

---

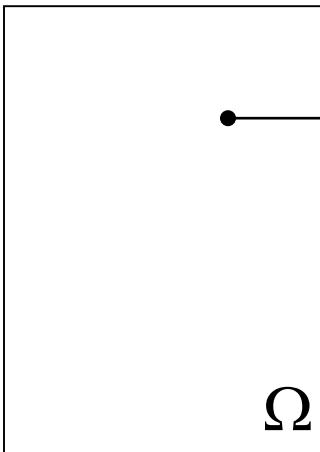
# Estatística: Aplicação ao Sensoriamento Remoto

SER 204 - ANO 2024

Componentes Principais

Camilo Daleles Rennó  
camilo.renno@inpe.br  
<http://www.dpi.inpe.br/~camilo/estatistica/>

# Informação X Redundância



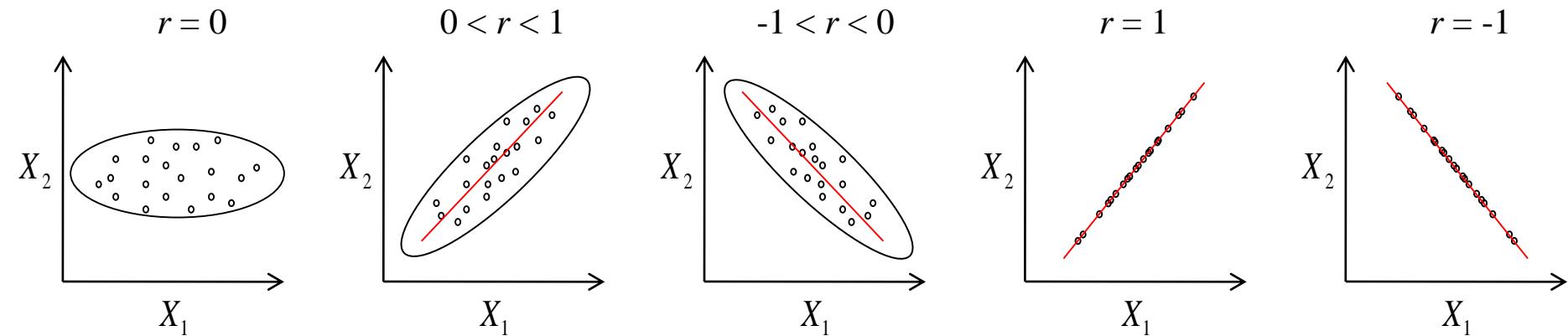
Quando pretendemos estudar uma população qualquer, em geral, coletamos um grande número de dados (atributos) com o intuito de descrever seu comportamento e/ou entender o relacionamento existente entre os diversos aspectos característicos desta população.

Cada atributo (ou variável) traz em si uma informação a respeito de um determinado aspecto da população estudada.

Quanto maior a **variância**, maior é a variabilidade e portanto maior a **informação** contida nesta variável. Num caso extremo, se a variância é zero, a variável não apresenta nenhuma informação a respeito do fenômeno por ela representada.

# Informação X Redundância

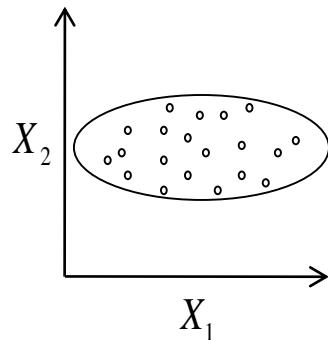
Quando as medidas adquiridas são feitas sobre o mesmo indivíduo (ou mesma posição geográfica), espera-se que possa haver alguma relação entre elas. Essas relações podem ser fortes, fracas ou nem existir.



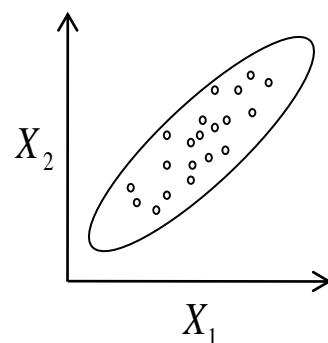
$$r = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}$$

Dessa forma, a **covariância** (e a **correlação**) pode ser interpretada como uma **redundância** em análises múltiplas (2 ou mais variáveis), já que a informação contida numa variável está parcialmente representada em outra. Num caso extremo, para que utilizar duas variáveis perfeitamente correlacionadas ( $|r| = 1$ ), uma vez que, conhecendo-se o valor de uma, pode-se inferir com precisão o valor da outra?

# Informação X Redundância



Quando a relação entre variáveis não é o objetivo da pesquisa em si, o ideal é trabalharmos com variáveis independentes pois assim cada variável trará uma informação nova a respeito da população estudada.

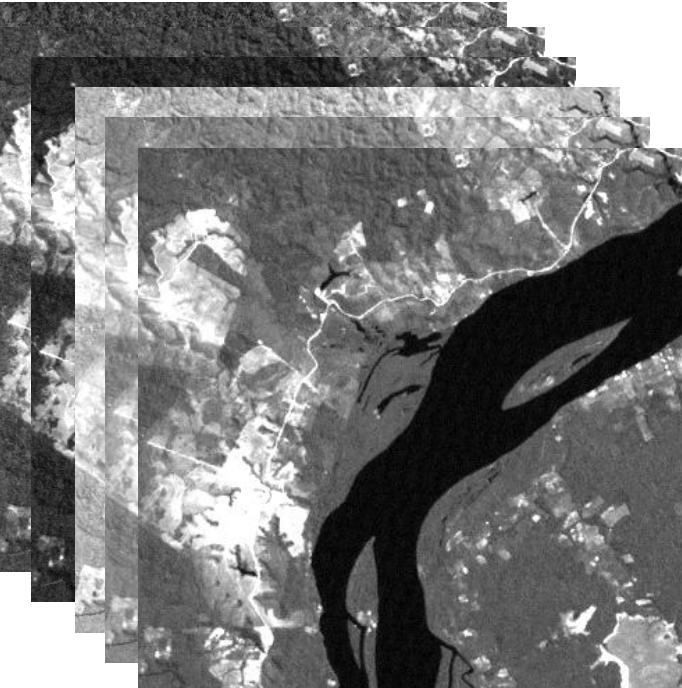


Se as variáveis guardam alguma relação entre si, podemos pensar que parte do esforço de coleta de uma segunda medida está sendo desperdiçada pois a “nova informação” já está parcialmente representada na primeira medida.



Por exemplo, para estudar a biomassa de uma vegetação, obtém-se a altura e diâmetro do tronco de várias árvores. Mas já não é esperado que árvores mais grossas sejam também as mais altas?

# Informação X Redundância



TM/LANDSAT (Bandas 1 a 5 e 7)  
228/63 08/07/1999

Numa classificação de uso e cobertura do solo, a utilização de muitas bandas espectrais pode prejudicar o desempenho do classificador devido ao volume de dados analisados

Imagens hiperespectrais podem conter centenas de bandas!

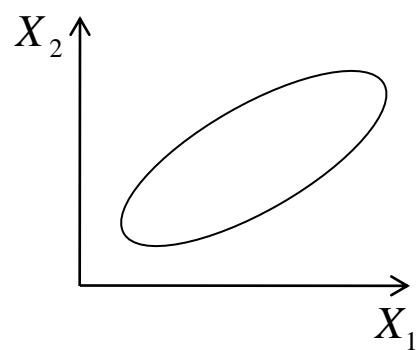
Será mesmo necessário processar todos esses dados para extrair a informação necessária?

É possível reduzir a redundância de informação e “condensar” esta informação num menor volume de dados?

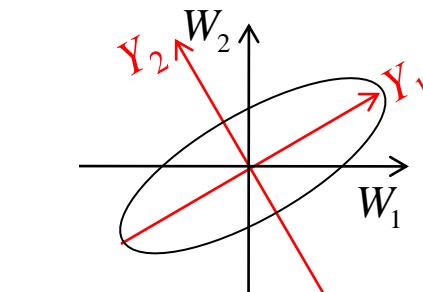
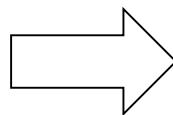
« Transformação por Componentes Principais »

# Transformação por Componentes Principais (exemplo 2D)

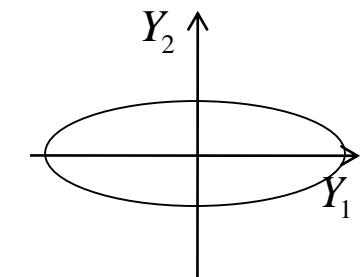
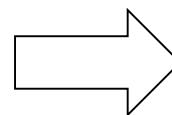
Como poderíamos eliminar a redundância entre  $X_1$  e  $X_2$ ?



$$Cov(X_1, X_2) > 0$$

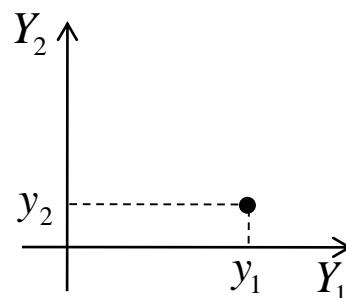
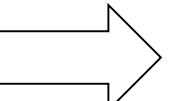
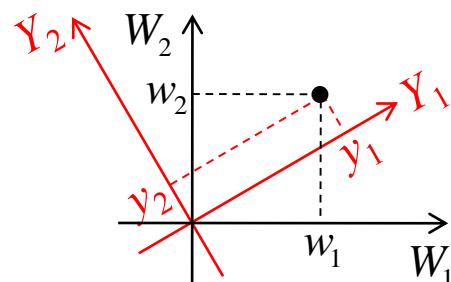


$$Cov(W_1, W_2) = Cov(X_1, X_2)$$



$$Cov(Y_1, Y_2) = 0$$

**rotação com preservação da ortogonalidade dos eixos**



$$Y_1 = \alpha_{11}W_1 + \alpha_{12}W_2 \quad Var(Y_1) = \lambda_1$$

$$Y_2 = \alpha_{21}W_1 + \alpha_{22}W_2 \quad Var(Y_2) = \lambda_2$$

1ª CP

$\alpha$  são chamados de autovetores

2ª CP

$\lambda$  são chamados de autovalores ( $\lambda_1 > \lambda_2$ )

$$Var(X_1) + Var(X_2) = \lambda_1 + \lambda_2 \quad \text{informação total é preservada}$$

# Autovalores e Autovetores

Os autovetores representam as transformações lineares aplicadas em  $m$  variáveis correlacionadas de modo a obter  $m$  variáveis não correlacionadas e podem ser representados na forma de uma matriz  $\alpha$  ( $m \times m$ ):

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{m1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1m} & \alpha_{2m} & \cdots & \alpha_{mm} \end{bmatrix}$$

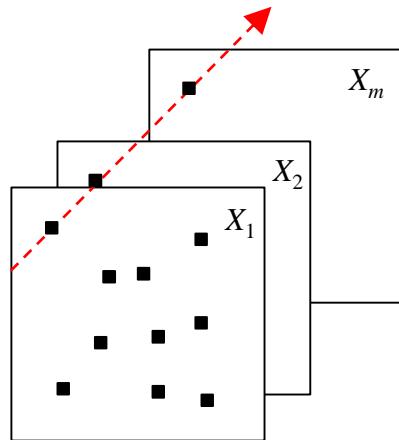
Nesta matriz, os autovetores estão organizados em colunas, ou seja, o 1º autovetor está na 1ª coluna, o 2º autovetor está na 2ª coluna e assim por diante.

Os autovalores representam a variância de cada componente (variável transformada) e podem ser representadas na forma de uma matriz diagonal  $\lambda$  ( $m \times m$ ):

$$\lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \quad (\lambda_1 > \lambda_2 > \dots > \lambda_m)$$

Por ser uma matriz diagonal, usualmente os autovalores são apresentados como um vetor  
Mas como são calculados os autovalores e autovetores?

# Relacionamento entre Variáveis



Amostra	$X_1$	$X_2$
1	3,5	18,9
2	7,5	31,5
3	4,4	22,2
4	1,1	8,7
5	4,4	19,2
6	4,7	21,3
7	7,2	27,0
8	3,6	16,8
9	9,2	33,6
10	3,1	15,9

...

$X_m$
143,2
138,6
142,7
145,5
143,7
141,2
141,5
145,1
139,0
145,3

← mesma posição geográfica

Supondo que  $n$  amostras sejam avaliadas segundo  $m$  variáveis diferentes (atributos), podemos representar o conjunto de valores observados por um matriz  $\mathbf{X}$ , onde cada elemento  $x_{ij}$  representa o valor da  $i$ -ésima amostra para a  $j$ -ésima variável.

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 & & X_m \\ x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{nm} & \cdots & x_{nm} \end{bmatrix}$$

# Matriz de Variância-Covariância

A variância de cada variável e as covariâncias entre todos os pares de variáveis podem ser representadas através da matriz de variância-covariância  $\Sigma_X$  ( $m \times m$ ):

$$\Sigma_X = \begin{bmatrix} X_1 & X_2 & \dots & X_m \\ Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_m) \\ Cov(X_1, X_2) & Var(X_2) & \dots & Cov(X_2, X_m) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_1, X_m) & Cov(X_2, X_m) & \dots & Var(X_m) \end{bmatrix} \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{matrix}$$

$$Cov(X_1, X_2) = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{n-1}$$

$$Cov(X_k, X_l) = Cov(X_l, X_k)$$

$$Cov(X_k, X_k) = Var(X_k)$$

# Cálculo dos Autovalores e Autovetores

Os autovalores ( $\lambda$ ) e autovetores ( $\alpha$ ) são obtidos de modo a satisfazer a seguintes condições:

$$\Sigma_x \alpha = \lambda \alpha \quad \text{ou} \quad (\Sigma_x - \lambda I) \alpha = 0 \quad \text{sendo} \quad \alpha' \alpha = I \quad (\text{autovetores ortogonais})$$

Como os autovetores  $\alpha$  são não nulos, então

$|\Sigma_x - \lambda I| = 0$  gera um polinômio de grau  $m$  cujas raízes representam os  $m$  autovalores  $\lambda$

Ordenam-se os autovalores do maior para o menor e para cada autovalor  $\lambda_k$ , calcula-se o autovetor  $\alpha_k$  de modo que:

$$(\Sigma_x - \lambda_k I) \alpha_k = 0$$

Assim, os valores transformados para a componente principal  $k$  são calculados a partir da matriz  $X$ :

$$Y_k = \alpha'_k X'$$

# Exemplo

Suponha que  $\Sigma_x = \begin{bmatrix} 3,4 & -1,1 & 0,3 \\ -1,1 & 5,2 & -0,3 \\ 0,3 & -0,3 & 0,5 \end{bmatrix}$

Então  $|\Sigma_x - \lambda I| = 0$

$$\begin{aligned}\Sigma_x - \lambda I &= \begin{bmatrix} 3,4 & -1,1 & 0,3 \\ -1,1 & 5,2 & -0,3 \\ 0,3 & -0,3 & 0,5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} 3,4 - \lambda & -1,1 & 0,3 \\ -1,1 & 5,2 - \lambda & -0,3 \\ 0,3 & -0,3 & 0,5 - \lambda \end{bmatrix}\end{aligned}$$

$$\begin{aligned}|\Sigma_x - \lambda I| &= (3,4 - \lambda)(5,2 - \lambda)(0,5 - \lambda) + 2(-1,1)(-0,3)(0,3) - (3,4 - \lambda)(-0,3)^2 - (5,2 - \lambda)(0,3)^2 - (0,5 - \lambda)(-1,1)^2 \\ &= (3,4 - \lambda)(5,2 - \lambda)(0,5 - \lambda) + 0,198 - 0,306 + 0,09\lambda - 0,468 + 0,09\lambda - 0,605 + 1,21\lambda \\ &= (3,4 - \lambda)(5,2 - \lambda)(0,5 - \lambda) + 1,39\lambda - 1,181 \\ &= (17,68 - 8,6\lambda + \lambda^2)(0,5 - \lambda) + 1,39\lambda - 1,181 \\ &= -\lambda^3 + 9,1\lambda^2 - 20,59\lambda + 7,659 = 0\end{aligned}$$

Encontrando as raízes do polinômio, tem-se

$$\lambda_1 = 5,7517 \quad \lambda_2 = 2,8871 \quad \lambda_3 = 0,4612$$

# Exemplo

Agora, para cada autovalor, calcula-se o autovetor correspondente de modo que  $\Sigma_x \alpha = \lambda \alpha$   
Demonstrando para o primeiro autovalor  $\lambda_1 = 5,7517$

$$\begin{bmatrix} 3,4 & -1,1 & 0,3 \\ -1,1 & 5,2 & -0,3 \\ 0,3 & -0,3 & 0,5 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \end{bmatrix} = 5,7517 \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \end{bmatrix} \quad \begin{aligned} -2,3517\alpha_{11} - 1,1\alpha_{12} + 0,3\alpha_{13} &= 0 \\ -1,1\alpha_{11} - 0,5517\alpha_{12} - 0,3\alpha_{13} &= 0 \\ 0,3\alpha_{11} - 0,3\alpha_{12} - 5,2517\alpha_{13} &= 0 \end{aligned}$$
$$\alpha_{11} = 5,6657\alpha_{13}$$
$$\alpha_{12} = -11,8400\alpha_{13}$$

Repetindo para os demais autovalores, chega-se a:

$$\alpha = \begin{bmatrix} 5,6657\alpha_{13} & 15,4202\alpha_{23} & -0,0858\alpha_{33} \\ -11,8400\alpha_{13} & 7,4633\alpha_{23} & 0,0434\alpha_{33} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{bmatrix}$$

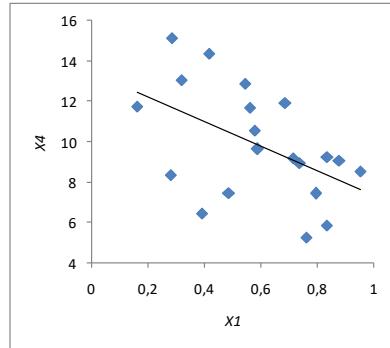
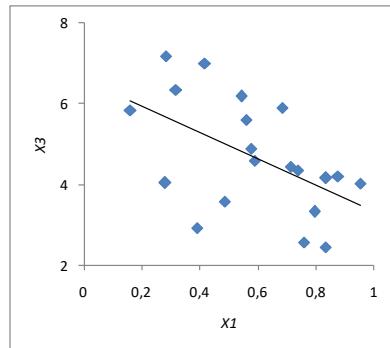
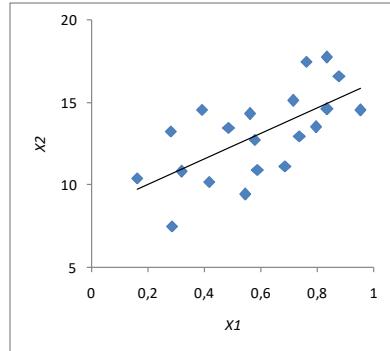
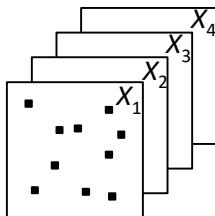
Como os autovetores são ortogonais então  $\alpha' \alpha = I$

Assim  $\alpha_{13} = 0,0760$   $\alpha_{23} = 0,0583$   $\alpha_{33} = 0,9954$

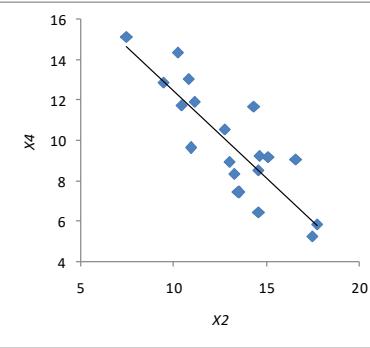
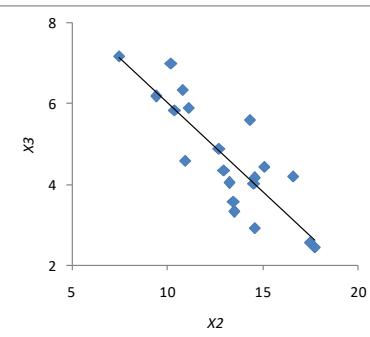
$$\alpha = \begin{bmatrix} 0,43039784 & 0,89858539 & -0,08545169 \\ -0,89943698 & 0,43491207 & 0,04318105 \\ 0,07596583 & 0,05827338 & 0,99540615 \end{bmatrix}$$

# Análise de Componentes Principais

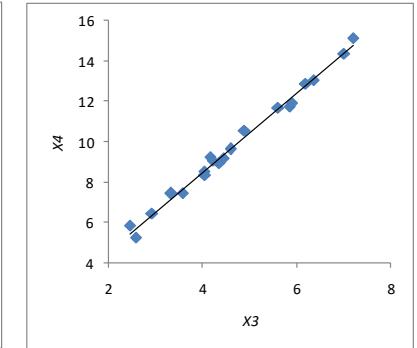
$X_1$	$X_2$	$X_3$	$X_4$
0,1596	10,4043	5,8457	11,7544
0,4160	10,1971	6,9976	14,3787
0,6841	11,1176	5,8942	11,9172
0,5421	9,4291	6,1867	12,8827
0,7133	15,0980	4,4443	9,1690
0,7587	17,4789	2,5728	5,2574
0,2824	7,4640	7,1848	15,1390
0,8336	14,5949	4,1581	9,2457
0,5752	12,7215	4,8841	10,5356
0,8308	17,7273	2,4392	5,8491
0,9510	14,5368	4,0345	8,5466
0,2796	13,2739	4,0397	8,3507
0,7357	12,9786	4,3470	8,9358
0,8759	16,5915	4,2100	9,0698
0,3904	14,5826	2,9220	6,4501
0,5584	14,3165	5,5950	11,6657
0,4849	13,4711	3,5743	7,4431
0,3169	10,8015	6,3426	13,0677
0,5869	10,9385	4,5867	9,6230
0,7940	13,5060	3,3238	7,4271



Matriz de Variância-Covariância				
0,0516	0,4024	-0,1683	-0,3171	
0,4024	7,3698	-3,2402	-6,3600	
-0,1683	-3,2402	1,9737	3,8806	
-0,3171	-6,3600	3,8806	7,6995	

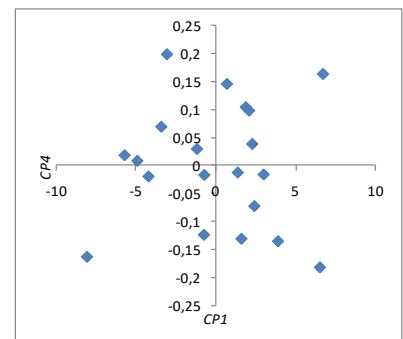
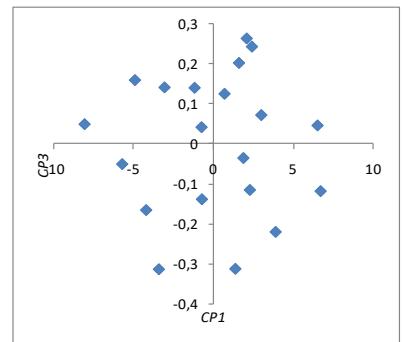
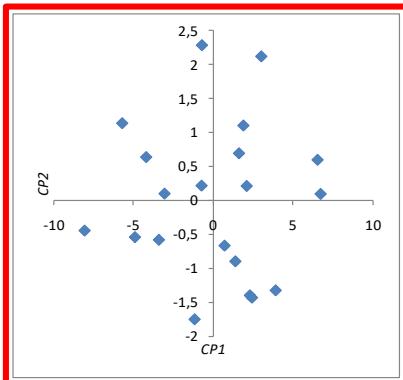


Autovalores			
15,7615	1,2900	0,0310	0,0121
Autovetores			
0,0340	0,0583	0,9481	0,3108
0,6485	0,7582	-0,0670	-0,0087
-0,3437	0,2787	-0,2837	0,8507
-0,6784	0,5865	0,1272	-0,4238

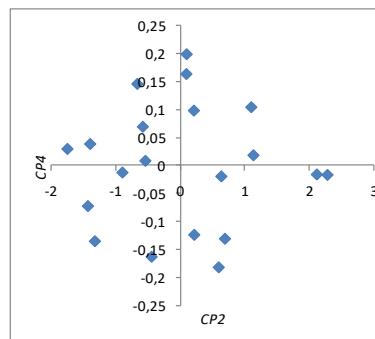
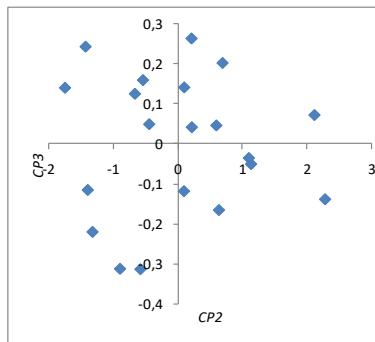


# Análise de Componentes Principais

$CP_1$	$CP_2$	$CP_3$	$CP_4$
-3,4404	-0,5891	-0,3154	0,0689
-5,7422	1,1290	-0,0515	0,0181
-3,0872	0,0914	0,1410	0,1980
-4,9424	-0,5494	0,1594	0,0082
1,8577	1,0951	-0,0362	0,1037
6,7000	0,0872	-0,1194	0,1628
-8,0993	-0,4530	0,0486	-0,1626
1,5819	0,6859	0,2024	-0,1305
-0,7664	0,2092	0,0411	-0,1235
6,5079	0,5895	0,0455	-0,1814
2,0649	0,2042	0,2638	0,0977
1,3543	-0,9060	-0,3144	-0,0126
0,6757	-0,6745	0,1249	0,1452
2,9795	2,1135	0,0716	-0,0161
3,8802	-1,3335	-0,2217	-0,1349
-0,7436	2,2786	-0,1397	-0,0168
2,2647	-1,4066	-0,1165	0,0383
-4,2393	0,6300	-0,1669	-0,0195
-1,2009	-1,7602	0,1400	0,0294
2,3949	-1,4413	0,2433	-0,0723

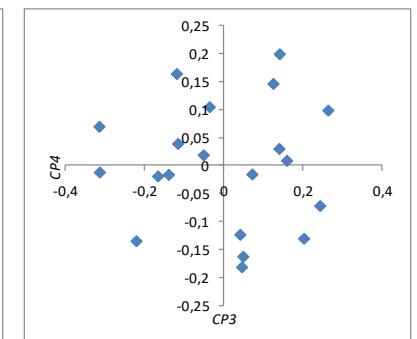


Matriz de Variância-Covariância				
15,7615	0	0	0	0
0	1,2900	0	0	0
0	0	0,0310	0	0
0	0	0	0,0121	



A primeira componente guarda 92,2% da variação total

As 2 primeiras CP, acumulam 99,7% da variação total



# Análise de Componentes Principais no R

```
x1<-c(0.1596,0.416,0.6841,0.5421,0.7133,0.7587,0.2824,0.8336,0.5752,0.8308,0.951,0.2796,0.7357,0.8759,0.3904,  
0.5584,0.4849,0.3169,0.5869,0.794)  
x2<-c(10.4043,10.1971,11.1176,9.4291,15.098,17.4789,7.464,14.5949,12.7215,17.7273,14.5368,13.2739,12.9786,  
16.5915,14.5826,14.3165,13.4711,10.8015,10.9385,13.506)  
x3<-c(5.8457,6.9976,5.8942,6.1867,4.4443,2.5728,7.1848,4.1581,4.8841,2.4392,4.0345,4.0397,4.347,4.21,2.922,  
5.595,3.5743,6.3426,4.5867,3.3238)  
x4<-c(11.7544,14.3787,11.9172,12.8827,9.169,5.2574,15.139,9.2457,10.5356,5.8491,8.5466,8.3507,8.9358,9.0698,  
6.4501,11.6657,7.4431,13.0677,9.623,7.4271)  
x<-cbind(x1,x2,x3,x4)
```

```
round(cov(x),4)
```

- x1 x2 x3 x4
- x1 0.0516 0.4024 -0.1683 -0.3171
- x2 0.4024 7.3698 -3.2402 -6.3599
- x3 -0.1683 -3.2402 1.9737 3.8805
- x4 -0.3171 -6.3599 3.8805 7.6995

```
sum(diag(cov(x)))
```

- [1] 17.09458

```
pccov<-prcomp(x)  
pccov
```

- Standard deviations:  
[1] 3.9700671 1.1358016 0.1759476 0.1101858  $\sqrt{\lambda}$
- Rotation:  

	PC1	PC2	PC3	PC4
x1	0.03398365	-0.05831431	0.94807140	-0.310813742
x2	0.64847514	-0.75822916	-0.06702796	0.008705756
x3	-0.34373410	-0.27866559	-0.28371255	-0.850705320
x4	-0.67835977	-0.58654366	0.12718124	0.423815393

```
sum(pccov$sdev^2)
```

- [1] 17.09458

# Análise de Componentes Principais no R

```
summary(pccov)
```

- Importance of components:

	PC1	PC2	PC3	PC4
• Standard deviation	3.970	1.13580	0.17595	0.11019
• Proportion of Variance	0.922	0.07547	0.00181	0.00071
• Cumulative Proportion	0.922	0.99748	0.99929	1.00000

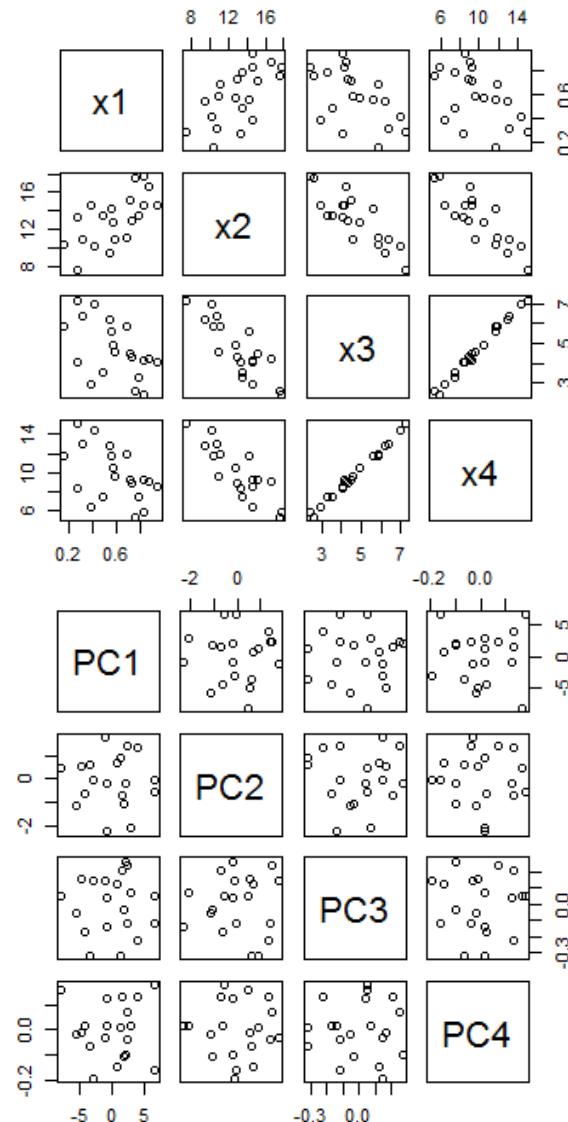
```
pc<-pccov$x #Calculando as componentes principais
```

```
round(cov(pc),4)
```

	PC1	PC2	PC3	PC4
• PC1	15.7614	0.00	0.000	0.0000
• PC2	0.0000	1.29	0.000	0.0000
• PC3	0.0000	0.00	0.031	0.0000
• PC4	0.0000	0.00	0.000	0.0121

```
pairs(x)
```

```
pairs(pc)
```



# Observações

- A transformação prioriza as variáveis com maior variância

**Atenção:** isso pode ocorrer quando as variáveis possuem diferentes grandezas

Para neutralizar o efeito das variâncias de cada variável de modo que todas tenham o mesmo peso, os autovalores e autovetores devem ser calculados a partir da **matriz de correlação** (que equivale realizar a transformação sobre as variáveis padronizadas)

$$r_{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_m \\ 1 & r_{X_1, X_2} & \cdots & r_{X_1, X_m} \\ r_{X_1, X_2} & 1 & \cdots & r_{X_2, X_m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{X_1, X_m} & r_{X_2, X_m} & \cdots & 1 \end{bmatrix} \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{matrix}$$
$$r_{X_1, X_2} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}$$
$$r_{X_k, X_l} = r_{X_l, X_k}$$
$$r_{X_k, X_k} = 1$$

Obs: em R, para computar a transformação por componentes principais utilizando a matriz de correlação, pode-se usar a função `pccor<-prcomp(x, scale=TRUE)`

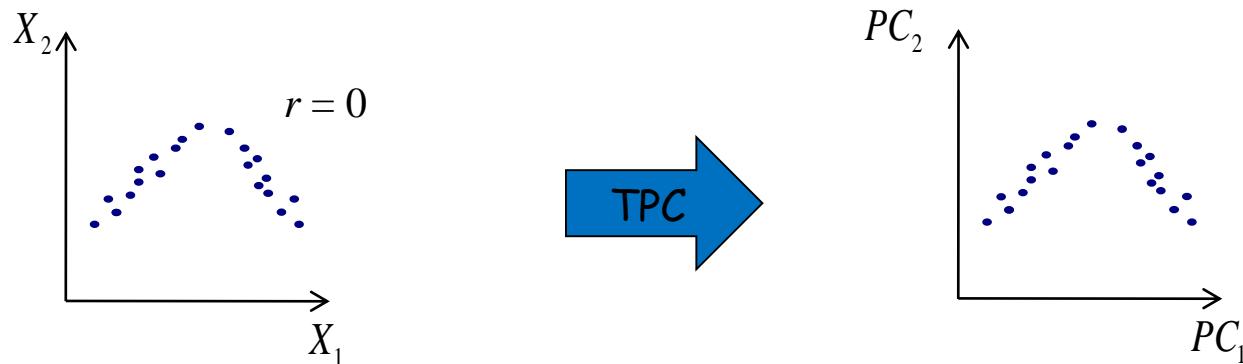
# Observações

- A transformação prioriza as variáveis com maior variância

**Atenção:** isso pode ocorrer quando as variáveis possuem diferentes grandezas

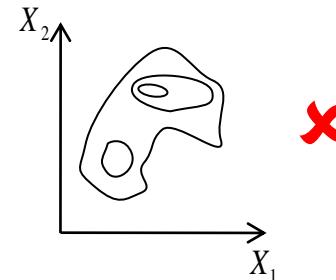
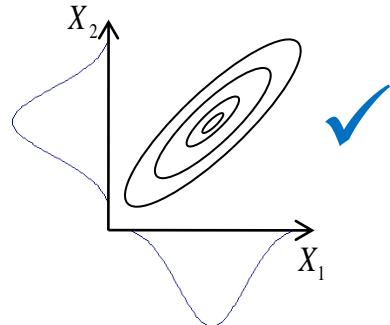
Para neutralizar o efeito das variâncias de cada variável de modo que todas tenham o mesmo peso, os autovalores e autovetores devem ser calculados a partir da matriz de correlação (que equivale realizar a transformação sobre as variáveis padronizadas)

- Por usar covariância (ou correlação), esta transformação pressupõe **relações lineares** entre variáveis

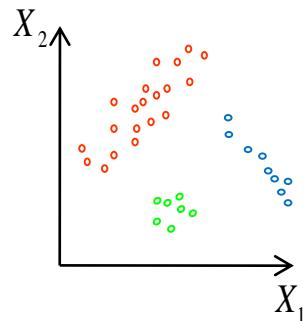


# Observações

- Por usar a matriz de covariância ou correlação, os resultados desta transformação são mais adequados quando os dados podem ser representados por uma distribuição multigaussiana (gaussiana multidimensional)



- A transformação pode não ser eficiente se diversas sub-populações com diferentes relações forem analisadas simultaneamente



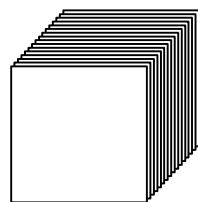
- A interpretação dos valores de uma determinada componente principal pode ser bastante difícil, necessitando a avaliação do autovetor correspondente

# Aplicações

- Diminuição da dimensionalidade do problema

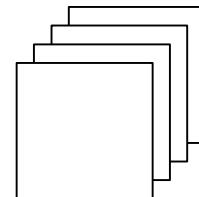
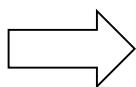
Ao invés de se trabalhar com 50 dimensões, escolhem-se as primeiras componentes que guardam a maior parte da informação total

Isso melhora o desempenho de classificadores que se baseiam em inversões de matrizes



$$\mu_X [50]$$

$$\Sigma_X [50 \times 50]$$



$$\mu_{PC} [4]$$

$$\Sigma_{PC} [4 \times 4]$$

(matriz diagonal!)

(97% da variação total, p.ex.)

# Aplicações

---

- **Diminuição da dimensionalidade do problema**

Ao invés de se trabalhar com 50 dimensões, escolhem-se as primeiras componentes que guardam a maior parte da informação total

Isso melhora o desempenho de classificadores que se baseiam em inversões de matrizes

- **Visualização de Dados**

A informação contida numa imagem hiperespectral pode ser visualizada usando-se as 3 primeiras componentes numa composição colorida RGB (obs: as cores podem ser de difícil interpretação)

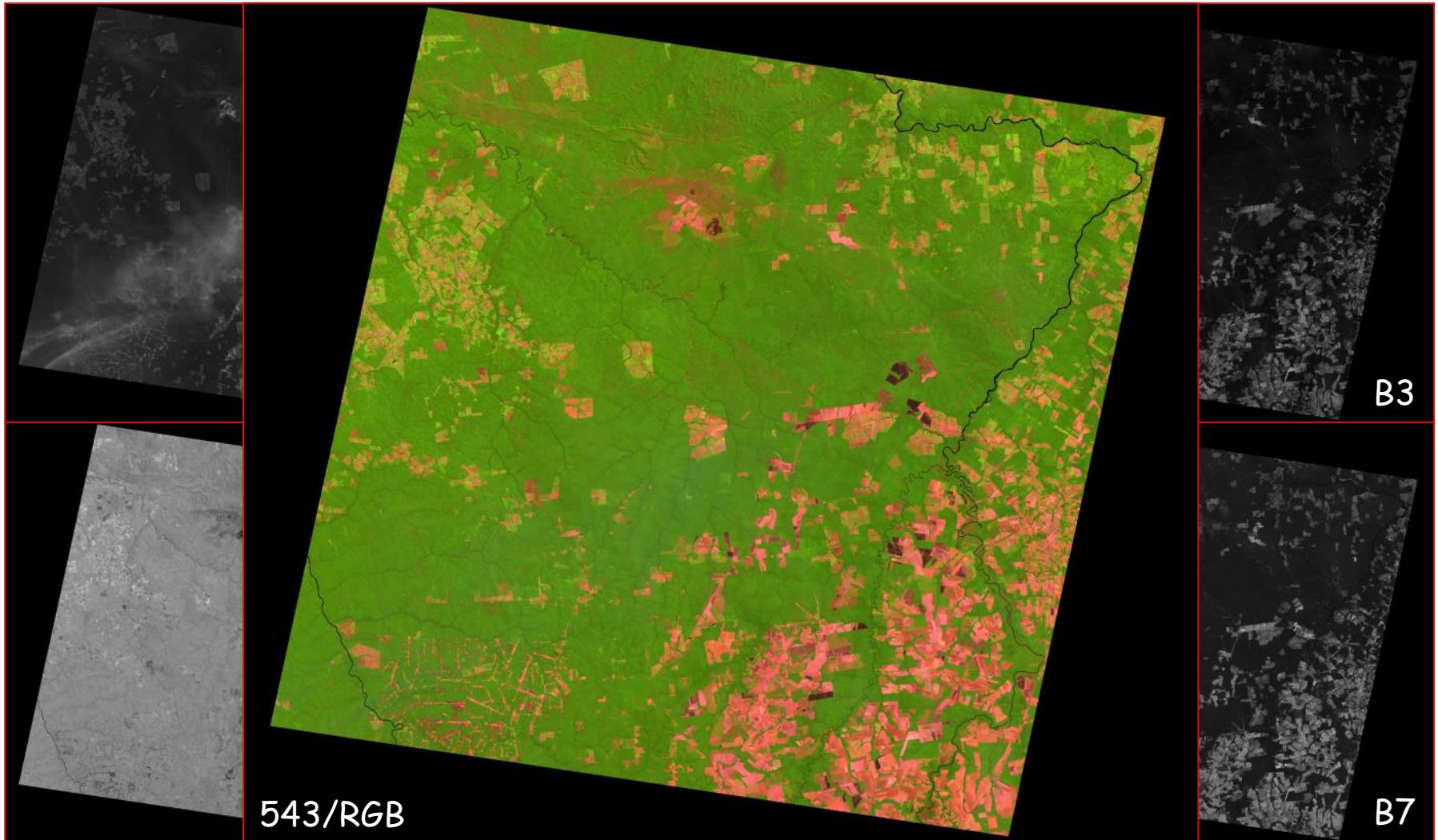
# Visualização de Dados

TM/LANDSAT (Bandas 1 a 5 e 7) 227/68 ano 1999



# Visualização de Dados

TM/LANDSAT (Bandas 1 a 5 e 7) 227/68 ano 1999 (ganho 2,45 e offset variável)



543/RGB

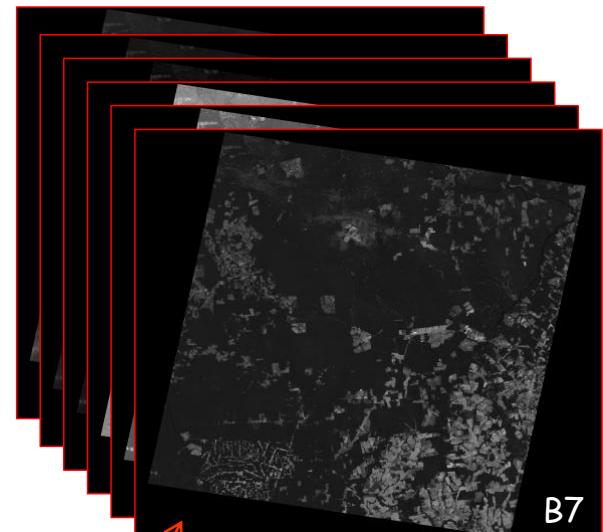
B3

B7

# Visualização de Dados

TM/LANDSAT (Bandas 1 a 5 e 7) 227/68 ano 1999 (ganho 2,45 e offset variável)

	Média	D.Padrão				
B1	45,556	19,300				
B2	28,236	11,600				
B3	25,484	20,837				
B4	130,698	18,811				
B5	108,052	44,184				
B7	39,701	25,810				
<b>Covariância</b>						
	B1	B2	B3	B4	B5	B7
B1	372,489	195,992	308,969	-62,672	493,634	295,327
B2	195,992	134,554	228,131	-33,144	419,773	245,247
B3	308,969	228,131	434,163	-107,410	837,561	499,226
B4	-62,672	-33,144	-107,410	353,868	-161,005	-161,894
B5	493,634	419,773	837,561	-161,005	1952,258	1094,010
B7	295,327	245,247	499,226	-161,894	1094,010	666,157
<b>Autovetor</b>						
<b>Autovalor</b>		B1	B2	B3	B4	B5
3235,547	CP1	0,230	0,180	0,351	-0,088	0,766
345,967	CP2	0,079	-0,022	0,055	-0,966	-0,215
272,900	CP3	0,842	0,291	0,241	0,132	-0,328
35,210	CP4	-0,333	0,169	0,549	0,167	-0,487
19,145	CP5	0,312	-0,291	-0,563	0,107	-0,150
4,720	CP6	0,157	-0,877	0,444	0,047	0,010
						-0,085

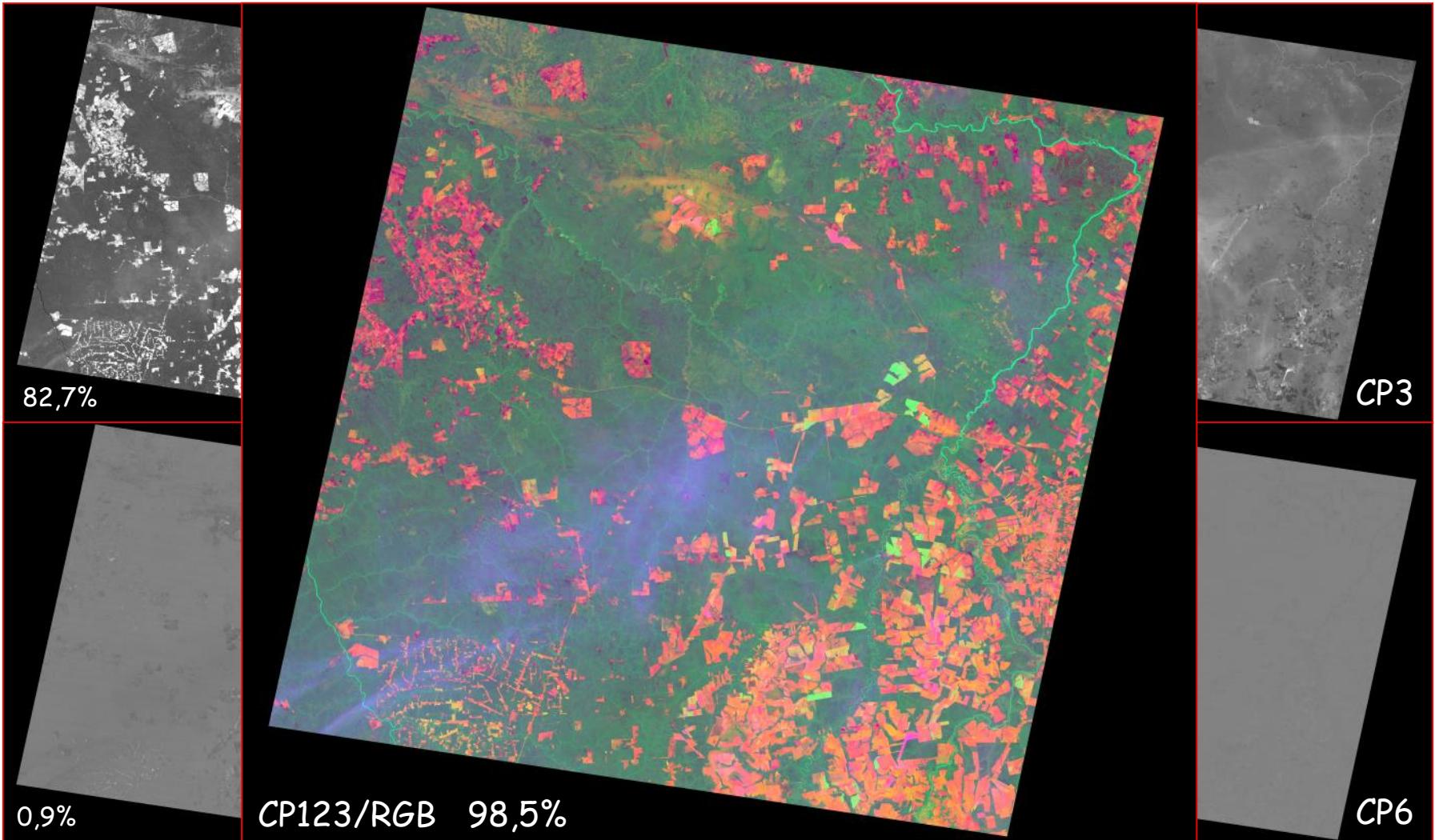


**Importante:** Para obtenção das estatísticas é necessário desconsiderar a região "sem informação"

OBS: Em geral, as imagens resultantes não são representadas em bytes (0 a 255), tendo valores positivos e negativos

# Visualização de Dados

## Componentes Principais

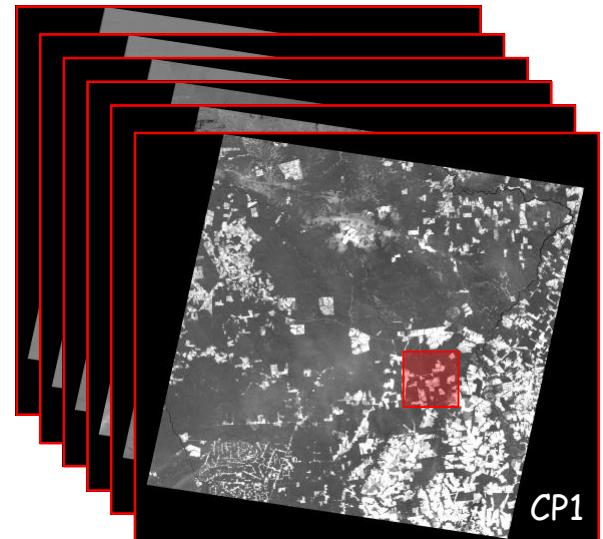


# Visualização de Dados

	Média	D.Padrão				
CP1	0	56,882				
CP2	0	18,600				
CP3	0	16,520				
CP4	0	5,934				
CP5	0	4,375				
CP6	0	2,172				

Covariância	CP1	CP2	CP3	CP4	CP5	CP6
CP1	3235,55	0	0	0	0	0
CP2	0	345,967	0	0	0	0
CP3	0	0	272,900	0	0	0
CP4	0	0	0	35,210	0	0
CP5	0	0	0	0	19,145	0
CP6	0	0	0	0	0	4,720



(ganho e offset arbitrários)

# Aplicações

- **Diminuição da dimensionalidade do problema**

Ao invés de se trabalhar com 50 dimensões, escolhem-se as primeiras componentes que guardam a maior parte da informação total

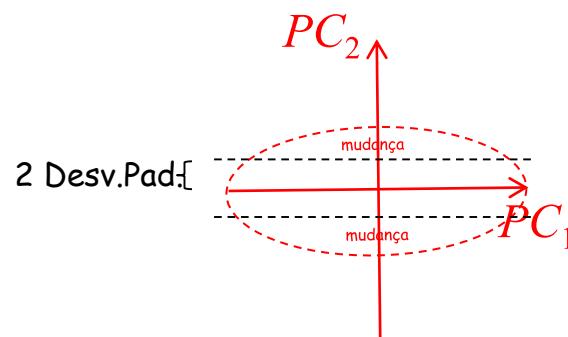
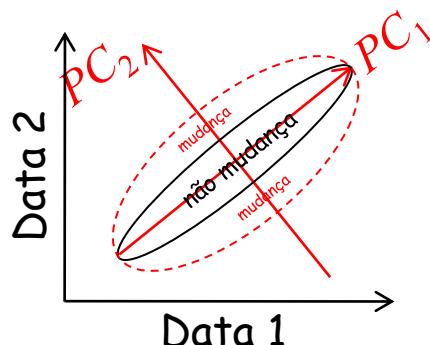
Isso melhora o desempenho de classificadores que se baseiam em inversões de matrizes

- **Visualização de Dados**

A informação contida numa imagem hiperespectral pode ser visualizada usando-se as 3 primeiras componentes numa composição colorida RGB (obs: as cores podem ser de difícil interpretação)

- **Detecção de Mudanças**

Numa comparação de 2 datas (mesma banda), os valores extremos da segunda componente evidenciam onde ocorreu mudança entre datas



# Aplicações

- **Diminuição da dimensionalidade do problema**

Ao invés de se trabalhar com 50 dimensões, escolhem-se as primeiras componentes que guardam a maior parte da informação total

Isso melhora o desempenho de classificadores que se baseiam em inversões de matrizes

- **Visualização de Dados**

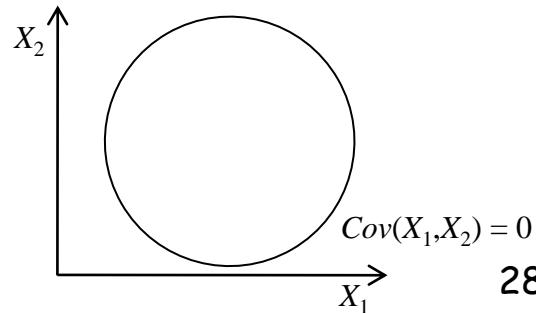
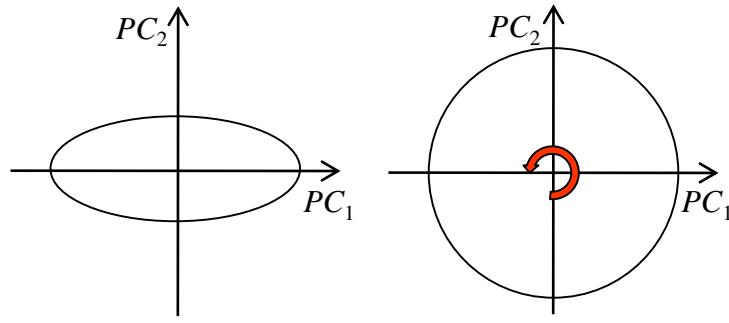
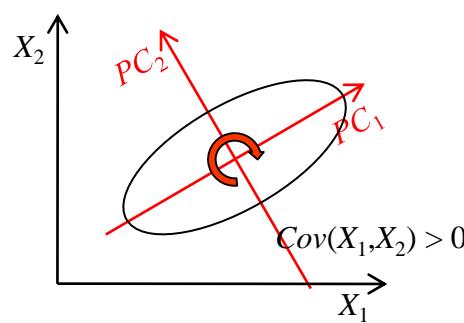
A informação contida numa imagem hiperespectral pode ser visualizada usando-se as 3 primeiras componentes numa composição colorida RGB (obs: as cores podem ser de difícil interpretação)

- **Detecção de Mudanças**

Numa comparação de 2 datas (mesma banda), os valores extremos da segunda componente evidenciam onde ocorreu mudança entre datas

- **Aumento de Contraste por Decorrelação**

Aplica-se a transformada por componentes principais, faz-se um aumento de contraste das componentes de modo que todas fiquem com a mesma variância e, por fim, faz-se a transformação inversa, restabelecendo-se as variáveis originais



# Aplicações

- **Diminuição da dimensionalidade do problema**

Ao invés de se trabalhar com 50 dimensões, escolhem-se as primeiras componentes que guardam a maior parte da informação total

Isso melhora o desempenho de classificadores que se baseiam em inversões de matrizes

- **Visualização de Dados**

A informação contida numa imagem hiperespectral pode ser visualizada usando-se as 3 primeiras componentes numa composição colorida RGB (obs: as cores podem ser de difícil interpretação)

- **Detecção de Mudanças**

Numa comparação de 2 datas (mesma banda), os valores extremos da segunda componente evidenciam onde ocorreu mudança entre datas

- **Aumento de Contraste por Decorrelação**

Aplica-se a transformada por componentes principais, faz-se um aumento de contraste das componentes de modo que todas fiquem com a mesma variância e, por fim, faz-se a transformação inversa, restabelecendo-se as variáveis originais

- **Simulação de dados correlacionados**

Calculam-se os autovalores e autovetores da matriz de variância-covariância.

Simulam-se dados não-correlacionados com variância igual aos autovalores e faz-se a transformação inversa

## Exemplo: Simulação de dados correlacionados

Simulação de 3 variáveis correlacionadas sendo que

$$\begin{aligned} X_1 &\sim N(\mu_{X_1} = 10,3; \sigma_{X_1}^2 = 3,5) & \boldsymbol{\mu} &= [10,3 \quad 35,1 \quad 107,8] \\ X_2 &\sim N(\mu_{X_2} = 35,1; \sigma_{X_2}^2 = 7,8) & \boldsymbol{\Sigma_X} &= \begin{bmatrix} 3,5 & -2,2 & -0,5 \\ -2,2 & 7,8 & 0,9 \\ -0,5 & 0,9 & 1,2 \end{bmatrix} \Leftrightarrow \mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma_X}) \\ X_3 &\sim N(\mu_{X_3} = 107,8; \sigma_{X_3}^2 = 1,2) \end{aligned}$$

Deseja-se simular 500 valores destas 3 variáveis respeitando suas correlações.

A simulação conjunta, mesmo para uma distribuição gaussiana, não é muito simples.

Por outro lado, a simulação de variáveis independentes é bastante simples de ser implementada quando se conhece a distribuição destas variáveis.

## Exemplo: Simulação de dados correlacionados

$$\Sigma_x = \begin{bmatrix} 3,5 & -2,2 & -0,5 \\ -2,2 & 7,8 & 0,9 \\ -0,5 & 0,9 & 1,2 \end{bmatrix} \xrightarrow{\text{TCP}} \lambda = [8,8629 \quad 2,5819 \quad 1,0552] \\ \mathbf{a} = \begin{bmatrix} 0,3867 & 0,9146 & 0,1182 \\ -0,9126 & 0,3980 & -0,0934 \\ -0,1324 & -0,0717 & 0,9886 \end{bmatrix}$$

Por definição, todas as componentes principais tem média zero e a variância de cada componente é dada pelo autovalor correspondente.

Além disso, no caso de distribuições gaussianas, a forma da distribuição é preservada após a transformação, ou seja, as componentes também têm distribuição gaussiana:

$$PC_1 \sim N(\mu = 0; \sigma_{PC_1}^2 = \lambda_1 = 8,8629)$$

$$PC_2 \sim N(\mu = 0; \sigma_{PC_2}^2 = \lambda_2 = 2,5819)$$

$$PC_3 \sim N(\mu = 0; \sigma_{PC_3}^2 = \lambda_3 = 1,0552)$$

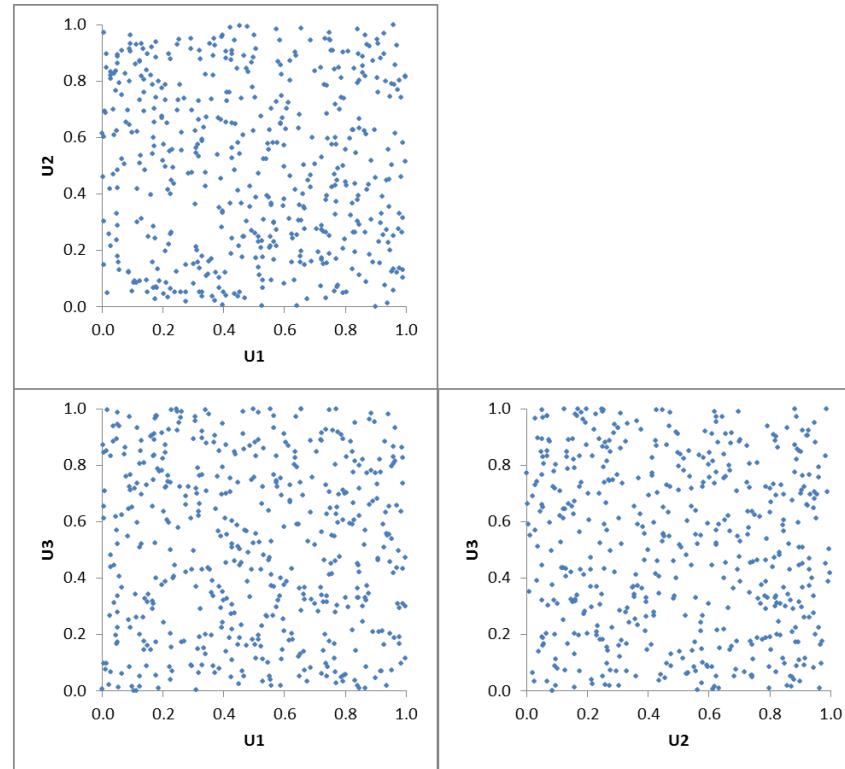
$$\mathbf{PC}' = \mathbf{a}' (\mathbf{X} - \mathbf{1}_{(500 \times 1)} \boldsymbol{\mu})' \quad \mathbf{X}' = (\mathbf{a}^{-1})' \mathbf{PC}' + (\mathbf{1}_{(500 \times 1)} \boldsymbol{\mu})' \quad \mathbf{a}^{-1} = \mathbf{a}' \quad \mathbf{X}' = \mathbf{a} \mathbf{PC}' + (\mathbf{1}_{(500 \times 1)} \boldsymbol{\mu})'$$

# Exemplo: Simulação de dados correlacionados

Neste exemplo, deseja-se gerar 500 valores para cada variável  $X$ .

O processo inicia-se com a geração de números aleatórios de uma distribuição uniforme contínua entre 0 e 1:

<b>U1</b>	<b>U2</b>	<b>U3</b>
0.9789	0.1352	0.1669
0.7383	0.7855	0.3370
0.9224	0.3914	0.2114
0.5247	0.0030	0.6637
0.9010	0.0007	0.7718
.	.	.
.	.	.
.	.	.
0.7961	0.7392	0.0535
0.4524	0.9962	0.5041
0.5987	0.3815	0.5824
0.6906	0.6003	0.7568
0.2216	0.5514	0.0853

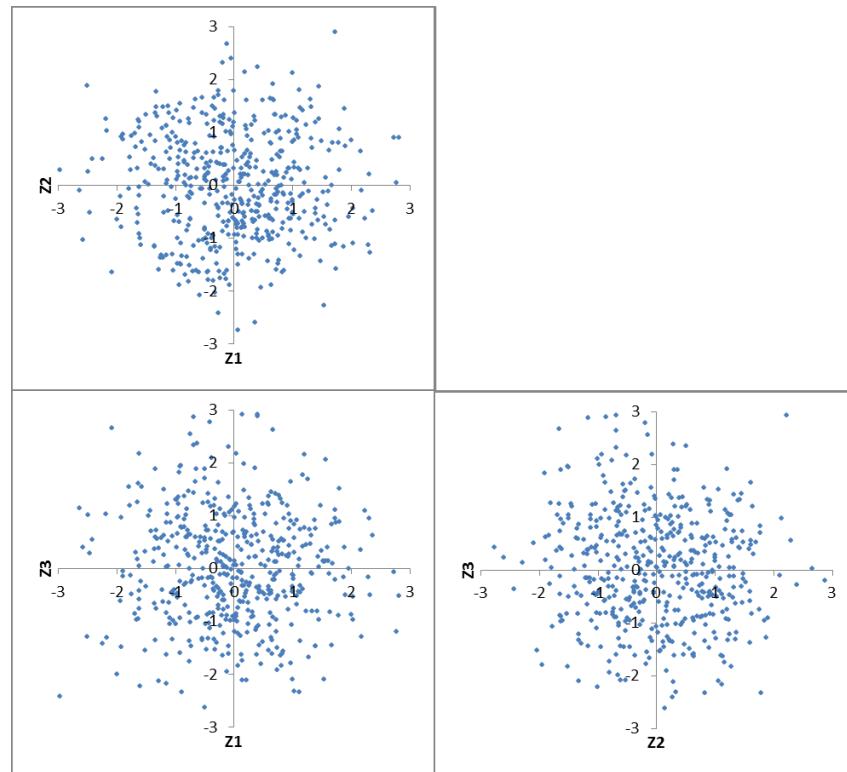


# Exemplo: Simulação de dados correlacionados

A partir de  $U$  calcula-se  $Z$  através da relação:  $P(Z < z_{ij}) = u_{ij}$

Exemplo:  $P(Z < z_{1,1}) = 0,9789 \Rightarrow z_{1,1} = 2,0312$

Z1	Z2	Z3
2.0312	-1.1020	-0.9663
0.6381	0.7909	-0.4208
1.4215	-0.2756	-0.8016
0.0619	-2.7474	0.4226
1.2872	-3.1950	0.7448
.	.	.
.	.	.
.	.	.
0.8277	0.6408	-1.6114
-0.1195	2.6664	0.0103
0.2500	-0.3016	0.2081
0.4975	0.2542	0.6960
-0.7667	0.1293	-1.3702

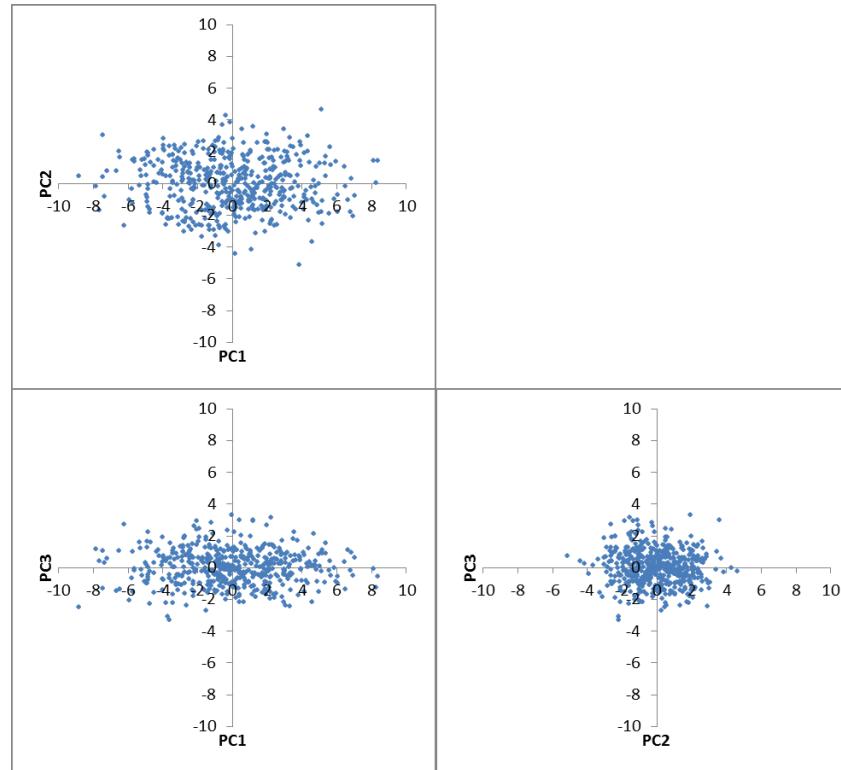


# Exemplo: Simulação de dados correlacionados

A partir de  $Z$  calcula-se  $PC$  multiplicando cada  $z_{ij}$  pelo desvio padrão da componente principal  $i$ , ou seja, a raiz quadrada do autovalor  $i$ :  $pc_{ij} = z_{ij} \times \sqrt{\lambda_i}$

Exemplo:  $pc_{1,1} = z_{1,1} \times \sqrt{\lambda_1} = 2,0312 \times \sqrt{8,8629} = 6,0470$

PC1	PC2	PC3
6.0470	-1.7708	-0.9926
1.8996	1.2708	-0.4322
4.2320	-0.4428	-0.8234
0.1842	-4.4145	0.4341
3.8320	-5.1337	0.7651
.	.	.
.	.	.
.	.	.
2.4641	1.0296	-1.6553
-0.3558	4.2844	0.0106
0.7443	-0.4846	0.2138
1.4812	0.4085	0.7149
-2.2824	0.2077	-1.4075



# Exemplo: Simulação de dados correlacionados

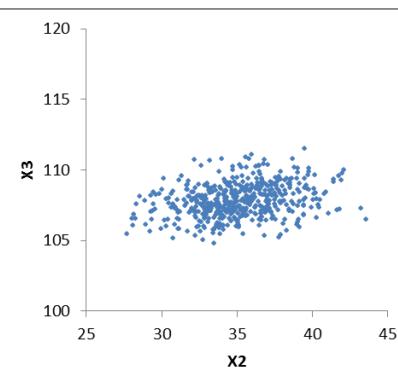
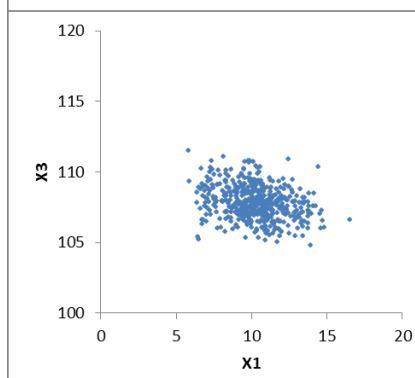
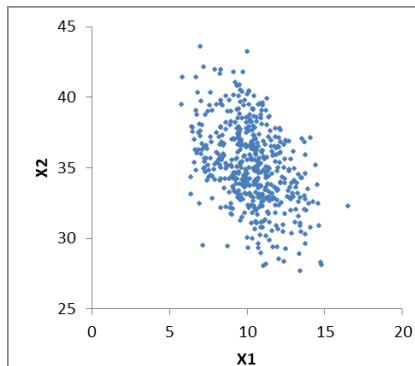
Finalmente, aplica-se a rotação inversa definida pela matriz de autovetores e soma-se a média correspondente a cada uma das variáveis  $X$ :

$$x_{ij} = \alpha_{1,1}pc_{1,j} + \alpha_{1,2}pc_{2,j} + \alpha_{1,3}pc_{3,j} + \mu_i \quad \text{onde } \alpha_{kw} \text{ é o } k\text{-ésimo elemento do autovetor } w$$

Exemplo:  $x_{1,1} = \alpha_{1,1}pc_{1,1} + \alpha_{1,2}pc_{2,1} + \alpha_{1,3}pc_{3,1} + \mu_1 =$

$$= 0,3867 \times 6,0470 - 0,9146 \times 1,7708 - 0,1182 \times 0,9926 + 10,3 = 10,9015$$

X1	X2	X3
10.9015	28.9694	106.1450
12.1458	33.9125	107.0300
11.4342	31.1386	106.4574
6.3850	33.1343	108.5212
7.1769	29.4882	108.4171
.	.	.
.	.	.
.	.	.
11.9989	33.4157	105.7635
14.0822	37.1289	107.5504
10.1698	34.2079	107.9476
11.3309	33.8440	108.2814
9.4410	37.3971	106.6958



## Exemplo: Simulação de dados correlacionados

---

É importante notar que, por se tratar de uma simulação, os valores de média, variância e covariância estimados a partir dos valores simulados não são exatamente os que se desejava (e variam a cada simulação!):

$$\mu = [10,3 \quad 35,1 \quad 107,8]$$

$$\bar{\mathbf{X}} = [10,28 \quad 35,11 \quad 107,83]$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} 3,5 & -2,2 & -0,5 \\ -2,2 & 7,8 & 0,9 \\ -0,5 & 0,9 & 1,2 \end{bmatrix}$$

$$\hat{\Sigma}_{\mathbf{x}} = \begin{bmatrix} 3,37 & -2,16 & -0,59 \\ -2,16 & 8,19 & 0,93 \\ -0,59 & 0,93 & 1,40 \end{bmatrix}$$

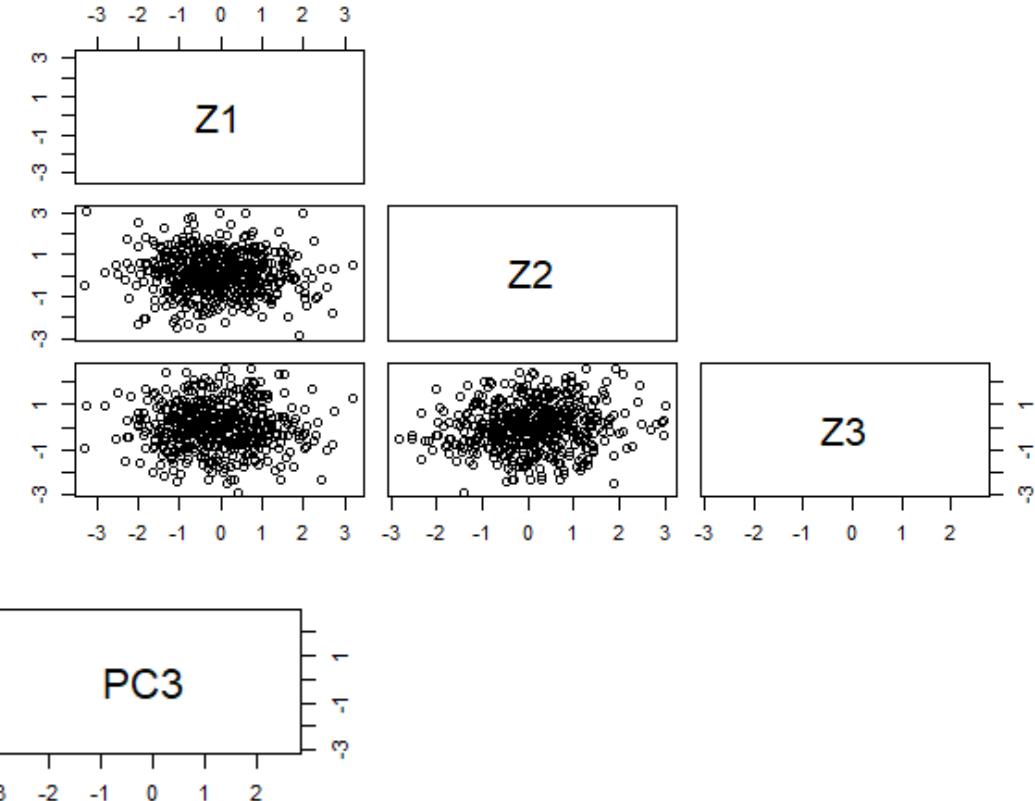
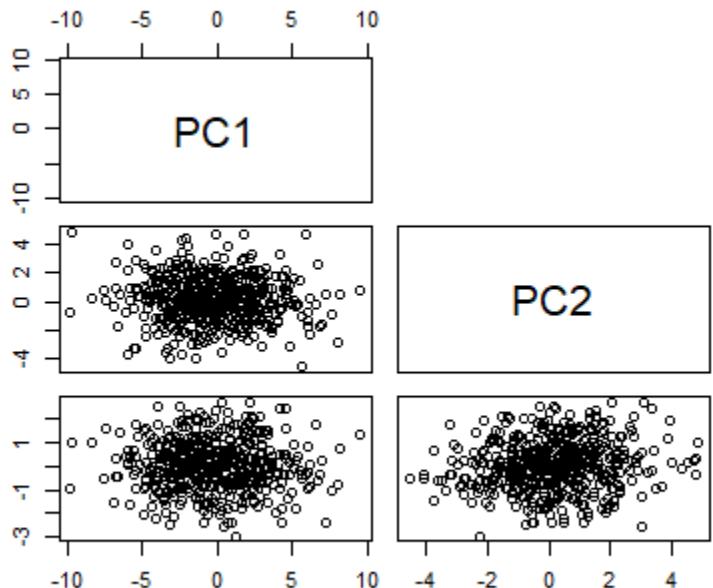
# Exemplo: Simulação de dados correlacionados em R

```
mX<-c(10.3,35.1,107.8)
covX<-cbind(c(3.5,-2.2,-0.5),c(-2.2,7.8,0.9),c(-0.5,0.9,1.2))
auto<-eigen(covX) #gera autovalores e autovetores a partir da matriz de covariância
```

```
n<-500
```

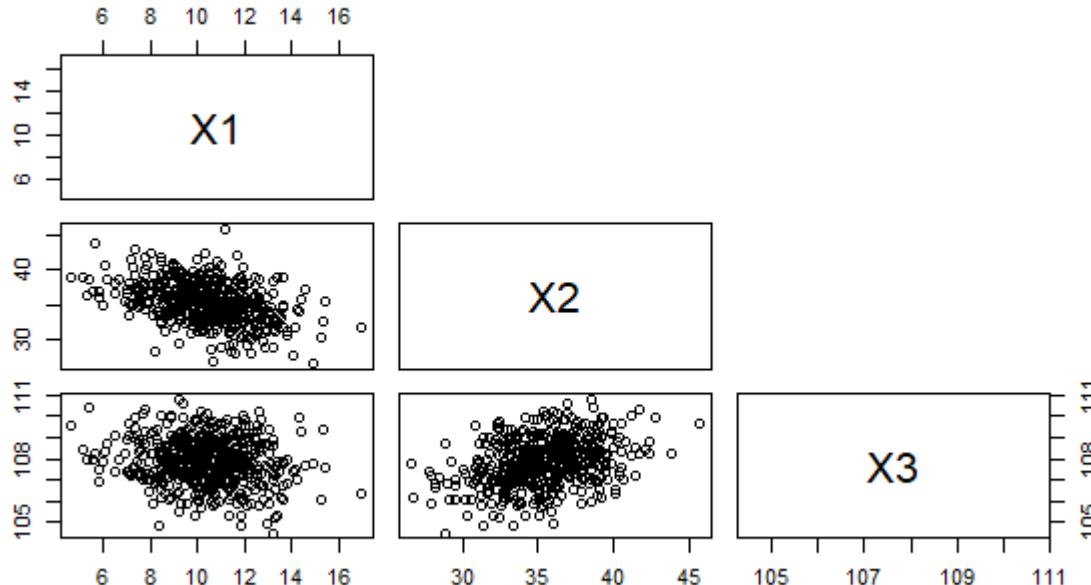
```
Z<-matrix(rnorm(3*n),nrow=n,ncol=3) #as colunas representam amostras de uma normal padrão (independentes)
colnames(Z)<-c("Z1","Z2","Z3")
pairs(Z,upper.panel = NULL)
```

```
PC<-Z*(matrix(rep(1,n))%*%t(sqrt(auto$values)))
colnames(PC)<-c("PC1", "PC2", "PC3")
pairs(PC,upper.panel = NULL)
```



# Exemplo: Simulação de dados correlacionados em R

```
X<-t(auto$vectors%*%t(PC)+t(matrix(rep(1,n))%*%mX))
colnames(X)<-c("X1", "X2", "X3")
pairs(X,upper.panel = NULL)
```



```
round(colMeans(X),2)
```

- $\begin{matrix} X1 & X2 & X3 \\ 10.34 & 35.39 & 107.86 \end{matrix} \leftarrow \bar{X}$   $\mu = [10,3 \quad 35,1 \quad 107,8]$

```
round(cov(X),2)
```

- $\begin{matrix} X1 & X2 & X3 \\ X1 & 3.36 & -2.16 & -0.31 \\ X2 & -2.16 & 8.29 & 1.07 \\ X3 & -0.31 & 1.07 & 1.20 \end{matrix} \leftarrow \hat{\Sigma}_x$   $\Sigma_x = \begin{bmatrix} 3,5 & -2,2 & -0,5 \\ -2,2 & 7,8 & 0,9 \\ -0,5 & 0,9 & 1,2 \end{bmatrix}$