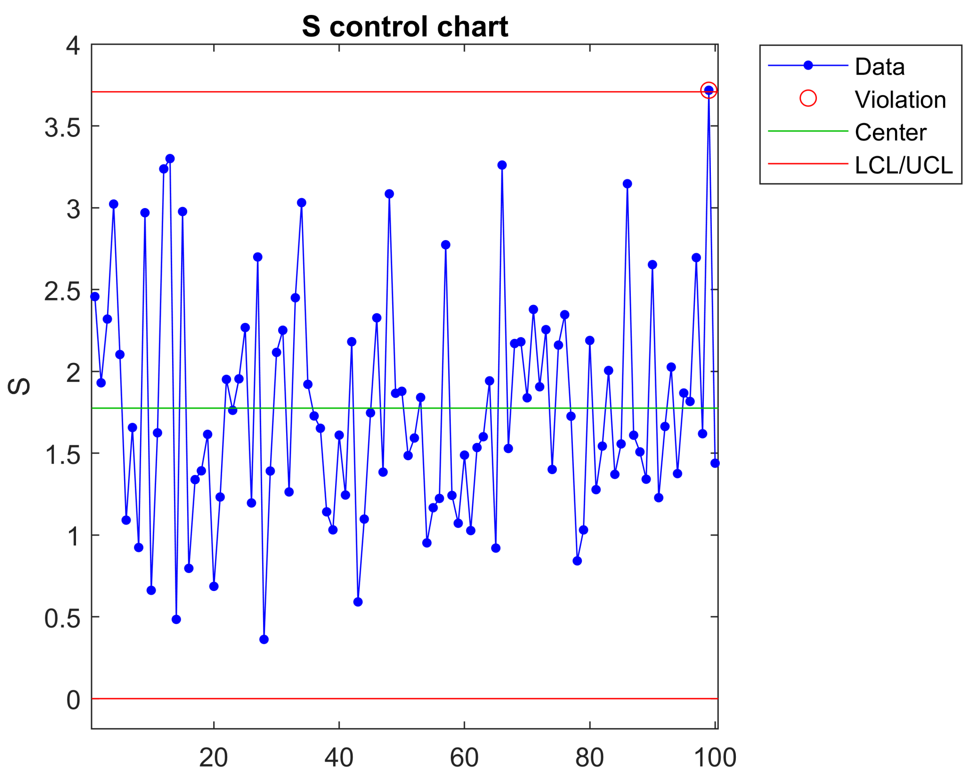
• Faça o gráfico de controle de média de x_1 e meça TD e recall.

• Faça o gráfico de controle de desvio padrão de x_1 e meça TD e recall.

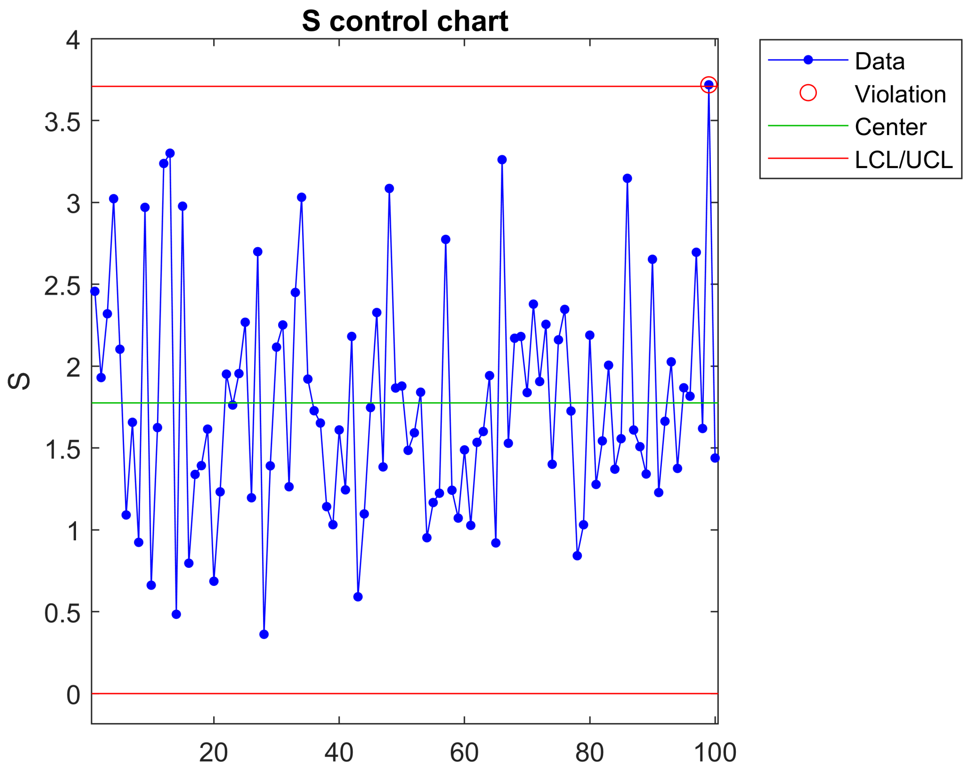
Como foi discutido na conclusão da atividade 2 e 3, temos uma intersecção na escolhe para n = 5. Logo, será escolhido n = 5.

Sendo assim, os valores de UCL, LCL e CL para média e desvio padrão, são obtidos usando as amostras de 1 até 500 de x1. Em seguida, esses limites são aplicados nos dados de teste x2. Dessa forma, são calculados os valores de recall e TD para a média e para o desvio padrão, os gráficos utilizados para o cálculo estão logo abaixo.

```
amostra_atvid4=x1(1:500,1);
subgrup_5_x1=reshape(amostra_atvid4,5,[]);

[stats10_xbar,plotdata10_xbar] = controlchart(subgrup_5_x1',"charttype","s");
```




```
[stats5_s,plotdata5_s] = controlchart(subgrup_5_x1',"charttype","s");
```




```
ucl_5_xbar=plotdata10_xbar.ucl(1);
lcl_5_xbar=plotdata10_xbar.lcl(1);
cl_5_xbar=plotdata10_xbar.cl(1);
ucl_5_s=plotdata5_s.ucl(1);
lcl_5_s=plotdata5_s.lcl(1);
cl_5_s=plotdata5_s.cl(1);

subgrup_5_x2=reshape(x2,5,[]);
medias_x2=mean(subgrup_5_x2);
std_x2=std(subgrup_5_x2);

out_lim_5_xbar=find(medias_x2>ucl_5_xbar|medias_x2<lcl_5_xbar);
out_lim_5_s=find(std_x2>ucl_5_s|std_x2<lcl_5_s);

TD_5_xbar=5*(out_lim_5_xbar(1)-1)
```

```
TD_5_xbar = 0
```

```
recall_5_xbar=numel(out_lim_5_xbar)/numel(medias_x2)
```

```
recall_5_xbar = 1
```

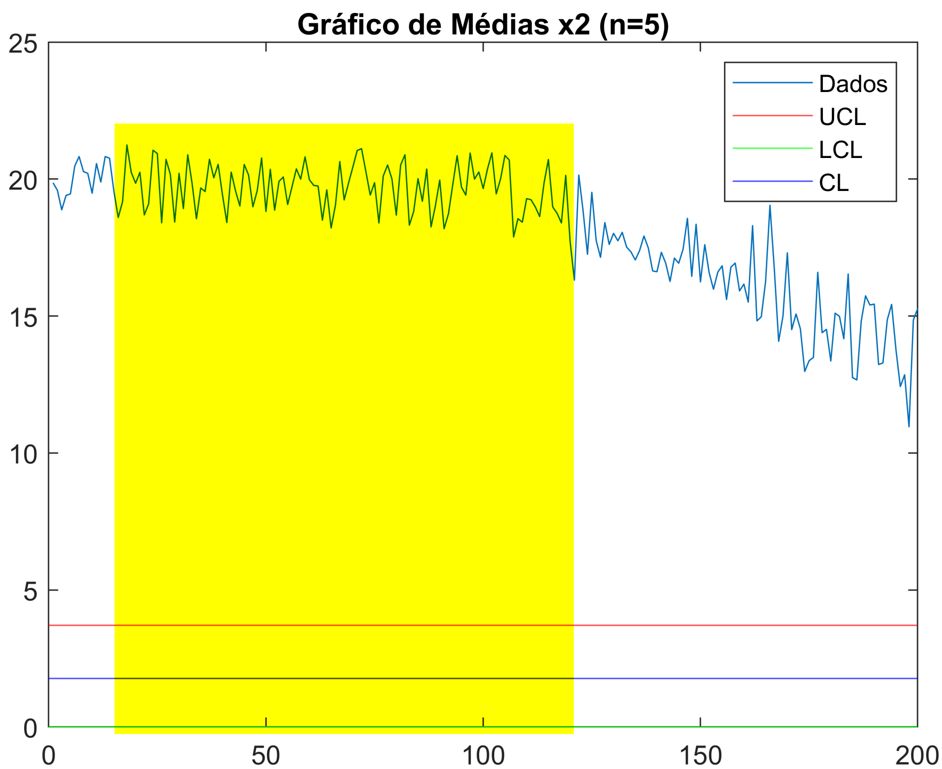
```
TD_5_s=5*(out_lim_5_s(1)-1)
```

```
TD_5_s = 800
```

```
recall_5_s=numel(out_lim_5_s)/numel(std_x2)
```

```
recall_5_s = 0.0450
```

```
figure
plot(medias_x2)
yline(ucl_5_xbar, "color", "r")
yline(lcl_5_xbar,"color", "g")
yline(cl_5_xbar, "color", "b")
title('Gráfico de Médias x2 (n=5)')
legend('Dados','UCL','LCL','CL')
```




```
figure
plot(std_x2)
yline(ucl_5_s,"color", "r")
yline(lcl_5_s,"color", "g")
yline(cl_5_s, "color", "b")
title('Gráfico de Desvio Padrão x2 (n=5)')
legend('Dados','UCL','LCL','CL')
```

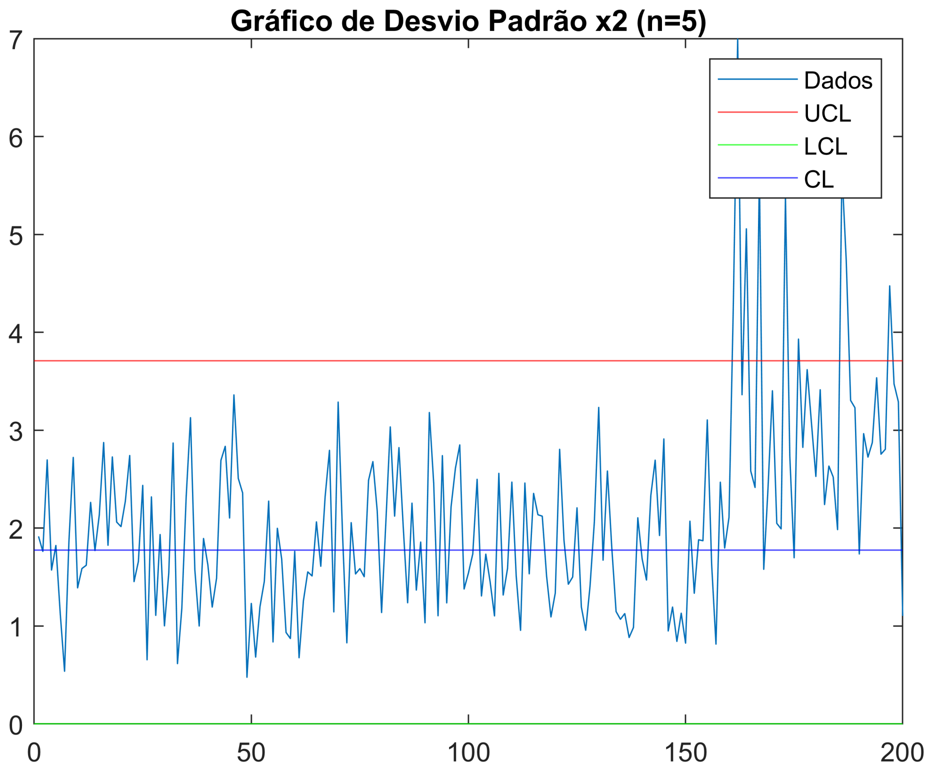


Gráfico de Desvio Padrão x2 (n=5)
Dados
UCL
LCL
CL
0
1
2
3
4
5
6
7
0
50
100
150
200