

Pós-Graduação em Estatística com 

Análise Multivariada Aplicada

Jorge Cadima

Secção de Matemática (DCEB) - Instituto Superior de Agronomia (UL)

2015-16

Programa

- Alguns resultados de teoria de matrizes: decomposição espectral e decomposição em valores singulares
- Análise em Componentes Principais (ACP)
- Análise Discriminante Linear (ADL)
- Análises Classificatórias

Bibliografia

- Jolliffe, I.T. (2002) *Principal Component Analysis*, 2d. ed., Springer (Springer Series in Statistics)
- Krzanowski, W.J. (1998) *Principles of Multivariate Analysis: A User's Perspective*, Oxford Science Publications.
- Morrison, D.F. (1990) *Multivariate Statistical Methods*, 3rd.ed., McGraw-Hill.

Multivariada no :

- Everitt, B. & Hothorn, T. (2011) *An Introduction to Applied Multivariate Analysis with R*. Springer (Use R! Series).
- Zelterman, D. (2015). *Applied Multivariate Statistics with R*. Springer (Statistics for Biology and Health Series).

Conceitos introdutórios - ACP e ADL

Matéria prima: Matriz de dados $\mathbf{Y}_{n \times p}$: conjuntos de observações de p variáveis (**quantitativas**) em n indivíduos, ou unidades experimentais.

Nota: Ao contrário do que sucede na modelação, aqui **todas as variáveis estão em pé de igualdade**.

Quer numa Análise em Componentes Principais (ACP), quer numa Análise Discriminante Linear (ADL) **procuram-se novas variáveis**, construídas a partir das p variáveis observadas **que salientem**:

- numa ACP: a **variabilidade** entre indivíduos;
- numa ADL: a **separação** entre subgrupos conhecidos dos indivíduos.

Em ambos os casos, as **novas variáveis** são **combinações lineares das p variáveis observadas**.

Conceitos introdutórios - Análises Classificatórias

Matéria prima: Matriz de dissemelhanças $\mathbf{Y}_{n \times n}$: dissemelhanças (talvez distâncias) entre n indivíduos, ou unidades experimentais.

Nota: Na base das dissemelhanças podem, ou não, existir medições dos indivíduos em p variáveis.

Numa Análise Classificatória (AC) procura-se agrupar os indivíduos em classes internamente homogêneas e heterogêneas entre si.

Diferentes critérios de dissemelhanças entre indivíduos, bem como de homogeneidade interna e heterogeneidade externa, produzem diferentes métodos de AC.

A representação usual dos dados

Na **representação tradicional**, a uma **matriz de dados** $Y_{n \times p}$ corresponde uma **nuvem de n pontos em $\mathbb{R}^p$** :

$$\begin{array}{lll} p \text{ eixos} & \longleftrightarrow & p \text{ variáveis} \\ n \text{ pontos} & \longleftrightarrow & n \text{ indivíduos} \end{array}$$

Esta nuvem de pontos **não é visualizável para $p > 3$** .

A **ACP** pode ser vista como uma **técnica de redução “óptima” da dimensionalidade**. No caso duma redução para $k=2$ ou $k=3$ dimensões, ter-se-á uma aproximação visualizável da nuvem de pontos.

Um exemplo: os lavagantes de Somers

Dados: $p=13$ variáveis morfométricas sobre $n=63$ lavagantes

```
> lavagantes
```

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
1	29.42	21.43	14.91	12.58	12.85	10.57	1.76	6.45	6.67	9.14	24.54	10.38	15.37
2	30.06	22.05	14.81	12.54	12.96	10.75	1.73	6.11	7.04	8.76	26.21	11.00	11.92
3	30.30	21.95	15.10	12.97	13.05	11.11	2.05	6.46	7.14	9.35	26.55	11.84	16.50
4	30.75	21.91	15.89	12.85	13.75	10.75	1.71	6.62	6.84	9.53	25.35	11.60	15.47
5	31.06	20.37	15.83	13.15	13.37	11.50	2.15	5.96	7.09	9.15	26.88	11.92	17.24
6	31.27	24.04	17.45	14.49	14.77	12.64	2.06	6.59	7.43	10.75	31.60	14.32	18.95
7	31.39	21.91	15.96	13.41	13.74	11.79	2.03	6.40	6.89	9.82	28.16	12.53	16.90
8	31.51	23.63	15.95	13.14	13.89	11.74	1.94	6.26	6.81	9.36	26.09	11.15	15.48
9	32.12	22.81	16.06	13.29	13.80	12.14	2.02	6.47	7.00	9.70	27.01	11.22	16.65
10	32.40	22.96	16.69	13.82	14.30	12.06	2.03	6.14	7.27	9.53	29.34	12.59	17.90
.....													
56	33.44	24.72	17.06	14.25	16.74	12.42	2.04	6.52	7.25	10.21	26.92	11.40	16.23
57	33.48	25.32	17.50	14.15	17.20	12.40	2.17	6.94	7.54	10.37	26.85	11.40	16.34
58	33.57	25.00	16.74	14.10	16.49	12.43	1.95	7.27	7.37	10.15	25.13	11.23	14.98
59	33.74	25.30	17.11	14.26	16.35	12.37	2.26	6.82	7.41	11.14	26.43	10.91	16.02
60	34.37	25.35	17.98	14.49	16.95	12.69	2.02	7.04	7.35	10.33	27.97	11.75	17.19
61	34.66	25.32	18.50	14.16	17.37	12.60	2.32	6.88	7.59	11.00	27.76	11.87	17.58
62	34.93	26.77	18.00	14.13	16.89	12.67	2.04	7.14	7.79	10.36	26.98	11.55	17.20
63	35.73	25.79	18.35	15.06	17.15	13.14	2.15	7.09	7.83	10.59	28.29	12.30	17.45

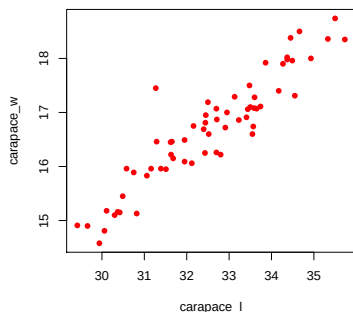


A representação gráfica de dados multivariados

É fácil representar graficamente n observações em $p = 2$ variáveis:

Eixos – Variáveis

Pontos – Indivíduos



Com bom *software* ainda se representam nuvens de pontos com $p=3$ variáveis.

Mas temos uma nuvem de $n = 63$ pontos num espaço a $p = 13$ dimensões.

Como tentar visualizar?

Os comandos `plot` e `pairs` no

Pode-se construir *todas* as nuvens de pontos entre pares de variáveis. São as *projeções ortogonais* da nuvem de n pontos sobre os planos coordenados de $\mathbb{R}^p$. É um paliativo, mas *não é o melhor paliativo*.

Para dados numa *data frame*, o comando `plot` constrói essa “matriz de gráficos”. Para dados numa estrutura da classe *matrix*, o comando `plot` *não produz* o mesmo resultado (produz a nuvem de pontos relativa às duas primeiras colunas da matriz: experimente!).

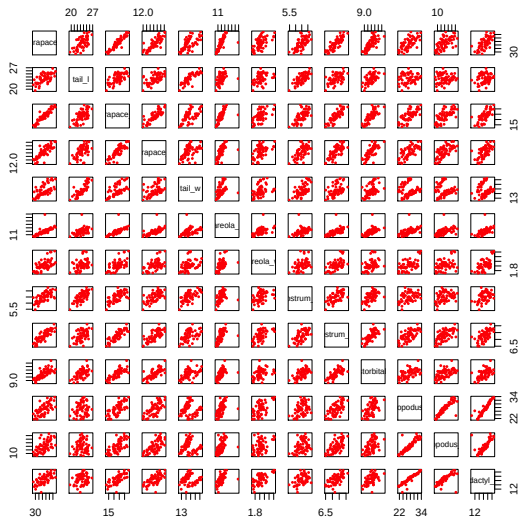
O comando `pairs` produz, para objectos da classe *matrix*, a “matriz de gráficos”.

```
> class(lavagantes)
[1] "matrix"
```

Há duas *alternativas*:

```
> pairs(lavagantes, pch=16, cex=0.6, col='red')
> plot(as.data.frame(lavagantes), pch=16, cex=0.6, col='red')
```

O comando pairs nos lavagantes



Com $p = 3$ variáveis ainda é possível visualizar

Uma ferramenta para visualização de dados tri-dimensionais está disponível no módulo `car`, através da função `scatter3d`. A função `Identify3d` permite identificar pontos interactivamente.

```
> library(car)
> scatter3d(lavagantes[,1],lavagantes[,2],lavagantes[,3],surface=FALSE)
> Identify3d(lavagantes[,1],lavagantes[,2],lavagantes[,3],offset=0.1)
```

Mantendo carregado o botão direito do rato, pode seleccionar um ou vários pontos no gráfico, que serão identificados.

Um clique no botão direito do rato interrompe a execução do comando.

Uma ferramenta alternativa de visualização tri-dimensional encontra-se no módulo `rggobi`.

Este módulo necessita que esteja previamente instalado no sistema (exterior ao R) o *software* `GGobi` (gratuito e de código aberto, www.ggobi.org).

Projectões ortogonais

Mas continuamos a ter visões parciais, resultantes de projectar ortogonalmente a nuvem de $n = 63$ pontos em $\mathbb{R}^{13}$ sobre espaços (neste caso tri-dimensionais) coordenados.

Qualquer projecção empobrece a representação: ficamos com uma visão parcial. Distâncias podem ficar camufladas.

Mas,

- Porquê só planos coordenados? Porque não outros planos?
- Qual o plano onde a projecção é mais fidedigna?
- Qual o critério de “projecção fidedigna”?

Ideia intuitiva: estar globalmente mais perto da generalidade dos pontos. Esta é a abordagem que conduz à Análise em Componentes Principais.

Ferramentas: Teoria de Matrizes

$$\mathbf{A}_{n \times p} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1,p} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2,p} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{(n-1),1} & a_{(n-1),2} & a_{(n-1),3} & \cdots & a_{(n-1),p} \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,p} \end{bmatrix} \equiv [a_{ij}] ,$$

($i = 1, 2, \dots, n$ indica no. de linha e $j = 1, 2, \dots, p$ indica no. de coluna)

Se $n = p$ a matriz diz-se **quadrada**.

Se $n \neq p$ a matriz diz-se **rectangular**.

A matriz **transposta** de $\mathbf{A}$, $\mathbf{A}^t$, é uma matriz $p \times n$, obtida escrevendo cada linha de $\mathbf{A}$ como a coluna correspondente de $\mathbf{A}^t$ (ou, analogamente, cada coluna de $\mathbf{A}$ como a linha correspondente de $\mathbf{A}^t$). De forma equivalente, o elemento genérico da transposta, a_{ij}^t , é dado por:

$$a_{ij}^t = a_{ji} \quad (i = 1, \dots, p \quad , \quad j = 1, \dots, n).$$

Operações com matrizes

Definem-se as seguintes operações sobre matrizes (**atenção às dimensões das matrizes**):

- **Soma:** $\mathbf{C}_{n \times p} = \mathbf{A}_{n \times p} + \mathbf{B}_{n \times p} \Leftrightarrow c_{ij} = a_{ij} + b_{ij}, \quad \forall i, j.$
- **Multiplicação escalar:** $\alpha \mathbf{A} \equiv [\alpha a_{ij}] \quad \forall i, j.$
- **Produto de Hadamard:** Não é o produto de matrizes habitual!

$$\mathbf{C}_{n \times p} = \mathbf{A}_{n \times p} \circ \mathbf{B}_{n \times p} \Leftrightarrow c_{ij} = a_{ij} \cdot b_{ij}, \quad \forall i, j.$$

- **Produto matricial:**

$$\mathbf{C}_{n \times m} = \mathbf{A}_{n \times p} \mathbf{B}_{p \times m} \Leftrightarrow c_{ij} = (\vec{\mathbf{a}}_i^{\text{linha}})^t (\vec{\mathbf{b}}_j^{\text{coluna}}) = \sum_{k=1}^p a_{ik} b_{kj}, \quad \forall i, j,$$

onde

- ▶ $\vec{\mathbf{a}}_i^{\text{linha}}$ representa a i -ésima linha da matriz $\mathbf{A}$
- ▶ $\vec{\mathbf{b}}_j^{\text{coluna}}$ representa a j -ésima coluna da matriz $\mathbf{B}$

Notas sobre o produto matricial

Considere duas matrizes $\mathbf{A}_{n \times p}$ e $\mathbf{B}_{m \times q}$:

- A multiplicação de matrizes $\mathbf{AB}$ só é possível se as matrizes forem **compatíveis**: $p = m$.
- O produto matricial *não* é comutativo, isto é, em geral $\mathbf{AB} \neq \mathbf{BA}$ (podendo um dos produtos nem ser possível).
- Se $\vec{x}$ é um vector $p \times 1$, o produto $\mathbf{A}\vec{x}$ é uma combinação linear das colunas da matriz $\mathbf{A}$, $\mathbf{A}\vec{x} = \sum_{i=1}^p x_i \vec{a}_i^{\text{coluna}}$.
- Se $\vec{x}$ é um vector $n \times 1$, o produto $\vec{x}^t \mathbf{A}$ é uma combinação linear das linhas da matriz $\mathbf{A}$, $\vec{x}^t \mathbf{A} = \sum_{i=1}^n x_i \vec{a}_i^{\text{linha}^t}$.
- Se $\mathbf{B}_{n \times p}$, $\vec{a} \in \mathbb{R}^n$ e $\vec{c} \in \mathbb{R}^p$, tem-se

$$\vec{a}^t \mathbf{B} \vec{c} = \sum_{i=1}^n \sum_{j=1}^p a_i b_{ij} c_j .$$

Matrizes quadradas

Eis alguns tipos importantes de matrizes quadradas, $\mathbf{A}_{p \times p}$:

A Diagonal	$a_{ij} = 0$ se $i \neq j$ (e existe i tal que $a_{ii} \neq 0$)
A Simétrica	$\mathbf{A}^t = \mathbf{A} \iff a_{ij} = a_{ji}, \forall i, j$
$\mathbf{I}_p$ Identidade	$\mathbf{A} = \mathbf{I}_p \iff a_{ij} = \begin{cases} 0 & \text{para } i \neq j \\ 1 & \text{para } i = j \end{cases}$
$\mathbf{A}^{-1}$ Inversa de A	$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_p$ (nem sempre existe, mas quando existe é única)
A Ortogonal	$\mathbf{A}^{-1} = \mathbf{A}^t \iff \mathbf{A}^t\mathbf{A} = \mathbf{A}\mathbf{A}^t = \mathbf{I}_p$
A Idempotente	$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$

Produto interno e norma usual em $\mathbb{R}^n$

Os conceitos geométricos nos espaços $\mathbb{R}^n$ estão associados à definição dum **produto interno**.

O produto interno usual nos espaços $\mathbb{R}^n$: $\langle \vec{x}, \vec{y} \rangle = \vec{x}^t \vec{y} = \sum_{i=1}^n x_i y_i$.

O **comprimento** ou **norma** usual de um vector de $\mathbb{R}^n$ é:

Em $\mathbb{R}^n$, a **norma habitual** (euclideana) é:

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\vec{x}^t \vec{x}} = \sqrt{\sum_{i=1}^n x_i^2}$$

Mas **existem outros produtos internos e outras normas** em $\mathbb{R}^n$.

Ortogonalidade e ângulos

O conceito de **ortogonalidade** (perpendicularidade) e, de forma mais geral, o conceito de **ângulo** entre dois vectores dum espaço linear também se introduzem através de produtos internos:

Dois vectores dum espaço linear com produto interno, $\vec{x}, \vec{y} \in L$, dizem-se **ortogonais** se $\langle \vec{x}, \vec{y} \rangle = 0$, escrevendo-se $\vec{x} \perp \vec{y}$

Nota: O vector $\vec{0}$ é, por definição, ortogonal a qualquer outro vector.

Sejam $\vec{x}, \vec{y}$ dois vectores **não nulos** dum espaço linear L com produto interno. O **ângulo** θ entre $\vec{x}$ e $\vec{y}$ define-se como:

$$\theta = \arccos \left(\frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \cdot \|\vec{y}\|} \right) \Rightarrow \cos \theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \cdot \|\vec{y}\|} .$$

Valores e vectores próprios

Definição

Dada uma matriz real $\mathbf{A}_{p \times p}$, um **vector não-nulo** $\vec{\mathbf{x}} \in \mathbb{C}^p$ diz-se um **vector próprio** de $\mathbf{A}$, e $\lambda \in \mathbb{C}$ diz-se o **valor próprio** correspondente se:

$$\mathbf{A}\vec{\mathbf{x}} = \lambda \vec{\mathbf{x}} .$$

Nota: Se $\mathbf{A}_{p \times p}$ for uma matriz **simétrica**, os seus valores/vectores próprios têm boas propriedades:

- Os valores e vectores próprios são sempre **reais**.
- Vectores próprios associados a valores próprios diferentes são sempre **ortogonais**.
- Mesmo que haja valores próprios repetidos, **é possível determinar um conjunto ortonormado de p vectores próprios** (todos de norma 1 e ortogonais entre si).

Teorema da Decomposição Espectral

Teorema (Decomposição espectral)

Seja $\mathbf{A} \in \mathbb{M}_{p \times p}$ uma matriz **simétrica**. Então existe uma **matriz ortogonal** $\mathbf{V} \in \mathbb{M}_{p \times p}$ (com colunas $\vec{\mathbf{v}}_i$) e uma **matriz diagonal** $\mathbf{\Lambda} \in \mathbb{M}_{p \times p}$ (com elementos diagonais λ_i), tal que:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t \iff \mathbf{A} = \sum_{i=1}^p \lambda_i \vec{\mathbf{v}}_i \vec{\mathbf{v}}_i^t.$$

Notas:

- Os λ_i são **valores próprios** e os vectores $\vec{\mathbf{v}}_i$ são um **conjunto ortonormado** de **vectores próprios** de $\mathbf{A}$.
- Se todos os valores próprios forem diferentes, os vectores próprios $\vec{\mathbf{v}}_i$ são únicos, a menos de troca de sinal (tanto se pode usar $\vec{\mathbf{v}}_i$ como $-\vec{\mathbf{v}}_i$). Ordenando $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, a decomposição é única (a menos de trocas de sinais nos vectores).

Potências de matrizes simétricas

Seja $\mathbf{A}$ uma matriz simétrica, $p \times p$, e seja $\mathbf{A}_{p \times p} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t$ a sua decomposição espectral.

- Para qualquer $k \in \mathbb{N}$, tem-se $\mathbf{A}^k = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^t$, com $\mathbf{\Lambda}^k \equiv [\lambda_i^k]$.
- Se $\mathbf{A}$ invertível,
 - ▶ define-se $\mathbf{A}^0 = \mathbf{V}\mathbf{\Lambda}^0\mathbf{V}^t = \mathbf{V}\mathbf{V}^t = \mathbf{I}_{p \times p}$;
 - ▶ tem-se, para a inversa, que $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^t$, com $\mathbf{\Lambda}^{-1} \equiv \left[\frac{1}{\lambda_i}\right]$;
 - ▶ para $k \in \mathbb{N}$, $\mathbf{A}^{-k} = \mathbf{V}\mathbf{\Lambda}^{-k}\mathbf{V}^t = (\mathbf{A}^k)^{-1} = (\mathbf{A}^{-1})^k$
- Se $\mathbf{A}$ definida positiva, define-se $\mathbf{A}^k = \mathbf{V}\mathbf{\Lambda}^k\mathbf{V}^t$ para qualquer $k \in \mathbb{R}$.

Para os expoentes que faz sentido em cada caso, definiu-se uma álgebra de potências análoga à dos números reais:

$$\mathbf{A}^k \mathbf{A}^m = \mathbf{A}^{k+m}.$$

Valores próprios e matrizes simétricas

Teorema de Rayleigh-Ritz

Seja $\mathbf{A}_{p \times p}$ matriz simétrica, com valores próprios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p-1} \geq \lambda_p$.

- λ_1 e o maior valor do quociente de Rayleigh-Ritz $\frac{\vec{x}^t \mathbf{A} \vec{x}}{\vec{x}^t \vec{x}}$, atingido quando $\vec{x} = \vec{v}_1$, o vector próprio correspondente.
- Os restantes valores (λ_i)/vectores($\vec{v}_i$) próprios de $\mathbf{A}$ são sucessivos máximos dos quocientes de Rayleigh-Ritz, sujeitos à ortogonalidade de $\vec{x}$ com anteriores soluções.

Definição

Seja $\mathbf{A}$ uma matriz simétrica, de tipo $p \times p$. Então:

$\mathbf{A}$ definida positiva	$\iff \lambda_i > 0, \forall i$
$\mathbf{A}$ semi-definida positiva	$\iff \lambda_i \geq 0, \forall i$ (pelo menos um zero)
$\mathbf{A}$ definida negativa	$\iff \lambda_i < 0, \forall i$
$\mathbf{A}$ semi-definida negativa	$\iff \lambda_i \leq 0, \forall i$ (pelo menos um zero)
$\mathbf{A}$ é indefinida	$\iff \exists i : \lambda_i < 0, \quad \exists j : \lambda_j > 0.$

ACP: Uma introdução estatística

Uma introdução frequente à ACP usa sobretudo conceitos estatísticos. Seja dada a **matriz de dados** $n \times p$, $\mathbf{X}$.

Pretende-se a **combinação linear das p variáveis de variância máxima**. Isto é, pretende-se determinar o vector $\vec{\mathbf{v}} = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$ tal que

$$\mathbf{X}\vec{\mathbf{v}} = v_1 \vec{\mathbf{x}}_1 + v_2 \vec{\mathbf{x}}_2 + v_3 \vec{\mathbf{x}}_3 + \dots + v_p \vec{\mathbf{x}}_p$$

tenha variância máxima (sendo $\vec{\mathbf{x}}_j \in \mathbb{R}^n$ o vector das observações da variável j , ou seja, a j -ésima coluna de $\mathbf{X}$).

A **variância duma combinação linear $\mathbf{X}\vec{\mathbf{v}}$** é $\vec{\mathbf{v}}^t \mathbf{S} \vec{\mathbf{v}}$, sendo $\mathbf{S}$ a **matriz de (co)variâncias dos dados**. Logo, quer-se $\vec{\mathbf{v}}$ que maximize $\vec{\mathbf{v}}^t \mathbf{S} \vec{\mathbf{v}}$.

Sem outras restrições, o problema não tem solução (pode-se escolher $\vec{\mathbf{v}}$ de elementos arbitrariamente grandes). **Impõe-se a restrição $\vec{\mathbf{v}}^t \vec{\mathbf{v}} = 1$** (corresponde à soma dos quadrados dos coeficientes v_i ser igual a 1).

Introdução estatística (cont.)

Assim, pretende-se maximizar o quociente de Rayleigh-Ritz de $\mathbf{S}$:

$$\max_{\vec{\mathbf{v}} \in \mathbb{R}^p} \frac{\vec{\mathbf{v}}^t \mathbf{S} \vec{\mathbf{v}}}{\vec{\mathbf{v}}^t \vec{\mathbf{v}}}$$

A solução é dada pelo vector próprio $\vec{\mathbf{v}}_1$, associado ao maior valor próprio de $\mathbf{S}$, λ_1 .

Nota: Se $\vec{\mathbf{v}}$ é vector próprio, $-\vec{\mathbf{v}}$ também é. As soluções do problema definem direcções, mas não definem sentidos.

A primeira Componente Principal é a combinação linear $\pm \mathbf{X} \vec{\mathbf{v}}_1$ (também definida a menos duma multiplicação por -1).

O valor próprio λ_1 é a variância desta primeira CP.

ACP: introdução estatística (cont.)

Fixada a 1a. CP, procura-se nova combinação linear $\mathbf{X}\vec{\mathbf{v}}$, com $\vec{\mathbf{v}}^t\vec{\mathbf{v}} = 1$, de variância máxima, **não-correlacionada com a anterior**.

Uma **correlação nula** equivale a uma **covariância nula**, e a **covariância de duas combinações lineares das colunas duma matriz $\mathbf{X}$** , no nosso caso **$\mathbf{X}\vec{\mathbf{v}}_1$ e $\mathbf{X}\vec{\mathbf{v}}$** , é dada por **$\vec{\mathbf{v}}^t\mathbf{S}\vec{\mathbf{v}}_1$** , sendo **$\mathbf{S}$** a matriz de covariâncias associada aos dados em **$\mathbf{X}$** .

Mas **$\vec{\mathbf{v}}_1$** é um vector próprio de **$\mathbf{S}$** , associado ao valor próprio **λ_1** . Logo:

$$\text{cov}(\mathbf{X}\vec{\mathbf{v}}, \mathbf{X}\vec{\mathbf{v}}_1) = \vec{\mathbf{v}}^t\mathbf{S}\vec{\mathbf{v}}_1 = 0 \quad \Leftrightarrow \quad \lambda_1\vec{\mathbf{v}}^t\vec{\mathbf{v}}_1 = 0 \quad \Leftrightarrow \quad \vec{\mathbf{v}} \perp \vec{\mathbf{v}}_1.$$

Logo, maximizar a variância de **$\mathbf{X}\vec{\mathbf{v}}$** sujeito à não correlação de **$\mathbf{X}\vec{\mathbf{v}}$** com **$\mathbf{X}\vec{\mathbf{v}}_1$** equivale a **maximizar $\frac{\vec{\mathbf{v}}^t\mathbf{S}\vec{\mathbf{v}}}{\vec{\mathbf{v}}^t\vec{\mathbf{v}}}$** , sujeito a **$\vec{\mathbf{v}} \perp \vec{\mathbf{v}}_1$** .

É o problema associado aos quocientes de Rayleigh-Ritz.

ACP: introdução estatística (cont.)

Assim, maximizar a variância de $\mathbf{X}\vec{\mathbf{v}}$ sujeito à não correlação de $\mathbf{X}\vec{\mathbf{v}}$ com $\mathbf{X}\vec{\mathbf{v}}_1$ corresponde a tomar $\vec{\mathbf{v}} = \pm\vec{\mathbf{v}}_2$, o vector próprio de $\mathbf{S}$ associado ao seu segundo maior valor próprio, λ_2 .

$\pm\mathbf{X}\vec{\mathbf{v}}_2$ é a segunda componente principal, com variância λ_2 .

As CPs são soluções do problema de determinar sucessivas combinações lineares de variância máxima, não-correlacionadas entre si.

A j -ésima componente principal é dada por $\pm\mathbf{X}\vec{\mathbf{v}}_j$, onde $\pm\vec{\mathbf{v}}_j$ é o vector próprio de $\mathbf{S}$ associado ao j -ésimo maior valor próprio $\lambda_j = \text{var}(\mathbf{X}\vec{\mathbf{v}}_j)$.

O comando usual para efectuar uma ACP no R é o comando `prcomp`.

O comando `prcomp` tem um único argumento obrigatório (embora adiante veremos a oportunidade de incluir mais argumentos): o nome do objecto contendo os dados, que deve ser da classe `matrix` ou da classe `data.frame` (com cada coluna associada a uma variável).

Como noutros comandos R, o resultado é um objecto de classe `list`, contendo informação vária sobre o resultado da análise.

Nota: Existe também um comando `princomp`, mas por várias razões (incluindo a precisão numérica no caso de matrizes de (co-)variâncias quase singulares), é preferível a utilização do comando `prcomp`.

O comando `prcomp`

```
> lav.acp <- prcomp(lavagantes)
> lav.acp
```

Standard deviations:

```
[1] 4.4171243 2.1583124 0.9617894 0.7071970 0.6163559 0.4992560 0.4639879 <- desvios padrões
[8] 0.3848417 0.3362918 0.2500701 0.2060563 0.1770375 0.1405790 de cada CP
```

Rotation:

	PC1	PC2	PC3	PC4	PC5	
carapace_l	0.28762060	0.36935786	0.08475822	-0.31404094	-0.454639049	<- cada coluna é um vector próprio v_j da matriz de (co)variâncias dos dados. Estes vectores contêm os coeficientes das combinações lineares que definem as CPs.
tail_l	0.10615292	0.61487598	-0.01728674	0.46421995	0.550775374	
carapace_w	0.19089393	0.22112280	0.09978650	-0.10987953	-0.186701149	
carapace_d	0.13951311	0.14784642	0.13138041	0.01598041	0.105009202	
tail_w	0.04682070	0.49290700	-0.05172379	0.06592005	-0.405755003	
areola_l	0.13858508	0.15588574	-0.03136931	-0.78849399	0.514893584	
areola_w	0.02862658	0.02088959	-0.05104427	-0.01123927	-0.005062728	
rostrum_l	0.04321132	0.10238463	-0.00534869	0.10538116	-0.015312405	
rostrum_w	0.06381638	0.06445436	0.05636521	-0.02008425	-0.071806372	
postorbital_w	0.08947075	0.12850014	0.07576734	-0.01777992	0.021872310	
propodus_l	0.70705994	-0.28621233	0.04885310	0.16407517	0.077728529	
propodus_w	0.31334632	-0.14849063	0.69820134	0.07580938	0.026674997	
dactyl_l	0.46456390	-0.10926197	-0.67839228	0.05350023	-0.040805783	
[...]	<- omitidos, por razão de espaço, os restantes vectores de coeficientes.					

Os coeficientes (colunas do objecto `Rotation`) são designados *loadings* em inglês.

Vectores de scores

Não são mostrados por omissão os **scores** de cada individuo numa CP, ou seja, o valor que cada individuo toma na combinação linear $X\vec{v}_j$. Mas esses valores são guardados num **objecto de nome x**, na lista produzida pelo comando `prcomp`:

```
> names(lav.acp)
```

```
[1] "sdev"      "rotation"  "center"    "scale"     "x"
```

```
> lav.acp$x
```

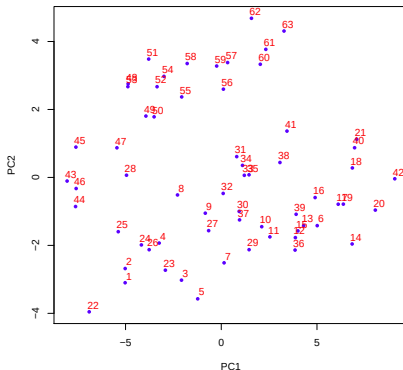
	PC1	PC2	PC3	PC4	PC5	PC6
1	-5.0216041	-3.09975004	-0.93638716	0.590170762	0.34242883	-0.311295721
2	-5.0199046	-2.68138921	1.93090666	0.652936303	0.71306147	2.411219117
3	-2.0772687	-3.02373521	-0.44934354	0.613510708	0.54941375	-0.365822245
[...]						
62	1.5767872	4.68339718	-0.49231884	0.246787192	-0.11313707	0.138658304
63	3.2782407	4.30830749	0.15373020	-0.562657698	-0.73379507	0.200035217
[...]						

São estas as coordenadas nas **representações gráficas a baixa dimensão** que melhor preservam a variabilidade dos dados.

A melhor representação bidimensional

```
> plot(lav.acp$x[,1:2],col="blue",pch=16,cex=0.8)
```

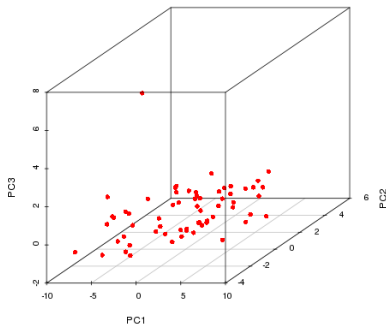
```
> text(lav.acp$x[,1:2]+0.2,label=rownames(lavagantes),col="red")
```



Os indivíduos 43 a 63 são fêmeas, o resto machos. A ACP não usou essa informação: usou apenas o seu reflexo na variabilidade das variáveis morfométricas.

A melhor representação tridimensional

Pode ser obtida pelas funções do módulo `car` (embora o gráfico abaixo seja obtido com o comando `scatterplot3d`, no módulo do mesmo nome):

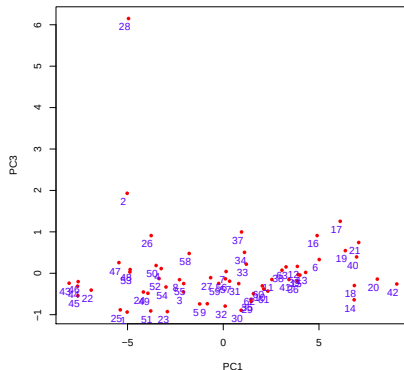


No gráfico verifica-se que a terceira CP separa uma observação discrepante (*outlier*) das restantes. Acontece com alguma frequência em CPs posteriores.

A observação 28 e a terceira CP

```
> plot(lav.acp$x[,c(1,3)],col="red", pch=16, cex=0.8)
```

```
> text(lav.acp$x[,c(1,3)]-0.2, label=rownames(lavagantes), col="blue")
```



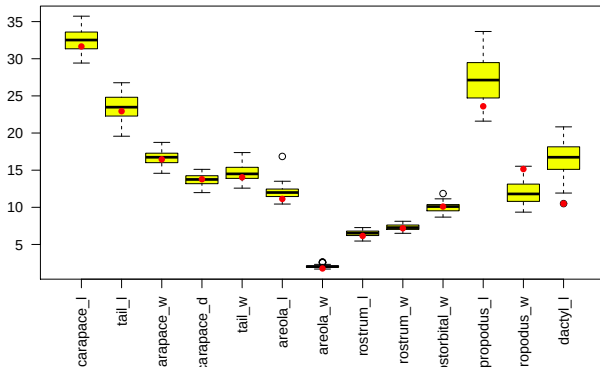
O individuo 28 contribui decisivamente para a terceira direcção ortogonal de maior variabilidade.

Revisitando o individuo 28

Vamos procurar perceber o que tem de diferente o individuo 28.

```
> boxplot(lavagantes, col="yellow", las=2)
```

```
> points(1:13,lavagantes[28,], pch=16, col="red")
```



Decomposição em valores/vectores próprios

A informação acima indicada poderia ser obtida através da decomposição espectral da matriz de (co-)variâncias dos dados, utilizando o comando **eigen**:

> eigen(var(lavagantes))

\$values

```
[1] 19.51098705  4.65831240  0.92503887  0.50012760  0.37989465  0.24925657
[7]  0.21528474  0.14810313  0.11309220  0.06253506  0.04245919  0.03134228
[13]  0.01976246
```

\$vectors

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	-0.28762060	-0.36935786	-0.08475822	0.31404094	-0.454639049	0.272071976
[2,]	-0.10615292	-0.61487598	0.01728674	-0.46421995	0.550775374	0.088028646
[3,]	-0.19089393	-0.22112280	-0.09978650	0.10987953	-0.186701149	-0.178125878
[4,]	-0.13951311	-0.14784642	-0.13138041	-0.01598041	0.105009202	-0.171612241
[5,]	-0.04682070	-0.49290700	0.05172379	-0.06592005	-0.405755003	-0.046182873
[6,]	-0.13858508	-0.15588574	0.03136931	0.78849399	0.514893584	-0.004876079
[7,]	-0.02862658	-0.02088959	0.05104427	0.01123927	-0.005062728	0.026873555
[8,]	-0.04321132	-0.10238463	0.00534869	-0.10538116	-0.015312405	-0.029408152
[9,]	-0.06381638	-0.06445436	-0.05636521	0.02008425	-0.071806372	0.007891374
[10,]	-0.08947075	-0.12850014	-0.07576734	0.01777992	0.021872310	-0.276900583
[11,]	-0.70705994	0.28621233	-0.04885310	-0.16407517	0.077728529	0.541197594
[12,]	-0.31334632	0.14849063	-0.69820134	-0.07580938	0.026674997	-0.476061633
[13,]	-0.46456390	0.10926197	0.67839228	-0.05350023	-0.040805783	-0.506989966
[...]						

Decomposição em valores/vectores próprios (cont.)

```
> sqrt(eigen(var(lavagantes))$val)
```

```
[1] 4.4171243 2.1583124 0.9617894 0.7071970 0.6163559 0.4992560 0.4639879  
[8] 0.3848417 0.3362918 0.2500701 0.2060563 0.1770375 0.1405790
```

```
> summary(lav.acp)$sdev
```

```
[1] 4.4171243 2.1583124 0.9617894 0.7071970 0.6163559 0.4992560 0.4639879  
[8] 0.3848417 0.3362918 0.2500701 0.2060563 0.1770375 0.1405790
```

```
> eigen(var(lavagantes))$vec
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	-0.28762060	-0.36935786	-0.08475822	-0.31404094	-0.454639049	-0.272071976
[2,]	-0.10615292	-0.61487598	0.01728674	0.46421995	0.550775374	-0.088028646
[3,]	-0.19089393	-0.22112280	-0.09978650	-0.10987953	-0.186701149	0.178125878
[4,]	-0.13951311	-0.14784642	-0.13138041	0.01598041	0.105009202	0.171612241
[...]						

```
> summary(lav.acp)$rot
```

	PC1	PC2	PC3	PC4	PC5
carapace_l	0.28762060	0.36935786	0.08475822	-0.31404094	-0.454639049
tail_l	0.10615292	0.61487598	-0.01728674	0.46421995	0.550775374
carapace_w	0.19089393	0.22112280	0.09978650	-0.10987953	-0.186701149
carapace_d	0.13951311	0.14784642	0.13138041	0.01598041	0.105009202
[...]					

Nota: Repare-se como alguns vectores próprios **diferem num factor de -1** .

Propriedades de CPs

- A soma das variâncias (inércia) das p componentes principais é igual à soma das variâncias das p variáveis originais.
- Logo, pode afirmar-se que a j -ésima CP explica uma proporção da variabilidade (inércia) total igual a $\pi_j = \frac{\lambda_j}{\sum_{i=1}^p \lambda_i}$.
- Esta medida é extensível a subconjuntos de componentes principais. Às primeiras q CPs corresponde

$$\sum_{i=1}^q \pi_i \times 100\% = \frac{\sum_{i=1}^q \lambda_i}{\sum_{j=1}^p \lambda_j} \times 100\%$$

da variabilidade total (inércia) do conjunto de dados.

Propriedades de CPs (cont.)

- A correlação entre a i -ésima variável $\vec{\mathbf{x}}_i$ e a j -ésima CP $\mathbf{X}\vec{\mathbf{v}}_j$ é:

$$\text{cor}(\vec{\mathbf{x}}_i, \mathbf{X}\vec{\mathbf{v}}_j) = \sqrt{\lambda_j} \cdot \frac{v_{ij}}{s_i}$$

- s_i — desvio padrão da variável $\vec{\mathbf{x}}_i$
- v_{ij} — coeficiente (*loading*) de $\vec{\mathbf{x}}_i$ na CP j
- λ_j — variância da j -ésima CP

Nota: A covariância entre duas combinações lineares das colunas de $\mathbf{X}$, como o são $\mathbf{X}\vec{\mathbf{v}}_j$ e $\vec{\mathbf{x}}_i = \mathbf{X}\vec{\mathbf{e}}_i$, (com $\vec{\mathbf{e}}_i$ o i -ésimo vector da base canónica de $\mathbb{R}^p$) é dada por $\vec{\mathbf{e}}_i^t \mathbf{S} \vec{\mathbf{v}}_j$, onde $\mathbf{S}$ é a matriz de (co-)variâncias dos dados. Logo,

$$\text{cor}(\vec{\mathbf{x}}_i, \mathbf{X}\vec{\mathbf{v}}_j) = \frac{\vec{\mathbf{e}}_i^t \mathbf{S} \vec{\mathbf{v}}_j}{s_i \cdot \sqrt{\lambda_j}} = \frac{\lambda_j \vec{\mathbf{e}}_i^t \vec{\mathbf{v}}_j}{s_i \cdot \sqrt{\lambda_j}} = \sqrt{\lambda_j} \frac{v_{ij}}{s_i} .$$

O comando `summary`

```
> summary(prcomp(lavagantes))
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Std. Dev.	4.417	2.158	0.9618	0.7072	0.6164	0.49926	0.46399	0.38484	0.33629	0.25007	0.20606	0.17704	0.14058
Prop.Var.	0.727	0.173	0.0344	0.0186	0.0141	0.00928	0.00802	0.00551	0.00421	0.00233	0.00158	0.00117	0.00074
Cum.Prop.	0.727	0.900	0.9344	0.9530	0.9672	0.97645	0.98446	0.98998	0.99419	0.99652	0.99810	0.99926	1.00000

Na recta definida pela **primeira componente principal** preservamos **72,7%** da variabilidade total dos dados.

No plano definido pelas **duas primeiras componentes principais** preservamos **90,0%** da variabilidade total dos dados.

No espaço a 3 dimensões definido pelas **três primeiras CPs** preservamos **93,4%** da variabilidade total.

Apenas não visualizamos na representação tri-dimensional cerca de **6,6%** da variabilidade total.

Interpretação de CPs

É tentador procurar interpretar o significado de CPs (que não são variáveis directamente observáveis).

As correlações entre variáveis originais e CPs são úteis nesse sentido. Pode-se usar a fórmula dada acima, ou calculá-las através de:

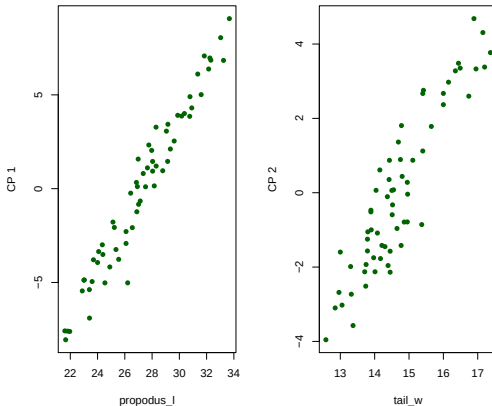
```
> round(cor(lavagantes, lav.acp$x),d=2)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
carapace_l	0.81	0.51	0.05	-0.14	-0.18	0.09	-0.15	0.06	-0.03	0.02	-0.02	0.01
tail_l	0.31	0.89	-0.01	0.22	0.23	0.03	-0.04	0.06	0.02	-0.01	-0.01	0.00
carapace_w	0.83	0.47	0.09	-0.08	-0.11	-0.09	0.02	-0.02	0.09	-0.14	0.12	0.05
carapace_d	0.78	0.41	0.16	0.01	0.08	-0.11	-0.06	-0.25	-0.33	-0.03	0.03	-0.01
tail_w	0.18	0.91	-0.04	0.04	-0.21	-0.02	0.28	-0.04	0.00	0.02	-0.04	-0.01
areola_l	0.64	0.35	-0.03	-0.58	0.33	0.00	0.11	0.01	0.01	0.03	0.00	0.00
areola_w	0.60	0.21	-0.23	-0.04	-0.01	0.06	0.10	0.09	-0.03	-0.14	0.08	-0.37
rostrum_l	0.50	0.58	-0.01	0.20	-0.02	-0.04	0.03	0.03	0.01	0.49	0.35	0.04
rostrum_w	0.76	0.38	0.15	-0.04	-0.12	0.01	-0.13	-0.03	0.12	-0.02	0.12	-0.40
postorbital_w	0.65	0.45	0.12	-0.02	0.02	-0.23	-0.23	-0.40	0.29	0.08	-0.09	0.01
propodus_l	0.98	-0.19	0.01	0.04	0.01	0.08	0.03	-0.03	0.01	0.00	0.00	0.00
propodus_w	0.87	-0.20	0.42	0.03	0.01	-0.15	0.03	0.08	0.00	0.01	-0.02	0.00
dactyl_l	0.94	-0.11	-0.30	0.02	-0.01	-0.12	-0.01	0.03	-0.01	0.00	-0.01	0.00

A primeira CP está muito fortemente correlacionada com as medições da tenaz, em particular propodus_l. A segunda CP está fortemente correlacionada com as medições da cauda, em particular tail_w.

Correlações entre CPs e variáveis

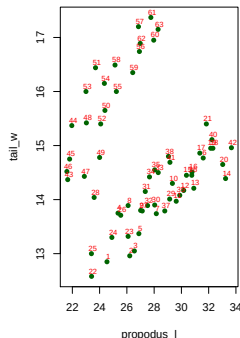
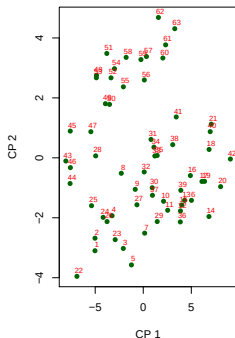
```
> par(mfrow=c(1,2))      <- cria uma "matriz 1x2 de gráficos"  
> plot(lavagantes[,11], lav.acp$x[,1], xlab="propodus_l", ylab="CP 1", pch=16, col="darkgreen")  
> plot(lavagantes[,5], lav.acp$x[,2], xlab="tail_w", ylab="CP 2", pch=16, col="darkgreen")  
> par(mfrow=c(1,1))      <- re-estabelece a situação original
```



Correlações entre CPs e variáveis (cont.)

As fortes correlações sugerem uma análise da nuvem de pontos definida pelas duas variáveis originais:

```
> plot(lav.acp$x[,1:2], xlab="CP 1", ylab="CP 2", pch=16, col="darkgreen")  
> text(lav.acp$x[,1:2]+0.2, label=rownames(lavagantes), col="red", cex=0.7)  
> plot(lavagantes[,c(11,5)], xlab="propodus_l", ylab="tail_w", pch=16, col="darkgreen")  
> text(lavagantes[,c(11,5)]+0.1, label=rownames(lavagantes), col="red", cex=0.7)
```



ACP sobre a matriz de correlações

Uma característica pouco simpática da ACP (que a distingue, por exemplo, da regressão linear), é que os resultados duma ACP mudam se houver mudanças de escala diferenciadas nas várias variáveis.

Esta sensibilidade da ACP é natural, dadas as características do critério que se pretende otimizar: a variância.

Para tornar este problema, e sendo a generalidade das mudanças de escala transformações lineares, é hábito normalizar os dados antes de efectuar uma ACP:

$$x_{ij} \longrightarrow z_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{s_j},$$

onde

- x_{ij} é a observação do individuo i na variável j ;
- $\bar{x}_{.j}$ é a média das n observações na variável j ;
- s_j é o desvio padrão das n observações na variável j ;
- z_{ij} é a observação normalizada do individuo i na variável j .

ACP sobre a matriz de correlações (cont.)

A matriz de (co)variâncias dos dados normalizados é a **matriz de correlações $\mathbf{R}$ dos dados originais** (ou normalizados). Assim, uma ACP sobre os dados normalizados é conhecida por **ACP sobre a matriz de correlação**.

Numa ACP sobre a matriz de correlação,

- as Componentes Principais são **combinações lineares dos dados normalizados**;
- Os vectores dos coeficientes (*loadings*) das comb. lineares são os sucessivos **vectores próprios da matriz de correlações $\mathbf{R}$** ;
- As variâncias de sucessivas CPs são dadas pelos **valores próprios de $\mathbf{R}$** , cuja soma tem de ser $\text{tr}(\mathbf{R}) = p$.

ACP sobre a matriz de correlações no R

No R, há duas formas alternativas de efectuar esta Análise:

```
> prcomp(scale(lavagantes))  
> prcomp(lavagantes,scale=TRUE)
```

Standard deviations:

```
[1] 2.8298571 1.4518966 0.8481395 0.7315674 0.6117634 0.5371346 0.5119344 0.4730480 0.4106900  
[10] 0.3761469 0.3016251 0.2178130 0.1793918
```

Rotation:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
carapace_l	0.3336487	-0.051654918	0.002147496	-0.05337901	0.05903158	-0.25593010	0.13991163
tail_l	0.2328489	-0.455025510	-0.004432513	0.02919494	-0.06389168	0.06642917	-0.32471231
carapace_w	0.3399357	-0.026168964	0.042817387	-0.05649310	0.11876996	-0.18817081	0.02954496
carapace_d	0.3161771	-0.001543245	0.174339992	-0.06927295	-0.01269919	0.02103474	-0.65346959
tail_w	0.1963703	-0.522307992	-0.097172600	0.02943249	0.06817824	-0.29897195	-0.06638706
areola_l	0.2625765	0.014998718	-0.203444780	-0.78727388	-0.41920392	0.00605338	0.19498049
areola_w	0.2320279	0.063340777	-0.813027317	0.19646231	0.26234962	0.17992496	-0.10423247
rostrum_l	0.2559610	-0.260192772	0.122258123	0.50436942	-0.58565962	0.13677260	0.30765231
rostrum_w	0.3122279	0.011301755	0.084409773	0.06116672	0.43328915	-0.24980467	0.49425052
postorbital_w	0.2883485	-0.080276403	0.361940139	-0.14548391	0.36223013	0.71927271	0.11234877
propodus_l	0.2741268	0.405235606	0.006549232	0.13377738	-0.13020525	-0.02606551	-0.05259459
propodus_w	0.2474141	0.398376708	0.281998129	0.09065523	0.00717611	-0.33417966	-0.19598386
dactyl_l	0.2740158	0.339649079	-0.152524450	0.15373369	-0.22361974	0.24824297	0.03271971
[...]							

As duas variantes de ACP

Os resultados duma e outra ACP não são directamente comparáveis.

```
> lav.acpR <- prcomp(lavagantes,scale=TRUE)
> summary(lav.acpR)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Std.dev	2.830	1.4519	0.84814	0.73157	0.61176	0.53713	0.51193	0.47305	0.41069	0.37615	0.3016	0.21781	0.17939
Prp.Var	0.616	0.1621	0.05533	0.04117	0.02879	0.02219	0.02016	0.01721	0.01297	0.01088	0.0070	0.00365	0.00248
Cum.Prp	0.616	0.7782	0.83350	0.87466	0.90345	0.92565	0.94581	0.96302	0.97599	0.98688	0.9939	0.99752	1.00000

```
> summary(lav.acp)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Std.dev	4.4171	2.1583	0.96179	0.70720	0.61636	0.49926	0.46399	0.38484	0.33629	0.25007	0.20606	0.17704	0.14058
Prp.Var	0.7265	0.1734	0.03444	0.01862	0.01415	0.00928	0.00802	0.00551	0.00421	0.00233	0.00158	0.00117	0.00074
Cum.Prp	0.7265	0.9000	0.93440	0.95302	0.96716	0.97645	0.98446	0.98998	0.99419	0.99652	0.99810	0.99926	1.00000

Em geral, numa ACP sobre a matriz de correlações são precisas mais CPs para alcançar uma mesma proporção da inércia total explicada.

As duas variantes de ACP (cont.)

Mudam também os vectores de *loadings* (vectores próprios de **S** e **R** são diferentes), bem como os vectores de *scores* a que dão origem.

Vejamos as correlações entre os dois tipos de CPs:

> round(cor(lav.acp\$x, lav.acpR\$x), d=2)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
PC1	0.89	0.44	0.00	0.06	-0.07	-0.01	-0.02	-0.09	-0.04	0.04	0.01	0.04	0.03
PC2	0.44	-0.88	-0.03	-0.10	0.05	-0.05	-0.04	-0.01	-0.04	0.03	-0.05	-0.02	-0.04
PC3	0.05	0.05	0.53	-0.09	0.24	-0.42	-0.16	0.56	0.27	0.18	-0.13	0.07	0.07
PC4	-0.04	-0.10	0.19	0.79	0.09	0.11	-0.36	-0.05	-0.25	0.26	0.17	0.06	0.08
PC5	0.00	0.02	-0.03	-0.38	-0.37	0.34	-0.28	0.45	-0.42	0.31	0.21	-0.01	0.08
PC6	-0.05	-0.03	-0.21	0.02	-0.05	-0.26	0.11	0.01	-0.31	-0.05	-0.33	0.48	0.66
PC7	-0.02	-0.06	-0.26	-0.01	-0.26	-0.33	-0.12	-0.10	0.45	0.18	0.64	0.16	0.21
PC8	-0.05	0.04	-0.28	0.14	-0.27	-0.53	0.17	0.04	-0.22	0.45	-0.24	-0.15	-0.44
PC9	0.01	-0.02	0.10	-0.04	0.25	0.29	0.61	-0.08	0.10	0.64	0.08	-0.01	0.21
PC10	0.02	-0.10	0.22	0.27	-0.54	0.21	0.37	0.25	0.20	-0.17	-0.05	0.46	-0.24
PC11	0.05	-0.04	-0.04	0.27	-0.24	-0.05	0.26	0.32	0.04	-0.23	0.03	-0.68	0.40
PC12	-0.07	-0.03	0.33	-0.11	-0.47	0.07	-0.27	-0.45	0.26	0.23	-0.42	-0.18	0.21
PC13	-0.03	-0.02	0.56	-0.16	-0.13	-0.32	0.25	-0.30	-0.46	-0.15	0.38	-0.03	0.00

As duas variantes de ACP (cont.)

Vejam as correlações entre as CPs sobre os dados normalizados e as variáveis originais:

```
> round(cor(lavagantes, lav.acpR$x), d=2)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
carapace_l	0.94	-0.07	0.00	-0.04	0.04	-0.14	0.07	-0.11	-0.03	-0.06	-0.24	0.05	-0.04
tail_l	0.66	-0.66	0.00	0.02	-0.04	0.04	-0.17	0.05	-0.24	0.19	-0.01	0.00	0.00
carapace_w	0.96	-0.04	0.04	-0.04	0.07	-0.10	0.02	-0.10	0.08	0.07	-0.03	-0.16	0.06
carapace_d	0.89	0.00	0.15	-0.05	-0.01	0.01	-0.33	0.07	-0.02	-0.24	0.02	-0.02	0.00
tail_w	0.56	-0.76	-0.08	0.02	0.04	-0.16	-0.03	-0.18	0.17	0.00	0.12	0.07	-0.01
areola_l	0.74	0.02	-0.17	-0.58	-0.26	0.00	0.10	0.09	0.02	0.01	0.04	0.01	0.00
areola_w	0.66	0.09	-0.69	0.14	0.16	0.10	-0.05	0.14	0.08	0.01	-0.02	0.00	0.00
rostrum_l	0.72	-0.38	0.10	0.37	-0.36	0.07	0.16	0.14	0.07	-0.05	-0.01	-0.01	0.01
rostrum_w	0.88	0.02	0.07	0.04	0.27	-0.13	0.25	0.12	-0.16	-0.09	0.10	0.00	-0.01
postorbital_w	0.82	-0.12	0.31	-0.11	0.22	0.39	0.06	-0.01	0.10	0.04	-0.01	0.03	0.00
propodus_l	0.78	0.59	0.01	0.10	-0.08	-0.01	-0.03	-0.08	-0.05	0.04	0.02	0.10	0.12
propodus_w	0.70	0.58	0.24	0.07	0.00	-0.18	-0.10	0.16	0.12	0.16	0.02	0.02	-0.07
dactyl_l	0.78	0.49	-0.13	0.11	-0.14	0.13	0.02	-0.26	-0.09	-0.01	0.07	-0.03	-0.08

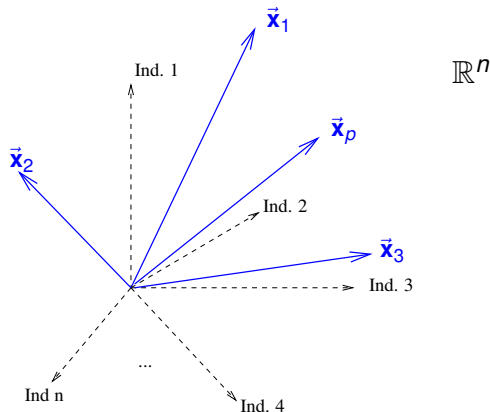
Notas:

- Em relação à ACP sobre dados originais, não só mudam as correlações entre CPs e variáveis, como também as interpretações possíveis.
- CP1 é agora essencialmente uma medida da **dimensão geral do animal**.
- CP2, mais difícil de interpretar, mas **contrasta dimensões de caudas e tenazes**.

A representação em $\mathbb{R}^n$, o espaço das variáveis

Representação alternativa dum matriz de dados $\mathbf{X}$, no espaço das variáveis:

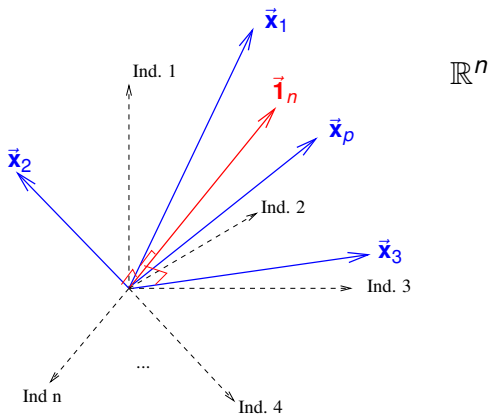
- cada eixo corresponde a um indivíduo observado;
- cada vector corresponde a uma variável.



Variáveis centradas no espaço das variáveis

A representação mais interessante no espaço das variáveis é a das **variáveis centradas**, porque os conceitos geométricos introduzidos pelo habitual produto interno em $\mathbb{R}^n$ têm interpretações estatísticas.

Centrar as colunas de $\mathbf{X}$ corresponde a tornar os vectores que representam as variáveis ortogonais ao vector $\vec{\mathbf{1}}_n$ dos n uns:



Geometria e estatística no espaço das variáveis

O elemento genérico da matriz centrada dos dados, $\mathbf{X}^c$, é:

$$x_{ij}^c = x_{ij} - \bar{x}_{.j},$$

- x_{ij} indica a observação do i -ésimo indivíduo na variável j ;
- $\bar{x}_{.j}$ indica a média das n observações na variável j .

Logo,

- a **norma usual** duma coluna de $\mathbf{X}^c$ é proporcional ao **desvio padrão** dessa variável: $\|\vec{\mathbf{x}}_j^c\| = \sqrt{n-1} s_j$.
- o **produto interno** usual de duas diferentes colunas de $\mathbf{X}^c$ é proporcional à **covariância** das respectivas variáveis: $\langle \vec{\mathbf{x}}_j^c, \vec{\mathbf{x}}_k^c \rangle = (n-1) \text{cov}_{i,j}$.
- o **cosseno do ângulo** entre os vectores representando duas diferentes colunas de $\mathbf{X}^c$ é o **coeficiente de correlação** das respectivas variáveis:
$$\cos \theta = \frac{\langle \vec{\mathbf{x}}_j^c, \vec{\mathbf{x}}_k^c \rangle}{\|\vec{\mathbf{x}}_j^c\| \cdot \|\vec{\mathbf{x}}_k^c\|} = \frac{\text{cov}_{i,j}}{s_i \cdot s_j} = r_{i,j}$$
- Vectores centrados **ortogonais** correspondem a variáveis **não correlacionadas**.

Interpretação de CPs no espaço das variáveis

A representação dos dados no espaço das variáveis ($\mathbb{R}^n$) associa cada variável a um vector. Combinações lineares de variáveis são combinações lineares de vectores, logo novos vectores.

Para vectores centrados, o quadrado do comprimento do vector é proporcional à variância da variável respectiva.

O critério da ACP corresponde a procurar combinações lineares dos vectores de comprimento máximo (com soma 1 de quadrados dos coeficientes).

É geometricamente intuitivo que variáveis com variância muito maior que as restantes tenham grande influência na definição da primeira CP (“dominam a primeira CP”).

Interpretação de CPs em $\mathbb{R}^n$ (cont.)

Transformações lineares (afins) nas variáveis correspondem a re-escalar os vectores que as representam (mantendo a direcção):

- constantes aditivas desaparecem na centragem, logo não alteram o vector correspondente em $\mathbb{R}^n$.
- constantes multiplicativas c correspondem a multiplicar o vector correspondente por c , logo:
 - ▶ preservam a direcção do vector representativo;
 - ▶ mudam o sentido se $c < 0$;
 - ▶ esticam o vector se $|c| > 1$;
 - ▶ encolhem o vector se $|c| < 1$.

A normalização dos dados associada à ACP sobre a matriz de correlações corresponde a tornar todos os vectores, representativos das variáveis centradas, do mesmo tamanho.

Logo, o critério da ACP é sensível a mudanças de escalas diferenciadas nas p variáveis.

A matriz centrada, $\mathbf{X}^c$, na representação em $\mathbb{R}^p$

Qual o efeito de centrar a matriz $\mathbf{X}$ na nuvem de pontos associada aos dados, em $\mathbb{R}^p$?

A transformação de $\mathbf{X}$ em $\mathbf{X}^c$ apenas altera a média de cada variável, que passa a ser zero. Geometricamente, o centro de gravidade da nuvem de n pontos em $\mathbb{R}^p$ passa a ser a origem, ou seja, há uma translação do centro de gravidade:

$$(\overline{x_1}, \overline{x_2}, \dots, \overline{x_p}) \longrightarrow (0, 0, \dots, 0) .$$

É habitual fazer a representação das CPs em $\mathbb{R}^p$ com esta centragem (ou seja, a nuvem dos scores ter centro de gravidade na origem), que corresponde a considerar a combinação linear das variáveis centradas (com os vectores de *loadings* usuais).

Advertências sobre ACPs (em geral)

- A redução da dimensionalidade associada à ACP **não significa redução no número de variáveis originais** com que se trabalha: cada CP é combinação linear de **todas** as variáveis observadas.
- É frequente procurar **interpretar** cada CP ignorando as variáveis cujos coeficientes (*loadings*) na combinação linear que define a CP são “próximos de zero”. Isto **pode induzir em erro**, e convém **utilizar informação complementar para validar as interpretações baseadas nos coeficientes**.
- Outra prática frequente, mas discutível, em ACP é a **rotação** das CPs: modificam-se os coeficientes da combinação linear para os aproximar de zero ou um, visando “simplificar a interpretação”. Mas esse objectivo pode ser **ilusório** (como vimos) e **sacrifica a optimalidade** das soluções.
- Alguns autores também chamam **componentes principais** aos vectores próprios de **$\mathbf{S}$** (vectores de *loadings*), gerando confusão.
- **Não** faz sentido que qualquer das variáveis originais seja uma variável **qualitativa (categórica)**.

Ainda a ACP sobre matriz de correlações

Em termos geométricos,

- Em $\mathbb{R}^n$, cada um dos p vectores é re-dimensionado e fica com comprimento (norma) comum.
- Em $\mathbb{R}^p$, cada eixo é esticado ou encolhido, com factores de alteração das escalas diferenciados para cada eixo.
Muda a forma da nuvem de pontos.

Observações:

- A variabilidade total é $\text{tr}(\mathbf{R}) = p$ (o número de variáveis).
- A correlação entre a variável $\tilde{\mathbf{x}}_i$ e a j -ésima CP é agora $\sqrt{\lambda_j} v_{ij}$.
- Por vezes, os coeficientes das componentes numa ACP sobre a matriz de correlações são re-escalados de forma a que $\tilde{\mathbf{v}}_j^t \tilde{\mathbf{v}}_j = \lambda_j$. Nesse caso, os coeficientes da combinação linear são as correlações entre a variável e a CP.

Outro critério

As CPs da matriz de correlações são também solução óptima de outro problema: determinar a combinação linear que maximiza a soma de quadrados das p correlações com cada variável original.

A soma de quadrados das correlações entre esta nova variável e as p variáveis originais é o maior valor próprio λ_1 , associado ao vector próprio $\vec{v}_1$, da matriz de correlações $\mathbf{R}$.

As restantes CPs da matriz de correlações são as combinações lineares que sucessivamente maximizam o quociente, sujeitas às restrições de correlação nula com todas as soluções anteriores.

O critério é invariante a mudanças lineares de escala nas variáveis, porque envolve coeficientes de correlação.

O critério alternativo (cont.)

Interpretação geométrica do novo critério, em $\mathbb{R}^n$: procuram-se as combinações lineares das variáveis que maximizam a soma dos quadrados dos cossenos dos ângulos com os vectores das variáveis originais. Como transformações lineares (afins) das variáveis não alteram esses ângulos, o critério fica invariante.

As combinações lineares centradas das variáveis reduzidas $\mathbf{Z}\vec{v}$ (onde $\vec{v}$ é vector próprio de $\mathbf{R}$):

- só são componentes principais das variáveis reduzidas;
- mas são as combinações lineares sucessivamente “globalmente mais correlacionadas com as variáveis originais”, independentemente das unidades de medida originais.

Matrizes da forma $\mathbf{X}^t\mathbf{X}$

Seja $\mathbf{X}$ uma matriz $n \times p$ de dados e $\mathbf{X}^c$ a correspondente matriz de colunas centradas. A matriz de (co)variâncias é $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^{ct}\mathbf{X}^c$.

As matrizes de correlações são análogas, mas partindo de matrizes $\mathbf{Z}$ de dados centrados e reduzidos: $\mathbf{R} = \frac{1}{n-1}\mathbf{Z}^t\mathbf{Z}$.

Além de simétricas, as matrizes de covariâncias (e de correlações) são necessariamente semi-definidas positivas. São definidas positivas (isto é, sem valores próprios nulos) se e só se as p colunas da matriz $\mathbf{X}^c$ (ou $\mathbf{Z}$) são linearmente independentes (i.e., se e só se não há multicolinearidade).

Decomposição em Valores Singulares

Teorema

Seja $\mathbf{X}$ uma matriz $n \times p$. Então, é sempre possível decompôr $\mathbf{X}$ da seguinte forma:

$$\mathbf{X} = \mathbf{W}\mathbf{\Delta}\mathbf{V}^t \iff \mathbf{X} = \sum_{i=1}^r \delta_i \vec{\mathbf{w}}_i \vec{\mathbf{v}}_i^t,$$

onde

r é a **característica** da matriz, com $r \leq \min\{n, p\}$

$\mathbf{\Delta}_{r \times r}$ matriz diagonal

$\mathbf{V}_{p \times r}$ matriz com colunas ortonormadas ($\mathbf{V}^t \mathbf{V} = \mathbf{I}_r$)

$\mathbf{W}_{n \times r}$ matriz com colunas ortonormadas ($\mathbf{W}^t \mathbf{W} = \mathbf{I}_r$)

δ_i elementos diagonais de $\mathbf{\Delta}$ (**valores singulares** de $\mathbf{X}$)

$\vec{\mathbf{w}}_i$ colunas de $\mathbf{W}$ (**vectores singulares esquerdos** de $\mathbf{X}$)

$\vec{\mathbf{v}}_i$ colunas de $\mathbf{V}$ (**vectores singulares direitos** de $\mathbf{X}$)

Observações sobre a DVS

- A Decomposição em Valores Singulares (DVS) é válida para qualquer matriz.
- Tem-se $\mathbf{X}^t\mathbf{X} = \mathbf{V}\Delta^2\mathbf{V}^t$, logo $\mathbf{V}$ é matriz cujas colunas são um conjunto ortonormado de vectores próprios de $\mathbf{X}^t\mathbf{X}$, associados a valores próprios não-nulos, dados por δ_i^2 .
- $\mathbf{W}$ é a matriz análoga, de vectores próprios de $\mathbf{X}\mathbf{X}^t$ associados a valores próprios não nulos δ_i^2 .
- Δ é a matriz diagonal das raízes quadradas dos valores próprios não-nulos de $\mathbf{X}^t\mathbf{X}$ (iguais aos de $\mathbf{X}\mathbf{X}^t$). Admite-se que os valores singulares δ_i estão ordenados por ordem decrescente.
- A DVS duma matriz é sempre possível, mas não é única (pelo menos a troca de sinal em cada par de vectores singulares).
- Se $\mathbf{X}$ tem Decomposição em Valores Singulares $\mathbf{X} = \mathbf{W}\Delta\mathbf{V}^t$, $\mathbf{X}^t$ tem Decomposição em Valores Singulares $\mathbf{X}^t = \mathbf{V}\Delta\mathbf{W}^t$.

A DVS e a ACP

A ACP é apenas uma **Decomposição em Valores Singulares** da **matriz centrada dos dados $\mathbf{X}^c$** , a dividir por $\sqrt{n}$ ou $\sqrt{n-1}$, (consoante a convenção usada para criar a matriz de covariâncias):

$$\frac{1}{\sqrt{n-1}} \mathbf{X}^c = \mathbf{U} \mathbf{\Delta} \mathbf{V}^t,$$

onde

$\mathbf{V}$ - matriz cujas **colunas** são os vectores próprios de $\mathbf{S} = \frac{1}{n-1} \mathbf{X}^{ct} \mathbf{X}^c$, ou seja, os **loadings** das CPs.

$\mathbf{\Delta}$ - matriz cujos **elementos diagonais** são as **raízes quadradas dos valores próprios de $\mathbf{S}$** , ou seja, os **desvios padrão das CPs**;

$\mathbf{X}^c \mathbf{V} = \sqrt{n-1} \mathbf{U} \mathbf{\Delta}$ - matriz cujas **colunas** são os **scores centrados** de cada individuo em cada CP.

$\mathbf{U} = \frac{1}{\sqrt{n-1}} \mathbf{X}^c \mathbf{V} \mathbf{\Delta}^{-1}$ - matriz dos **vectores singulares esquerdos**, que corresponde aos **vectores de scores normalizados**.

DVS e ACP (cont.)

Confirmemos, efectuando a DVS da matriz $\frac{1}{\sqrt{n-1}} \mathbf{X}^c$ no R.

O comando `scale` faz **simultaneamente** a **centragem** (subtracção da média) e **divisão pelo desvio padrão** das colunas duma matriz.

Cada uma destas operações é controlada por um argumento, respectivamente `center` e `scale`.

Por omissão, estes argumentos são TRUE. Qualquer das operações pode ser omitida dando ao correspondente argumento o valor `FALSE`.

Assim, a **centragem** (apenas) da matriz de dados pode fazer-se da seguinte forma:

```
> lav.centrado <- scale(lavagantes, scale=FALSE)
```

ACP e DVS (cont.)

No R uma Decomposição em Valores Singulares faz-se através do comando **svd**. Vejamos uma ACP feita com este comando:

```
> nm1 <- dim(lav.centrado)[1]-1  
> svd(lav.centrado/sqrt(nm1))
```

\$d

```
[1] 4.4171243 2.1583124 0.9617894 0.7071970 0.6163559 0.4992560 0.4639879  
[8] 0.3848417 0.3362918 0.2500701 0.2060563 0.1770375 0.1405790
```

\$u

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.144379990	-0.182396510	-0.123645871	0.1059842750	0.070557452
[2,]	-0.144331125	-0.157779185	0.254967864	0.1172558607	0.146926297
[3,]	-0.059725146	-0.177923620	-0.059333869	0.1101757182	0.113206688
[4,]	-0.093246935	-0.113657051	0.014976742	0.0804924915	-0.069971697
[5,]	-0.035380664	-0.210254166	-0.097758921	-0.1206499751	-0.146049537
[...]					

\$v

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.28762060	0.36935786	0.08475822	-0.31404094	-0.454639049	0.272071976
[2,]	0.10615292	0.61487598	-0.01728674	0.46421995	0.550775374	0.088028646
[3,]	0.19089393	0.22112280	0.09978650	-0.10987953	-0.186701149	-0.178125878
[4,]	0.13951311	0.14784642	0.13138041	0.01598041	0.105009202	-0.171612241
[5,]	0.04682070	0.49290700	-0.05172379	0.06592005	-0.405755003	-0.046182873
[...]						

Atenção: As componentes **\$u** e **\$v** são, respectivamente, as **matrizes U e V** referidas acima. Mas a componente **\$d** é um **vector**, com os **elementos diagonais da matriz Δ** .

ACP e DVS (cont.)

O produto das matrizes $\mathbf{U}$ e Δ :

```
> DVS <- svd(lav.centrado/sqrt(nm1))
```

```
> U <- DVS$u
```

```
> D <- diag(DVS$d)
```

```
> U %*% D * sqrt(nm1)
```

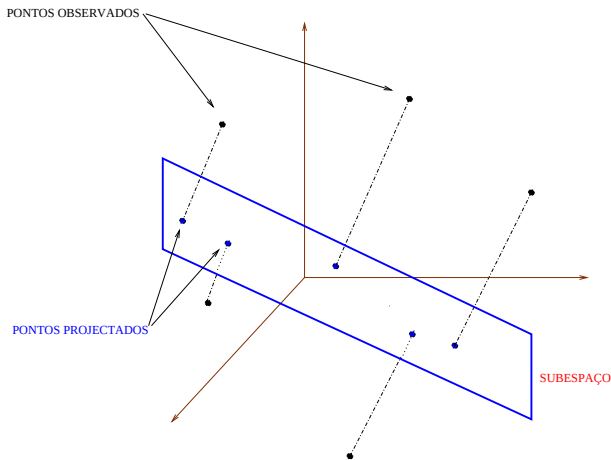
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	-5.0216041	-3.09975004	-0.93638716	0.590170762	0.34242883	-0.311295721
[2,]	-5.0199046	-2.68138921	1.93090666	0.652936303	0.71306147	2.411219117
[3,]	-2.0772687	-3.02373521	-0.44934354	0.613510708	0.54941375	-0.365822245
[...]						
[62,]	1.5767872	4.68339718	-0.49231884	0.246787192	-0.11313707	0.138658304
[63,]	3.2782407	4.30830749	0.15373020	-0.562657698	-0.73379507	0.200035217
[...]						

```
> prcomp(lavagantes)$x
```

	PC1	PC2	PC3	PC4	PC5	PC6
1	-5.0216041	-3.09975004	-0.93638716	0.590170762	0.34242883	-0.311295721
2	-5.0199046	-2.68138921	1.93090666	0.652936303	0.71306147	2.411219117
3	-2.0772687	-3.02373521	-0.44934354	0.613510708	0.54941375	-0.365822245
[...]						
62	1.5767872	4.68339718	-0.49231884	0.246787192	-0.11313707	0.138658304
63	3.2782407	4.30830749	0.15373020	-0.562657698	-0.73379507	0.200035217
[...]						

As projecções em $\mathbb{R}^p$ e em $\mathbb{R}^n$

Para ambas as representações dos dados em $\mathbf{X}^c$, a ACP é a solução do problema de determinar o subespaço de dimensão m tal que a projecção ortogonal dos dados nesse subespaço minimiza a soma de quadrados das distâncias (na perpendicular) entre pontos originais e pontos projectados.



Biplots

- Intimamente relacionada com a Decomposição em Valores Singulares numa matriz centrada de dados (logo, com uma ACP).
- Ideia fundamental do *biplot*: obter uma boa **representação simultânea (aproximada)** dos indivíduos e das variáveis, em baixa dimensão.
- Nessa representação simultânea as **principais características estatísticas** numa ACP são visíveis em termos geométricos.

Biplots (cont.)

- Seja $\mathbf{X}^c$ matriz centrada de dados, com SVD: $\frac{1}{\sqrt{n-1}}\mathbf{X}^c = \mathbf{U}\mathbf{\Delta}\mathbf{V}^t$.
- Definindo:

$$\begin{aligned}\mathbf{G} &= \mathbf{U} \\ \mathbf{H} &= \mathbf{V}\mathbf{\Delta}\end{aligned}$$

tem-se: $\frac{1}{\sqrt{n-1}}\mathbf{X}^c = \mathbf{GH}^t$.

- Se $\mathbf{X}^c$ é de característica r ,
 - ▶ $\mathbf{G}$ é $n \times r$, e existe correspondência entre linhas de $\mathbf{G}$ e indivíduos.
 - ▶ $\mathbf{H}$ é $p \times r$, e existe correspondência entre linhas de $\mathbf{H}$ e variáveis.
- As linhas de $\mathbf{G}$ e $\mathbf{H}$ são marcadores de, respectivamente, indivíduos e variáveis, mas vivem no mesmo espaço ($\mathbb{R}^r$) e podem ser representadas em simultâneo.

Os marcadores de variáveis e de indivíduos

Consideremos as propriedades dos marcadores de variáveis, que são vectores de $\mathbb{R}^r$. Tem-se:

$$\mathbf{H}\mathbf{H}^t = \mathbf{S} \quad (\text{ou } \mathbf{R}).$$

- O produto interno entre cada par de linhas da matriz $\mathbf{H}$ é a **covariância** entre as variáveis correspondentes.
- A **norma** de cada linha da matriz $\mathbf{H}$ é o **desvio padrão** da variável correspondente.
- O **cosseno do ângulo** entre cada par de linhas da matriz $\mathbf{H}$ é o **coeficiente de correlação** entre as variáveis correspondentes.

Quanto aos marcadores de indivíduos:

- A distância euclidiana entre cada par de linhas da matriz $\mathbf{G}$ é a distância de Mahalanobis entre os indivíduos correspondentes:

$$M_{ij} = (\vec{\mathbf{x}}_{[i]} - \vec{\mathbf{x}}_{[j]})^t \mathbf{S}^{-1} (\vec{\mathbf{x}}_{[i]} - \vec{\mathbf{x}}_{[j]}) ,$$

Biplots (cont.)

A visualização dum *biplot* exige que se reduza a representação a um espaço a $k = 2$ (ou $k = 3$) dimensões.

Tal é feito retendo apenas as coordenadas dos marcadores associadas às duas (ou três) primeiras dimensões:

- $\mathbf{G}^{(k)}$ submatriz $n \times k$ com as k primeiras colunas de $\mathbf{G}$.
- $\mathbf{H}^{(k)}$ submatriz $p \times k$ com as k primeiras colunas de $\mathbf{H}$.

As linhas de $\mathbf{G}^{(k)}$ e $\mathbf{H}^{(k)}$ são **marcadores de indivíduos e variáveis** que satisfazem aproximadamente as propriedades acima referidas.

Biplots (cont.)

Tomando $k = 2$, obtemos representação gráfica bidimensional, com

- marcadores de indivíduos representados por pontos; e
- marcadores de variáveis representados por vetores.

Tem-se, aproximadamente:

- o cosseno do ângulo entre marcadores de variáveis é o coeficiente de correlação entre as variáveis;
- o comprimento do marcador de cada variável é proporcional ao seu desvio padrão;
- a distância euclidiana entre marcadores de indivíduos é a distância de Mahalanobis entre esses indivíduos.

A qualidade da aproximação mede-se como na ACP.

Biplots (cont.)

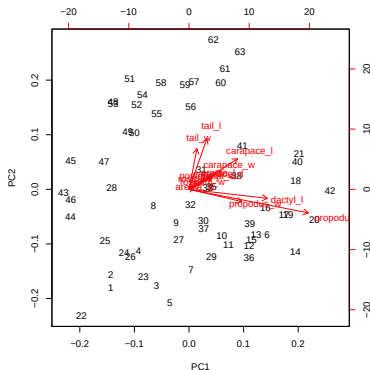
Verificam-se ainda as seguintes propriedades *aproximadas* (por ser a $k=2$ dimensões):

- o *cosseno do ângulo* entre cada vector e o eixo horizontal é aproximadamente o *coeficiente de correlação* entre a respectiva variável e a CP 1;
- o *cosseno do ângulo* entre cada vector e o eixo vertical é aproximadamente o *coeficiente de correlação* entre a respectiva variável e a CP 2;
- A *projectão ortogonal* de cada ponto sobre a direcção definida por um vector é aproximadamente o *valor* do respectivo individuo na variável correspondente.

Os biplots no

Função **biplot**, aplicada ao resultado duma ACP:

```
> biplot(lav.acp)
```

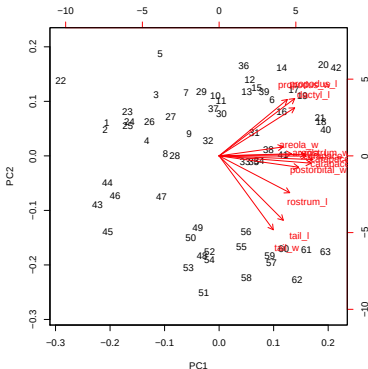


Os indivíduos do grupo 43-63 (fêmeas) têm tenazes mais pequenas e caudas maiores do que os machos.

Os *biplots* no (cont.)

Eis o *biplot* dos dados normalizados:

```
> biplot(lav.acpR)
```



A separação machos/fêmeas continua a ser visível. A primeira CP aponta agora na direcção dum grupo de variáveis altamente correlacionadas entre si (tamanho?).

Análise Discriminante Linear

Análises Discriminantes

Conjunto de técnicas multivariadas em que:

- Parte-se do conhecimento de que n indivíduos observados pertencem a k subgrupos ou classes.
- Procura-se determinar funções das p variáveis observadas que melhor permitam distinguir ou discriminar esses subgrupos.

Análise Discriminante Linear (ou de Fisher)

Procuram-se combinações lineares das p variáveis observadas que melhor permitam distinguir ou discriminar esses subgrupos.

NOTA: Só trabalharemos num contexto descritivo.

Análise Discriminante Linear (cont.)

Ponto de partida: uma matriz de dados $\mathbf{X}_{n \times p}$ e a correspondente matriz centrada $\mathbf{X}^c$.

Os n indivíduos (linhas de $\mathbf{X}$) definem uma partição em k subgrupos, que é conhecida.

Objectivo informal: determinar a melhor combinação linear $\mathbf{X}^c \vec{a}$ das variáveis observadas centradas, tal que:

- indivíduos duma mesma classe tenham valores próximos, e
- indivíduos de classes diferentes tenham valores diferentes.

Soluções: combinações lineares $\mathbf{X}^c \vec{a}$, chamadas eixos discriminantes (ou variáveis canónicas).

ADL (cont.)

Os k subgrupos de indivíduos podem ser associados a k níveis dum factor.

A solução do problema será dada pela combinação linear $\mathbf{X}^c \vec{\mathbf{a}}$ que maximiza a estatística F numa ANOVA sobre esse factor definidor dos k subgrupos.

A solução vai envolver projecções ortogonais sobre o subespaço gerado pelas colunas indicatrizes da constituição de cada subgrupo.

A matriz da classificação

Matriz da classificação **G**, cuja coluna i é indicatriz do subgrupo i :

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ \hline 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \hline 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \dots & 0 \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Nota: Semelhante à matriz do modelo na **ANOVA a um factor**, mas na 1a. coluna indicatriz do nível 1 do factor. O subespaço de $\mathbb{R}^n$ gerado pelas duas matrizes é igual.

A matriz da classificação (cont.)

- As colunas de $\mathbf{G}$ são ortogonais entre si.
- A soma das k colunas de $\mathbf{G}$ é o vector $\vec{\mathbf{1}}_n$, logo $\vec{\mathbf{1}}_n \in \mathcal{C}(\mathbf{G})$.
- O quadrado da norma da coluna j é n_j (no. indivíduos da classe j).
- Os vectores do espaço das colunas da matriz $\mathbf{G}$ têm valor igual nos elementos de cada subgrupo, isto é, $\vec{\mathbf{z}} \in \mathcal{C}(\mathbf{G})$ são da forma:

$$\vec{\mathbf{z}}^t = \left[\underbrace{z_1 \ z_1 \ \dots \ z_1}_{n_1 \text{ vezes}} \mid \underbrace{z_2 \ z_2 \ \dots \ z_2}_{n_2 \text{ vezes}} \mid \dots \mid \underbrace{z_k \ z_k \ \dots \ z_k}_{n_k \text{ vezes}} \right]$$

Logo, vectores em $\mathcal{C}(\mathbf{G})$ são homogéneos no seio das classes.

Mas não necessariamente heterogéneos entre classes: $\mathcal{C}(\mathbf{G})$ inclui também o vector $\vec{\mathbf{1}}_n$, que não discrimina subgrupos.

É desejável que a combinação linear esteja o mais longe possível de $\mathcal{C}(\vec{\mathbf{1}}_n) \subset \mathcal{C}(\mathbf{G})$, ou seja, **que seja ortogonal ao vector $\vec{\mathbf{1}}_n$** .

Formulação do problema

Os vectores ortogonais ao vector $\vec{1}_n$ são os vectores **centrados**.

Consideremos apenas as **combinações lineares centradas**: $\mathbf{X}^c \vec{a}$

A projecção ortogonal de qualquer combinação linear centrada no espaço das colunas da matriz de classificação é dada por $\mathbf{P}_G \mathbf{X}^c \vec{a}$, onde $\mathbf{P}_G = \mathbf{G}(\mathbf{G}^t \mathbf{G})^{-1} \mathbf{G}^t$.

A projecção ortogonal cria um triângulo rectângulo. Pelo Teorema de Pitágoras, tem-se:

$$\begin{aligned}\|\mathbf{X}^c \vec{a}\|^2 &= \|\mathbf{P}_G \mathbf{X}^c \vec{a}\|^2 + \|(\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c \vec{a}\|^2 \\ \vec{a}^t \mathbf{X}^{ct} \mathbf{X}^c \vec{a} &= \vec{a}^t \mathbf{X}^{ct} \mathbf{P}_G \mathbf{X}^c \vec{a} + \vec{a}^t \mathbf{X}^{ct} (\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c \vec{a}\end{aligned}$$

Vamos ver que a combinação linear $\mathbf{X}^c \vec{a}$ **desejável** é a que, nesta decomposição, **maximiza** (em termos relativos) a primeira parcela do lado direito.

Os vectores projectados $\mathbf{P}_G \vec{y}$ e $(\mathbf{I}_n - \mathbf{P}_G) \vec{y}$

Consideremos vectores com dupla indexação i, j , sendo i subgrupo e j repetição.

Se $\vec{y}^c = \mathbf{X}^c \vec{a}$ é um vector **centrado** ($\vec{y} \perp \vec{1}_n$):

$$\mathbf{P}_G \vec{y}^c = \begin{bmatrix} \bar{y}_{1.} - \bar{y}_{..} \\ \vdots \\ \frac{\bar{y}_{1.} - \bar{y}_{..}}{\bar{y}_{2.} - \bar{y}_{..}} \\ \vdots \\ \bar{y}_{2.} - \bar{y}_{..} \\ \vdots \\ \frac{\bar{y}_{k.} - \bar{y}_{..}}{\bar{y}_{k.} - \bar{y}_{..}} \\ \vdots \\ \bar{y}_{k.} - \bar{y}_{..} \end{bmatrix}$$

$$(\mathbf{I}_n - \mathbf{P}_G) \vec{y}^c = \begin{bmatrix} y_{11} - \bar{y}_{1.} \\ \vdots \\ \frac{y_{1n_1} - \bar{y}_{1.}}{y_{21} - \bar{y}_{2.}} \\ \vdots \\ \frac{y_{2n_2} - \bar{y}_{2.}}{\bar{y}_{k.} - \bar{y}_{..}} \\ \vdots \\ \frac{y_{k1} - \bar{y}_{k.}}{\bar{y}_{k.} - \bar{y}_{..}} \\ \vdots \\ y_{kn_k} - \bar{y}_{k.} \end{bmatrix}$$

$\|\mathbf{P}_G \vec{y}^c\|^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_{..})^2 = SQF$ mede dispersão das médias de cada classe em torno da média geral de $\vec{y}$.

É a **variabilidade inter-classes**.

$\|(\mathbf{I}_n - \mathbf{P}_G) \vec{y}^c\|^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{ij})^2 = SQRE$

mede a dispersão das observações individuais em torno da respectiva média de nível.

É a **variabilidade intra-classes**.

De novo a equação resultante de Pitágoras

A equação final do acetato 79 simplifica definindo as matrizes:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^{ct} \mathbf{X}^c$$

$$\mathbf{B} = \frac{1}{n-1} \mathbf{X}^{ct} \mathbf{P}_G \mathbf{X}^c$$

$$\mathbf{W} = \frac{1}{n-1} \mathbf{X}^{ct} (\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c$$

Matriz de variâncias-covariâncias de $\mathbf{X}$

Matriz da variabilidade inter-classes

Matriz da variabilidade intra-classes

Tem-se:

$$\vec{a}^t \underbrace{\mathbf{X}^{ct} \mathbf{X}^c}_{=(n-1) \cdot \mathbf{S}} \vec{a} = \vec{a}^t \underbrace{\mathbf{X}^{ct} \mathbf{P}_G \mathbf{X}^c}_{=(n-1) \cdot \mathbf{B}} \vec{a} + \vec{a}^t \underbrace{\mathbf{X}^{ct} (\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c}_{=(n-1) \cdot \mathbf{W}} \vec{a}$$

$$\iff \vec{a}^t \mathbf{S} \vec{a} = \vec{a}^t \mathbf{B} \vec{a} + \vec{a}^t \mathbf{W} \vec{a}$$

Quer-se a maximização relativa da primeira parcela do lado direito.

Formulação de Fisher

Uma formulação do problema (de Fisher): de entre as combinações lineares $\mathbf{X}^c \vec{\mathbf{a}}$, escolher a que maximiza:

$$\frac{\vec{\mathbf{a}}^t \mathbf{B} \vec{\mathbf{a}}}{\vec{\mathbf{a}}^t \mathbf{W} \vec{\mathbf{a}}}$$

Essa será a primeira função discriminante.

Do ponto de vista de matrizes, é o problema de valores/vectores próprios generalizados.

Solução: Se $\mathbf{W}$ for definida positiva, tomar $\vec{\mathbf{a}} = \vec{\mathbf{a}}_1$, o vector próprio do maior valor próprio de $\mathbf{W}^{-1} \mathbf{B}$.

O valor próprio $\lambda_1 = \frac{\vec{\mathbf{a}}_1^t \mathbf{B} \vec{\mathbf{a}}_1}{\vec{\mathbf{a}}_1^t \mathbf{W} \vec{\mathbf{a}}_1}$ é a capacidade discriminante do eixo: razão entre as variabilidades entre- e intra-grupos nesse eixo.

Formulação de Fisher (cont.)

Se houver mais valores próprios não nulos de $\mathbf{W}^{-1}\mathbf{B}$, podem procurar-se novas combinações lineares discriminantes.

Sucessivas soluções são as combinações lineares $\mathbf{X}\vec{\mathbf{a}}$ de capacidade discriminante máxima, sujeitas à exigência de não correlação com soluções anteriores.

Os vectores de coeficientes $\vec{\mathbf{a}}_j$ são os restantes vectores próprios da matriz $\mathbf{W}^{-1}\mathbf{B}$ associados a valores próprios não-nulos. A capacidade discriminante destes novos eixos é dada pelo valor próprio λ_j associado.

Contrariamente à ACP, eixos discriminantes $\mathbf{X}\vec{\mathbf{a}}_j$ não-correlacionados não significa vectores de coeficientes $\vec{\mathbf{a}}_j$ ortogonais em $\mathbb{R}^p$: os vectores próprios de $\mathbf{W}^{-1}\mathbf{B}$ são $\mathbf{W}$ -ortogonais.

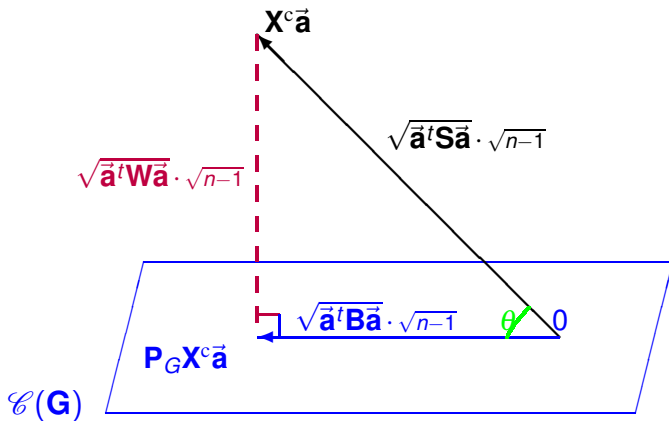
Observações

- As matrizes usadas na ADL verificam a relação $\mathbf{S} = \mathbf{B} + \mathbf{W}$.
De facto,

$$\begin{aligned}\mathbf{I}_n &= \mathbf{P}_G + (\mathbf{I}_n - \mathbf{P}_G) \Rightarrow \mathbf{X}^{ct} \mathbf{I}_n \mathbf{X}^c = \mathbf{X}^{ct} \mathbf{P}_G \mathbf{X}^c + \mathbf{X}^{ct} (\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c \\ &\Leftrightarrow \mathbf{S} = \mathbf{B} + \mathbf{W}\end{aligned}$$

- Sucessivos eixos discriminantes são não correlacionados entre si. Diferentes eixos discriminantes são da forma $\mathbf{X}^c \vec{\mathbf{a}}_i$ e $\mathbf{X}^c \vec{\mathbf{a}}_j$, sendo $\vec{\mathbf{a}}_i$ e $\vec{\mathbf{a}}_j$ dois diferentes vectores próprios da matriz $\mathbf{W}^{-1} \mathbf{B}$. Sabemos ainda que $\vec{\mathbf{a}}_i$ e $\vec{\mathbf{a}}_j$ são $\mathbf{W}$ -ortogonais. Logo, se $i \neq j$:

$$\begin{aligned}\mathbf{W}^{-1} \mathbf{B} \vec{\mathbf{a}}_j &= \lambda_j \vec{\mathbf{a}}_j \Rightarrow \mathbf{B} \vec{\mathbf{a}}_j = \lambda_j \mathbf{W} \vec{\mathbf{a}}_j \\ &\Rightarrow \mathbf{W} \vec{\mathbf{a}}_j + \mathbf{B} \vec{\mathbf{a}}_j = \mathbf{W} \vec{\mathbf{a}}_j + \lambda_j \mathbf{W} \vec{\mathbf{a}}_j \\ &\Rightarrow \mathbf{S} \vec{\mathbf{a}}_j = (1 + \lambda_j) \mathbf{W} \vec{\mathbf{a}}_j \\ &\Rightarrow \vec{\mathbf{a}}_i^t \mathbf{S} \vec{\mathbf{a}}_j = (1 + \lambda_j) \vec{\mathbf{a}}_i^t \mathbf{W} \vec{\mathbf{a}}_j = 0\end{aligned}$$



Maximizar $\frac{\vec{a}^t \mathbf{B} \vec{a}}{\vec{a}^t \mathbf{W} \vec{a}}$ é maximizar $\text{ctg}^2(\theta)$. Em cada eixo, $\lambda_j = \text{ctg}^2(\theta)$.

Observações

- Se $k > n - p$, $\mathbf{W}$ não pode ser invertível. Em geral, para $k \leq n - p$ haverá invertibilidade.
- havendo k subgrupos, **não pode haver mais do que $k - 1$ eixos discriminantes** (com capacidade de discriminação não nula).
- O critério que define a ADL equivale a pedir $\vec{\mathbf{a}}$ tal que a estatística do teste F da ANOVA aos efeitos do factor, com variável resposta $\vec{\mathbf{y}} = \mathbf{X}^c \vec{\mathbf{a}}$, seja máxima:

$$F = \frac{QMF}{QMRE} = \frac{SQF}{SQRE} \cdot \frac{n-k}{k-1} = \frac{\|\mathbf{P}_G \vec{\mathbf{y}}\|^2}{\|(\mathbf{I}_n - \mathbf{P}_G) \vec{\mathbf{y}}\|^2} \cdot \frac{n-k}{k-1} = \frac{\vec{\mathbf{a}}^t \mathbf{B} \vec{\mathbf{a}}}{\vec{\mathbf{a}}^t \mathbf{W} \vec{\mathbf{a}}} \cdot \frac{n-k}{k-1}.$$

- as soluções duma Análise Discriminante **são** invariantes a transformações lineares das escalas das variáveis.

Formulações alternativas

Maximizar a cotangente do ângulo θ quer dizer minimizar θ .

Formulações alternativas que minimizam o ângulo:

- 1 Minimizar o quadrado do seno de θ .
i.e., minimizar a proporção da variabilidade total da combinação linear $\mathbf{X}^c \vec{\mathbf{a}}$ que corresponde à variabilidade intra-classes

$$\frac{\vec{\mathbf{a}}^t \mathbf{W} \vec{\mathbf{a}}}{\vec{\mathbf{a}}^t \mathbf{S} \vec{\mathbf{a}}}$$

- 2 Maximizar o quadrado do cosseno do ângulo θ
ou seja, maximizar a proporção da variabilidade total da combinação linear $\mathbf{X}^c \vec{\mathbf{a}}$ que corresponde à variabilidade inter-classes

$$\frac{\vec{\mathbf{a}}^t \mathbf{B} \vec{\mathbf{a}}}{\vec{\mathbf{a}}^t \mathbf{S} \vec{\mathbf{a}}}$$

Relações entre formulações alternativas

Mas problema igual (minimizar θ) $\Rightarrow$ solução igual.

É fácil de verificar que são iguais:

- Vectores próprios de $\mathbf{W}^{-1}\mathbf{B}$;
- Vectores próprios de $\mathbf{S}^{-1}\mathbf{W}$;
- Vectores próprios de $\mathbf{S}^{-1}\mathbf{B}$;

Logo, as combinações lineares $\mathbf{X}^c \vec{\mathbf{a}}$ obtidas com as formulações alternativas são iguais.

Os correspondentes valores próprios **não** são iguais porque correspondem às diferentes funções trigonométricas. Mas, seja $\vec{\mathbf{a}}$ vector próprio comum às três matrizes.

- Se λ é o valor próprio associado em $\mathbf{W}^{-1}\mathbf{B}$;
- $\frac{1}{\lambda+1}$ é o valor próprio em $\mathbf{S}^{-1}\mathbf{W}$ (queremos minimizar);
- $\frac{\lambda}{\lambda+1}$ é o valor próprio em $\mathbf{S}^{-1}\mathbf{B}$;

Classificação de novos indivíduos num eixo

Podemos classificar novos indivíduos, de “filiação” desconhecida.

Seja $\vec{x}$ vector de observações de novo indivíduo nas p variáveis. O respectivo valor (*score*) no eixo discriminante 1 é $y^* = \vec{x}^t \vec{a}_1$. Comparando este valor com as k médias de classe nesse eixo, $\bar{y}^{(1)}, \bar{y}^{(2)}, \dots, \bar{y}^{(k)}$, podemos classificar na classe cujo centro de gravidade:

- lhe está mais próxima, na habitual distância euclidiana:

fica na classe i se $|y^ - \bar{y}^{(i)}| < |y^* - \bar{y}^{(j)}|, \forall j \neq i$.*

- lhe está mais próxima, numa distância euclidiana inversamente ponderada pelo desvio padrão da classe:

fica na classe i se $\frac{|y^ - \bar{y}^{(i)}|}{\sigma_y^{(i)}} < \frac{|y^* - \bar{y}^{(j)}|}{\sigma_y^{(j)}}, \forall j \neq i$,*

onde $\sigma_y^{(i)}$ indica o desvio padrão dos *scores* do grupo i .

Classificação em q eixos

Com q eixos discriminantes, um indivíduo tem vector de scores:

$$\vec{y}^* = \vec{x}^t \mathbf{A}_q,$$

com $\mathbf{A}_q$ matriz $p \times q$ cujas colunas são os vectores $\vec{a}_1, \dots, \vec{a}_q$.

Pode-se classificar o indivíduo na classe cujo centro de gravidade $\vec{y}_{(i)}$:

- esteja mais próximo de $\vec{y}^*$, na habitual distância euclidiana:

$$\text{associar à classe } i \text{ se } \|\vec{y}^* - \vec{y}_{(i)}\| < \|\vec{y}^* - \vec{y}_{(j)}\|, \forall j \neq i$$

- esteja mais próximo na distância de Mahalanobis definida pela matriz de variâncias-covariâncias das observações dessa classe:

$$\text{classe } i \text{ se } \|\vec{y}^* - \vec{y}_{(i)}\|_{\mathbf{S}_i^{-1}} < \|\vec{y}^* - \vec{y}_{(j)}\|_{\mathbf{S}_j^{-1}}, \forall j \neq i,$$

onde $\mathbf{S}_i$ é matriz de (co)variâncias dos scores do grupo i .

ADL no R

A função `lda`, do módulo `MASS`, fornece as informações básicas para uma **Análise Discriminante Linear (de Fisher)**.

A função `lda` foi concebida pensando num contexto inferencial (que não é o nosso e não é necessário para uma ADL). No entanto permite obter a informação necessária.

Exemplifiquemos com os dados dos lavagantes, em que as 21 primeiras observações são de **machos reprodutores** (grupo MR); as 21 seguintes de **machos não-reprodutores** (grupo MN); e as 21 observações finais são de **fêmeas** (grupo F).

É necessário **construir o factor dos grupos e carregar o módulo MASS**:

```
> lav.grupos <- factor(rep(c("MR","MN","F"),c(21,21,21)))  
> library(MASS)
```

O exemplo dos lavagantes

Eis o comando, e principais resultados para a ADL dos lavagantes:

```
> lda(lav.grupos ~ . , data=as.data.frame(lavagantes))
```

```
[...]  
Coefficients of linear discriminants:                                <- vectores de loadings
```

	LD1	LD2
carapace_l	-0.0005163473	-1.19955746
tail_l	-0.1736612417	0.33191555
carapace_w	0.1866238904	-0.90101141
carapace_d	-0.3521185558	-0.23124418
tail_w	-2.6055856004	1.28663805
areola_l	0.3588957427	-0.06043209
areola_w	-2.1123185437	-0.03550332
rostrum_l	1.2415578489	1.22874815
rostrum_w	-0.3314912527	1.39715849
postorbital_w	0.1940959791	-1.59005854
propodus_l	0.6321803333	0.17783018
propodus_w	0.4297842346	0.71193763
dactyl_l	-0.0850563760	0.36615202

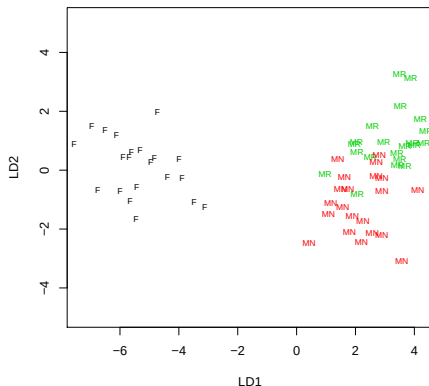
```
Proportion of trace:                                                <- proporção dos valores próprios não nulos de  $W^{(-1)}B$ 
```

LD1	LD2
0.9501	0.0499

Exemplos dos lavagantes (cont.)

Existe um método `plot` para objectos da classe `lda`, como os que são produzidos pelos comando `lda`:

```
> plot(lav.lda, col=as.numeric(lav.grupos))
```



Exemplos dos lavagantes (cont.)

Os **vectores de scores** usados para construir o gráfico podem ser obtidos com o comando **predict**, estando no argumento de saída **x**:

```
> predict(lav.lda)
```

\$x

	LD1	LD2
1	2.9590031	0.9654792
2	3.6848954	0.8131683
3	3.5259200	2.1811447
4	2.0462745	0.6083346
[...]		
60	-4.9547011	0.2934347
61	-6.7592582	-0.6571673
62	-5.6927267	0.4566755
63	-5.4276951	-0.5692571

O comando **predict** pode ser usado para **determinar as coordenadas nos eixos discriminantes de um novo indivíduo**, de forma semelhante ao que foi visto para os modelos lineares.

Exemplo dos lavagantes (cont.)

Definamos três novos indivíduos com todas as observações iguais aos máximos de cada variável em cada subgrupo, e coloque-mo-los nos novos eixos discriminantes:

```
> lxm1 <- apply(lavagantes[1:21,],2,max)
> lxm2 <- apply(lavagantes[22:42,],2,max)
> lxm3 <- apply(lavagantes[43:63,],2,max)
> predict(lav.lda, new=data.frame(rbind(lxm1, lxm2, lxm3)))
```

[...]

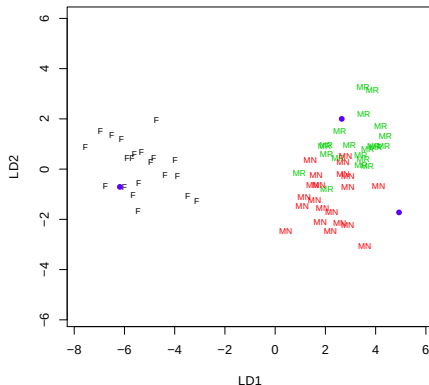
\$x

	LD1	LD2
lxm1	2.650187	1.9990716
lxm2	4.931230	-1.7232999
lxm3	-6.183037	-0.7078646

Exemplos dos lavagantes (cont.)

Existe um método `plot` para objectos da classe `lda`, como os que são produzidos pelos comando `lda`:

```
> lxm1 <- predict(lav.lda, new=data.frame(rbind(lxm1,lxm2,lxm3)))$x  
> plot(lav.lda, col=as.numeric(lav.grupos), xlim=c(-8,6))  
> points(lxm1, col="blue", pch=16)
```



Observações sobre a função lda do R

Na função lda:

- a matriz **W** é definida como $\mathbf{W} = \frac{1}{n-k} \mathbf{X}^{ct} (\mathbf{I}_n - \mathbf{P}_G) \mathbf{X}^c$;
- a matriz **B** é definida como $\mathbf{B} = \frac{1}{k-1} \mathbf{X}^{ct} \mathbf{P}_G \mathbf{X}^c$;
- logo, o quadrado de cada valor próprio de $\mathbf{W}^{-1} \mathbf{B}$ calculado pela função lda é o valor da estatística do teste *F* na ANOVA de cada eixo discriminante sobre o factor. Mas deixa de ser válida a decomposição $\mathbf{S} = \mathbf{W} + \mathbf{B}$.
- as proporções representadas por cada valor próprio (dadas na listagem) não são afectadas por esta diferente definição.
- a **W**-ortogonalidade dos *loadings* dados na listagem também é preservada (embora a norma ao quadrado dos vectores de *loadings* seja afectada, sendo $\frac{n-k}{n-1}$ quando medida com a norma da matriz **W** definida no acetato 81).

A classificação de novos indivíduos

O método `predict` da função `lda` classifica indivíduos nos grupos, com critérios baseados em conceitos inferenciais, mas próximos da classificação baseada em distâncias de Mahalanobis. Essas classificações são guardadas no objecto `class` da função `predict`:

```
> predict(lav.lda)$class
[1] MR MR MR MR MR MR MR MN MN MR MR MR MR MR MR MR MR MR MR MR MN MR MN MN
[26] MN MR MN MN MN MN MN MN MN MN MN MN MN MN MR MN MN MN F F F F F F F F
[51] F F F F F F F F F F F F F F

> predict(lav.lda, new=data.frame(rbind(lxm1, lxm2, lxm3)))$class
[1] MR MN F
```

Tabelas de classificação podem ser criadas com a função `table`:

```
> lav.pred <- predict(lav.lda)$class
> table(lav.pred, lav.grupos)
      lav.grupos
lav.pred  F  MN  MR
      F   21   0   0
      MN   0  18   2
      MR   0   3  19
```