

## EPC 7 - Reposição

Sejam as equações para fazer os gráficos de controle abaixo:

Gráfico de controle da média

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

Gráfico de controle do desvio padrão

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

Gráfico de controle do range

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

$$\bar{r} = \frac{1}{m} \sum_{i=1}^m r_i \quad \bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i \quad \bar{s} = \frac{1}{m} \sum_{i=1}^m s_i, \text{ sendo } r_i, s_i, \bar{x}_i \text{ calculados com } n \text{ amostras.}$$

Os parâmetros  $c_4, D_3, D_4$  são obtidos da Tabela X do livro texto, em função de  $n$ .

O parâmetro  $m$  é escolhido para caracterizar adequadamente os valores de  $\bar{r}, \bar{s}, \bar{\bar{x}}$  da população.

O parâmetro  $n$  é escolhido para definir o intervalo de confiança e assim permitir a detecção de pequenas mudanças na variável monitorada do processo.

Pode-se usar as funções do Matlab ou Python para facilitar a obtenção dos resultados, sem a necessidade de usar os parâmetros da tabela acima.

Python: `ShewhartControlModel`

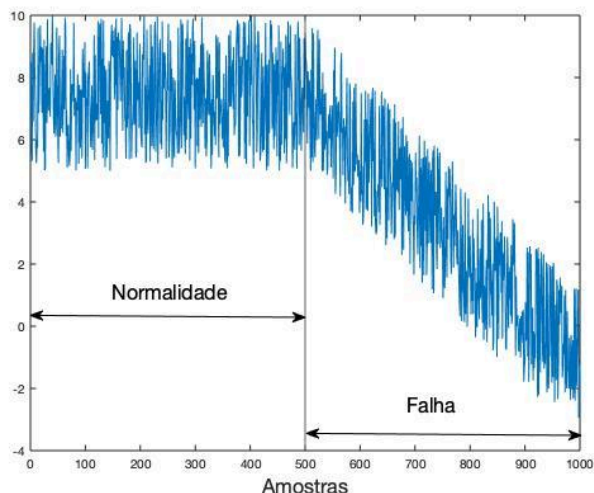
Matlab: `controlchart`

É fornecida uma série temporal  $x$  para construir os gráficos de controle, da seguinte forma:

$x_1$ : 1000 amostras para a etapa de treino e validação, ou seja, para extrair as informações  $\bar{r}, \bar{s}, \bar{\bar{x}}$  e selecionar o valor de  $n$ .

$x_2$ : 1000 amostras para teste, visam avaliar as métricas do monitoramento usando um novo conjunto de dados.

Tanto  $x_1$  quanto  $x_2$  têm as primeiras 500 amostras de normalidade e as 500 amostras seguintes de falha, conforme abaixo:



Para as métricas de desempenho, medem-se os verdadeiros positivos (TP) e falsos negativos (FN) conforme abaixo:

Para a média:

Medição de TP: amostras de 501 a 1000 nas quais a estatística de teste foi maior que UCL ou menor que LCL.

Medição de FN: amostras de 501 a 1000 nas quais a estatística de teste ficou entre UCL e LCL.

Para o desvio padrão e range: (o desvio padrão aumenta a partir da amostra 801)

Medição de TP: amostras de 801 a 1000 nas quais a estatística de teste foi maior que UCL ou menor que LCL.

Medição de FN: amostras de 801 a 1000 nas quais a estatística de teste ficou entre UCL e LCL.

**Métricas:** usaremos as métricas abaixo para medir o desempenho:

$Recall = \frac{TP}{TP + FN}$ , onde TP=True positives, FN=false negatives. [Ver detalhes](#)

**TD:** Tempo para detecção: quantas amostras foram necessárias para detectar o mudança. Por exemplo, caso  $n=5$  e após a terceira média calculada  $\bar{\bar{x}}$  ultrapassou o UCL, então o tempo foi de  $3*5=15$  amostras.

**Importante: Cálculo de  $\bar{r}$ ,  $\bar{s}$ ,  $\bar{\bar{x}}$ :**

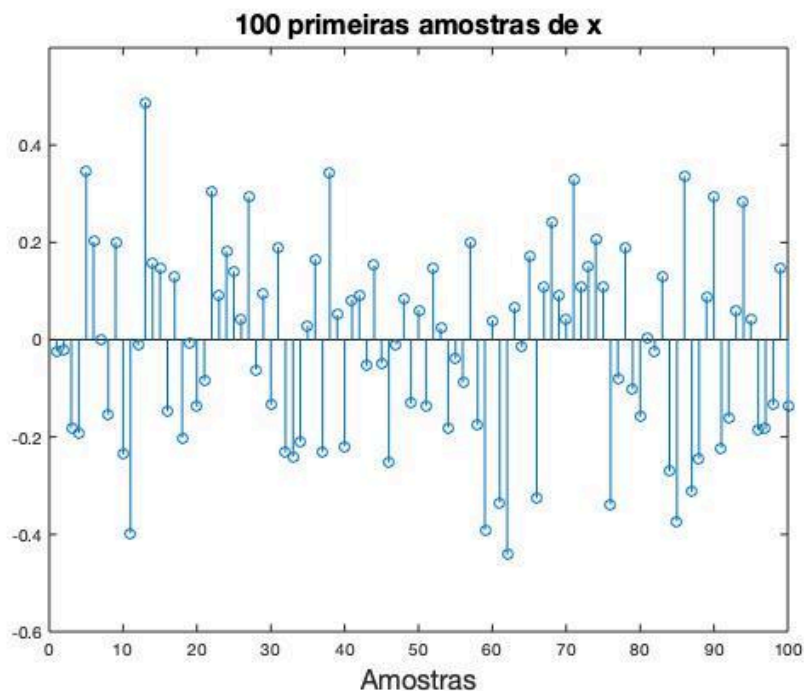
Como  $x$  é uma série temporal, essas estatísticas são calculadas usando  $n$  amostras em sequência. Seja a figura abaixo, e o cálculo da média  $\bar{x}_i$ , assumindo  $n = 10$ . As primeiras 10 amostras (1 a 10) permitem calcular  $\bar{x}_1$ . Para calcular  $\bar{x}_2$ , usa-se as próximas amostras (11 a 20).

O uso de janelas deslizantes cria correlação entre os valores calculados, e deve ser evitado.

Por fim, por ser uma série temporal, não se amostra aleatoriamente os valores para fazer os cálculos de  $\bar{r}$ ,  $\bar{s}$ ,  $\bar{x}$ .

Tanto para o treino quanto para o monitoramento, os valores de  $\bar{r}$ ,  $\bar{s}$ ,  $\bar{x}$  serão calculados usando  $n$  amostras.

No caso de vocês, há 500 amostras de normalidade, e podem calcular 100 valores de  $\bar{r}$ ,  $\bar{s}$ ,  $\bar{x}$  usando  $n = 5$ , e 25 valores usando  $n = 20$ .



Os dados estão nos formatos abaixo, sendo que tanto X1 (treino/validação) quanto X2(teste) contêm 1000 amostras e 21 colunas. Use a coluna de X1 e de X2 referente a seu número I.

### 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  $\bar{r}$ ,  $\bar{s}$ ,  $\bar{x}$  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  $n = 20$ .

### 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.

### **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 de  $x_1$ .
- Mostre cada um dos gráficos e informe o valor de recall e TD, comentando os resultados.

### **Atividade 4: Fazer os gráficos de controle para média e desvio padrão nos dados de teste $x_2$ .**

- 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  $x_2$  e meça TD e recall.
- Faça o gráfico de controle de desvio padrão de  $x_2$  e meça TD e recall.