Controle e Monitoramento de Processos Multivariados

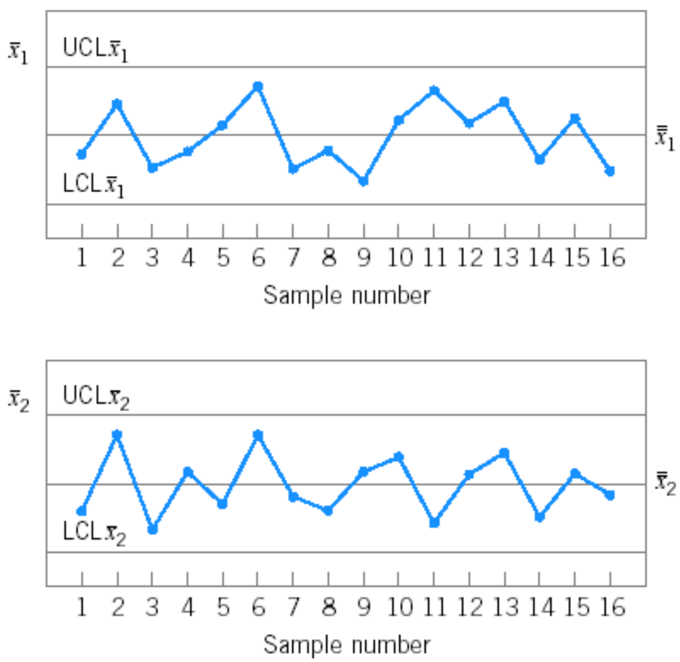
Profa. Carmela Maria Polito Braga – DELT/EEUFMG

O PROBLEMA DE CONTROLE MULTIVARIADO

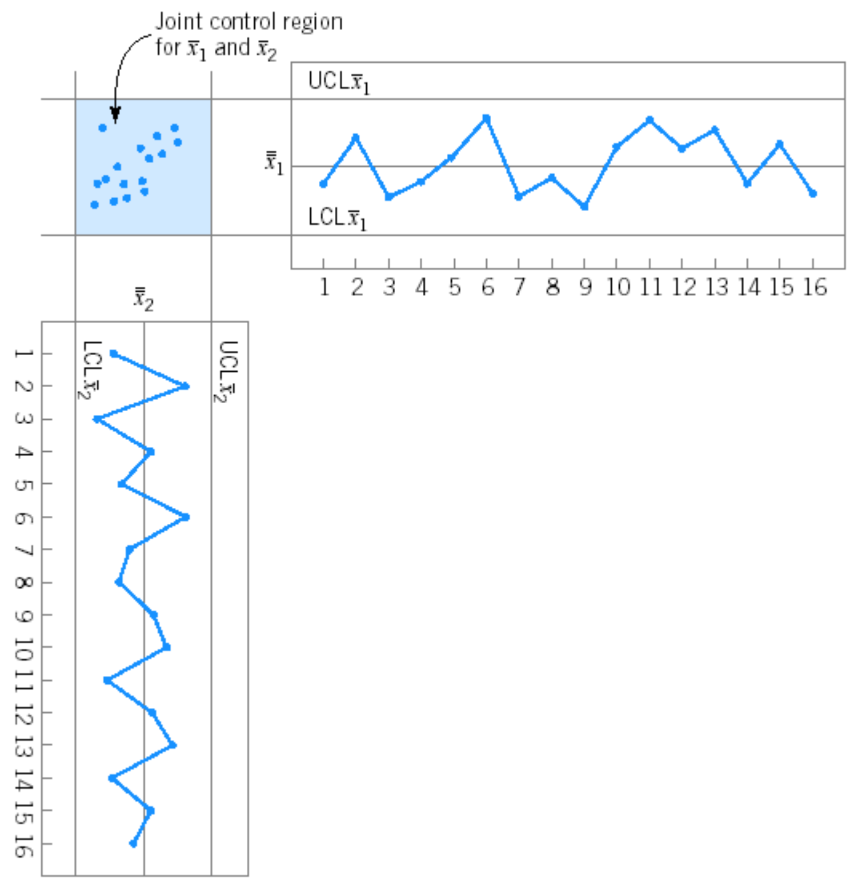
- Muitas vezes o monitoramento ou o controle simultâneo de duas ou mais características de qualidade relacionadas é necessário.
  - P.ex., considere a necessidade de um rolamento possuir os diâmetros interno e externo em conformidade para ser adequado ao uso.
  - O monitoramento de cada um destes diâmetros independentemente pode levar a resultados equivocados.
- O uso de múltiplas cartas univariadas de $\bar{x}$ independentes distorce o monitoramento simultâneo de médias, conforme mostram as figuras 1 e 2.

O PROBLEMA DE CONTROLE MULTIVARIADO

Uso de múltiplas cartas univariadas de $\overline{x}$



Cartas $\overline{x}$ para os diâmetros interno (x_1) e externo (x_2)



Região de controle usando limites de controle independentes para x_1 e .x_2

O PROBLEMA DE CONTROLE MULTIVARIADO

- Erro do tipo I e probabilidade de um ponto plotado corretamente em controle não são iguais aos níveis de advertência para as cartas de controle individuais.
  - A distorção nos procedimentos de monitoramento do processo aumentam à medida em que o número de características de qualidade aumenta para:

$$P\{\text{todas } p \text{ médias plotadas em controle}\} = (1-\alpha)^p$$

  - Lembrando que:

$$\alpha = \mathrm{P}\{\text{erro tipo I}\} = \mathrm{P}\{\text{rejeitar } \mathrm{H_o} \mid \mathrm{H_o} \text{ é verdadeira}\}$$

DESCRIÇÃO DE DADOS MULTIVARIADOS

Distribuição Normal Multivariada

- A função de densidade de probabilidade normal univariada é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

- Sendo a média da distribuição normal μ e a variância σ^2.
- Observe que, a menos do sinal -, o termo no expoente da distribuição normal pode ser escrito como:

$$(x-\mu)\left(\sigma^2\right)^{-1}(x-\mu)$$

DESCRIÇÃO DE DADOS MULTIVARIADOS

- A mesma abordagem pode ser usada no caso multivariado. Suponha o caso de p variáveis dadas por $x_1, x_2, \ldots.., x_p$. Estas variáveis podem ser arranjadas em um vetor de p componentes como $x' = [x_1, x_2, \ldots\ldots, x_p]$.
- Seja $\mu' = [\mu_1, \mu_2, \ldots., \mu_p]$ o vetor de médias de x's e sejam as variâncias e covariâncias de x contidas em uma matriz de convariâncias Σ_{pxp}.
- A diagonal principal de Σ corresponde às variâncias dos x's e os demais elementos as covariâncias. Pode-se expressar a **distância quadrada padronizada (generalizada)** de **x a μ** como:

$$(x-\mu)'\sum^{-1}(x-\mu)$$

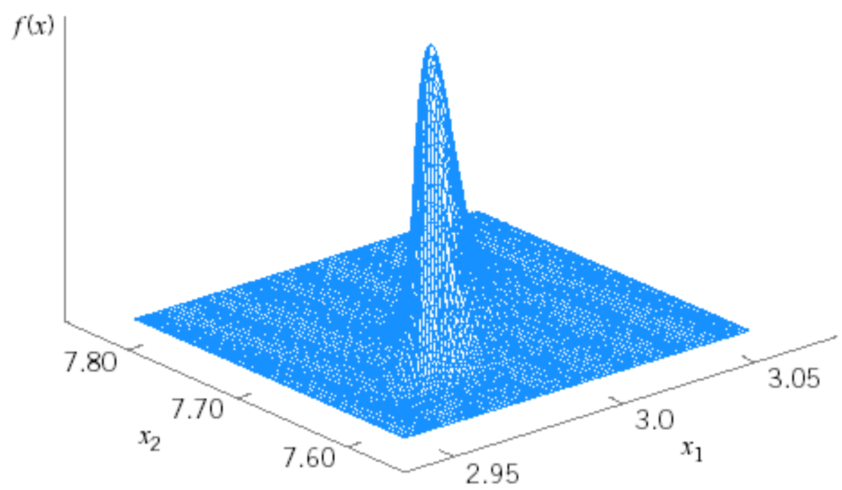
DESCRIÇÃO DE DADOS MULTIVARIADOS

- Então, a função de densidade de probabilidade normal multivariada é:

$$f(x) = \frac{1}{2\pi^{p/2}\left|\sum\right|^{1/2}} e^{-\frac{1}{2}(x-\mu)' S^{-1}(x-\mu)} \qquad -\infty < xj < \infty, j = 1, 2, \ldots, p.$$

- A distribuição normal multivariada para p=2 variáveis é chamada normal bivariada, como a figura abaixo. Note que a função densidade é uma superfície.

O Coeficiente de correlação entre as duas variáveis neste exemplo é 0.8, o que produz uma probabilidade concentrada ao longo de uma linha.



DESCRIÇÃO DE DADOS MULTIVARIADOS

Vetor de Média Amostral e Matriz de Covariância

- Suponha que tenhamos uma amostra aleatória de uma distribuição normal multivariada: $x_1, x_2, \ldots, x_n$.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad S = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)'$$

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2 \qquad S_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)\left(x_{ik} - \overline{x_k}\right)$$

DESCRIÇÃO DE DADOS MULTIVARIADOS

Dados Subgrupados

- Suponha que duas características de qualidade x_1 e x_2 são conjuntamente distribuidas de acordo com a distribuição normal bivariada.
  - Sejam μ_1 e μ_2 os valores médios das características de qualidade e σ_1 e σ_2 seus desvios padrão.
  - A covariância entre x_1 e x_2 é denotada por σ_{12}.
- A estatística:

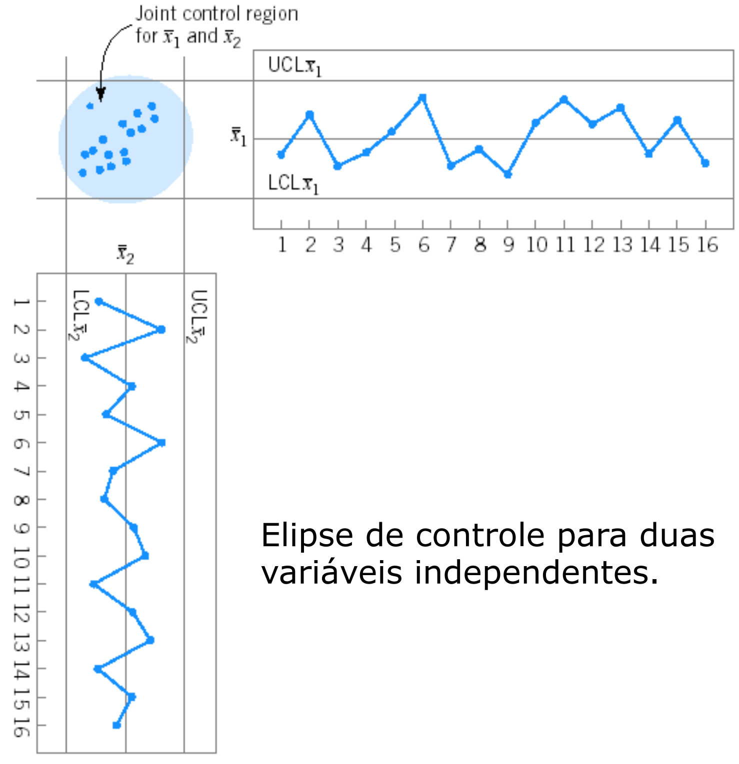
$$\chi_0^2 = \frac{n}{\sigma_1\sigma_2 - \sigma_{12}^2}\left[\sigma_2^2\left(\overline{x}_1 - \mu_1\right)^2 + \sigma_1^2\left(\overline{x}_2 - \mu_2\right)^2 - 2\sigma_{12}\left(\overline{x}_1 - \mu_1\right)\left(\overline{x}_2 - \mu_2\right)\right]$$

tem distribuição χ^2 com dois graus de liberdade e pode ser usada como base para uma carta de controle para as médias do processo μ_1 e μ_2.

DESCRIÇÃO DE DADOS MULTIVARIADOS

- Os valores de χ_0^2 devem ser menores que o limite de controle superior $LCS = \chi^2_{\alpha,2}$, onde $\chi^2_{\alpha,2}$ é o ponto percentual α superior da distribuição qui-quadrada, com dois graus de liberdade.

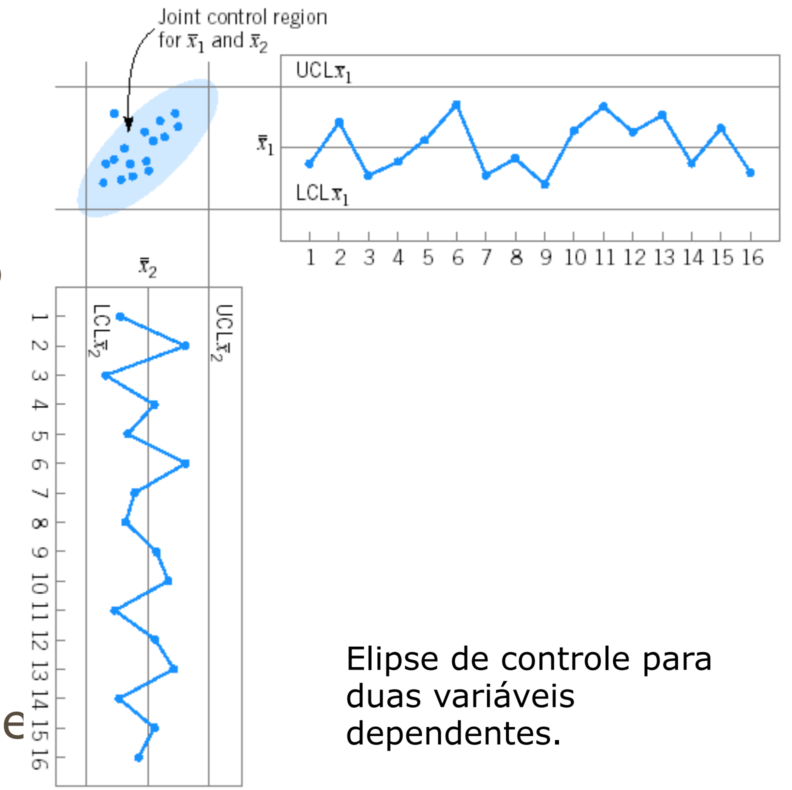
Observe o gráfico ao lado. Como x_1 e x_2 são independentes, $\sigma_{12} = 0$. Com isso a equação da distribuição de probabilidade define uma elipse centrada em (μ_1 e μ_2), com o eixo principal paralelo aos eixos $\bar{x}_1$ e $\bar{x}_2$.



Elipse de controle para duas variáveis independentes.

DESCRIÇÃO DE DADOS MULTIVARIADOS

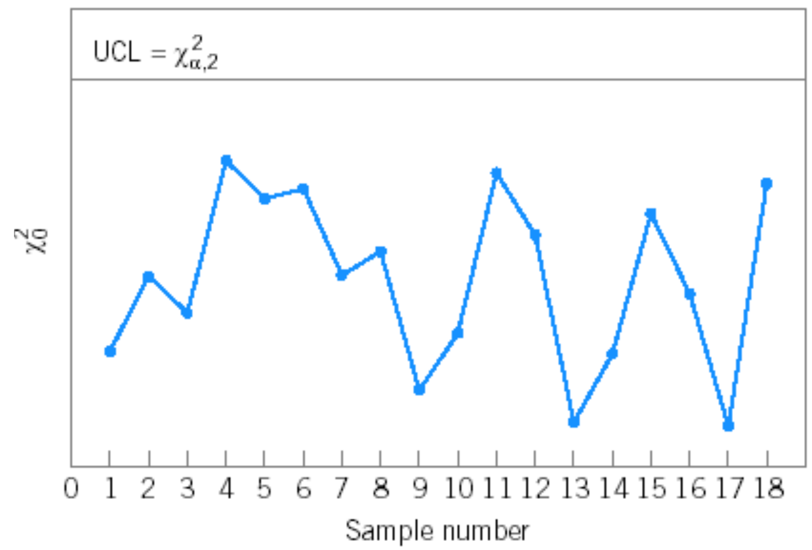
- No caso de duas características de qualidade dependentes, $\sigma_{12} \neq 0$ e a elipse de controle é aquela mostrada na figura abaixo.
  - Os eixos principais da elipse não são mais paralelos aos eixos das médias de x_1 e x_2.
- Desvantagens do monitoramento com elipses de controle:
  - A sequência temporal dos pontos plotados não é preservada.
  - É difícil construir elipses para mais de duas características de qualidade.



Elipse de controle para duas variáveis dependentes.

DESCRIÇÃO DE DADOS MULTIVARIADOS

- Para evitar estas desvantagens é usual plotar-se os valores da estatística χ_0^2 para cada amostra em uma carta de controle com um único limite de de controle superior, como na **carta de controle chi-quadrado** abaixo.
  - Agora o estado do processo é representado por um único número (a estatística χ_0^2).



DESCRIÇÃO DE DADOS MULTIVARIADOS

Carta de Controle Chi-Quadrado

- O conjunto de médias de p características de qualidade pode ser representado por:

$$\overline{X} = \begin{bmatrix} \overline{x_1} \\ \overline{x_2} \\ \vdots \\ \overline{x_p} \end{bmatrix}$$

- A estatística de teste plotada em cada amostra é:

$$\chi_0^2 = n\left[\left(\overline{X} - \mu\right)' \Sigma^{-1} \left(\overline{X} - \mu\right)\right]$$

em que $\mu' = [\mu_1, \mu_2, \ldots, \mu_p]$ é o vetor de médias em controle para cada x e Σ é a matriz de covariância.

- O limite de controle superior é: $LCS = \chi^2_{\alpha,p}$

DESCRIÇÃO DE DADOS MULTIVARIADOS

Estimando μ e Σ

- Os vetores de médias e variâncias são calculados de cada amostra como:

$$\bar{x}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} x_{ijk} \quad \text{em que} \quad \begin{matrix} j = 1,2,\ldots,p \\ k = 1,2,\ldots,m \end{matrix}$$

$$S_{jk}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ijk} - \overline{x_{jk}}\right)^2 \quad \text{em que} \quad \begin{matrix} j = 1,2,\ldots,p \\ k = 1,2,\ldots,m \end{matrix}$$

em que $\mathbf{x_{ijk}}$ é a **ith observação** da **jth característica de qualidade** na **kth amostra**.

- A covariância entre as características de qualidade j e h, na kth amostra é:

$$S_{jhk} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ijk} - \overline{x_{jk}}\right)\left(x_{ihk} - \overline{x_{hk}}\right) \quad \text{em que} \quad \begin{matrix} k = 1,2,\ldots,m \\ j \neq h \end{matrix}$$

DESCRIÇÃO DE DADOS MULTIVARIADOS

- As estatísticas $\bar{x}_{jk}$, s^2_{jk} e s_{jhk} são calculadas a partir de todas as **m** amostras para obter:

$$\bar{\bar{x}}_j = \frac{1}{m}\sum_{k=1}^{m}\overline{x_{jk}} \qquad j=1,2,\ldots,p$$

$$\bar{s}^2_j = \frac{1}{m}\sum_{k=1}^{m} s^2_{jk} \qquad j=1,2,\ldots,p$$

$$\bar{s}_{jh} = \frac{1}{m}\sum_{k=1}^{m} s_{jhk} \qquad j \neq h$$

em que $\bar{\bar{x}}_j$ são os elementos do vetor $\bar{\bar{X}}$ e a matriz de covariâncias **S** é formada como:

$$S = \begin{bmatrix} \bar{s}^2_1 & \bar{s}_{12} & \cdots & \bar{s}_{1p} \\ & \bar{s}^2_2 & \cdots & \bar{s}_{2p} \\ & & \ddots & \vdots \\ & & & \bar{s}^2_p \end{bmatrix}$$

CARTA DE CONTROLE T^2 DE HOTELLING

- Suponha que o **S**, da equação matricial anterior é usado para estimar Σ e que o vetor seja assumido como sendo os valores em controle do vetor de médias do processo.
- Se substituirmos o valor de μ por $\overline{\overline{X}}$ e Σ por S, a estatística de teste torna-se:

$$T^2 = n\left(\overline{X} - \overline{\overline{X}}\right)' S^{-1} \left(\overline{X} - \overline{\overline{X}}\right)$$

- Nesta forma, o procedimento de controle estatístico é usualmente chamado de **Carta de Controle T^2 de Hotelling**.

CARTA DE CONTROLE T^2 DE HOTELLING

Estatística *t Student* x Estatística T^2

(Mason and Young, 2002)

- Estatística *t Student* é computada a partir de amostras aleatórias, tamanho n, tomadas de uma **população com distribuição normal de média** μ **e variância** σ^2, como:

$$t = \frac{(\bar{x} - \mu)}{s / \sqrt{n}}$$

em que $\bar{x}$ é a média amostral e *s* é o seu desvio padrão amostral.

- O quadrado da estatística t Student é dado por:

$$t^2 = \frac{(\bar{x} - \mu)^2}{s^2 / n} = n(\bar{x} - \mu)(s^2)^{-1}(\bar{x} - \mu)$$

e seu valor é definido como sendo a **distância estatística quadrada entre a média amostral e média da população**.

CARTA DE CONTROLE T^2 DE HOTELLING

Estatística *t Student* x Estatística T^2

- O numerador da equação anterior, $(\bar{x} - \mu)^2$, corresponde à distância Euclidiana entre $\bar{x}$ e μ (que expressa a proximidade da média amostral à média da população).
  - A divisão desta distância Euclidiana quadrática pelo estimador da variância de $\bar{x}$ (s^2/n) produz a distância estatística quadrada.
- Hotelling estendeu a estatística t univariada para o caso multivariado, usando a forma da estatística T^2 baseada em estimativas amostrais da matriz de covariância, dada por:

$$T^2 = n\left(\overline{X} - \overline{\overline{X}}\right)' S^{-1}\left(\overline{X} - \overline{\overline{X}}\right)$$

e seu valor é definido como sendo a **distância estatística quadrada entre a média amostral e média da população**.

CARTA DE CONTROLE T^2 DE HOTELLING

Estatística *t Student* x Estatística T^2

- A premissa básica do uso da estatística T^2 de hotelling é que as observações multivariadas consideradas são resultado de uma amostragem aleatória de uma população normal p-variada, com vetor de média μ e matriz de covariância Σ.
  - Se os parâmetros da distribuição são desconhecidos, devem ser estimados a partir de um conjunto de dados históricos (HDS), que tenha sido coletado sob condições de estado estacionário, quando o processo operava sob controle.
  - A estatística T^2 será calculada a partir das observações amostrais p-variadas e, desde que as variáveis originais são aleatórias, os valores T^2 também o são e podem ser descritos por uma distribuição adequada.

CARTA T²
LIMITES DE CONTROLE

- Os limites de controle para a fase I de uma Carta de Controle T^2 são dados por:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$LCL = 0$$

Distribuição F com 1 e (n-1) graus de liberdade. Para o caso da distribuição de T^2 de um vetor X com observações independentes de $\overline{X}$ e S.

- Na fase II, quando a carta é usada para monitoramento da produção futura, os limites de controle são calculados como:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$LCL = 0$$

CARTA T^2
LIMITES DE CONTROLE

- **Exemplo1:** A força de resistência e o diâmetro de uma fibra têxtil são características importantes e demandam controle conjunto. Um engenheiro decidiu usar n=10 espécimes de fibras em cada amostras e obteve 20 amostras preliminares. A partir destes dados calculou:

$$\overline{X_1} = 115.59psi, \overline{X_2} = 1.06(x10^{-2})inch, \overline{S_1^2} = 1.23, \overline{S_2^2} = 0.83, \overline{S_{12}} = 0.79$$

e a estatística que usará para o controle e monitoramento do processo é:

$$T^2 = \frac{10}{(1.23)(0.83) - (0.79)^2}[0.83(\overline{x_1} - 115.59)^2 + 1.23(\overline{x_2} - 1.06)^2 - 2(0.79)(\overline{x_1} - 115.59)(\overline{x_2} - 1.06)]$$

CARTA T^2
LIMITES DE CONTROLE

- Tabela de dados e estatísticas do exemplo 1.

Sample Number k	(a) Means		(b) Variances and Covariances			(c) Control Chart Statistics	
	Tensile Strength ($\bar{x}_{1k}$)	Diameter ($\bar{x}_{2k}$)	s_{1k}^2	s_{2k}^2	s_{12k}	T_k^2	$\lvert \mathbf{S}_k \rvert$
1	115.25	1.04	1.25	0.87	0.80	2.16	0.45
2	115.91	1.06	1.26	0.85	0.81	2.14	0.41
3	115.05	1.09	1.30	0.90	0.82	6.77	0.50
4	116.21	1.05	1.02	0.85	0.81	8.29	0.21
5	115.90	1.07	1.16	0.73	0.80	1.89	0.21
6	115.55	1.06	1.01	0.80	0.76	0.03	0.23
7	114.98	1.05	1.25	0.78	0.75	7.54	0.41
8	115.25	1.10	1.40	0.83	0.80	3.01	0.52
9	116.15	1.09	1.19	0.87	0.83	5.92	0.35
10	115.92	1.05	1.17	0.86	0.95	2.41	0.10
11	115.75	0.99	1.45	0.79	0.78	1.13	0.54
12	114.90	1.06	1.24	0.82	0.81	9.96	0.36
13	116.01	1.05	1.26	0.55	0.72	3.86	0.17
14	115.83	1.07	1.17	0.76	0.75	1.11	0.33
15	115.29	1.11	1.23	0.89	0.82	2.56	0.42
16	115.63	1.04	1.24	0.91	0.83	0.08	0.44
17	115.47	1.03	1.20	0.95	0.70	0.19	0.65
18	115.58	1.05	1.18	0.83	0.79	0.00	0.36
19	115.72	1.06	1.31	0.89	0.76	0.35	0.59
20	115.40	1.04	1.29	0.85	0.68	0.62	0.63
Averages	$\bar{\bar{x}}_1 = 115.59$	$\bar{\bar{x}}_2 = 1.06$	$\bar{s}_1^2 = 1.23$	$\bar{s}_2^2 = 0.83$	$\bar{s}_{12} = 0.79$		

CARTA T^2
LIMITES DE CONTROLE

- Para o projeto da carta T^2 na fase I, o limite superior de controle foi calculado pela equação indicada para esta condição, considerando $\alpha = 0.001$:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$UCL = \frac{2(19)(9)}{20(10)-20-2+1} F_{0.001,2,20(10)-20-2+1}$$

$$UCL = \frac{342}{179} F_{0.001,2,179}$$

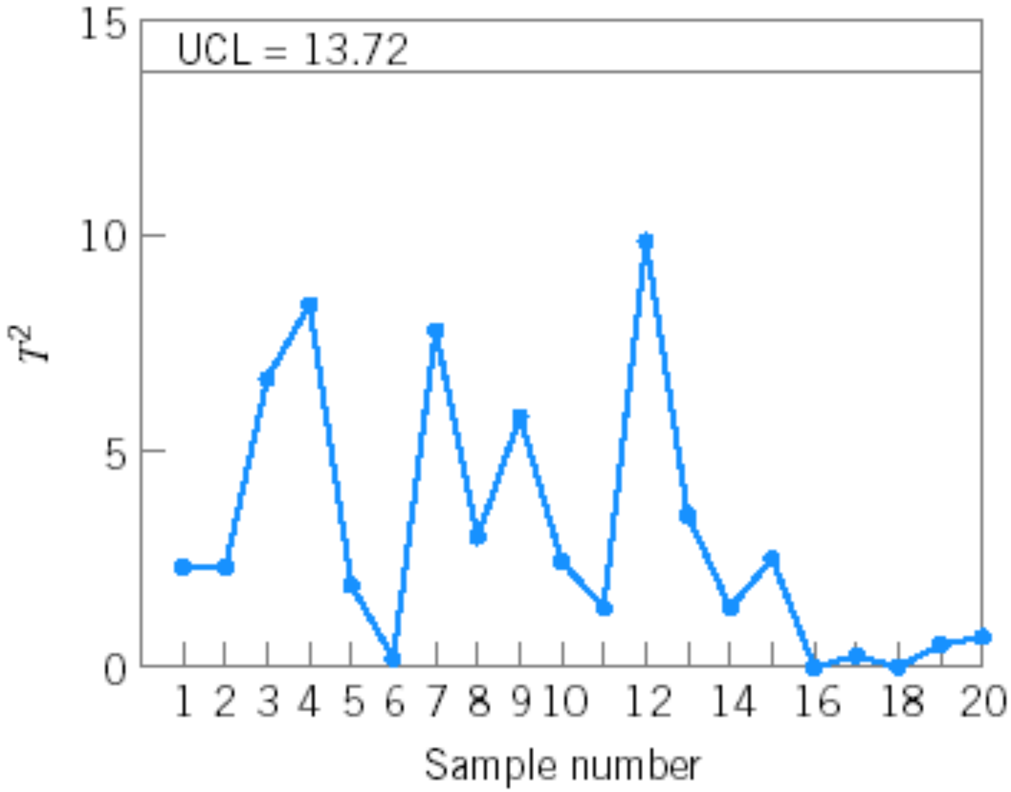
$$UCL = (1.91)7.18$$

$$UCL = 13.72$$

- Observa-se na carta de controle que não há pontos excedendo os limites de controle, donde conclui-se que o processo opera em controle. O limite de controle superior para a Fase II foi calculado como sendo 15.16.
- Se fosse usado o limite calculado por $\chi^2_{0.001,2}$ = 3,816, seria pequeno para a fase II.

CARTA DE CONTROLE T^2 DE HOTELLING

- Carta de Controle T^2 de Hotelling para qualidade da fibra têxtil (força de resistência e diâmetro):



CARTA DE CONTROLE T^2 DE HOTELLING

Observações Individuais

- Considere m amostras de tamanho n=1 para p características de qualidade observadas em cada amostra. Sejam x, o vetor de média amostral e S a matriz de covariância destas observações. A estatística T2 de Hotelling é expressa por:

$$T^2 = \left(X - \overline{X}\right)' S^{-1} \left(X - \overline{X}\right)$$

- Os limites de controle para esta estatística são:

$$LCS = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p}$$

$$LCI = 0$$

CARTA DE CONTROLE T^2 DE HOTELLING

- Tracy, Young and Mason (1992) indicam o cálculo dos limites de controle para a **Fase I** baseado na **distribuição Beta**
  - quando se assume que o vetor de observações X não é independente dos estimadores de $\overline{X}$ e S (mas sim, incluidos no seu cálculo):

$$LCS = \frac{(m-1)^2}{m}\beta_{\alpha,p/2,(m-p-1)/2}$$

$$LCI = 0$$

MÉTODOS DE ESTRUTURA LATENTE

Análise de Componentes Principais – PCA

- As componentes principais de um conjunto de variáveis x_1, x_2, …,x_p são um conjunto particular de combinações lineares destas variáveis:

$$z_1 = c_{11}x_1 + c_{12}x_2 + \ldots + c_{1p}x_p$$

$$z_2 = c_{21}x_1 + c_{22}x_2 + \ldots + c_{2p}x_p$$

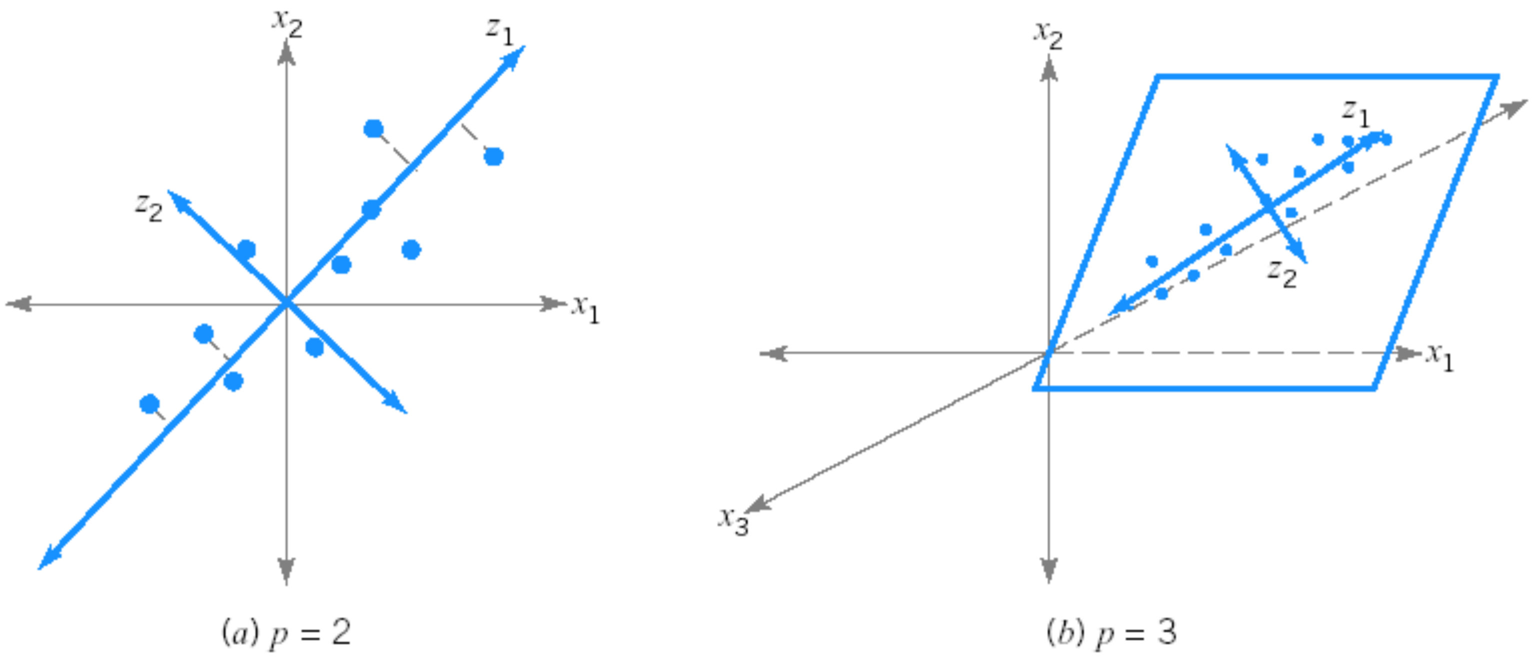
$$\vdots$$

$$z_p = c_{p1}x_1 + c_{p2}x_2 + \ldots + c_{pp}x_p$$

em que os cij's são constantes a serem determinadas.

- Geometricamente, as componentes principais z_1, z_2, …, z_p são os eixos do novo sistema de coordenadas obtidas por meio da rotação dos eixos do sistema original (de x's).
- Os novos eixos representam as direções de máxima variabilidade.

ANÁLISE DE COMPONENTES PRINCIPAIS PCA



Componentes principais para p=2 e p=3 variáveis de processo.

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

- Seja C uma matriz, cujas colunas são os autovetores:

$$C^{'} \sum C = \Lambda$$

E Λ é uma matriz diagonal pxp, cujos elementos da diagonal principal são os autovalores: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

A **variância da $\mathbf{i_{th}}$** componente principal é o **$\mathbf{i_{th}}$ autovalor $\boldsymbol{\lambda_i}$**. Assim,a proporção de variabilidade do dado original, explicada pela **$\mathbf{i_{th}}$** componente principal é dada por:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

Assim, pode-se saber quanto da variabilidade é explicada por cada componente e decidir quantas e quais serão retidas (poucas, digamos r) dentre as p componentes principais (calculando a soma dos autovalores da r retidas e comparando com a soma total dos autovalores).

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

- Uma vez calculadas as **p** componentes principais e selecionado o subconjunto de **r** componentes, pode-se obter novas observações das componentes principais retidas Z_{ij}, simplesmente substituindo as observações originais x_{ij} no conjunto das componentes principais retidas. Por ex.:

$$z_{i1} = c_{11}x_{i1} + c_{12}x_{i2} + \ldots + c_{1p}x_{ip}$$

$$z_{i2} = c_{21}x_{i1} + c_{22}x_{i2} + \ldots + c_{2p}x_{ip}$$

$$\vdots$$

$$z_{ir} = c_{r1}x_{i1} + c_{r2}x_{i2} + \ldots + c_{rr}x_{ip}$$

- Esses **z_{ij}'s** são usualmente chamados **SCORES** das componentes principais.

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

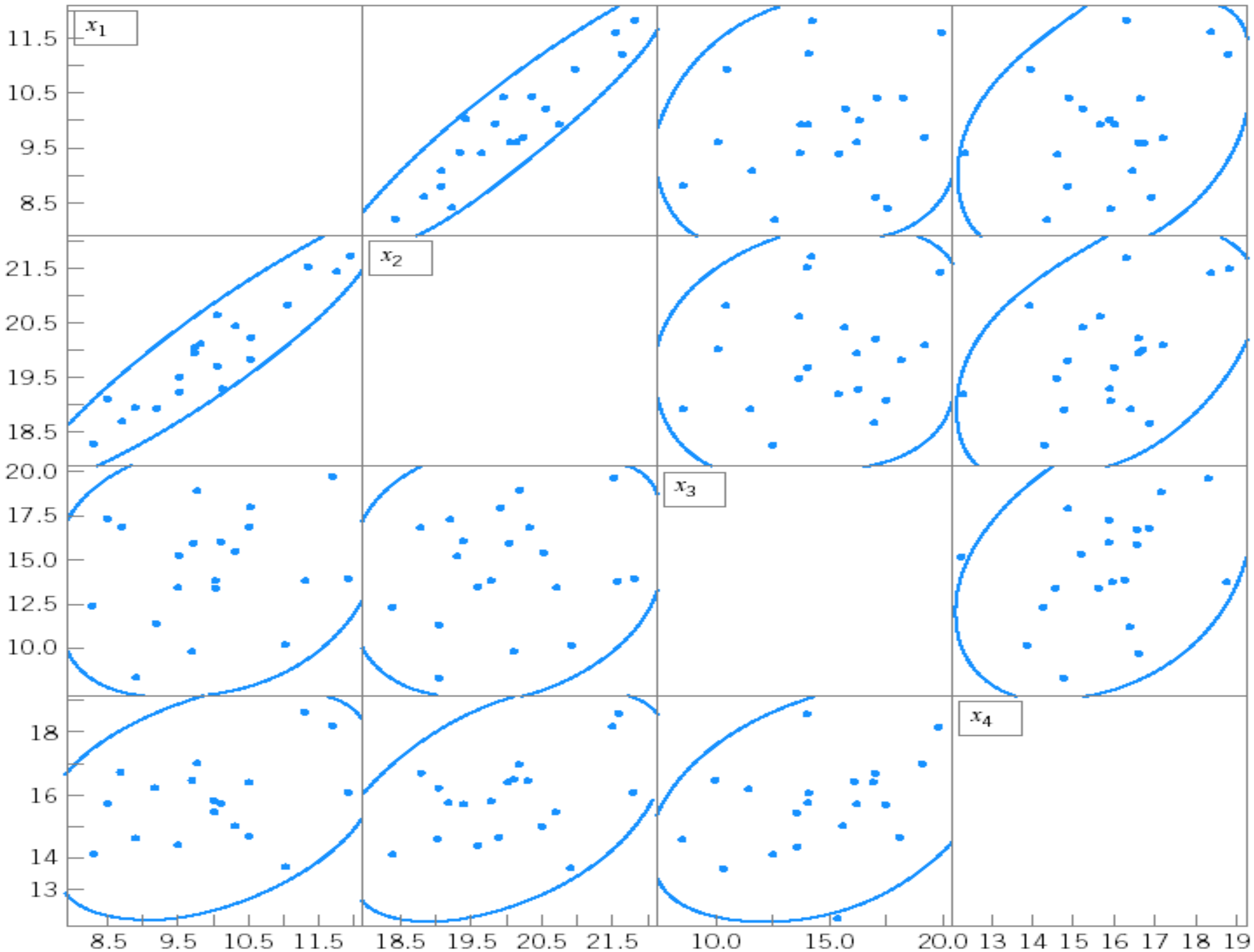
Este procedimento é ilustrado a seguir por meio da análise de componentes principais (PCA) usando os dados das p=4 variáveis x_1, x_2, x_3 e x_4 mostrados na tabela seguinte e provenientes de um processo químico.

Original Data

Observation	x_1	x_2	x_3	x_4	z_1	z_2
1	10	20.7	13.6	15.5	0.291681	−0.6034
2	10.5	19.9	18.1	14.8	0.294281	0.491533
3	9.7	20	16.1	16.5	0.197337	0.640937
4	9.8	20.2	19.1	17.1	0.839022	1.469579
5	11.7	21.5	19.8	18.3	3.204876	0.879172
6	11	20.9	10.3	13.8	0.203271	−2.29514
7	8.7	18.8	16.9	16.8	−0.99211	1.670464
8	9.5	19.3	15.3	12.2	−1.70241	−0.36089
9	10.1	19.4	16.2	15.8	−0.14246	0.560808
10	9.5	19.6	13.6	14.5	−0.99498	−0.31493
11	10.5	20.3	17	16.5	0.944697	0.504711
12	9.2	19	11.5	16.3	−1.2195	−0.09129
13	11.3	21.6	14	18.7	2.608666	−0.42176
14	10	19.8	14	15.9	−0.12378	−0.08767
15	8.5	19.2	17.4	15.8	−1.10423	1.472593
16	9.7	20.1	10	16.6	−0.27825	−0.94763
17	8.3	18.4	12.5	14.2	−2.65608	0.135288
18	11.9	21.8	14.1	16.2	2.36528	−1.30494
19	10.3	20.5	15.6	15.1	0.411311	−0.21893
20	8.9	19	8.5	14.7	−2.14662	−1.17849

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

As 20 observações da tabela do slide anterior, são plotadas para cada variável contra a outra e mostradas na figura seguinte na forma de **scatter plots (diagramas de espalhamento)**.



ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

A matriz de convariância amostral das primeiras 20 observações de x’s na forma de correlação é dada por:

$$\Sigma = \begin{bmatrix} 1.0000 & 0.9302 & 0.2060 & 0.3595 \\ 0.9302 & 1.0000 & 0.1669 & 0.4502 \\ 0.2060 & 0.1669 & 1.0000 & 0.3439 \\ 0.3595 & 0.4502 & 0.3439 & 1.0000 \end{bmatrix}$$

Note que o coeficiente de correlação entre x_1 e x_2 é 0.9302, o que confirma a impressão visual obtida da matriz de scatter plots da figura anterior.

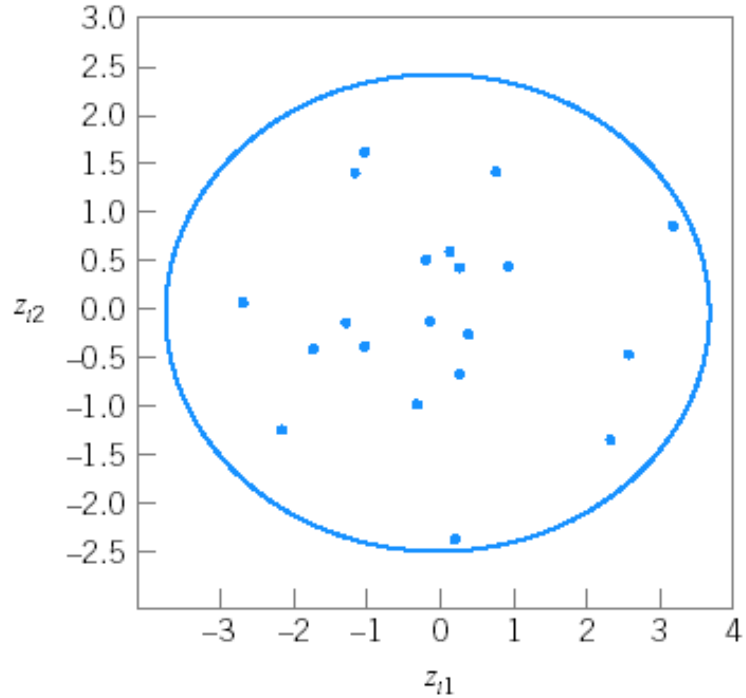
ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

A tabela seguinte apresenta os resultados do PCA (calculado usando minitab) para as 20 observações de x_1, x_2, x_3 e x_4 da tabela do slide 31, mostrando os autovalores e autovetores, tanto como a porcentagem cumulativa da variablidade explicada por cada componente principal:

Eigenvalues:	2.3181	1.0118	0.6088	0.0613
Percent:	57.9516	25.2951	15.2206	1.5328
Cumulative Percent:	57.9516	83.2466	98.4672	100.0000
Eigenvectors				
x_1	0.59410	–0.33393	0.25699	0.68519
x_2	0.60704	–0.32960	0.08341	–0.71826
x_3	0.28553	0.79369	0.53368	–0.06092
x_4	0.44386	0.38717	–0.80137	0.10440

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

As duas últimas colunas da tabela do slide 31 mostram os valores calculados para os scores das componentes principais z_1 e z_2 para as primeiras 20 observações. A figura abaixo é um scatter plot dos 20 scores das componentes principais dentro de um contorno com intervalo de confiança de aproximadamente 95%.



ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

A tabela seguinte mostra outras 10 novas observações para as variáveis de processo x_1, x_2, x_3 e x_4 que não foram utilizados para o cálculo das componentes principais. Os scores das componentes principais calculados para estas novos observações também são mostrados nas duas últimas colunas. A figura seguinte, no slide 37, mostra a representação destes novos scores com símbolo distinto dos anteriores (x).

New Data

Observation	x_1	x_2	x_3	x_4	z_1	z_2
21	9.9	20	15.4	15.9	0.074196	0.239359
22	8.7	19	9.9	16.8	–1.51756	–0.21121
23	11.5	21.8	19.3	12.1	1.408476	–0.87591
24	15.9	24.6	14.7	15.3	6.298001	–3.67398
25	12.6	23.9	17.1	14.2	3.802025	–1.99584
26	14.9	25	16.3	16.6	6.490673	–2.73143
27	9.9	23.7	11.9	18.1	2.738829	–1.37617
28	12.8	26.3	13.5	13.7	4.958747	–3.94851
29	13.1	26.1	10.9	16.8	5.678092	–3.85838
30	9.8	25.8	14.8	15	3.369657	–2.10878

ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

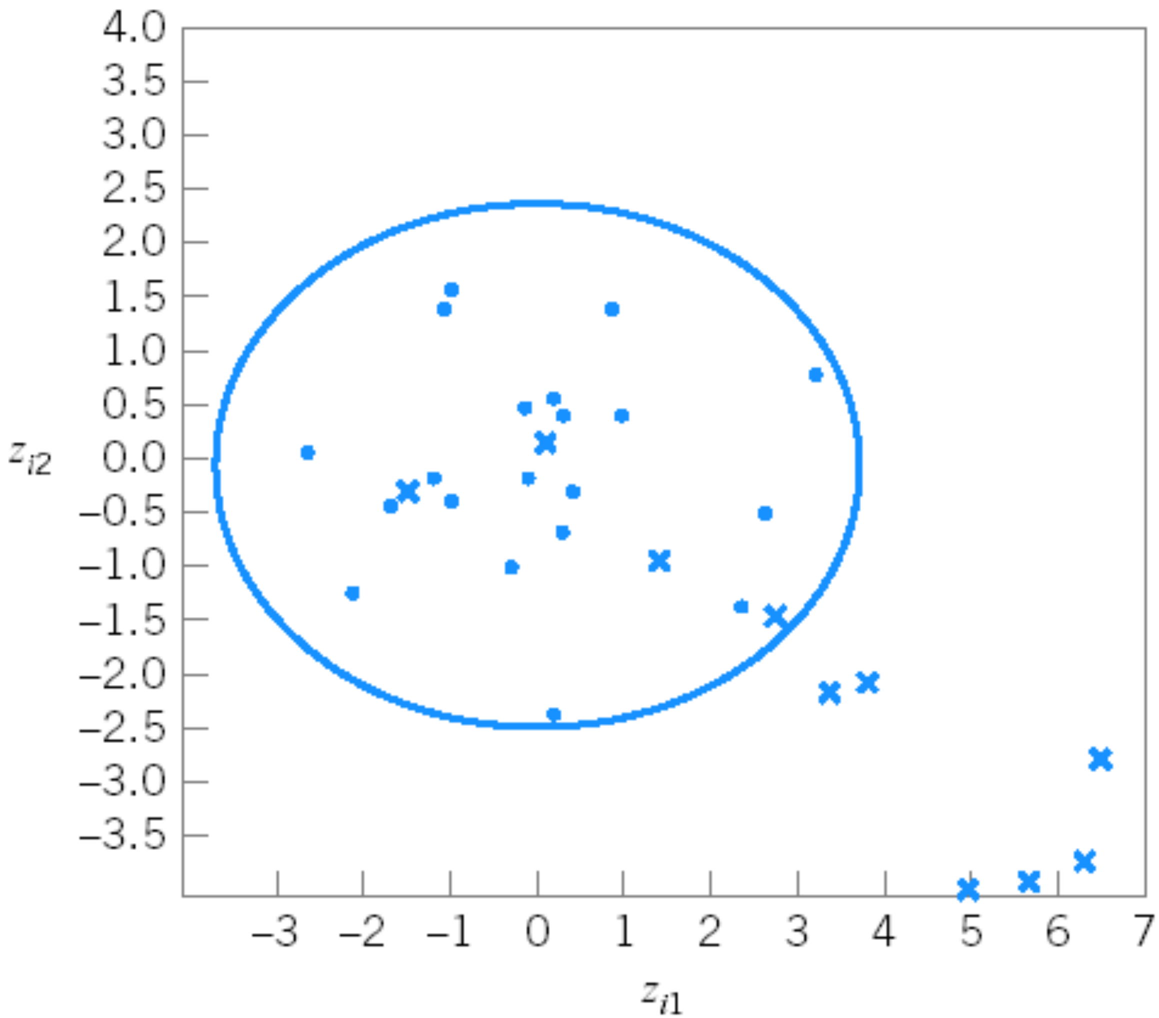


Gráfico das componentes principais incluindo os novos 10 scores.

BIBLIOGRAFIA

1. **Douglas C. Montgomery**: *Introduction to Statistical Quality Control*, 4th Edition.
2. Robert L. Mason, John C. Young: Multivariate Statistical Process Control with Industrial Applications, ASA-SIAM , 2002.