# Análise de Componentes Principais

# Introdução

Um problema central na análise de dados multivariados é a **redução da dimensionalidade:** é possível descrever com precisão a informação contida nos dados mensurados em $p$ variáveis utilizando um conjunto $r < p$ de novas variáveis, perdendo a menor quantidade de informação possível?

# Introdução

Um problema central na análise de dados multivariados é a **redução da dimensionalidade:** é possível descrever com precisão a informação contida nos dados mensurados em $p$ variáveis utilizando um conjunto $r < p$ de novas variáveis, perdendo a menor quantidade de informação possível?

A **análise de componentes principais** tem este objetivo: dadas $n$ observações de $p$ variáveis, se analisa se é possível representar adequadamente esta informação com um número menor de variáveis construídas como **combinações lineares** das variáveis originais.

# O Problema...

Dado um conjunto de variáveis $\mathbf{x} = [X_1\ X_2\ \cdots\ X_p]^t$, podemos encontrar outro conjunto de variáveis $\mathbf{y} = [Y_1\ Y_2\ \cdots\ Y_r]^t$, dadas por

$$Y_i = \sum_{j=1}^{p} a_{ij} X_j,\ \ i = 1, \cdots, r < p$$

de tal forma que a informação contida em $\mathbf{x}$ esteja sendo bem representada por $\mathbf{y}$?

# Algumas questões

Vamos encontrar combinações lineares para representar informação.

# Algumas questões

Vamos encontrar combinações lineares para representar informação.

O que é **informação**?

# Algumas questões

Vamos encontrar combinações lineares para representar informação.

O que é **informação**?

**Informação $\Longrightarrow$ Variância: quanto maior a variabilidade, maior a informação contida nos dados, maior a variância dos dados**

# Algumas questões

Outra questão importante:

# Algumas questões

Outra questão importante:

O que é **uma boa representação da informação**?

# Algumas questões

Outra questão importante:

O que é **uma boa representação da informação**?

**Boa representação da informação** $\Longrightarrow$ **tomar as componentes de** $\mathbf{y}$ **que assegurem uma variância similar à de** $\mathbf{x}$

# Esquematicamente

Nestas condições, temos que buscar **combinações lineares** $\mathbf{y}$ das variáveis $\mathbf{x}$ de forma que se *maximize* a **variância**

# Esquematicamente

Nestas condições, temos que buscar **combinações lineares** $\mathbf{y}$ das variáveis $\mathbf{x}$ de forma que se *maximize* a **variância**

| **Variáveis Originais** | | **Combinações Lineares** |
|---|---|---|
| $X_1$ | | $Y_1$ |
| $X_2$ | | $Y_2$ |
| $\vdots$ | | $\vdots$ |
| $X_r$ | $\Longrightarrow$ | $Y_r$ |
| $\vdots$ | | $\vdots$ |
| $X_p$ | | $Y_p$ |

# Esquematicamente

Nestas condições, temos que buscar **combinações lineares** $\mathbf{y}$ das variáveis $\mathbf{x}$ de forma que se *maximize* a **variância**

| **Variáveis Originais** | | **Combinações Lineares** |
|---|---|---|
| $X_1$ | | $Y_1$ |
| $X_2$ | | $Y_2$ |
| $\vdots$ | | $\vdots$ |
| $X_r$ | $\Longrightarrow$ | $Y_r$ |
| $\vdots$ | | $\vdots$ |
| $X_p$ | | $Y_p$ |

$\mathrm{Var}[\mathbf{y}]$ : Máxima

# Esquematicamente

Ideia básica da técnica de Análise de Componentes Principais:

# Esquematicamente

Ideia básica da técnica de Análise de Componentes Principais:

| **Variáveis Originais** | | **Componentes Principais** |
|---|---|---|
| $X_1$ | ACP | $Y_1$ |
| $X_2$ | $\Longrightarrow$ | $Y_2$ |
| $\vdots$ | | $\vdots$ |
| $X_p$ | | $Y_r$ |
| | | $\vdots$ |
| | | $Y_p$ |

# Esquematicamente

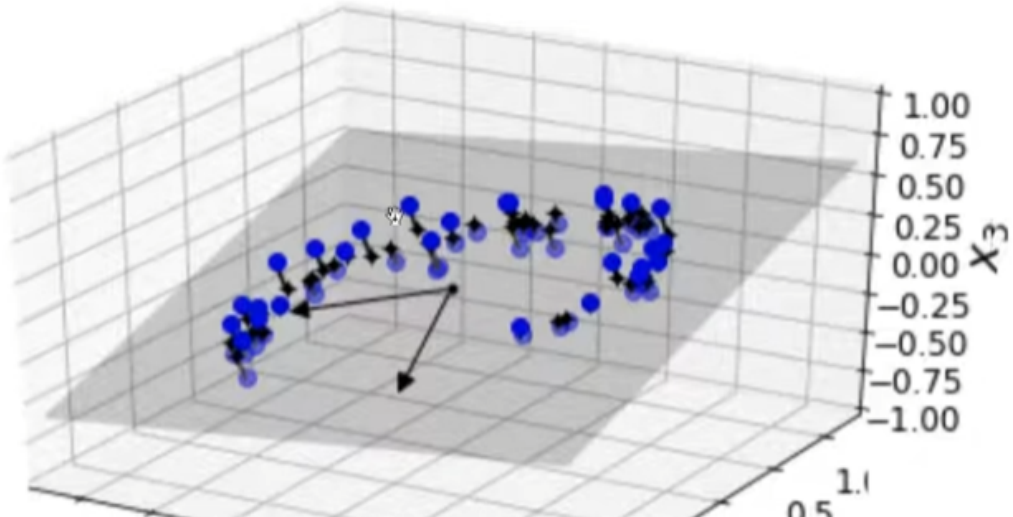
Ideia básica da técnica de Análise de Componentes Principais:

| **Variáveis Originais** | | **Componentes Principais** |
|---|---|---|
| $X_1$ | ACP | $Y_1$ |
| $X_2$ | $\Longrightarrow$ | $Y_2$ |
| $\vdots$ | | $\vdots$ |
| $X_p$ | | $Y_r$ |
| | | $\vdots$ |
| | | $Y_p$ |

As $r$ primeiras componentes resumem, por exemplo, $80\%$ do comportamento geral das $p$ variáveis originais.

# Principais objetivos

- Redução da dimensionalidade dos dados, projetando-os em uma dimensão $r < p$;



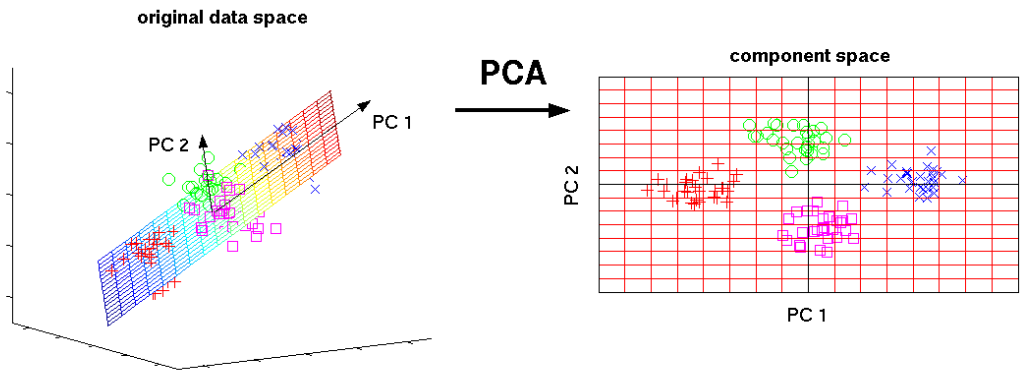
## Principais objetivos

- Obtenção de combinações interpretáveis: determinar índices e produzir escores com base nos resultados avaliados para as $p$ variáveis;



# Principais objetivos

- Descrição e entendimento da estrutura de correlação entre as variáveis, através de algumas combinações lineares das mesmas.



# Componentes Principais: o que são?

**Algebricamente:** são combinações lineares das $p$ variáveis originais, $X_1, X_2, \cdots, X_p$.

# Componentes Principais: o que são?

**Algebricamente:** são combinações lineares das $p$ variáveis originais, $X_1, X_2, \cdots, X_p$.

**Geometricamente:** são as coordenadas dos pontos amostrais em um sistema de eixos obtido pela rotação do sistema de eixos original, na direção de variabilidade máxima.

# Componentes Principais: alguns comentários

- Não pressupõe normalidade dos dados, embora componentes derivadas de populações normais tenham interpretações úteis.

# Componentes Principais: alguns comentários

- Não pressupõe normalidade dos dados, embora componentes derivadas de populações normais tenham interpretações úteis.
- Com frequência, revela relações insuspeitas. Pode permitir interpretações que não seriam obtidas preliminarmente.

# Componentes Principais: alguns comentários

- Não pressupõe normalidade dos dados, embora componentes derivadas de populações normais tenham interpretações úteis.
- Com frequência, revela relações insuspeitas. Pode permitir interpretações que não seriam obtidas preliminarmente.
- Em algumas aplicações, os componentes da ACP configuram o objetivo final do estudo. Em outras, servem como passo intermediário para realização de outras análises, como regressão, classificação, agrupamento, etc...

# Componentes Principais: como obtê-los?

- Sejam $X_1,\ X_2,\ \cdots,\ X_p$ as variáveis originais

# Componentes Principais: como obtê-los?

- Sejam $X_1,\ X_2,\ \cdots,\ X_p$ as variáveis originais
- A ideia é encontrar um novo conjunto de variáveis $Y_1,\ Y_2,\ \cdots,\ Y_p$, tais que:

$$\mathrm{Var}[Y_1] \geqslant \mathrm{Var}[Y_2] \geqslant \cdots \geqslant \mathrm{Var}[Y_p]$$

# Componentes Principais: como obtê-los?

- Sejam $X_1,\ X_2,\ \cdots,\ X_p$ as variáveis originais
- A ideia é encontrar um novo conjunto de variáveis $Y_1,\ Y_2,\ \cdots,\ Y_p$, tais que:

$$\mathrm{Var}[Y_1] \geqslant \mathrm{Var}[Y_2] \geqslant \cdots \geqslant \mathrm{Var}[Y_p]$$

- Vamos tomar cada nova variável $Y_i$, $i = 1, \cdots, p$, como uma combinação linear das variáveis originais $\mathbf{x}$:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p = \mathbf{a}_i^t\mathbf{x}$$

# Componentes Principais: como obtê-los?

- Para fixar problemas de escala, adicionamos uma primeira restrição aos vetores $\mathbf{a}_i$:

$$\mathbf{a}_i^t \mathbf{a}_i = \sum_{j=1}^{p} a_{ij}^2 = 1$$

# Componentes Principais: como obtê-los?

- Para fixar problemas de escala, adicionamos uma primeira restrição aos vetores $\mathbf{a}_i$:

$$\mathbf{a}_i^t \mathbf{a}_i = \sum_{j=1}^{p} a_{ij}^2 = 1$$

- Para evitar que duas variáveis $Y_i$ e $Y_k$, $i \neq k$, $i, k = 1, \cdots, p$, compartilhem informação, adicionamos uma segunda restrição aos vetores $\mathbf{a}_i$:

$$\mathbf{a}_i^t \mathbf{a}_k = \sum_{j=1}^{p} a_{ij} a_{kj} = 0$$

# Componentes Principais: como obtê-los?

- Para fixar problemas de escala, adicionamos uma primeira restrição aos vetores $\mathbf{a}_i$:

$$\mathbf{a}_i^t \mathbf{a}_i = \sum_{j=1}^{p} a_{ij}^2 = 1$$

- Para evitar que duas variáveis $Y_i$ e $Y_k$, $i \neq k$, $i, k = 1, \cdots, p$, compartilhem informação, adicionamos uma segunda restrição aos vetores $\mathbf{a}_i$:

$$\mathbf{a}_i^t \mathbf{a}_k = \sum_{j=1}^{p} a_{ij} a_{kj} = 0$$

**Garantia:** ortogonalidade, componentes não correlacionadas, independência

# Componentes Principais: como obtê-los?

## Primeira Componente Principal

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = a_1^t \mathbf{x}$$

# Componentes Principais: como obtê-los?

## Primeira Componente Principal

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = a_1^t \mathbf{x}$$

**Objetivo:** Encontrar $a_1^t = [a_{11} \quad a_{12} \quad \cdots \quad a_{1p}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y}_1]$ seja máxima

# Componentes Principais: como obtê-los?

## Primeira Componente Principal

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = a_1^t \mathbf{x}$$

**Objetivo:** Encontrar $a_1^t = [a_{11} \quad a_{12} \quad \cdots \quad a_{1p}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y}_1]$ seja máxima

Sujeita à restrição:

$$a_1^t a_1 = a_{11}^2 + a_{12}^2 + \cdots + a_{1p}^2 = 1$$

## Segunda Componente Principal

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = a_2^t\mathbf{x}$$

## Segunda Componente Principal

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = a_2^t\mathbf{x}$$

**Objetivo:** Encontrar $a_2^t = [a_{21} \quad a_{22} \quad \cdots \quad a_{2p}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y}_2]$ seja máxima

## Segunda Componente Principal

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = a_2^t\mathbf{x}$$

**Objetivo:** Encontrar $a_2^t = [a_{21} \quad a_{22} \quad \cdots \quad a_{2p}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y}_2]$ seja máxima

Sujeita à restrição:

$$a_2^t a_2 = a_{21}^2 + a_{22}^2 + \cdots + a_{2p}^2 = 1$$

$$\mathrm{Cov}[\mathrm{Y}_1, \mathrm{Y}_2] = 0$$

# Componentes Principais: como obtê-los?

## i-ésima Componente Principal

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p = a_i^t \mathbf{x}$$

# Componentes Principais: como obtê-los?

## i-ésima Componente Principal

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p = a_i^t \mathbf{x}$$

**Objetivo:** Encontrar $a_i^t = [a_{i1} \quad a_{i2} \quad \cdots \quad a_{ip}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y_i}]$ seja máxima

# Componentes Principais: como obtê-los?

## i-ésima Componente Principal

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p = a_i^t \mathbf{x}$$

**Objetivo:** Encontrar $a_i^t = [a_{i1} \quad a_{i2} \quad \cdots \quad a_{ip}]^t$ tal que:

$\mathrm{Var}[\mathrm{Y_i}]$ seja máxima

Sujeita à restrição:

$$a_i^t a_i = a_{i1}^2 + a_{i2}^2 + \cdots + a_{ip}^2 = 1$$

$$\mathrm{Cov}[\mathrm{Y_, Y_k}] = 0, \text{para } \mathrm{k < i}$$

# A escolha dos vetores $a_i$

- Considere o vetor aleatório p-variado $\mathbf{x} = [X_1 \quad X_2 \quad \cdots \quad X_p]^t$ com vetor de médias $\mu$ e matriz de covariâncias $\Sigma$, positiva definida (todos os seus autovalores são positivos), sendo

$$\mu = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_p]^t \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

# A escolha dos vetores $a_i$

- Considere o vetor aleatório p-variado $\mathbf{x} = [X_1 \quad X_2 \quad \cdots \quad X_p]^t$ com vetor de médias $\mu$ e matriz de covariâncias $\Sigma$, positiva definida (todos os seus autovalores são positivos), sendo

$$\mu = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_p]^t \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

- Para determinação dos componentes principais, com base no que foi exposto, usaremos o seguinte teorema:

# A escolha dos vetores $a_i$

**Teorema - Maximização de formas quadráticas:** Seja $B$ uma matriz positiva definida com autovalores $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p > 0$ e autovetores associados normalizados $e_1, e_2, \cdots, e_p$. Então:

$$\max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^t B \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_1, \text{ obtido quando } \mathbf{x} = e_1;$$

$$\min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^t B \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_p, \text{ obtido quando } \mathbf{x} = e_p.$$

# A escolha dos vetores $a_i$

**Teorema - Maximização de formas quadráticas:** Seja $B$ uma matriz positiva definida com autovalores $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p > 0$ e autovetores associados normalizados $e_1, e_2, \cdots, e_p$. Então:

$$\max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^t B \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_1, \text{ obtido quando } \mathbf{x} = e_1;$$

$$\min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^t B \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_p, \text{ obtido quando } \mathbf{x} = e_p.$$

- Adicionalmente,

$$\max_{\mathbf{x} \perp e_1, e_1, \cdots, e_k} \frac{\mathbf{x}^t B \mathbf{x}}{\mathbf{x}^t \mathbf{x}} = \lambda_{k+1}, \text{ obtido quando } \mathbf{x} = e_{k+1}.$$

# A escolha dos vetores $a_i$

Assim, no contexto de componentes principais, seja $\mathbf{x} = [X_1 \quad X_2 \quad \cdots \quad X_p]^t$ um vetor aleatório. Seja $\Sigma$ a matriz de variâncias e covariâncias e $(\lambda_1, e_1)$, $(\lambda_2, e_2)$, ..., $(\lambda_p, e_p)$ seus autovalores e autovetores, tal que $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p > 0$. Então:

$$\max_{a \neq 0} \frac{a^t \Sigma a}{a^t a} = \max_{a \neq 0}(a^t \Sigma a) = \lambda_1, \text{ obtido quando } a = e_1;$$

$$\min_{a \neq 0} \frac{a^t \Sigma a}{a^t a} = \min_{a \neq 0}(a^t \Sigma a) = \lambda_p, \text{ obtido quando } a = e_p.$$

# A escolha dos vetores $a_i$

- Adicionalmente,

$$\max_{a \perp e_1, e_1, \cdots, e_k} \frac{a^t \Sigma a}{a^t a} = \max_{a \perp e_1, e_1, \cdots, e_k} (a^t \Sigma a) = \lambda_{k+1}, \text{ obtido quando } a = e_{k+1}.$$

# A escolha dos vetores $a_i$

- Uma escolha interessante para os vetores de constantes $a_i$, $i = 1, \cdots, p$ são os **autovetores normalizados** $e_i$ da matriz $\Sigma$.

# A escolha dos vetores $a_i$

- Uma escolha interessante para os vetores de constantes $a_i$, $i = 1, \cdots, p$ são os **autovetores normalizados** $e_i$ da matriz $\Sigma$.
- Dessa forma, podemos definir a $i$-ésima componente principal da matriz $\Sigma$, $i = 1, \cdots, p$ como sendo

# A escolha dos vetores $a_i$

- Uma escolha interessante para os vetores de constantes $a_i$, $i = 1, \cdots, p$ são os **autovetores normalizados** $e_i$ da matriz $\Sigma$.
- Dessa forma, podemos definir a $i$-ésima componente principal da matriz $\Sigma$, $i = 1, \cdots, p$ como sendo

$$Y_i = e_i^t \mathbf{x} = e_{i1} X_1 + e_{i2} X_2 + \cdots + e_{ip} X_p$$

# Componentes Principais: propriedades

- A esperança e a variância da componente $Y_i$ são respectivamente dadas por:

$$\begin{aligned} E[Y_i] &= E[e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p] \\ &= e_{i1}E[X_1] + e_{i2}E[X_2] + \cdots + e_{ip}E[X_p] \\ &= e_{i1}\mu_1 + e_{i2}\mu_2 + \cdots + e_{ip}\mu_p \\ &= e_i^{\mathsf{T}}\mu \end{aligned}$$

# Componentes Principais: propriedades

- A esperança e a variância da componente $Y_i$ são respectivamente dadas por:

$$\begin{aligned} E[Y_i] &= E[e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p] \\ &= e_{i1}E[X_1] + e_{i2}E[X_2] + \cdots + e_{ip}E[X_p] \\ &= e_{i1}\mu_1 + e_{i2}\mu_2 + \cdots + e_{ip}\mu_p \\ &= e_i^{\mathsf{T}}\mu \end{aligned}$$

$$\mathrm{Var}[Y_i] = \mathrm{Var}[e_i^t\mathbf{x}] = e_i^t\mathrm{Var}[\mathbf{x}]e_i = e_i^t\Sigma e_i = e_i^t\lambda_i e_i = e_i^t e_i\lambda_i = \lambda_i$$

# Na forma matricial

- Sejam $O$ a matriz dos autovetores normalizados da matriz $\Sigma$, isto é,

$$O = \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{12} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix} = [e_1 \quad e_2 \quad \cdots \quad e_p]$$

e $y$ o vetor das componentes principais. Então, $y = O^t\mathbf{x}$ e a matriz de covariâncias de $y$ será:

$$\mathrm{Var}[y] = \mathrm{Var}[O^t\mathbf{x}] = O^t\mathrm{Var}[\mathbf{x}]O = O^t\Sigma O = \Lambda$$

# Na forma matricial

sendo

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

ou ainda, $\Sigma = O\Lambda O^t = \sum_{i=1}^{p} \lambda_i e_i e_i^t$, uma vez que $O$ é uma matriz ortogonal tal que $OO^t = O^tO = I$. Estes resultados são conhecidos como **Teorema da decomposição espectral**.

# Variabilidade explicada

| **Variável** | **Variância** | **Componente** | **Variância** |
|---|---|---|---|
| $X_1$ | $\sigma_{11}$ | $Y_1$ | $\lambda_1$ |
| $X_2$ | $\sigma_{22}$ | $Y_2$ | $\lambda_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $X_p$ | $\sigma_{pp}$ | $Y_p$ | $\lambda_p$ |
| Total | $\sigma_T^2 = \sum_{j=1}^{p} \sigma_{jj} = \mathrm{tr}(\Sigma)$ | Total | $\lambda_T = \sum_{j=1}^{p} \lambda_j = \mathrm{tr}(\Lambda)$ |

# Variabilidade explicada

| **Variável** | **Variância** | **Componente** | **Variância** |
|---|---|---|---|
| $X_1$ | $\sigma_{11}$ | $Y_1$ | $\lambda_1$ |
| $X_2$ | $\sigma_{22}$ | $Y_2$ | $\lambda_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $X_p$ | $\sigma_{pp}$ | $Y_p$ | $\lambda_p$ |
| Total | $\sigma_T^2 = \sum_{j=1}^{p} \sigma_{jj} = \mathrm{tr}(\Sigma)$ | Total | $\lambda_T = \sum_{j=1}^{p} \lambda_j = \mathrm{tr}(\Lambda)$ |

$$\mathrm{tr}(\Sigma) = \mathrm{tr}(\mathrm{O}\Lambda\mathrm{O}^{\mathrm{t}}) = \mathrm{tr}(\Lambda\mathrm{O}^{\mathrm{t}}\mathrm{O}) = \mathrm{tr}(\Lambda\mathrm{I}) = \mathrm{tr}(\Lambda)$$

# Variabilidade explicada

| **Variável** | **Variância** | **Componente** | **Variância** |
|---|---|---|---|
| $X_1$ | $\sigma_{11}$ | $Y_1$ | $\lambda_1$ |
| $X_2$ | $\sigma_{22}$ | $Y_2$ | $\lambda_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $X_p$ | $\sigma_{pp}$ | $Y_p$ | $\lambda_p$ |
| Total | $\sigma_T^2 = \sum_{j=1}^{p} \sigma_{jj} = \mathrm{tr}(\Sigma)$ | Total | $\lambda_T = \sum_{j=1}^{p} \lambda_j = \mathrm{tr}(\Lambda)$ |

$$\mathrm{tr}(\Sigma) = \mathrm{tr}(\mathrm{O}\Lambda\mathrm{O}^{\mathrm{t}}) = \mathrm{tr}(\Lambda\mathrm{O}^{\mathrm{t}}\mathrm{O}) = \mathrm{tr}(\Lambda\mathrm{I}) = \mathrm{tr}(\Lambda)$$

$$\sigma_T^2 = \lambda_T$$

# Variabilidade explicada

- Pode-se então concluir que a **j-ésima componente explica**

$$\frac{\text{Var}[Y_j]}{\text{Variância Total de X}} = \frac{\lambda_j}{\text{tr}(\Sigma)} = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i}$$

da variação total original, e ainda, que as $r$ **primeiras** componentes explicam

$$\frac{\sum_{j=1}^{r} \text{Var}[Y_j]}{\text{Variância Total de X}} = \frac{\sum_{j=1}^{r} \lambda_j}{\text{tr}(\Sigma)} = \frac{\sum_{j=1}^{r} \lambda_j}{\sum_{i=1}^{p} \lambda_i}$$

da variação total.

# Variabilidade explicada

- Busca-se analisar um conjunto menor de variáveis sem perder muita informação sobre a estrutura de variabilidade original

# Variabilidade explicada

- Busca-se analisar um conjunto menor de variáveis sem perder muita informação sobre a estrutura de variabilidade original
- Aproximação de $\Sigma$: Analisando as $r$ primeiras componentes principais

$$\Sigma \approx \sum_{i=1}^{r} \lambda_i e_i e_i^t$$

# Variabilidade explicada

- Busca-se analisar um conjunto menor de variáveis sem perder muita informação sobre a estrutura de variabilidade original
- Aproximação de $\Sigma$: Analisando as $r$ primeiras componentes principais

$$\Sigma \approx \sum_{i=1}^{r} \lambda_i e_i e_i^t$$

- Cada parcela da soma envolve uma matriz de dimensão $p \times p$ correspondente apenas à informação da $j$-ésima componente principal

# Correlação com as variáveis originais

- Os coeficientes de correlação entre a $j$-ésima variável e a $i$-ésima componente principal é dada por:

$$\rho_{Y_i, X_j} = \frac{e_{ij}\sqrt{\lambda_i}}{\sqrt{\sigma_{jj}}}$$

# Correlação com as variáveis originais

- Os coeficientes de correlação entre a $j$-ésima variável e a $i$-ésima componente principal é dada por:

$$\rho_{Y_i,X_j} = \frac{e_{ij}\sqrt{\lambda_i}}{\sqrt{\sigma_{jj}}}$$

- As correlações medem unicamente a importância de uma variável individual sem considerar a influência das demais. Não medem a importância de $X_i$ na presença de outras variáveis.

# Correlação com as variáveis originais

- Os coeficientes de correlação entre a $j$-ésima variável e a $i$-ésima componente principal é dada por:

$$\rho_{Y_i, X_j} = \frac{e_{ij}\sqrt{\lambda_i}}{\sqrt{\sigma_{jj}}}$$

- As correlações medem unicamente a importância de uma variável individual sem considerar a influência das demais. Não medem a importância de $X_i$ na presença de outras variáveis.
- Os coeficientes (**cargas**) dos componentes ($e_{ij}$), seus sinais e magnitudes, permitem interpretar os componentes e avaliar a importância das variáveis em sua constituição.

# Estimação das Componentes Principais

- Em geral, $\Sigma$ é estimada por $S$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

# Estimação das Componentes Principais

- Em geral, $\Sigma$ é estimada por $S$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

- Autovalores de $S$: $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_p$

# Estimação das Componentes Principais

- Em geral, $\Sigma$ é estimada por $S$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

- Autovalores de $S$: $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_p$
- Autovetores de $S$: $\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_p$

# Estimação das Componentes Principais

- Estimação da $j$-ésima componente principal de $S$:

$$\hat{Y}_j = \hat{e}_j^t \mathbf{x} = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \cdots + \hat{e}_{jp} X_p, \quad j = 1, 2, \cdots, p$$

# Estimação das Componentes Principais

- Estimação da $j$-ésima componente principal de $S$:

$$\hat{Y}_j = \hat{e}_j^t \mathbf{x} = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \cdots + \hat{e}_{jp} X_p, \quad j = 1, 2, \cdots, p$$

- Componentes principais amostrais - Propriedades
  - Variância: $\text{Var}(\hat{Y}_j) = \hat{\lambda}_j$
  - Covariância entre as componentes: $\text{Cov}(\hat{Y}_j, \hat{Y}_k) = 0, \ \ j \neq k$
  - Variância total estimada explicada pela componente:

$$\frac{\text{Var}(\hat{Y}_j)}{\text{Variância total estimada de } \mathbf{x}} = \frac{\hat{\lambda}_j}{\text{tr}(S)} = \frac{\hat{\lambda}_j}{\sum_{i=1}^{p} \hat{\lambda}_i}$$

# Estimação das Componentes Principais

- Correlação estimada entre componente e variável:

$$r_{\hat{Y}_j, \hat{X}_k} = \frac{\hat{e}_{jk}\sqrt{\hat{\lambda}_j}}{\sqrt{\sigma_{kk}}}$$

# Estimação das Componentes Principais

- Correlação estimada entre componente e variável:

$$r_{\hat{Y}_j, \hat{X}_k} = \frac{\hat{e}_{jk}\sqrt{\hat{\lambda}_j}}{\sqrt{\sigma_{kk}}}$$

- Decomposição espectral de $S$:

$$S = \sum_{j=1}^{p} \hat{\lambda}_j \hat{e}_j \hat{e}_j^t$$

# Estimação das Componentes Principais

- Aproximação de $S$ pelas $r$ primeiras componentes

$$S \approx \sum_{j=1}^{r} \hat{\lambda}_j \hat{e}_j \hat{e}_j^t$$

# Estimação das Componentes Principais

- Aproximação de $S$ pelas $r$ primeiras componentes

$$S \approx \sum_{j=1}^{r} \hat{\lambda}_j \hat{e}_j \hat{e}_j^t$$

- Escores das componentes
  - Valor das componentes para cada elemento amostral
  - Na prática, o uso das componentes relevantes se dá através dos escores

# Exemplo: (Mingoti,2007)

12 empresas, 3 variáveis: ganho bruto $(X_1)$, ganho líquido $(X_2)$ e patrimônio acumulado $(X_3)$

| **Empresa** | **Ganho bruto** $(X_1)$ | **Ganho líquido** $(X_2)$ | **Patrimônio** $(X_3)$ |
|---|---|---|---|
| *E1* | 9893 | 564 | 17689 |
| *E2* | 8776 | 389 | 17359 |
| *E3* | 13572 | 1103 | 18597 |
| *E4* | 6455 | 743 | 8745 |
| *E5* | 5129 | 203 | 14397 |
| *E6* | 5432 | 215 | 3467 |
| *E7* | 3807 | 385 | 4679 |
| *E8* | 3423 | 187 | 6754 |
| *E9* | 3708 | 127 | 2275 |
| *E10* | 3294 | 297 | 6754 |
| *E11* | 5433 | 432 | 5589 |
| *E12* | 6287 | 451 | 8972 |

# Primeiro Exemplo: (Mingoti,2007)

```
load <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg, dependencies = TRUE)
  sapply(pkg, require, character.only = TRUE)
}

## Pacotes utilizados nessa análise

packages = c("tidyverse", "factoextra", "psych", "gridExtra")
load(packages)
```

```
 tidyverse factoextra      psych  gridExtra
      TRUE       TRUE       TRUE       TRUE
```

```
dados <- read.table("https://raw.githubusercontent.com/tiagomartin/est014/
dados %>%
  str()
```

# Primeiro Exemplo: (Mingoti,2007)

```
x_barra = dados %>%
  colMeans()

x_barra
```

```
 Granho_Bruto Ganho_Liquido    Patrimonio
    6267.4167      424.6667     9606.4167
```

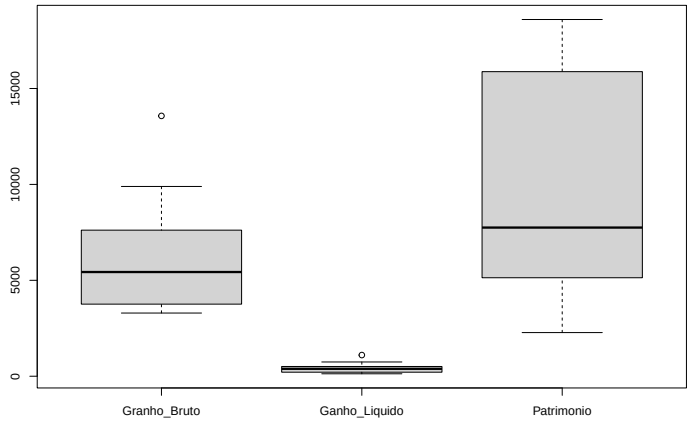
```
S = dados %>%
  var()

S
```

```
              Granho_Bruto Ganho_Liquido Patrimonio
Granho_Bruto     9550608.6     706121.06 14978232.5
Ganho_Liquido     706121.1      76269.52   933915.1
Patrimonio      14978232.5     933915.06 34408113.0
```

# Primeiro Exemplo: (Mingoti,2007)

```
dados %>%
  boxplot()
```



# Primeiro Exemplo: (Mingoti,2007)

```r
## Analise de Componentes Principais utilizando a matriz de covariancias (
acp_S = prcomp(dados)

## Proporcao da variacao explicada
summary(acp_S)
```

```
Importance of components:
                             PC1       PC2       PC3
Standard deviation     6440.0615 1.594e+03 145.23266
Proportion of Variance    0.9418 5.767e-02   0.00048
Cumulative Proportion     0.9418 9.995e-01   1.00000
```

```r
## Loadings (cargas)
acp_S$rotation
```

```
                      PC1         PC2         PC3
Granho_Bruto  -0.42509725 -0.89970680 -0.09909593
Ganho_Liquido -0.02766083 -0.09651661  0.99494694
```

# Primeiro Exemplo: (Mingoti,2007)

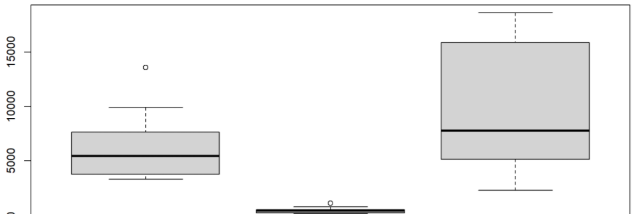
$Y_1 = 0,425 \times \text{GB} + 0,028 \times \text{GL} + 0,905 \times \text{PA} \Longrightarrow 94,18\%$ da informação total de $\mathbf{x}$

# Primeiro Exemplo: (Mingoti,2007)

$Y_1 = 0,425 \times \text{GB} + 0,028 \times \text{GL} + 0,905 \times \text{PA} \Longrightarrow 94,18\%$ da informação total de $\mathbf{x}$

**Qual a variável mais importante para $Y_1$?**

**Patrimônio**




**Será?**

# Componentes Principais: variáveis padronizadas

- Padronização do vetor aleatório $\mathbf{x}$:

$$z = D^{-1}(\mathbf{x} - \mu)$$

# Componentes Principais: variáveis padronizadas

- Padronização do vetor aleatório $\mathbf{x}$:

$$z = D^{-1}(\mathbf{x} - \mu)$$

- $D$: matriz diagonal de desvios-padrão

# Componentes Principais: variáveis padronizadas

- Padronização do vetor aleatório $\mathbf{x}$:

$$z = D^{-1}(\mathbf{x} - \mu)$$

- $D$: matriz diagonal de desvios-padrão
- Variável padronizada: $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$

# Componentes Principais: variáveis padronizadas

- Padronização do vetor aleatório $\mathbf{x}$:

$$z = D^{-1}(\mathbf{x} - \mu)$$

- $D$: matriz diagonal de desvios-padrão
- Variável padronizada: $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$
- Matriz de covariâncias de $z$

$$\text{Cov}(z) = D^{-1}\Sigma D^{-1} = P = \text{Cor}(\mathbf{x})$$

# Componentes Principais: variáveis padronizadas

- Padronização do vetor aleatório $\mathbf{x}$:

$$z = D^{-1}(\mathbf{x} - \mu)$$

- $D$: matriz diagonal de desvios-padrão
- Variável padronizada: $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}$
- Matriz de covariâncias de $z$

$$\text{Cov}(z) = D^{-1}\Sigma D^{-1} = P = \text{Cor}(\mathbf{x})$$

- Componentes principais de $z$: obtidas dos autovalores e autovetores de $P$

# Componentes Principais: variáveis padronizadas

- A $j$-ésima componente principal da matriz $P$:

$$Y_j = e_j^t z = e_j^t D^{-1}(\mathbf{x} - \mu) = e_{j1}Z_1 + e_{j2}Z_2 + \cdots + e_{jp}Z_p$$

sendo $e_j$, o $j$-ésimo autovetor da matriz $P$, $j = 1, \cdots p$.

# Componentes Principais: variáveis padronizadas

- A $j$-ésima componente principal da matriz $P$:

$$Y_j = e_j^t z = e_j^t D^{-1}(\mathbf{x} - \mu) = e_{j1}Z_1 + e_{j2}Z_2 + \cdots + e_{jp}Z_p$$

sendo $e_j$, o $j$-ésimo autovetor da matriz $P$, $j = 1, \cdots p$.

- Variância total de $P$

$$\sum_{j=1}^{p} \text{Var}(Y_j) = \sum_{j=1}^{p} \text{Var}(Z_j) = p$$

# Componentes Principais: variáveis padronizadas

- Proporção da variância populacional padronizada devido à $j$-ésima componente

$$\frac{\text{Var}(Y_j)}{\text{Variância total de } z} = \frac{\lambda_j}{\text{tr}(P)} = \frac{\lambda_j}{p}$$

# Componentes Principais: variáveis padronizadas

- Proporção da variância populacional padronizada devido à $j$-ésima componente

$$\frac{\text{Var}(Y_j)}{\text{Variância total de } z} = \frac{\lambda_j}{\text{tr}(P)} = \frac{\lambda_j}{p}$$

- Correlação entre componente $Y_j$ e a variável padronizada $Z_k$:

$$\rho_{Y_j, Z_k} = e_{jk}\sqrt{\lambda_j}$$

# Observações

- As componentes principais **não são invariantes à mudanças de escala**. Os resultados são **diferentes** quando se faz a análise utilizando a matriz de covariâncias e a matriz de correlações.

# Observações

- As componentes principais **não são invariantes à mudanças de escala**. Os resultados são **diferentes** quando se faz a análise utilizando a matriz de covariâncias e a matriz de correlações.
- **As componentes obtidas a partir da matriz de covariâncias são influenciadas pelas variáveis de maior variância**. A **matriz de correlações**, em geral, é a melhor opção quando as **variâncias são muito heterogêneas**.

# Observações

- As componentes principais **não são invariantes à mudanças de escala**. Os resultados são **diferentes** quando se faz a análise utilizando a matriz de covariâncias e a matriz de correlações.
- **As componentes obtidas a partir da matriz de covariâncias são influenciadas pelas variáveis de maior variância**. A **matriz de correlações**, em geral, é a melhor opção quando as **variâncias são muito heterogêneas**.
- Um valor **pequeno incomum** para o **último autovalor** da matriz de covariâncias (ou correlação) amostral pode indicar uma **dependência linear não detectada** no conjunto de dados.

# Observações

- As componentes principais **não são invariantes à mudanças de escala**. Os resultados são **diferentes** quando se faz a análise utilizando a matriz de covariâncias e a matriz de correlações.
- **As componentes obtidas a partir da matriz de covariâncias são influenciadas pelas variáveis de maior variância**. A **matriz de correlações**, em geral, é a melhor opção quando as **variâncias são muito heterogêneas**.
- Um valor **pequeno incomum** para o **último autovalor** da matriz de covariâncias (ou correlação) amostral pode indicar uma **dependência linear não detectada** no conjunto de dados.
- Valores grandes de autovalores (e correspondentes autovetores) são importantes em uma análise.

# Quantas componentes devem ser retidas?

- **Critério de Kaiser:** Trata-se de uma regra prática (heurística) para decidir quantas componentes principais manter em uma PCA extraída através da matriz de correlações. Ele diz:
  *Retenha apenas as componentes associadas a um autovalor maior que 1.*

- $\lambda = 1 \rightarrow$ a componente explica tanta variabilidade quanto 1 variável original
- $\lambda > 1 \rightarrow$ explica mais variabilidade do que qualquer variável isolada
- $\lambda < 1 \rightarrow$ explica menos variabilidade do que uma única variável $\rightarrow$ então não "vale a pena"

# Quantas componentes devem ser retidas?

- **Média dos autovalores:** É um critério alternativo quando a PCA é feita sobre **matriz de covariâncias**.
  *Se PCA foi feita na matriz de covariâncias, retenha PCs com autovalores maiores que a média dos autovalores.*

Então, se uma componente tem autovalor:

$$\lambda_k > \frac{\sum_{j=1}^{p} \lambda_j}{p}$$

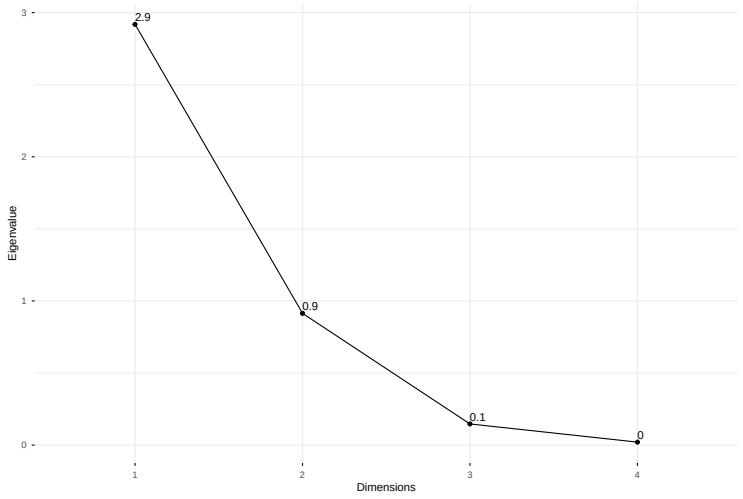
ela retém mais informação do que a variância média por dimensão do espaço original.

# Quantas componentes devem ser retidas?

- **Screeplot:** É um gráfico que coloca, no eixo X, o número da componente principal e, no eixo Y, o autovalor (ou a variância explicada).
  *O ponto onde a curva deixa de cair abruptamente e começa a "horizontalizar" é onde você para de manter componentes.*

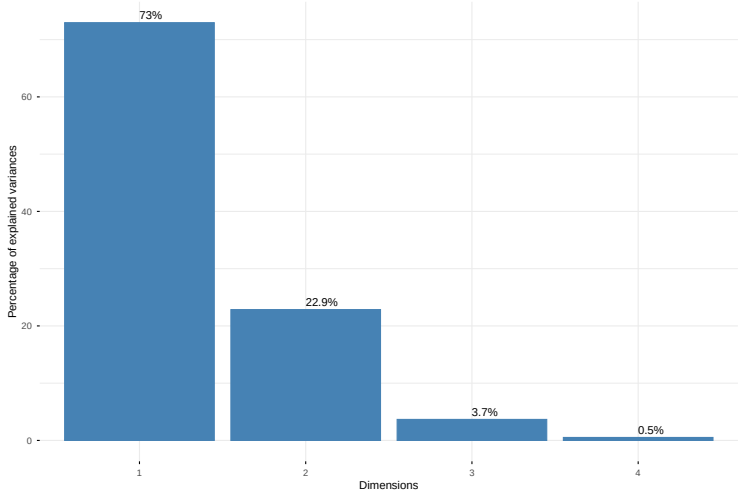
- A ideia é identificar o "cotovelo" da curva.
  - antes do cotovelo: cada componente adiciona bastante variância
  - depois do cotovelo: os autovalores ficam "quase uma linha reta" $\rightarrow$ só ruído

# Quantas componentes devem ser retidas?



# Quantas componentes devem ser retidas?

- Reter o número de componentes principais que acumulem pelo menos certa porcentagem da variabilidade total dos dados, digamos 70%.



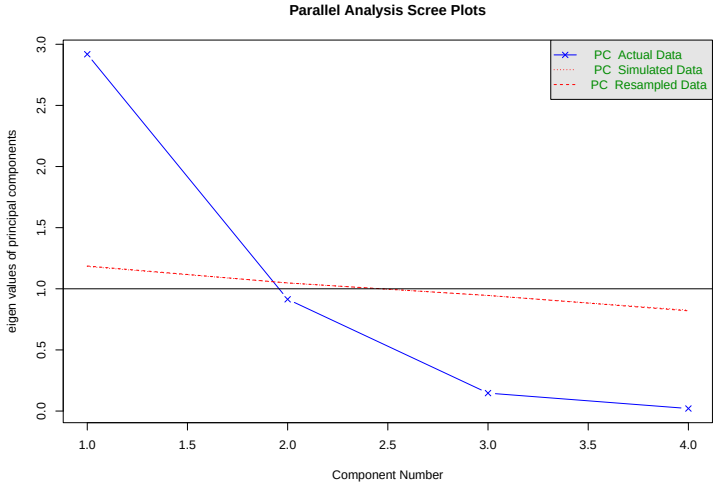
# Quantas componentes devem ser retidas?

- **Parallel Analysis (Horn, 1965):** É o método moderno mais recomendado para decidir quantos componentes reter em PCA.
  *Compare seus autovalores reais com autovalores esperados ao acaso.*

# Quantas componentes devem ser retidas?

- O procedimento envolve:
  - Calcular os autovalores dos dados originais.
  - Gerar dados aleatórios com a mesma estrutura dos dados originais.
  - Calcular autovalores para os dados aleatórios, repetindo o processo muitas vezes para obter uma distribuição amostral.
  - Comparar os autovalores reais com os autovalores médios ou percentis (como o 95º) dos dados simulados.
  - Reter os componentes cujos autovalores reais são maiores do que os autovalores aleatórios correspondentes.

# Quantas componentes devem ser retidas?




```
Parallel analysis suggests that the number of factors =  NA  and the numbe
```

# Quantas componentes devem ser retidas?

- **Interpretação desse gráfico**
  - linha azul = autovalores reais dos seus dados
  - linhas vermelha / pontilhada = autovalores esperados pelo acaso (parallel analysis)

# Quantas componentes devem ser retidas?

- **Interpretação desse gráfico**
  - linha azul = autovalores reais dos seus dados
  - linhas vermelha / pontilhada = autovalores esperados pelo acaso (parallel analysis)
- **Regra:** retenha apenas os componentes cuja linha azul está acima da linha vermelha.
  - Parallel Analysis (Horn) está dizendo: retenha 1 componente principal.

# Interpretação das componentes principais

- Em geral, quando existe uma **alta correlação positiva** entre todas as variáveis, os **sinais associados às variáveis coincidem** na primeira componente principal.

# Interpretação das componentes principais

- Em geral, quando existe uma **alta correlação positiva** entre todas as variáveis, os **sinais associados às variáveis coincidem** na primeira componente principal.
- Neste caso, a primeira componente principal pode ser interpretada como um **índice global**, calculado como uma média ponderada de todas as variáveis.

# Interpretação das componentes principais

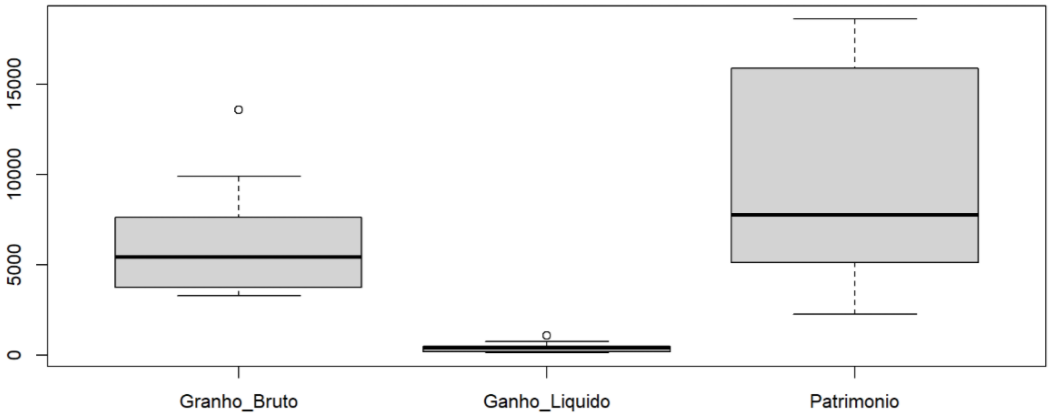
- Em geral, quando existe uma **alta correlação positiva** entre todas as variáveis, os **sinais associados às variáveis coincidem** na primeira componente principal.
- Neste caso, a primeira componente principal pode ser interpretada como um **índice global**, calculado como uma média ponderada de todas as variáveis.
- O restante das componentes, normalmente possuem **pesos negativos e positivos** e são interpretadas como um **contraste entre grupos de variáveis**.

# Voltando ao Exemplo

12 empresas, 3 variáveis: ganho bruto $(X_1)$, ganho líquido $(X_2)$ e patrimônio acumulado $(X_3)$

| **Empresa** | **Ganho bruto** $(X_1)$ | **Ganho líquido** $(X_2)$ | **Patrimônio** $(X_3)$ |
|---|---|---|---|
| *E1* | 9893 | 564 | 17689 |
| *E2* | 8776 | 389 | 17359 |
| *E3* | 13572 | 1103 | 18597 |
| *E4* | 6455 | 743 | 8745 |
| *E5* | 5129 | 203 | 14397 |
| *E6* | 5432 | 215 | 3467 |
| *E7* | 3807 | 385 | 4679 |
| *E8* | 3423 | 187 | 6754 |
| *E9* | 3708 | 127 | 2275 |
| *E10* | 3294 | 297 | 6754 |
| *E11* | 5433 | 432 | 5589 |
| *E12* | 6287 | 451 | 8972 |

# Voltando ao Exemplo



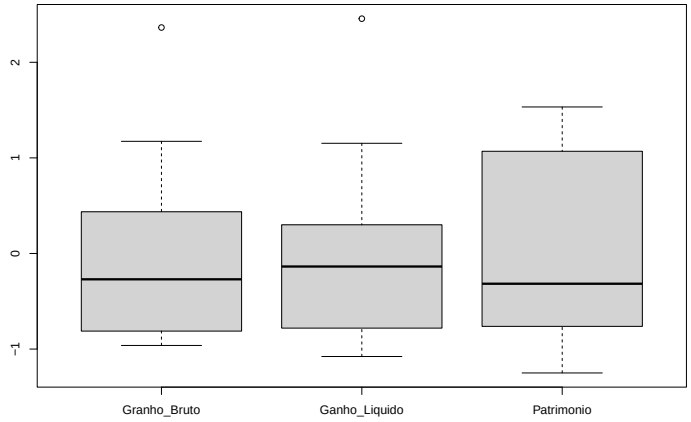
# Voltando ao Exemplo

```
dados %>%
  scale(center=TRUE, scale=TRUE)
```

```
    Granho_Bruto Ganho_Liquido Patrimonio
E1   1.173173832    0.50452133  1.3779065
E2   0.811732638   -0.12914781  1.3216486
E3   2.363632341    2.45622228  1.5327010
E4   0.060698607    1.15267434 -0.1468530
E5  -0.368371244   -0.80264757  0.8166914
E6  -0.270325871   -0.75919598 -1.0466384
E7  -0.796146767   -0.14363167 -0.8400185
E8  -0.920402290   -0.86058304 -0.4862757
E9  -0.828181394   -1.07784103 -1.2498488
E10 -0.962144379   -0.46227672 -0.4862757
E11 -0.270002289    0.02655375 -0.6848831
E12  0.006336816    0.09535212 -0.1081544
attr(,"scaled:center")
```

# Voltando ao Exemplo

```
dados %>%
  scale(center=TRUE, scale=TRUE) %>%
  boxplot()
```



# Voltando ao Exemplo

```r
## Analise de Componentes Principais utilizando a matriz de correlacoes (M
acp_R = prcomp(dados, scale. = TRUE)

## Proporcao da variacao explicada
summary(acp_R)
```

```
Importance of components:
                          PC1    PC2     PC3
Standard deviation     1.5788 0.6508 0.28978
Proportion of Variance 0.8308 0.1412 0.02799
Cumulative Proportion  0.8308 0.9720 1.00000
```

```r
## Loadings (cargas)
acp_R$rotation
```

```
                    PC1          PC2        PC3
Granho_Bruto  -0.6167027  0.001267206  0.7871952
Ganho_Liquido -0.5567945  0.706196936 -0.4373395
```

# Voltando ao Exemplo

```
## Escores das componentes principais
Yr = acp_R$x
Yr %>% head()
```

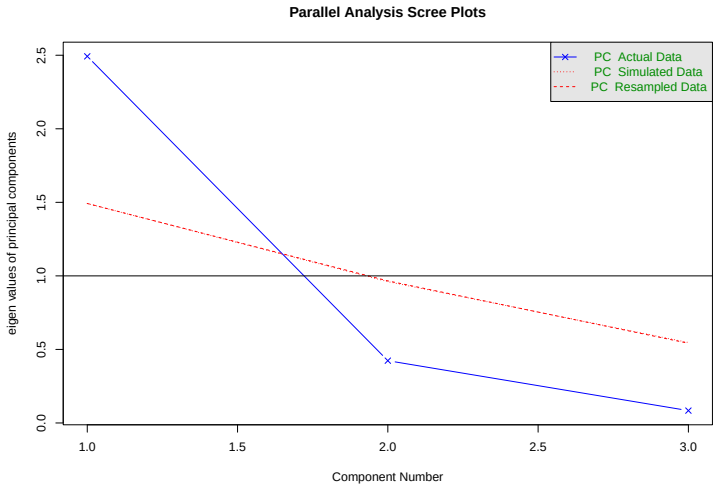
```
          PC1        PC2        PC3
E1  1.7711764  0.6177995 -0.1037449
E2  1.1641454  1.0259213 -0.1208101
E3  3.6781699 -0.6523976 -0.1200063
E4  0.5975165 -0.9180660  0.3924755
E5 -0.2196218  1.1455233  0.2940545
E6 -1.1718486 -0.2045506 -0.5743139
```

```
## Matriz de correlacoes entre variaveis originais e componentes principai
Ryx_R = cor(dados,Yr)
Ryx_R
```

```
                   PC1           PC2        PC3
Granho_Bruto  0.9736351 -0.0008246563 -0.2281098
```

# Voltando ao Exemplo

```
fa.parallel(dados %>% scale(center=TRUE, scale=TRUE), fa="pc", n.iter=1000)
```



# Voltando ao Exemplo

$$Z_{GB} = \frac{\text{Ganho Bruto} - \overline{\text{Ganho Bruto}}}{s_{\text{Ganho Bruto}}}$$

$$Z_{GL} = \frac{\text{Ganho Líquido} - \overline{\text{Ganho Líquido}}}{s_{\text{Ganho Líquido}}}$$

$$Z_{PA} = \frac{\text{Patrimônio} - \overline{\text{Patrimônio}}}{s_{\text{Patrimônio}}}$$

# Voltando ao Exemplo

$Y_1 = 0,617 \times Z_{GB} + 0,557 \times Z_{GL} + 0,556 \times Z_{PA} \Longrightarrow 83,08\%$ da informação total de $z$

# Voltando ao Exemplo

$Y_1 = 0,617 \times Z_{GB} + 0,557 \times Z_{GL} + 0,556 \times Z_{PA} \Longrightarrow 83,08\%$ da informação total de $z$

- **Interpretação:** É basicamente um índice de desempenho global da empresa. O coeficiente de maior grandeza numérica desta componente é relativo a ganho bruto enquanto que os demais coeficientes são aproximadamente iguais. Quanto maior os valores de ganhos brutos e líquido e patrimônio da empresa, maior será o valor numérico da componente. Além disso, todas as três variáveis possui alta correlação com essa componente, indicando serem importantes na composição da mesma.

# Voltando ao Exemplo

$$Y_2 = -0,001 \times Z_{GB} - 0,706 \times Z_{GL} + 0,708 \times Z_{PA} \Longrightarrow 14,12\% \text{ da informação total de } z$$

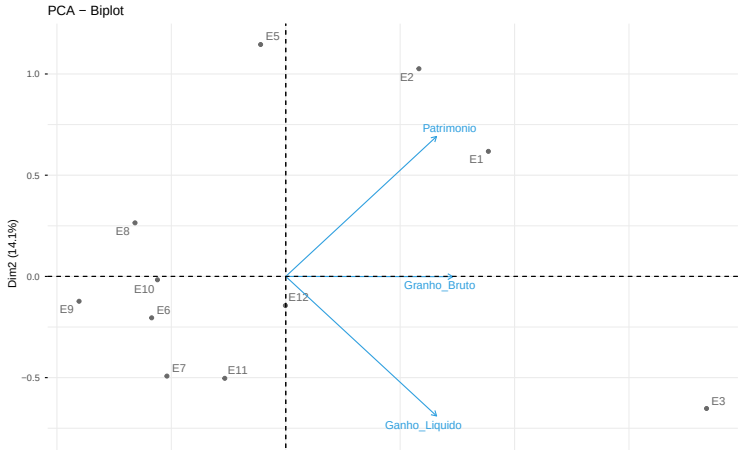
# Voltando ao Exemplo

$Y_2 = -0,001 \times Z_{GB} - 0,706 \times Z_{GL} + 0,708 \times Z_{PA} \Longrightarrow 14,12\%$ da informação total de $z$

- **Interpretação:** É uma comparação entre as variáveis ganho líquido e patrimônio, sendo que essas duas variáveis possuem igual importância na composição da mesma. Valores próximos de zero dessa componente indicam empresas com um certo equilíbrio entre ganho líquido e patrimônio acumulado no período.

# Voltando ao Exemplo

```
fviz_pca_biplot(acp_R, repel = TRUE,
                col.var = "#2E9FDF", # Variables color
                col.ind = "#696969"  # Individuals color
                )
```



# Voltando ao Exemplo

**Interpretação do biplot**

- As setas indicam em que direção a variável aumenta. os indivíduos que estão posicionados no sentido da seta são os que têm valores maiores naquela variável.

# Voltando ao Exemplo

**Interpretação do biplot**

- As setas indicam em que direção a variável aumenta. os indivíduos que estão posicionados no sentido da seta são os que têm valores maiores naquela variável.
- **Ângulo entre a seta (variável) e o eixo PC1:** O cosseno do ângulo entre a variável e a PC é exatamente o **loading**.

$$\cos(\theta_j, PC_i) = \text{loading}_{X_j, PC_i}$$

# Voltando ao Exemplo

então:

- se a seta está quase colada no eixo $PC_i$
  - loading próximo de $+1 \rightarrow$ variável altamente alinhada com a componente
  - é variável que define o componente principal
- se a seta faz um ângulo grande (perto de 90°) com $PC_i$
  - loading $\approx 0 \rightarrow$ essa variável não contribui para a componente
- se a seta aponta para o lado oposto da $PC_i$ (180°)
  - loading $\approx -1 \rightarrow$ variável altamente alinhada negativamente

# Voltando ao Exemplo

- **Observação:** Quando a PCA é feita na **matriz de correlações**, os loadings são exatamente as correlações entre as variáveis originais e as componentes principais. Quando usamos **matriz de covariâncias**, os loadings refletem contribuição em variância e o cosseno do ângulo não corresponde numericamente à correlação.

# Voltando ao Exemplo

- **PC1 (Dim1 = 83.1%):** Praticamente toda a informação relevante está aqui. *PC1 representa um eixo geral de nível financeiro / escala econômica. Quanto maior Patrimônio / Ganho Bruto / Ganho Líquido → mais à direita o indivíduo aparece.*
- Todas variáveis apontam para a direita, e com ângulos semelhantes → altíssima correlação entre elas

# Voltando ao Exemplo

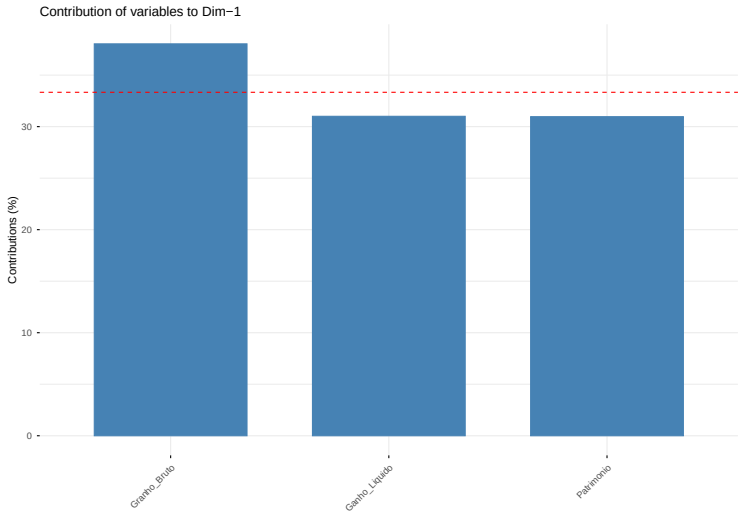
- Quem está mais pra direita (E1, E2, E3) $\rightarrow$ são aqueles com valores altos nessas variáveis
- Quem está mais pra esquerda (E8, E9, E10...) $\rightarrow$ são os com valores baixos
- A empresa E12 posiciona-se muito próxima à origem do plano principal, indicando um perfil mediano em todas as variáveis financeiras consideradas.
  - Ela não apresenta características extremas nem para valores altos, nem para valores baixos, sendo portanto uma empresa altamente representativa do centro da distribuição.

# Voltando ao Exemplo

- Quem está mais pra direita (E1, E2, E3) $\rightarrow$ são aqueles com valores altos nessas variáveis
- Quem está mais pra esquerda (E8, E9, E10...) $\rightarrow$ são os com valores baixos
- A empresa E12 posiciona-se muito próxima à origem do plano principal, indicando um perfil mediano em todas as variáveis financeiras consideradas.
  - Ela não apresenta características extremas nem para valores altos, nem para valores baixos, sendo portanto uma empresa altamente representativa do centro da distribuição.
- **PC2 (Dim2 = 14.1%):** quase não traz informação nova $\rightarrow$ setas não sobem muito, elas estão quase horizontais
  - não existe uma segunda dimensão "conceitual" forte

# Voltando a Exemplo

```
# Contribuições das variáveis para a PC1
fviz_contrib(acp_R, choice = "var", axes = 1)
```



# Voltando ao Exemplo

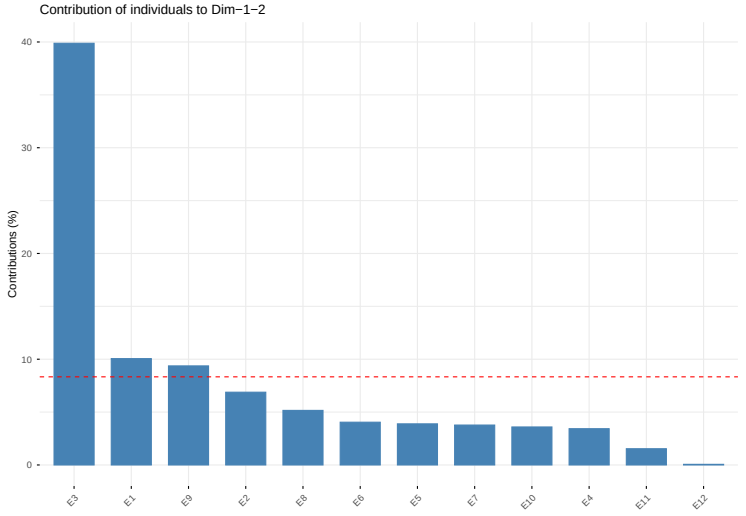
- Esse gráfico mostra o quanto cada variável explica/contribui para a formação do primeiro componente principal.
  - Ganho Bruto contribuiu um pouco mais
  - Ganho Liquido e Patrimonio contribuem praticamente igual e muito próximo

# Voltando ao Exemplo

- Esse gráfico mostra o quanto cada variável explica/contribui para a formação do primeiro componente principal.
  - Ganho Bruto contribuiu um pouco mais
  - Ganho Liquido e Patrimonio contribuem praticamente igual e muito próximo
- A linha vermelha tracejada é a contribuição média esperada (se todas contribuíssem igual).
  - Como temos 3 variáveis $\rightarrow$ contribuição média $= 100\%/3 \approx 33.33\%$

# Voltando ao Exemplo

```
fviz_contrib(acp_R, choice = "ind", axes = 1:2)
```



# Voltando ao Exemplo

- Esse gráfico mostra quais indivíduos contribuem mais para definir o plano principal da PC, isto é, quais observações estão orientando a direção das componentes principais.

# Voltando ao Exemplo

- Esse gráfico mostra quais indivíduos contribuem mais para definir o plano principal da PC, isto é, quais observações estão orientando a direção das componentes principais.
- A empresa E3 é a maior influenciadora da PCA: sua contribuição para o plano principal é muito superior à média (linha vermelha), indicando que ela é um caso extremo na direção da primeira componente. Esse ponto está orientando de forma dominante a estrutura da análise.