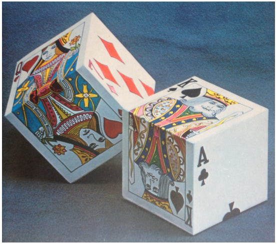
Fundamentos Estatísticos para Ciência dos Dados
Exemplos e Propriedades da Normal multivariada

Renato Martins Assunção

DCC, UFMG



Normal bivariada

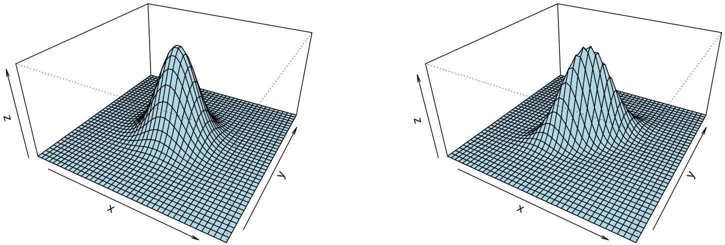
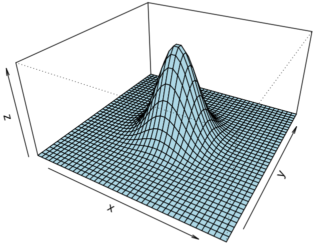


Figura: Densidade da normal bivariada

Normal bivariada

- Importante distribuição para um vetor aleatório $\boldsymbol{Y} = (Y_1, Y_2)$
- Cada uma das v.a's separadamente segue uma gaussiana com sua própria esperança μ_j e variância σ_j^2
- Isto é, $Y_1 \sim N(\mu_1, \sigma_1^2)$ e $Y_2 \sim N(\mu_2, \sigma_2^2)$
- As amostras de $\boldsymbol{Y}$ formam nuvens de pontos em forma de elipses centradas em (μ_1, μ_2).
- Além disso, elas não são (em geral) independentes: A distribuição de Y_2 MUDA SE SOUBERMOS O VALOR DE Y_1.
- Um único parâmetro $\rho \in [-1, 1]$ controla o grau de associação ou correlação entre Y_1 e Y_2.

Densidade e amostra



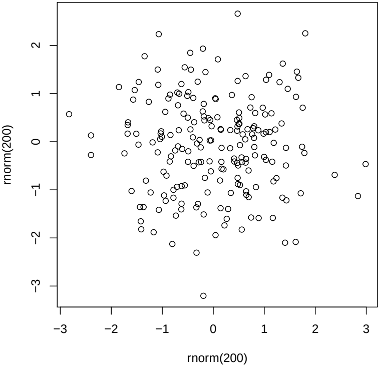


Figura: Normal bivariada com $Y_1 \sim N(\mu_1 = 0, \sigma_1^2 = 1)$ e $Y_2 \sim N(\mu_2 = 0, \sigma_2^2 = 1)$ e com correlação $\rho = 0$

Densidade e amostra

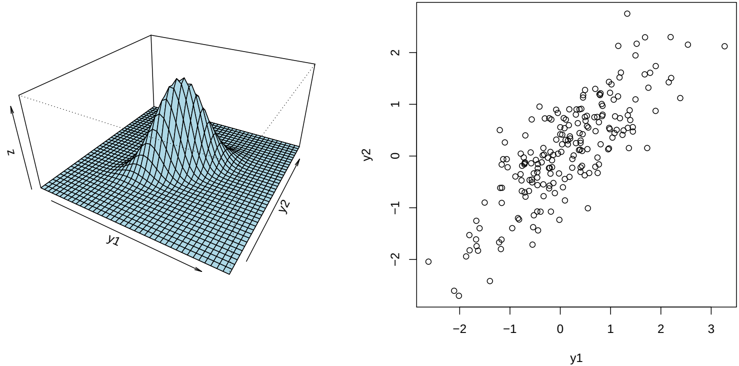


Figura: Normal bivariada com $Y_1 \sim N(\mu_1 = 0, \sigma_1^2 = 1)$ e $Y_2 \sim N(\mu_2 = 0, \sigma_2^2 = 1)$ e com correlação $\rho = 0.80$

Uma amostra de uma normal bivariada

- $n = 100$ instâncias do vetor aleatório $\boldsymbol{Y} = (Y_1, Y_2)$.
- Existem os valores MARGINAIS:
    - $\mathbb{E}(Y_1) = \mu_1$
    - $\sqrt{\mathbb{V}(Y_1)} = \sigma_1$
    - e também $\mathbb{E}(Y_2) = \mu_2$
    - e $\sqrt{\mathbb{V}(Y_2)} = \sigma_2$
- Estes valores são facilmente estimados a partir das MARGENS do gráfico.
- Por exemplo:
    - $\mathbb{E}(Y_1) = \mu_1 \approx 10$
    - $\sigma_1 \approx 2.5$
- Agora você: $\mathbb{E}(Y_2) = \mu_2 \approx ???$ e $\sigma_2 \approx ???$

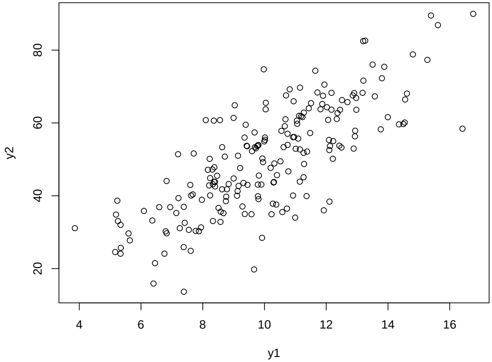


Figura: Amostra de normal bivariada

Uma amostra de uma normal bivariada

- Temos $\mathbb{E}(Y_2) = \mu_2 \approx 50$
- E $\sigma_2 \approx 15$
- Os valores de Y_1 e Y_2 medidos num mesmo ω não são independentes.
- O valor de Y_1 dá informação sobre o valor de Y_2.
- Como assim?
- Vamos ser mais específicos...

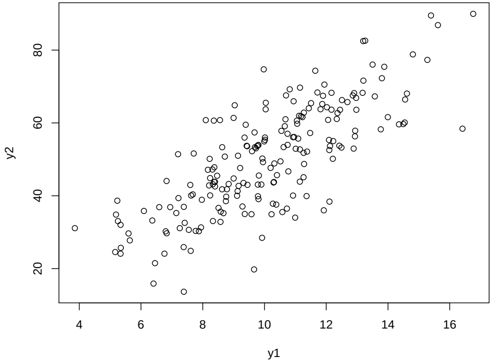


Figura: Amostra de normal bivariada

Distribuição de Y DADO QUE $Y_1 = 14$?

- Qual a distribuição de Y_2 DADO QUE $Y_1 = 14$?
- O que podemos dizer do valor esperado de Y_2 DADO QUE $Y_1 = 14$?
- Este valor esperado continua igual à esperança marginal $\mu_2 = 50$?

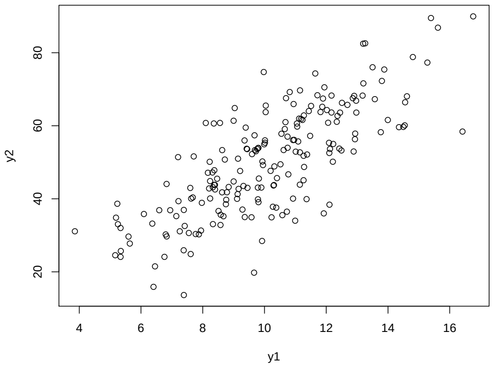


Figura: Amostra de normal bivariada

$\mathbb{E}(Y_2|Y_1 = 14)$

- Queremos ter uma ideia de $\mathbb{E}(Y_2|Y_1 = 14)$.
- Temos $\mu_2 = 50$ mas $\mathbb{E}(Y_2|Y_1 = 14)$ deve ser $> 50 = \mu_2$
- Qual a sua estimativa para $\mathbb{E}(Y_2|Y_1 = 14)$ no olhômetro?
- Suponha que um ponto aleatório será escolhido da distribuição CONDICIONAL de Y_2 dados $Y_1 = 14$.
- $\rightarrow$ o ponto estará na linha vertical $(14, y_2)$.

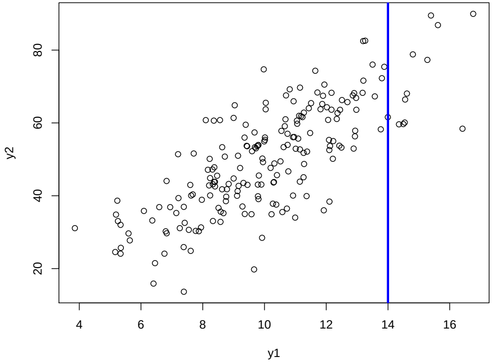


Figura: Amostra de normal bivariada

Distribuição de $(Y_2|Y_1 = 14)$

- Os pontos (y_1, y_2) da amostra que possuem $y_1 \approx 14$ indicam o que deve ser o comportamento probabilístico da v.a. Y_2 DADO que $Y_1 = 14$.
- Vemos que $\mathbb{E}(Y_2|Y_1 = 14) \approx 70$
- Veja que $70 >> 50 = \mathbb{E}(Y_2) = \mu_2$.
- μ_2 é a esperança MARGINAL de Y_2.
- A esperança condicional $\mathbb{E}(Y_2|Y_1 = 14)$ é bem maior que a marginal $\mathbb{E}(Y_2)$.

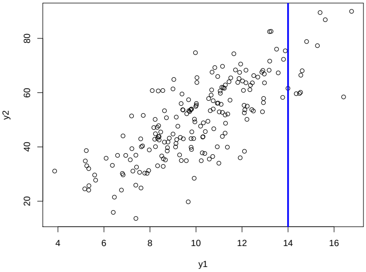


Figura: Amostra de normal bivariada

Distribuição de $(Y_2|Y_1 = 14)$

- Se $\mathbb{E}(Y_2|Y_1 = 14) \approx 70$, quanto é $\sqrt{\mathbb{V}(Y_2|Y_1 = 14)}$?
- Olhando os pontos (y_1, y_2) que possuem $y_1 \approx 14$, qual o tamanho médio dos desvios de Y_2 EM TORNO DE SUA ESPERANÇA $\mathbb{E}(Y_2|Y_1 = 14) \approx 70$??
- Os pontos estão no intervalo de $[50, 80]$ grosseiramente.
- Eu chutaria (ou estimaria) que $\sqrt{\mathbb{V}(Y_2|Y_1 = 14)} \approx (80 - 30)/4 = 7.5$
- Veja que $7.5 << 15 = \sqrt{\mathbb{V}(Y_2)}$, que é o DP marginal de Y_2.

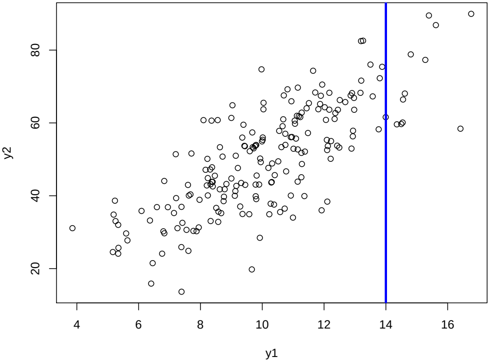


Figura: Amostra de normal bivariada

Dos momentos para a distribuição

- Estes são os dois primeiros MOMENTOS condicionais da v.a. $Y_2|Y_1 = 14$, a esperança e variância condicionais.
- Eles são RESUMOS da distribuição de probabilidade de $Y_2|Y_1 = 14$.
- E qual é a distribuição de probabilidade de $Y_2|Y_1 = 14$?
- Normal? Gama? Uniforme?
- É UMA NORMAL.
- Isto é, $(Y_2|Y_1 = 14) \approx N(70, 7.5^2)$

Dos momentos para a distribuição

- E que tal $Y_2|Y_1 = y$ com y genérico?
- Conseguimos obter uma fórmula geral para expressar qual é esta distribuição genérica.
- Ela depende do coeficiente de correlação ρ que neste exemplo vale $\rho = 0.8$.
- Temos

$$(Y_2|Y_1 = y) \sim N(\mu_{Y_2|Y_1=y}, \sigma^2_{Y_2|Y_1=y})$$

com

$$\mu_{Y_2|Y_1=y} = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(y - \mu_1)$$

e

$$\sigma_{Y_2|Y_1=y} = \sigma_2\sqrt{1-\rho^2}$$

- Por exemplo, com $y_2 = 14$ temos

$$\mu_{Y_2|Y_1=14} = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(14 - \mu_1) = 50 + \frac{0.8 * 15}{2.5}(14 - 10) = 69.2$$

e

$$\sigma_{Y_2|Y_1=y} = \sigma_2\sqrt{1-\rho^2} = 15\sqrt{1-0.8^2} = 9$$

e portanto

$$(Y_2|Y_1 = 14) \sim N(69.2, 9^2)$$

Como sabemos essa fórmula?

- Fazendo o cálculo matemático da densidade condicional:

$$f_{Y_2|Y_1}(y_2|y_1 = a) = \frac{f_{\mathbf{Y}}(a, y_2)}{f_{Y_1}(a)}$$

a partir da densidade conjunta da normal bivariada.

- Para entender esta importante expressão, vamos começar definindo a matriz 2 x 2 simétrica de covariância $\sum$ dada por

$$\sum = \begin{bmatrix} \sigma_1^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

- onde $\rho \in [-1, 1]$ e σ_x e σ_y são os desvios padrões de cada marginal.

Como sabemos essa fórmula?

- Seja o vetor-COLUNA 2×1 das esperanças marginais:

$$\boldsymbol{\mu} = (\mu_1, \mu_2)' = (\mathbb{E}(Y_1), \mathbb{E}(Y_2))$$

- A fórmula geral de uma normal bivariada é igual a

$$f_{\mathbf{Y}}(\mathbf{y}) = \text{ cte} \times \exp\left(- \frac{1}{2} \ d^2(\mathbf{y}, \boldsymbol{\mu}) \right)$$

  onde $d^2(\mathbf{y}, \boldsymbol{\mu})$ é uma medida de distância entre o ponto $\mathbf{y}$ e o vetor esperado μ.

- Esta medida de distância é MUITO importante e não é a distância euclidiana:

$$d^2(\mathbf{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

  Vamos estudá-la a seguir.

Resumo: densidade de normal multivariada

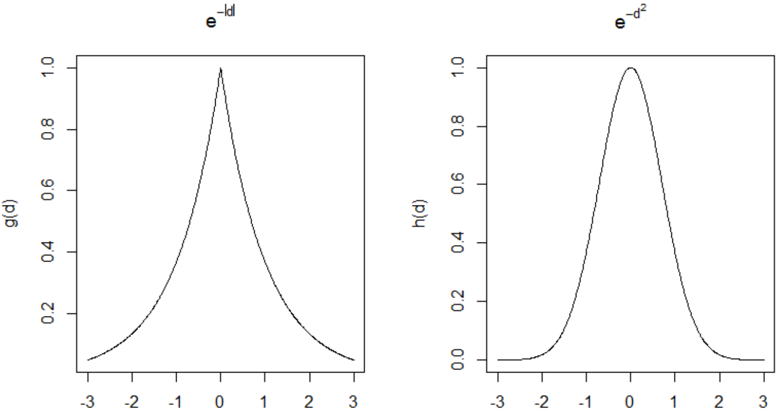
- Em resumo, um vetor normal multivariado tem uma densidade conjunta que é proporcional à exponencial de MENOS uma medida de distância ao quadrado.

$$f_{\mathbf{Y}}(\mathbf{y}) = \text{ cte} \times \exp\left(-\ \frac{1}{2}\ \ d^2(\mathbf{y}, \boldsymbol{\mu})\right)$$

- Densidade decai exponencialmente à medida que a distância AO QUADRADO entre $\mathbf{y}$ e $\boldsymbol{\mu}$ aumenta.

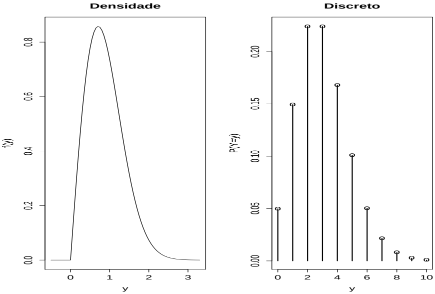
Decaimento exponencial

- Veja o efeito de decair eponencialmente e de decair com exponencialmente AO QUADRADO.
- Decai mais depressa com a distância e o pico é mais suave



Resumos teóricos

- Considere uma única variável aleatória Y.
- Sua distribuição de probabilidade fica determinada por:
  - **Caso Contínuo**: função densidade de probabilidade $f(y)$
  - **Caso Discreto**: função de probabilidade $p(y) = \mathbb{P}(Y = y)$

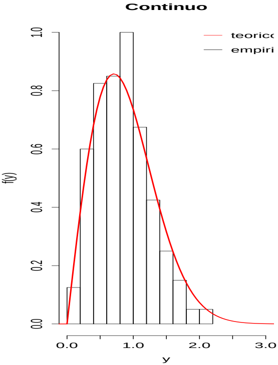


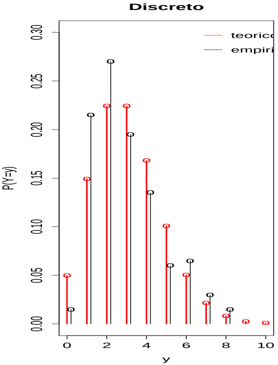
Resumos teóricos

- Podemos **resumir** a sua distribuição de probabilidade com os resumos numéricos (e teóricos) esperança $\mathbb{E}(Y)$ e desvio-padrão $DP_Y = \sqrt{\mathbb{V}(Y)}$.
- Os resumos $\mathbb{E}(Y)$ e $DP_Y = \sqrt{\mathbb{V}(Y)}$ não DEPENDEM de dados estatísticos.
- São resultados de cálculos matemáticos e resumem a DISTRIBUIÇÃO teórica de uma v.a.
- Vamos agora passar a olhar DADOS ESTATÍSTICOS.
- Suponha que temos uma amostra aleatória de Y.
- Isto é, v.a.'s $Y_1, Y_2, \ldots, Y_n$ i.i.d. com a mesma distribuição que Y.
- Estes n números ficam numa das colunas de nossa tabela de dados.

Resumos empíricos, a partir dos dados

- Para ter uma idéia de TODA A distribuição de probabilidade de Y:
  - **Caso Contínuo: histograma**. A altura do histograma em y é $\approx f(y)$.
  - **Caso Discreto: gráfico de barras** com as frequências empíricas n_k/n onde n_k é o número de elementos da amostra iguais a k. Temos $n_k/n \approx \mathbb{P}(Y = k)$.





Contraparte empírica dos resumos

- Podemos ESTIMAR os resumos TEÓRICOS $\mathbb{E}(Y)$ e $\sigma = DP_Y = \sqrt{\mathbb{V}(Y)}$ a partir dos dados.
- Pela Lei dos Grandes Números, se o tamanho n da amostra é grande, temos
  - A média aritmética $\overline{Y} = \overline{(Y_1 + \ldots + Y_n)} \approx \mathbb{E}(Y)$
  - O DP **amostral** $S = \sqrt{\sum_i \left(Y_i - \overline{Y}\right)^2 / n} \approx \sigma$
- Às vezes, define-se S usando $n - 1$ no denominador. A diferença é mínima a não ser que n seja muito pequeno.
- Note que $\overline{Y} \neq \mathbb{E}(Y)$
- E que $S \neq \sigma$
- $\overline{Y}$ e S dependem dos dados e variam de amostra para amostra, MESMO QUE O MECANISMO GERADOR DOS DADOS NÃO MUDE.

Desvio padronizado

- Um **desvio** em relação a $\mu = \mathbb{E}(Y)$: é a v.a. $Y - \mu$
- **Desvio padronizado**: $Z = (Y - \mu)/\sigma$
- Desvio padronizado é medido relativamente ao desvio-padrão σ da v.a. Y.
- Um desvio padronizado $Z = 2$ significa um afastamento de 2 DPs em relação a μ
- Qualquer que seja a distribuição de Y, termos $Z > 4$ é muito raro (pela desigualdade de Tchebyshev).

Duas variáveis

- Como medir a associação entre duas variáveis Y_1 e Y_2 medidas num MESMO item.
- Estas variáveis poderiam ser qualquer par de colunas da nossa tabela de dados.
- Seja $Z_1 = (Y_1 - \mu_1)/\sigma_1$ o desvio padronizado de Y_1
- e $Z_2 = (Y_2 - \mu_2)/\sigma_2$ o desvio padronizado de Y_2
- Quando Z_1 é grande existe alguma TENDÊNCIA de também termos Z_2 grande?
- Se sim, diremos que Y_1 e Y_2 possuem um grau de associação ou correlação.
- Como formalizar este conceito?

Duas variáveis

- Vamos começar com a versão EMPÍRICA da associação.
- Amostra de 147 pessoas (os itens) trabalhando em ocupações fisicamente demandantes.
- Em cada item, medimos o par de variáveis (Y_1, Y_2).
- Y_1 é a força do aperto de mão (ou *grip strength*)
- Y_2 é a força do braço (ou *arm strength*)

Duas variáveis: scatterplot

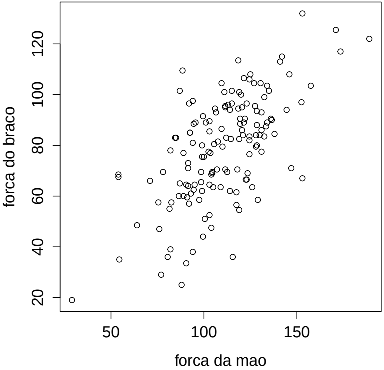


Figura: Relação entre força de preensão (do aperto de mão) e força do braço para 147 pessoas que trabalham em empregos fisicamente extenuantes.

Duas variáveis: scatterplot

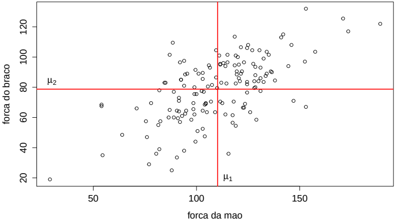


Figura: Linhas verticais indicando aproximadamente μ_1 e μ_2. A maioria dos pontos está nos quadrantes 1 e 3. Quando $Z_1 > 0$, em geral, temos $Z_2 > 0$.

Produto dos desvios

- Existem várias formas intuitivas de medir a associação entre Y_1 e Y_2.
- Uma forma não intuitiva mas que tem excelentes propriedades teóricas é o índice de correlação de Pearson.
- Considere o produto dos desvios padronizados:

$$Z_1 Z_2 = \frac{Y_1 - \mu_1}{\sigma_1} \times \frac{Y_2 - \mu_2}{\sigma_2}$$

- Se desvios grandes e positivos de Y_1 tendem a ocorrer com desvios grandes e positivos de Y_2, seu produto será maior ainda.
- Ao mesmo tempo, se os desvios grandes e negativos de Y_1 tendem a ocorrer com desvios grandes e positivos de Y_2, seu produto será maior ainda.

Produto dos desvios

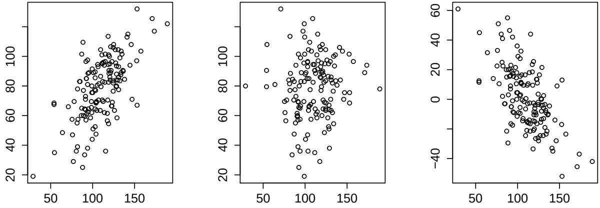


Figura: Tipicamente, em média, o produto dos desvios padronizados $Z_1 Z_2$ é positivo (esquerda), próximo de zero (centro) e negativo (direita).

Natureza de $Z_1 Z_2$

- (Y_1, Y_2) é um vetor aleatório: duas v.a.'s são medidas no mesmo item.
- Considere
$$Z_1 Z_2 = \frac{Y_1 - \mu_1}{\sigma_1} \times \frac{Y_2 - \mu_2}{\sigma_2}$$
- O que é μ_1? Uma constante? Uma v.a.?
- O mesmo vale para μ_2, σ_1 e σ_2.
- E o produto $Z_1 Z_2$?
- É uma constante?
- Uma v.a.?

Correlação

- O produto $Z_1 Z_2$ é uma v.a. !!
- Como resumir esta v.a. num único número?
- Já sabemos fazer isto com QUALQUER v.a.: tome o seu valor esperado.
- Isto é, vamos calcular

$$\rho = \text{Corr}(Y_1, Y_2) = \mathbb{E}\left(Z_1 Z_2\right) = \mathbb{E}\left(\frac{Y_1 - \mu_1}{\sigma_1} \times \frac{Y_2 - \mu_2}{\sigma_2}\right)$$

- Este resumo é o índice de correlação de Pearson.

Propriedades de ρ

- ρ está SEMPRE entre -1 e 1.
- Esta é uma das razões para usar ρ como medida de associação entre Y_1 e Y_2: uma escala fixa em qualquer problema.
- Além disso, pela definição,

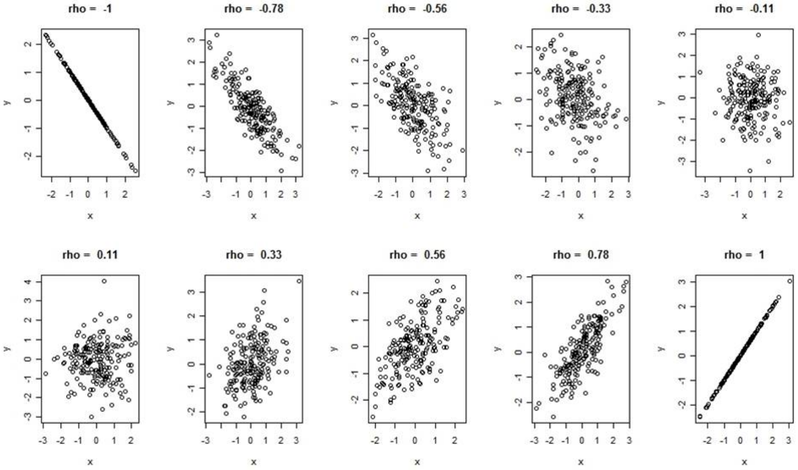
$$\text{Corr}(Y_1, Y_2) = \mathbb{E}\left(Z_1 Z_2\right) = \text{Corr}(Y_2, Y_1)$$

- Também temos que $\text{Corr}(Y, Y) = 1$: a correlação de uma v.a. consigo mesma é 1.

Propriedades de ρ

- Se Y_1 é uma v.a. independente da v.a. Y_2 então $\rho = 0$.
- Neste caso, uma amostra de valores do vetor (Y_1, Y_2) formará um gráfico de dispersão com forma indistinta, uma nuvem sem inclinação.
- Se $\rho \approx \pm 1$ então Y_2 é aproximadamente uma função linear perfeita de Y_1.
- Isto é, uma amostra de valores do vetor (Y_1, Y_2) formará uma gráfico de dispersão na forma aproximada de uma linha reta.

Amostras do vetor (Y_1, Y_2) com diferentes ρ



Correlation Matrix

- Correlação é uma medida de associação entre DUAS v.a.'s
- E quando tivermos p v.a.'s simultaneamente, todas medidas no mesmo item?
- Suponha que tenhamos um vetor $(Y_1, Y_2, \ldots, Y_p)$ de v.a.'s
- Podemos fazer uma matriz $p \times p$ de correlação.
- Na posição (i, j) teremos

$$\rho_{ij} = \text{Corr}(Y_i, Y_j) = \mathbb{E}\left(\frac{Y_i - \mu_i}{\sigma_i} \times \frac{Y_j - \mu_j}{\sigma_j}\right)$$

- Como $\text{Corr}(Y_i, Y_j) = \text{Corr}(Y_j, Y_i)$ a matriz é simétrica.
- E como $\text{Corr}(Y_i, Y_i) = 1$ a diagonal principal é toda de 1's.

Exemplo: Correlation Matrix

- Temos um vetor aleatório $(Y_1, Y_2, \ldots, Y_9)$.
- As 9 variáveis aleatórias são escores obtidos em 9 testes de habilidade cognitiva, todos aplicados num mesmo indivíduo.
- As v.a.'s são as seguintes:
  - 3 v.a.'s medindo habilidade verbal: Word Meaning, Sentence Completion, and Odd words;
  - 3 v.a.'s medindo habilidade quantitativa: Mixed Arithmetic, Remainders, and Missing numbers;
  - 3 v.a.'s medindo habilidade espacial: Gloves, Boots, and Hatchets.
- Como poderia ser a matriz de correlação 9×9 entre estas v.a's?

Matriz de correlação 9×9

Variable	1	2	3	4	5	6	7	8	9
WrdMean	1								
SntComp	0.75	1							
OddWrds	0.78	0.72	1						
MxdArit	0.44	0.52	0.47	1					
Remndrs	0.45	0.53	0.48	0.82	1				
MissNum	0.51	0.58	0.54	0.82	0.74	1			
Gloves	0.21	0.23	0.28	0.33	0.37	0.35	1		
Boots	0.30	0.32	0.37	0.33	0.36	0.38	0.45	1	
Hatchts	0.31	0.30	0.37	0.31	0.36	0.38	0.52	0.67	1

WrdMean, word meaning; SntComp, sentence completion; OddWrds, odd words; MxdArit, mixed arithmetic; Remndrs, remainders; MissNum, missing numbers, Hatchts, hatchets.

Figura: Correlações entre pares formados a partir de 9 medidas feitas num mesmo indivíduo em um teste de personalidade

Visualizando a matriz de correlação

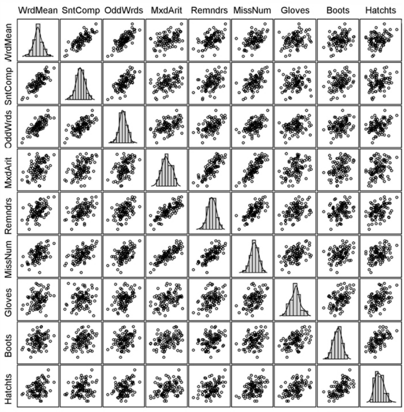


Figura: Amostra de 244 indivíduos e scatterplots dos pares de suas 9 medidas no teste de habilidade cognitiva

Visualizando outra matriz de correlação

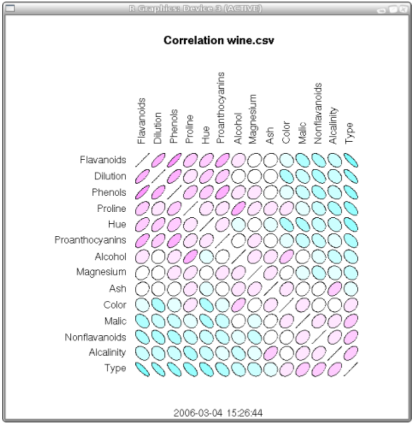


Figura: A partir de dados amostrais mostra-se o FORMATO da nuvem de pontos de uma amostra de VINHOS com 14 variáveis medidas em cada um dos vinhos. Gráfico em R + rattle

Mais uma visualização

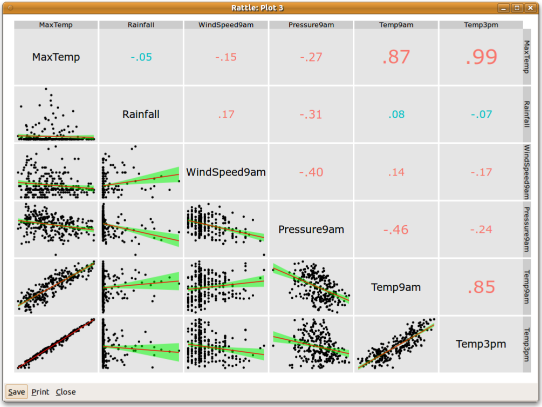


Figura: Mais uma visualização com R + rattle

Uma visualização com MUITAS v.a.'s: rede

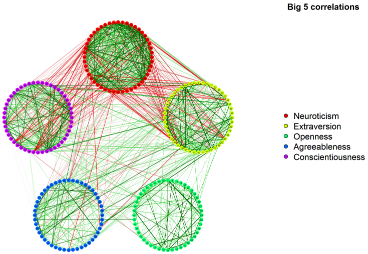


Figura: Uma visualização com qgraph: v.a.'s são vértices e correlações são arestas. Verde = correlação positiva e vermelha = negativa. As arestas mais grossas e saturadas tem $|\rho|$ grande.

Nem sempre, os gráficos são simples

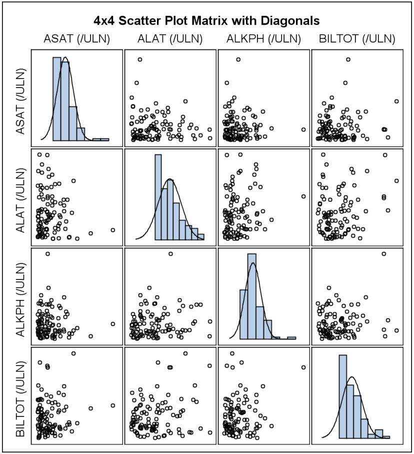


Figura: Scatterplot matrix of 4 lab variables to test liver functioning commonly used in clinical research

Misturas de diferentes populações

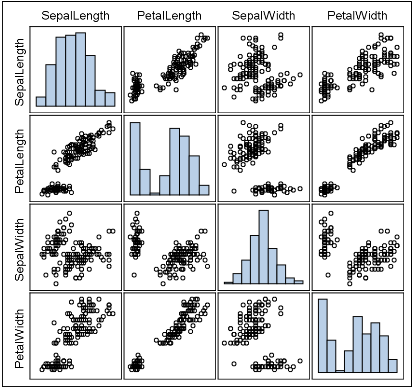


Figura: Scatterplot matrix de 4 variáveis medidas numa flor: comprimento de pétala, largura de pétala, comprimento de sétala, largura de sétala. Três espécies distintas misturadas. Relação entre as variáveis é diferente, ela depende da espécie.

Propriedades de ρ

- Se $\rho = -1$ ou $\rho = +1$, podemos predizer o valor de Y_2 como função linear de Y_1, sem erro, de forma perfeita.
- Isto é, se $\rho = \pm 1$, temos $Y_2 = \alpha + \beta Y_1$
- Se $\rho = 0$ *pode acontecer* que Y_1 seja fortemente relacionada a Y_2 de uma forma não-linear. São casos raros na prática.

Propriedades de ρ

- ρ é invariante por mudança linear de escala.
- Por ex, Y_1 é o estoque de café num certo mês e Y_2 é o preço do café em reais no mesmo mês.
- Seja $\rho = \text{Corr}(Y_1, Y_2)$.
- Suponha que outra variável seja usada: o preço Y_3 do café em dólares e que $Y_3 = 2.3 Y_2$.
- Neste caso,

$$\text{Corr}(Y_1, Y_3) = \text{Corr}(Y_1, 2.3 Y_2) = \text{Corr}(Y_1, Y_2)$$

- Do mesmo modo, se medirmos em graus centígrados (Y_2) ou em graus Farenheit ($Y_3 = 32 + 1.8 Y_2$), a correlação com outra variável Y_1 é

$$\text{Corr}(Y_1, Y_3) = \text{Corr}(Y_1, 32 + 1.8 Y_2) = \text{Corr}(Y_1, Y_2)$$

Estimando ρ

- ρ é um resumo teórico da distribuição CONJUNTA de duas v.a.'s
- Ele não depende de dados para ser obtido, é uma conta matemática.
- Relembre a definição:

$$\rho = \text{Corr}(Y_1, Y_2) = \mathbb{E}\left(\frac{Y_1 - \mu_1}{\sigma_1} \times \frac{Y_2 - \mu_2}{\sigma_2}\right)$$

- Precisa de $\mu_1 = \mathbb{E}(Y_1)$, $\sigma_1^2 = \mathbb{V}(Y_1)$, etc.
- Em seguida precisa calcular (usando teoria de probabilidade) o valor esperado do produto dos desvios.
- Para várias distribuições, esta conta matemática é inviável (não-analítica).
- No entanto, com dados, podemos estimar ρ.

Estimando ρ

- Como $\overline{Y} \approx \mathbb{E}(Y)$
- e como $S = \sqrt{\sum_i \left(Y_i - \overline{Y}\right)^2 / n} \approx \sigma$
- podemos aproximar

$$\rho = \mathbb{E}\left(\frac{Y_1 - \mu_1}{\sigma_1} \times \frac{Y_2 - \mu_2}{\sigma_2}\right) \approx \mathbb{E}\left(\frac{Y_1 - \overline{Y_1}}{S_1} \times \frac{Y_2 - \overline{Y_2}}{S_2}\right)$$

- onde $\overline{Y_1}$ é a média aritmética dos n valores da variável 1, etc.
- é, $\overline{Y_1}$ é média aritmética da coluna associada com a variável 1 na tabela de dados.
- Mas ainda precisaríamos calcular uma esperança matemática que é inviável na maioria das distribuições.

Desvio padronizado empírico

- Solução: calcule o desvio realizado de cada um dos n valores das duas variáveis.
- Para a variável 1 com os n valores $y_{11}, \ldots, y_{n1}$ da coluna 1 da tabela, calcule a nova coluna formada por

$$z_{i1} = \frac{y_{i1} - \overline{y_1}}{s_1}$$

- Faça o memso para a coluna 2, criando uma outra coluna de desvios padronizados empíricos:

$$z_{i2} = \frac{y_{i2} - \overline{y_2}}{s_2}$$

- A seguir, multiplique as duas colunas de desvios padronizados e tire a sua média aritmética calculando

$$r = \frac{1}{n} \sum_{i=1}^{n} z_{i1} z_{i2} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_{i1} - \overline{y_1}}{s_1} \right) \left(\frac{y_{i2} - \overline{y_2}}{s_2} \right)$$

Pressão sistólica

- A pressão sistólica mede a força do sangue nas artérias, à medida que o coração contrai para impulsionar o sangue através do corpo.
- Se alta, ela pode levar a doença de coração, angina e doenças vasculares nas pernas.
- Pressao sistólica saudável: entre 120 e 140 mm Hg
- Pressão sistólica $>$ 140 mm Hg: não saudável
- Pressão diastólica: deve ficar em torno de 80.
- Acima de 100 não é saudável.

Pressão sistólica e diastólica em amostra

- Amostra de 250 indivíduos (instâncias)
- Pressão em duas medições (atributos):
  - Diastólica
  - sistólica
- Como fica o gráfico dos atributos dessas 250 instâncias?

Pressão de 250 indivíduos

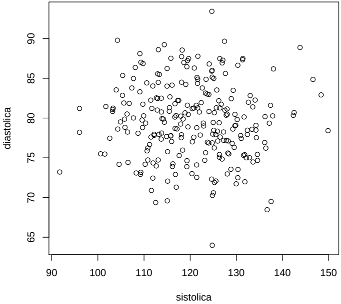


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Médias para referência

- 250 instâncias do vetor aleatáorio: $\boldsymbol{Y} = (Y_1, Y_2)$
- Vetor com os valores esperados de cada variável do vetor:

$$\mathbb{E}(\boldsymbol{Y}) = \mathbb{E}(Y_1, Y_2) = (\mathbb{E}(Y_1), \mathbb{E}(Y_2)) = (\mu_1, \mu_2) = \boldsymbol{\mu}$$

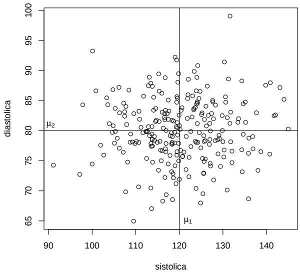
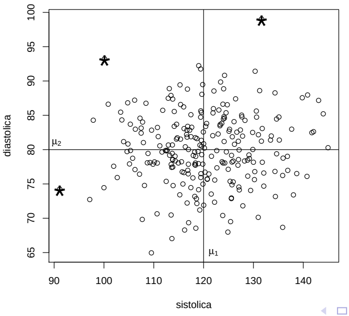


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Quem está distante do centro?

- Centro $\boldsymbol{\mu} = (\mu_1, \mu_2)$ é o perfil esperado ou típico.
- Quem está longe do perfil típico? Quem é anômalo?
- Medida baseada na distância euclidiana
  $d(y_1, y_2) = \sqrt{(y_1 - 120)^2 + (y_2 - 80)^2}$
- É razoável?



Exagerando um pouco...

- E se o segundo atributo for assim? Fazendo o aspect/raio $= 1$.
- Centro $\boldsymbol{\mu} = (\mu_1, \mu_2)$ continua o mesmo.
- Mas quem está distante do centro agora? Quem é anômalo?

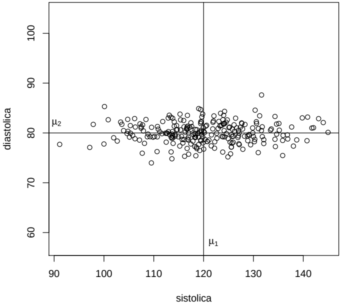


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Distantes são óbvios, não?

- Mas qualé a medida de distância que estamos usando implicitamente, sem nem mesmo perceber?
- Não é a distância euclidiana!

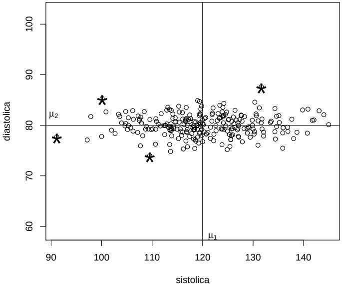
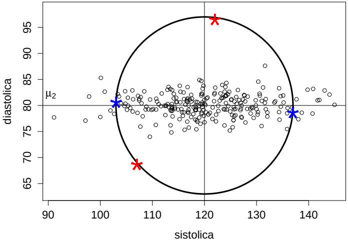


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Pontos à igual distância?

- Todos os pontos do círculo estão à mesma distância euclidiana do centro da nuvem de pontos.
- Queremos os dois pontos em vermelho à igual distância ESTATÍSTICA do centro queos pontos em azul?
- NÃO!!! Pontos vermelhos estão ESTATISTICAMENTE muito mais distantes do centro (μ_1, μ_2) do que os pontos azuis.



Pontos vermehos mais distantes

- Como fazer os pontos vermelhos mais distantes que os pontos azuis?
- Andar poucas unidades na direção norte-sul te leva para fora da nuvem de pontos (vira anomalia).
- Precisa andar MAIS unidades na direção leste-oeste para sair fora da nuvem de pontos.

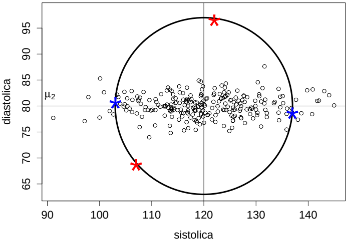


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Pontos vermehos mais distantes

- Então N unidades eucidlianas na direção leste-oeste VALEM O MESMO que N/k na direção norte-sul (onde $k > 1$).
- Como achar este k?
- Como equalizar as distâncias?
- RESPOSTA: Medindo distâncias em unidades de DESVIOS-PADRÃO.

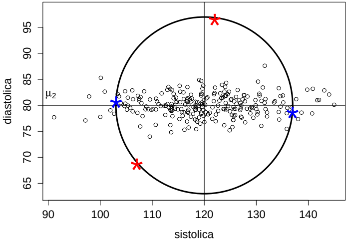


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Medida de dispersão

- Desvio-padrão DP: um para cada eixo, um DP para cada atributo.
- DP mede quanto, em média, um atributo aleatório desvia-se de seu valor esperados
- Por exemplo, $DP = 10$ significa:
  - Em geral, observações desviam-se de 10 unidades em torno de seu valor esperado
  - Às vezes mais de 10 unidades; às vezes, menos de 10 unidades
  - Em média, um afastamento de 10 unidades: isto é o DP.

Qual o desvio padrão de cada variável?

- Centro $\mathbb{E}(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1, \mu_2) = (120, 80)$
- $DP_1 = \sigma_1 = ??$
- $DP_2 = \sigma_2 = ??$

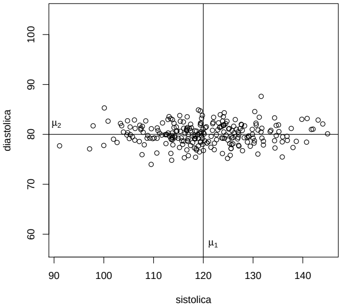


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Qual o desvio padrão de cada variável?

- Centro $\mathbb{E}(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1, \mu_2) = (120, 80)$
- $DP_1 = \sigma_1 = 10$
- $DP_2 = \sigma_2 = 2$

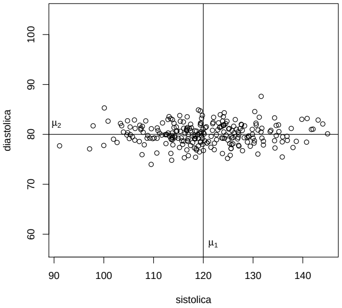
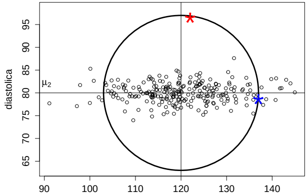


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Distância medida em DP

- $(\mu_1, \mu_2) = (120, 80)$ e $(\sigma_1, \sigma_2) = (10, 2)$
- AZUL: afastou-se do centro apenas ao longo do eixo 1 e afastou-se 15 unidades ou $1.5\sigma_1$)
- VERMELHO: afastou-se do centro apenas ao longo do eixo 2 e afastou-se 15 unidades ou $7.5\sigma_2$)
- O ponto VERMELHO está muito mais distante do centro em termos de DPs.
- Mas como fazer com pontos que afastam-se do centro não somente ao longo de um dos eixos?



Distância medida em DP

- $(\mu_1, \mu_2) = (120, 80)$ e $(\sigma_1, \sigma_2) = (10, 2)$
- Andar $n\sigma_1$ ao longo do eixo 1 É EQUIVALENTE a andar $n\sigma_2$ no eixo 2.
- Por exemplo, 20 unidades ao longo do eixo 1 (ou $2\sigma_1$) é ESTATISTICAMENTE EQUIVALENTE a 4 unidades ao (ou $2\sigma_2$) longo do eixo 2.

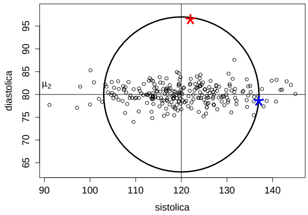


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Distância medida em DP

- Vamos medir o desvio em cada eixo EM UNIDADES DE SEU DESVIO-PADRÃO e calcular a distância com estes desvios padronizados.
- DESVIO PADRONIZADO ao longo do eixo 1: $z_1 = \frac{y_1 - \mu_1}{\sigma_1} = \frac{y_1 - 120}{10}$
- DESVIO PADRONIZADO ao longo do eixo 2: $z_2 = \frac{y_2 - \mu_2}{\sigma_2} = \frac{y_2 - 80}{2}$
- Distância:

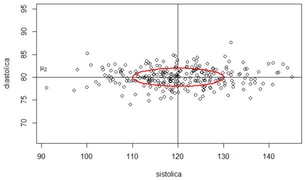
$$
\begin{aligned}
d(\boldsymbol{y}, \boldsymbol{\mu}) &= \sqrt{z_1^2 + z_2^2} \\
&= \sqrt{\left(\frac{y_1 - 120}{10}\right)^2 + \left(\frac{y_2 - 80}{2}\right)^2} \\
&= \sqrt{\left(\frac{y_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2 - \mu_2}{\sigma_2}\right)^2}
\end{aligned}
$$

Pontos a igual distância

- NESTA NOVA MÉTRICA, quais os pontos (y_1, y_2) que estão a uma MESMA distância do centro (μ_1, μ_2)?
- Tome uma distância fixa (por exemplo, 1).
- Eles formam uma ELIPSE centrada em (μ_1, μ_2) e com eixos paralelos aos eixos coordenados.

$$d(\mathbf{y}, \boldsymbol{\mu}) = \sqrt{\left(\frac{y_1 - 120}{10}\right)^2 + \left(\frac{y_2 - 80}{2}\right)^2} = 1$$

- Os pontos que satisfazem a equação acima formam uma elipse (esta é a equação de uma elipse).



Tamanhos dos eixos

- Distância $c > 0$ do centro: pontos satisfazem a equação

$$d(\boldsymbol{y}, \boldsymbol{\mu}) = \sqrt{\left(\frac{y_1 - 120}{10}\right)^2 + \left(\frac{y_2 - 80}{2}\right)^2} = c$$

- Os eixos têm comprimentos iguais a $c\sigma_1$ e $c\sigma_2$. O eixo maior da elipse: variável com maior DP.
- Quantas vezes maior é o eixo maior em relação ao eixo menor?
- Se σ_1 é o maior DP,

$$\frac{\text{eixo maior}}{\text{eixo menor}} = \frac{c\sigma_1}{c\sigma_2} = \frac{\sigma_1}{\sigma_2}$$

- Não depende da distância c: variando c, teremos elipses concêntricas.

Variando a distância

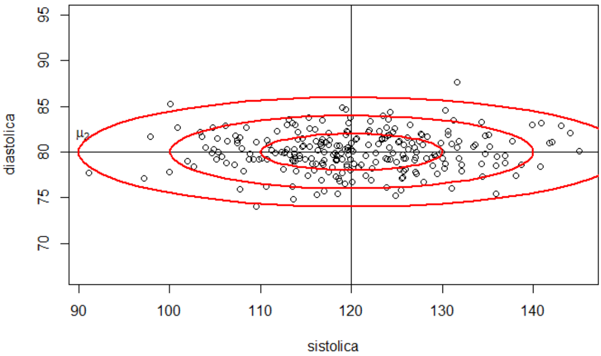


Figura: Pontos (y_1, y_2) que estão a uma distância c igual a 1, 2 ou 3 do centro (μ_1, μ_2). Isto é, os pontos de cada elipse satisfazem $d(\boldsymbol{y}, \boldsymbol{\mu}) = c$ para diferentes c's.

Jogando fora a raiz quadrada

- Preferimos trabalhar com a distância AO QUADRADO
- E se podemos complicar, por quê simplificar?

$$
\begin{aligned}
d^2(\boldsymbol{y}, \boldsymbol{\mu}) &= \left(\frac{y_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2 - \mu_2}{\sigma_2}\right)^2 \\
&= (y_1 - \mu_1, y_2 - \mu_2) \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix} \\
&= (y_1 - \mu_1, y_2 - \mu_2) \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix} \\
&= \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix} \\
&= (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})
\end{aligned}
$$

Elipses e distâncias

- Vimos que

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = \left(\frac{y_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2 - \mu_2}{\sigma_2}\right)^2 = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

 onde

$$\sum = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

 é a equção de uma elipse centrada no vetor $\boldsymbol{\mu} = (\mu_1, \mu_2)$.
- Quando a matriz $\sum$ é DIAGONAL com elementos positivos (com as variâncias σ_i's), então a elipse tem eixos paralelos aos eixos e o tamanho de cada eixo é porporcional ao σ_i da variável associada.

Caso mais realista

- Variáveis são associadas, não são independentes
- Dizemos que são correlacionadas: redundância da informação
- O valor de uma variável dá informação sobre o valor da variável
- Pode-se predizer (com algum erro) uma variável em função da outra

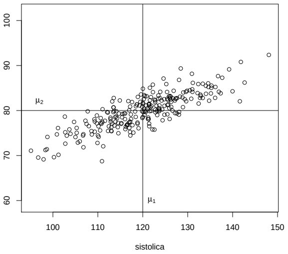
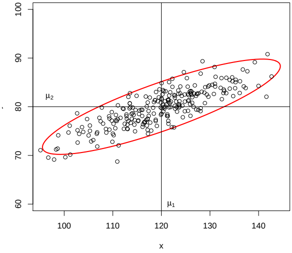


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Distância elíptica, e não circular

- Pelo mesmo raciocínio intuitivo que fizemos antes, os pontos na ELIPSE abaixo tendem a estar a igual distância do perfil esperado $\mathbb{E}(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1, \mu_2)$.
- Pontos estatisticamente equidistantes de $\boldsymbol{\mu}$ NÃO estão mais numa elipse paralela aos eixos.
- A elipse está inclinada seguindo a associação entre as variáveis.



Forma Quadrática

- Medida de distância é uma FORMA QUADRÁTICA.

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

- É a mesma expressão matricial de distância que usamos antes MAS...
- ...a matriz Σ não é mais DIAGONAL

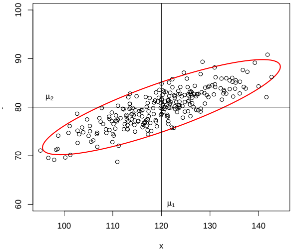


Figura: Amostra de (y_{i1}, y_{i2}) com $i = 1, 2, \ldots, 250$.

Quem é $\sum$?

- Medida de distância é uma FORMA QUADRÁTICA:

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$

- Matriz $\sum$ é matriz 2 x 2 simétrica chamada de matriz de covariância

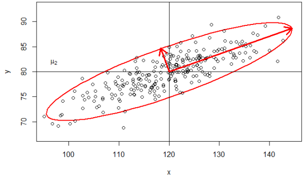
$$\sum = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

onde $\rho = \text{Corr}(Y_1, Y_2)$ é o índice de correlação de Pearson entre Y_1 e Y_2.

- Temos sempre $-1 \leq \rho \leq 1$.
- Os elementos fora da diagonal, $\rho\sigma_1\sigma_2$, são chamados de Covariância entre Y_1 e Y_2.
- Costumamos escrever $\text{Cov}(Y_1, Y_2) = \rho\sigma_1\sigma_2 = \sigma_{12}$

Relação entre $\sum$ e a elipse

- Distância é
$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$
onde a matriz $\sum$ é 2×2 simétrica e dada por
$$\sum = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$
- Pontos equidistantes de $\boldsymbol{\mu} = (\mu_1, \mu_2)$ estão numa elipse.
- Eixos da elipse: na direção dos AUTOVETORES da matriz Σ^{-1}.
- O tamanho de cada eixo é proporcional à raiz do AUTOVALOR correspondente.



Autovetor e autovalor de Σ^{-1}

- Definição: autovetor de uma matriz quadrada A é um vetor $\mathbf{y}$ tal que

$$A \cdot \mathbf{y} = \lambda \mathbf{y}$$

  onde λ é uma constante (pode até ser um número complexo).
- A constante λ é chamada de autovalor associado ao autovetor $\mathbf{y}$.
- Na nossa situação de distância estatística em que usamos a inversa da matriz de covariância como Σ^{-1} temos dois resultados especiais:
  - sempre temos dois autovetores ORTOGONAIS entre si.
  - autovalores são sempre REAIS E POSITVOS (e portanto podemos tomar sua raiz ou invertê-los).
- Voltaremos a este importante resultado daqui a pouco.

Autovetores de Σ e Σ^{-1}

- Os autovetores de Σ e Σ^{-1} são os mesmos.
- Prova: Suponha que **v**é autovetor de Σ com autovalor $\lambda > 0$:

$$\Sigma \cdot \mathbf{v} = \lambda \, \mathbf{v}$$

- Multiplique dos dois lados (pela esquerda) por Σ^{-1}:

$$\Sigma^{-1}\Sigma \cdot \mathbf{v} = \Sigma^{-1}(\lambda \, \mathbf{v})$$

ou seja

$$\mathbf{v} = \lambda\Sigma^{-1} \, \mathbf{v}$$

ou ainda, como $\lambda > 0$,

$$\frac{1}{\lambda} \, \mathbf{v} = \lambda\Sigma^{-1} \, \mathbf{v}$$

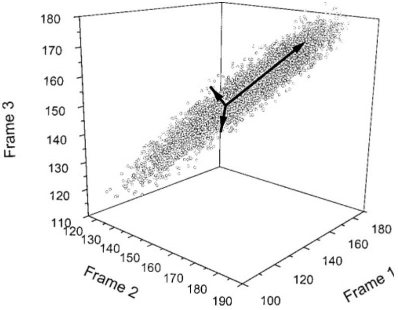
Distância estatística em k dimensões

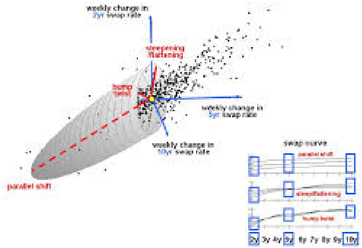
- Seja $\mathbf{Y} = (Y_1, \ldots, Y_k)$ um vetor aleatório de dimensão k.
- Seja $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ seu VETOR-COLUNA de valores esperados
- Seja $\sum$ a matriz $k \times k$ com a covariância $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ onde ρ_{ij} é a correlação entre Y_i e Y_j.
- Distância estatística:

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

- Pontos equidistantes de $\boldsymbol{\mu}$ formam um elipsóide em k dimensões com eixos nas direções dos autovetores de $\sum$ e com tamanhos proporcionais aos seus respectivos autovetores.

Caso 3-dim


Frame 3
180
170
160
150
140
130
120
110
Frame 2
120
130
140
150
160
170
180
190
Frame 1
100
120
140
160
180


weekly change in
2yr swap rate
steepening
flattening
weekly change in
5yr swap rate
weekly change in
10yr swap rate
parallel shift
swap curve
2y 3y 4y 5y 6y 7y 8y 9y 10y

Distância estatística em k dimensões

- $\sum$ é a matriz de covariância $k \times k$ de um vetor aleatório $\mathbf{Y}$ de dimensão k, temos:
  - sempre temos k autovetores ORTOGONAIS entre si.
  - autovalores são sempre REAIS E POSITVOS.
- Esta afirmação é uma consequência do teorema espectral de álgebra linear.
- Para todo ponto $\boldsymbol{y}$ que não seja o vetor esperado $\boldsymbol{\mu}$, queremos que a distância $d^2(\boldsymbol{y}, \boldsymbol{\mu})$ seja > 0.
- Uma matriz com esta propriedade é chamada de *positiva definida*.

Resumo de normal multivariada

- O vetor aleatório $\mathbf{Y} = (Y_1, \ldots, Y_k)$ segue uma distribuição normal (ou gaussiana) multivariada se sua densidade conjunta for da forma

$$f_{\mathbf{Y}}(\mathbf{y}) = \text{ cte} \times \exp\left(-\frac{1}{2}\ d^2(\mathbf{y}, \boldsymbol{\mu})\right)$$

  onde

$$d^2(\mathbf{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$

  e $\sum$ é a matriz de covariância entre as variáveis e $\boldsymbol{\mu}$ é o vetor de valores esperados.
- Notação: $\mathbf{Y} \sim N_k(\mu, \sum)$
- A densidade decresce com d^2. As superfícies de nível da densidade são elipsoides concêntricos centrados em $\boldsymbol{\mu}$.
- Os eixos do elipsóide estão na direçãodos autovetores de $\sum$ e com comprimentos proporcionaisa raiz do autovalor.

Simulando uma normal multivariada

- Sabemos gerar gaussiana univariada $N(0,1)$ com média $\mu = 0$ e DP $\sigma = 1$.
- Basta usar o algoritmo de Box-Muller (já vimos).
- Então sabemos gerar $\mathbf{Z} = (Z_1, \ldots, Z_k)$ INDEPENDENTES e IDENTICAMENTE DISTRIBUÍDOS onde cada $Z_i \sim N(0,1)$.
- Podemos passar de $\mathbf{Z}$ para um vetor $\mathbf{Y} \sim N_k(\mu, \sum)$ apenas manipulando matrizes.
- Seja L uma matriz $k \times k$ tal que $LL^t = \sum$.
- Calcule $\mathbf{Y} = \boldsymbol{\mu} + L\mathbf{Z}$.
- Temos $\mathbf{Y} \sim N_k(\mu, \sum)$

Como achar L

- Precisamos achar uma matriz L que seja $k \times k$ e tal que $LL^t = \sum$.
- É como se estivéssemos achando uma espécie de raiz quadrada de $\sum$.
- Quem é esta L? Qualquer uma que satisfaą $LL^t = \sum$.
- OK, mas como achar uma dessas?
- Pela *decomposição de Cholesky*: uma matriz *simétrica* e *postiva definida* $\sum$ possui uma matriz L triangular inferior tal que $LL^t = \sum$.
- Em R: `t(chol(A))` (precisa transpor pois a saída de `chol` é triangular superior)

Algoritmo para decomposição de Cholesky

$$
\begin{aligned}
A = LL^t \quad &= \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix} \\
&= \begin{pmatrix} l_{11}^2 & l_{21} l_{11} & l_{31} l_{11} \\ l_{21} l_{11} & l_{21}^2 + l_{22}^2 & l_{31} l_{21} + l_{32} l_{22} \\ l_{31} l_{11} & l_{31} l_{21} + l_{32} l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}
\end{aligned}
$$

- O que deve ser o valor l_{11}?
- Iguale o elemento 11 da matriz A e o elemento 11 da matriz produto LL^t.
- Vemos que $l_{11}^2 = a_{11}$, ou seja, $l_{11} = \sqrt{a_{11}}$.

Algoritmo para decomposição de Cholesky

$$
\begin{aligned}
A = LL^t &= \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix} \\
&= \begin{pmatrix} l_{11}^2 & * & * \\ l_{21} l_{11} & l_{21}^2 + l_{22}^2 & * \\ l_{31} l_{11} & l_{31} l_{21} + l_{32} l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}
\end{aligned}
$$

- Iguale o elemento 21 de A com o elemento 21 de $L^t L$.
- Temos $l_{21} l_{11} = a_{21}$. Como já obtivemos $l_{11} = \sqrt{a_{11}}$, encontramos $l_{21} = a_{21} / \sqrt{a_{11}}$.

Algoritmo para decomposição de Cholesky

$$
\begin{aligned}
A = LL^t &= \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix} \\
&= \begin{pmatrix} l_{11}^2 & & \\ l_{21} l_{11} & l_{21}^2 + l_{22}^2 & \\ l_{31} l_{11} & l_{31} l_{21} + l_{32} l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}
\end{aligned}
$$

- De maneira idêntica, obtemos $l_{31} = a_{31}/\sqrt{a_{11}}$.
- Primeira coluna de L está pronta. Vamos agora obter a segunda coluna de L.
- Iguale o elemento a_{22} e o elemento 22 do produto LL^t: $a_{22} = l_{21}^2 + l_{22}^2)$.
- Assim, $l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{a_{22} - a_{21}^2/a_{11}}$.

Algoritmo para decomposição de Cholesky

$$A = LL^t = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

$$= \begin{pmatrix} l_{11}^2 & & \\ l_{21} l_{11} & l_{21}^2 + l_{22}^2 & \\ l_{31} l_{11} & l_{31} l_{21} + l_{32} l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix}$$

- Tendo obtido $l_{11}, l_{21}, l_{31}, l_{22}$, passamos agora a l_{32}
- Iguale o elemento a_{32} e o elemento 32 do produto LL^t
- E assim sucessivamente.

$$\begin{pmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

- primeira coluna de L

$$\begin{pmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 3 & l_{22} & 0 \\ -1 & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} 5 & 3 & -1 \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

- conclusão:

$$\begin{pmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{pmatrix} \begin{pmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- Suponha que queremos gerar uma amostra de 200 instâncias de um vetor $\mathbf{Y}$ gaussiano multivariado de dimensão 4 tal que $\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \sum)$.
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_4) = (0, 10, 0, 1020)$
- A matriz de covariância:

$$\sum = \begin{pmatrix} 1.0 & -0.1 & 0 & -0.7 \\ -0.1 & 16 & -0.4 & 20 \\ 0 & -0.4 & 0.04 & -0.1 \\ -0.7 & 20 & -0.1 & 64 \end{pmatrix}$$

- Veja que a raiz quadrada da diagonal fornece os desvios-padrão:

$$\sqrt{\text{diag}(\sum)} = \sqrt{(1.0, 16, 0.04, 64)} = (1.0, 4, 0.2, 8)$$

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- Como a covariância está associada com a correlação e os desvios-padrão,

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

podemos escrever a matriz de covariância como resultado de manipular matricialmente a matriz de correlação:

$$\sum = \mathbf{V}^{1/2}\ \boldsymbol{\rho}\ \mathbf{V}^{1/2}$$

onde $\mathbf{V}$ é uma matriz diagonal com as variâncias $(\sigma_1^2, \ldots, \sigma_k^2)$ e $\boldsymbol{\rho}$ é uma matriz quadrada com as correlações ρ_{ij}.

- Portanto, podemos também escrever

$$\boldsymbol{\rho} = \left(\mathbf{V}^{1/2}\right)^{-1} \sum \left(\mathbf{V}^{1/2}\right)^{-1}$$

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- No nosso exemplo:

$$\sum = \begin{pmatrix} 1.0 & -0.1 & 0 & -0.7 \\ -0.1 & 16 & -0.4 & 20 \\ 0 & -0.4 & 0.04 & -0.1 \\ -0.7 & 20 & -0.1 & 64 \end{pmatrix} = \mathbf{V}^{1/2}\ \boldsymbol{\rho}\ \mathbf{V}^{1/2}$$

onde

$$\mathbf{V}^{1/2} = \begin{pmatrix} \sqrt{1} & 0 & 0 & 0 \\ 0 & \sqrt{16} & 0 & 0 \\ 0 & 0 & \sqrt{0.04} & 0 \\ 0 & 0 & 0 & \sqrt{64} \end{pmatrix}$$

e

$$\boldsymbol{\rho} = \begin{pmatrix} 1.0000 & -0.0250 & 0.0000 & -0.0875 \\ -0.0250 & 1.0000 & -0.5000 & 0.6250 \\ 0.0000 & -0.5000 & 1.0000 & -0.0625 \\ -0.0875 & 0.6250 & -0.0625 & 1.0000 \end{pmatrix}$$

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- Na direção inversa

$$\rho = \begin{pmatrix} 1.0000 & -0.0250 & 0.0000 & -0.0875 \\ -0.0250 & 1.0000 & -0.5000 & 0.6250 \\ 0.0000 & -0.5000 & 1.0000 & -0.0625 \\ -0.0875 & 0.6250 & -0.0625 & 1.0000 \end{pmatrix} = \mathbf{V}^{-1/2} \sum \mathbf{V}^{-1/2}$$

onde $\mathbf{V}^{-1/2} = \left(\mathbf{V}^{1/2}\right)^{-1}$ é dada por

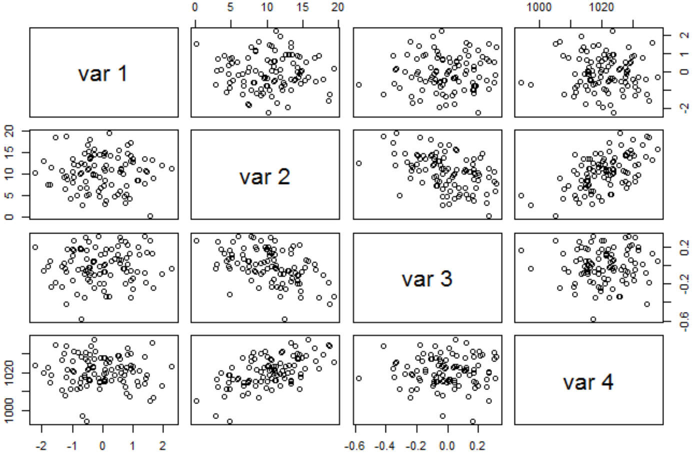
$$\begin{pmatrix} \sqrt{1} & 0 & 0 & 0 \\ 0 & \sqrt{16} & 0 & 0 \\ 0 & 0 & \sqrt{0.04} & 0 \\ 0 & 0 & 0 & \sqrt{64} \end{pmatrix}^{-1} = \begin{pmatrix} 1.000 & 0 & 0 & 0 \\ 0 & 0.250 & 0 & 0 \\ 0 & 0 & 5.000 & 0 \\ 0 & 0 & 0 & 0.125 \end{pmatrix}$$

- A partir da matriz de correlação ρ vemos que Y_1 é praticamente não-correlacionada com as outras três.
- Já Y_2 é possui uma correlação moderada com Y_3 (negativa) e com Y_4 (positiva).
- Apesar disso, curiosamente, Y_3 e Y_4 são praticamente não-correlacionadas.

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

```
nsims = 100
mu = matrix(c(0, 10, 0, 1020), ncol=1)
S = matrix(c(1, -0.1, 0, -0.7,
            -0.1, 16, -0.4, 20,
            0, -0.4, 0.04, -0.1,
            -0.7, 20, -0.1, 64), ncol=4)
L = t(chol(S))
Z = matrix(rnorm(4*nsims), nrow=4)
Y = mu + L %*% Z   # matriz 4 x nsims
pairs(t(Y)) # ver proximo slide
```

Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

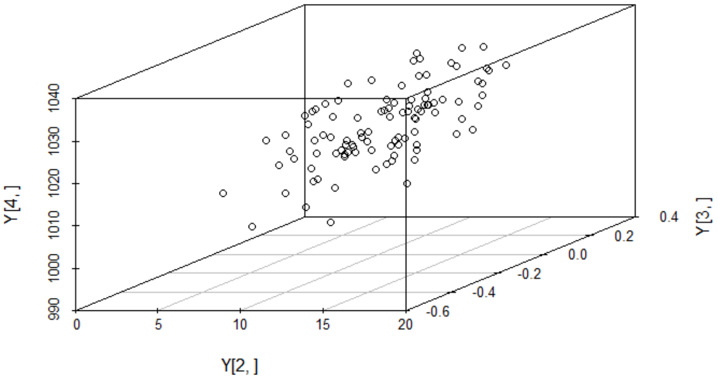


Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- Podemos criar scatterplots TRI-dimensionais com os vetores que simulamos.
- Como os vetores são 4-dim, vamos escolher três das 4 variáveis para fazer o plot.
- Precisamos do pacote `scatterplot3d`.
- Usamos a função `scatterplot3d(x, y, z)`.

```
library(scatterplot3d)
scatterplot3d(Y[2,], Y[3,], Y[4,])
```

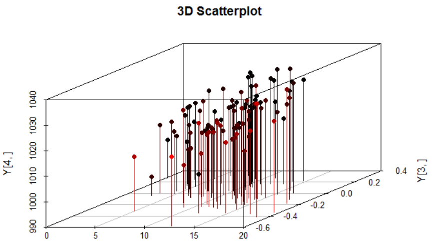
Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$



Exemplo - simulando $N_4(\boldsymbol{\mu}, \sum)$

- Scatterplot 3-dim com cores e linhas ajudando a localizar os pontos no espaço.
- Pontos são desenhados com cores diferentes de acordo com sua coordenada y (de outra forma, fica difícil ver que pontos estão mais à frente ou atrás no cubo 3-dim).

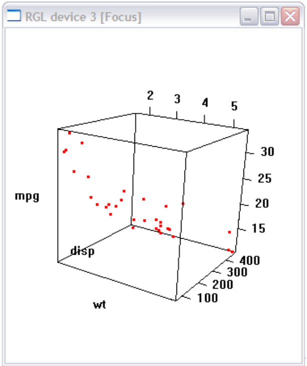
```
library(scatterplot3d)
scatterplot3d(Y[2,], Y[3,], Y[4,], pch=16,
  highlight.3d=TRUE, type="h", main="3D Scatterplot")
```



Exemplo - simulando $N_4(\mu, \sum)$

- Um scatter 3-dim dinâmico: `plot3D(x, y, z)` do pacote `rgl`.
- Cria um catter 3-dim que pode ser rotacionado com o mouse.
- `col=` e `size=` controlam a cor e tamanho dos pontos.

```
library(rgl)
plot3d(Y[2,], Y[3,], Y[4,], col="red", size=3)
```



Forma quadrática

- Seja $\mathbf{y} = (y_1, \ldots, y_k)$ um VETOR-COLUNA em $\mathbb{R}^k$.
- Seja A uma matriz $k \times k$.
- Forma quadrática é qualquer expressão assim:

$$\mathbf{y}' \ A \ \mathbf{y} = \sum_{ij} A_{ij} y_i y_j$$

- Por exemplo, se $\mathbf{y} = (y_1, y_2)$ e A for 2×2:

$$(y_1, y_2) \ A \ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

  que são iguais a

$$A_{11} y_1^2 + A_{12} y_1 y_2 + A_{21} y_2 y_1 + A_{22} y_2^2$$

- Envolvem combinações lineares dos produtos de pares de variáveis (produto de duas variáveis distintas ou produto de uma variável por ela mesma).

Exemplos de formas quadráticas

- Exemplos bi-dimensionais:

$$(y_1, y_2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1^2 + y_2^2$$

$$(y_1, y_2) \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 9y_1^2 + 4y_2^2$$

$$(y_1, y_2) \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 9y_1^2 + 4y_2^2 + 3y_1y_2 + 3y_2y_1 = 9y_1^2 + 4y_2^2 + 6y_1y_2$$

Formas quadráticas simétricas

- A matriz A na forma quadrática

$$\mathbf{y}' \ A \ \mathbf{y} = \sum_{ij} A_{ij} y_i y_j$$

  pode ser SEMPRE tomada como simétrica.
- Por exemplo, no caso bi-dimensional

$$\begin{aligned} \mathbf{y}' \ A \ \mathbf{y} &= (y_1, y_2) \begin{bmatrix} 9 & 2 \\ 4 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= 9y_1^2 + 4y_2^2 + 2y_1 y_2 + 4y_2 y_1 \\ &= 9y_1^2 + 4y_2^2 + 6y_1 y_2 \\ &= (y_1, y_2) \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \end{aligned}$$

- Caso geral: ver lista de exercícios.
- De agora em diante, A em formas quadráticas é sempre simétrica.

Matrizes positivas definidas

- Queremos que uma medida de distância mais geral que a euclidiana.
- Se a distância ao quadrado de $\mathbf{y}$ até a origem $\mathbf{0} = (0, \ldots, 0)$ for uma forma quadrática, precisamos garantir que, PARA TODO VETOR $\mathbf{y}$ que não seja nulo tenhamos

$$d^2(\mathbf{y}, \mathbf{0}) = \mathbf{y}' \ A \ \mathbf{y} = \sum_{ij} A_{ij} y_i y_j > 0$$

- Matrizes que atendem esta condição são chamadas de matrizes definidas positivas.

Exemplos de matrizes positivas definidas

- Exemplos de positiva defnida em que $\mathbf{y}' A \mathbf{y} > 0$ para todo $\mathbf{y} \neq \mathbf{0}$:

$$(y_1, y_2) \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 9y_1^2 + 4y_2^2 > 0$$

- Outro exemplo:

$$(y_1, y_2) \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 9y_1^2 + 4y_2^2 - 3y_1 y_2 - 3y_2 y_1 = 9y_1^2 + 4y_2^2 - 6y_1 y_2 > 0$$

- Não é óbvio que esta última matriz seja dp. E é apenas um caso 2-dim!!

Exemplos de matrizes que NÃO SÃO positiva definida

- Um exemplo:

$$(y_1, y_2) \begin{bmatrix} 9 & 0 \\ 0 & -4 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 9y_1^2 - 4y_2^2$$

- pois é menor que zero se $(y_1, y_2) = (0, 1)$, por exemplo.
- Outro exemplo:

$$(y_1, y_2) \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1^2 + y_2^2 - 4y_1y_2$$

- É menor que zero se tomarmos $(y_1, y_2) = (1, 1)$, por exemplo.

Checando se matriz é positiva definida

- Como verificar, em geral, se uma matriz simétrica A de dimensão $k \times k$ é definida positiva?
- Difícil se k for grande.
- Checando todos os infinitos $\mathbf{y}$??
- Não...
- A é definida positiva se, e somente se, todos os seus autovalores forem positivos.
- Algoritmos para encontrar autovalores são custosos, especialmente se a matriz é grande.
- A é definida positiva se, e somente se, existir a sua decomposição de Cholesky.
- Este é algoritmo simples e rápido.
- Vamos ver um exemplo

Nossa matriz é positiva definida

- Nossa matriz A será de um tipo especial: será uma matriz de covariância.
- Neste caso, A será sempre def pos (a não ser em exemplos patológicos, que não ocorrem na prática).

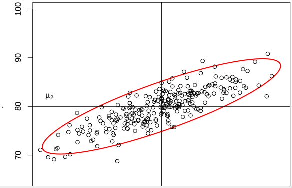
Qual a relação entre $\sum$ e a elipse?

- Suponha que a medida de distância do vator aleatório $\boldsymbol{y}$ até o perfil esperado $\boldsymbol{\mu}$ é dada por

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

onde Σ é uma matriz de covariância simétrica e definida positiva.

- Os pontos $\boldsymbol{y}$ que são equidistantes de $\boldsymbol{\mu}$ formam uma elipse centrada em $\boldsymbol{\mu}$ e com eixos na direção dos autovetores da matriz $\sum$. Os tamanhos dos semi-eixos da elipse são proporcionais a (raiz quadrada) dos seus respectivos autovalores.



Autovetores de matrizes simétricas

- Autovetor e autovalor de matriz QUADRADA A.
- Definição: $A\mathbf{v} = \lambda\mathbf{v}$
- Autovalor λ pode ser um número complexo.
- Se A for simétrica então λ é real.
- Só nos interessam as matrizes simétricas.

Imagine a matriz como uma transformação

- O que é um autovetor de uma matriz A de dimensaão $p \times p$?
- Olhe a definição de novo: $A\boldsymbol{v} = \lambda\boldsymbol{v}$
- Um autovetor é uma direção muito especial em $\mathbb{R}^p$.
- É uma direção $\boldsymbol{v}$ tal que, quando A é aplicado a $\boldsymbol{v}$, temos apenas $\boldsymbol{v}$ espichado (se $\lambda > 1$) ou encolhido (se $0 < \lambda < 1$)
- Se $\lambda < 0$, a direção muda de sentido.

Imagine a matriz como uma transformação

- Matriz real e simétrica A de dimensão $p \times p$.
- Pense na transformação de $\mathbb{R}^p$ para $\mathbb{R}^p$ pela aplicação da matriz A.
- Isto é, considere a transformação linear $T: \mathbb{R}^p \mapsto \mathbb{R}^p$ tal que $T(\boldsymbol{v}) = A\boldsymbol{v}$.
- Por exemplo, pense numa imagem em $\mathbb{R}^2$, um desenho feito com linhas e curvas.
- Cada ponto do deseho será identificado com um vetor de $\mathbb{R}^2$.
- Cada ponto será transformado através de uma matriz simétrica A de dimensão 2×2.
- O que será o novo desenho?

Imagine a matriz como uma transformação

- Em geral, a imagem $A\boldsymbol{v}$ de um ponto não tem uma relação geométrica simples com $\boldsymbol{v}$.
- Em geral, é difícil antecipar qual será o resultado de aplicar A em $\boldsymbol{v}$.
- A seguir, veremos o efeito de

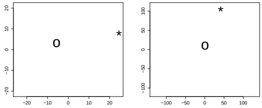
$$A = \begin{pmatrix} 1 & 1.2 \\ 1.2 & 5 \end{pmatrix}$$

- Em p dimensões não teremos uma imagem para olhar...

Imagine a matriz como uma transformação

- Matriz

$$A = \begin{pmatrix} 1 & 1.2 \\ 1.2 & 5 \end{pmatrix}$$



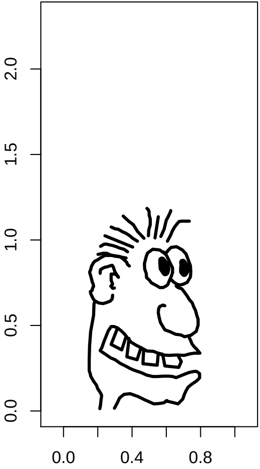
- O ponto-estrela da esquerda é levado por A no ponto-estrela da direita.
- O ponto-bolinha da direita é levado por A no ponto-bolinha da direita.
- Podemos ANTECIPAR o efeito de A num ponto arbitrário (x, y)?
- Onde ele será levado?
- Isto parece ser uma tarefa difícil.

Imagine a matriz como uma transformação

- Se A for $p \times p$, teremos vetores em p dimensões: não teremos uma imagem para olhar...
- Antecipar onde o vetor $x \in \mathbb{R}^p$ será levado por A parece uma tarefa impossível.
- E no entanto...

- Ao longo de ALGUMAS DIREÇÕES $\boldsymbol{v}$, o comportamento da transformação por A é fácilmente entendido.
- Pense num vetor que esteja numa destas direções especiais.
- Então A simplesmente espicha ou encolhe o ponto-vetor, SEM ALTERAR A SUA DIREÇÃO.

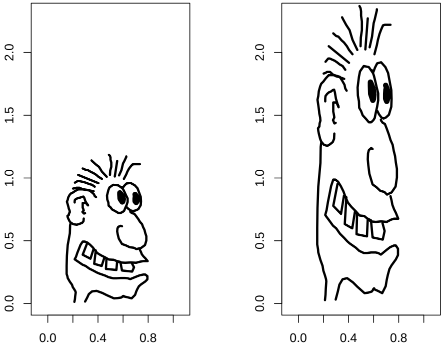
Cara Engraçada



Espichando verticalmente

- Espichando verticalmente com

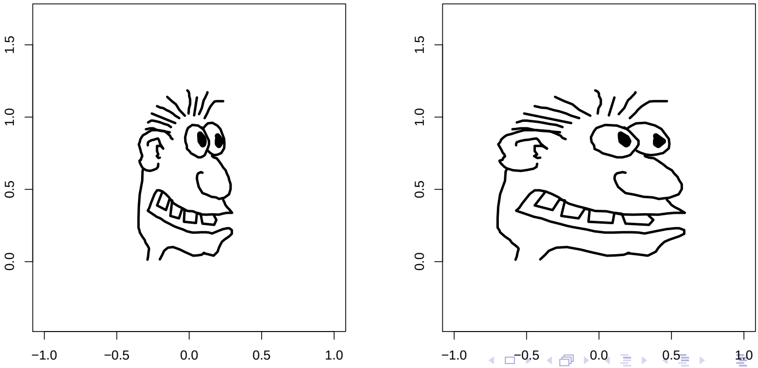
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$



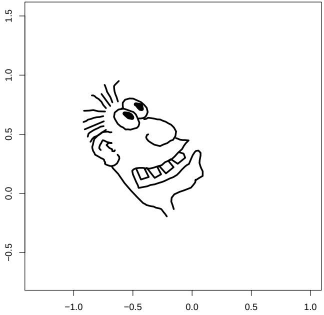
Espichando lateralmente

- Espichando lateralmente com

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$



E se o desenho estiver rotacionado?



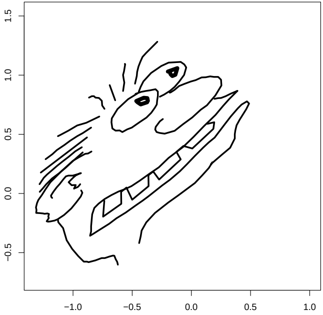
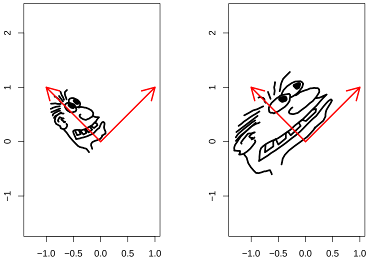


Figura: Queremos espichar apenas lateralmente o desenho da face mas mantendo a sua orientação. Como fazer?

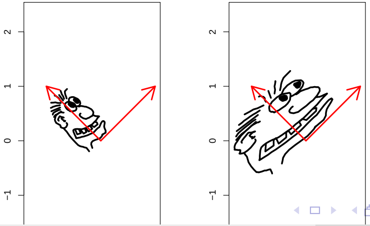
Direções especiais

- Existem duas direções especiais associadas com a transformação linear T que desejamos fazer na figura.
- Ao longo dessas duas direções especiais, basta espichar lateralmente ao longo de UMA delas para fazer a face ficar mais "gordinha".
- Basta espichar ou contrair a projeção do vetor ao longo dessa direção para obter o efeito desejado.



Direções especiais

- Então $A\boldsymbol{v}$ pode ser pensado assim: expresse $\boldsymbol{v}$ com coordenadas na base ortogonal formada pelas duas direções especiais.
- Espiche ou contraia cada uma das coordenadas dessas direções especiais.
- Volte para o sistema de coordenadas original.
- Se esta história de espichar a cara não ficou muito clara, não se preocupe. O que você REALMENTE precisa saber está resumido a seguir.



Autovetor = Direção especial

- O que é um autovetor $\boldsymbol{v}$ de uma matriz simétrica A de dimensão $p \times p$?
- Por definição: $A\boldsymbol{v} = \lambda\boldsymbol{v}$
- Um autovetor é uma direção muito especial em $\mathbb{R}^p$.
- É uma direção $\boldsymbol{v}$ tal que, quando A é aplicado a $\boldsymbol{v}$, temos apenas $\boldsymbol{v}$ espichado (se $\lambda > 1$) ou encolhido (se $0 < \lambda < 1$).

Teorema Espectral - 1

- Seja A uma matriz $p \times p$ simétrica e *positiva definida*.
- Existem p autovalores associados com A.
- Estes p autovalores são números reais pois A é simétrica
- Estes autovalores são POSITIVOS pois A é positiva definida.
- A cada autovalor corresponde um autovetor ou direção em $\mathbb{R}^p$.
- O que podemos falar desses autovetores?

Teorema Espectral - 2

- Os p autovetores são ortogonais entre si.
- Tomando todos esses autovetores com comprimento 1 e colocando-os como p colunas de uma matriz P, teremos $P^t P = I$ pois eles são ortogonais entre si.
- Seja D uma matriz diagonal $p \times p$ com os autovalores (na mesma ordem que as colunas de P).
- Teorema Espectral: $A = PDP^t$
- O que isto significa: A age simplesmente como uma matriz diagonal D (que é fácil de ser entendida) se trabalharmos no sistema de coordenadas dos autovetores (que são as colunas de P) !!
- Dizemos que A é diagonalizavel.

Coordenadas

- No sistema de coordenadas dos autovetores, a matriz A funciona como uma matriz diagonal.
- $\boldsymbol{x}$ no novo sistema de coordenadas dos autovetores é $\boldsymbol{x}^* = P\boldsymbol{x}$.
- Se $\boldsymbol{x}^*$ é o conjunto de coordenadas no sistema de autovetores, para voltar ao sistema original simplesmente multiplique pela inversa de P que é ... P^t.
- Lembre-se que $P^t P = I$.

Resumindo...

- Pontos na ELIPSE tendem a estar a igual distância do perfil esperado $\boldsymbol{\mu} = (\mu_1, \mu_2)$
- A maneira correta de medir distância ao perfil esperado $\boldsymbol{\mu} = (\mu_1, \mu_2)$ é pela forma quadrática

$$d^2(\boldsymbol{y}, \boldsymbol{\mu}) = (\boldsymbol{y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$$

- A elipse é determinada pelos autovetores e autovalores de Σ^{-1}, a inversa da matriz de covariância das v.a.'s envolvidas.
- Os autovetores de Σ^{-1} e de Σ são os mesmos
- Os autovalores de Σ^{-1} são os inversos $1/\lambda$ dos autovalores de λ

Um exemplo de normal multivariada

- O retorno diário de uma ação é a variação percentual no seu preço de um dia para o outro.
- Seja $S(t)$ o preço no dia t
- O retorno da ação no dia $t+1$ é dado por

$$Z(t+1) = \frac{S(t+1) - S(t)}{S(t)}$$

- É a diferença no preço da ação entre hoje e ontem RELATIVAMENTE ao preço de ontem.

Quatro ações

- Assim, se $Z(t+1) = 0.1$ isto siginifica que

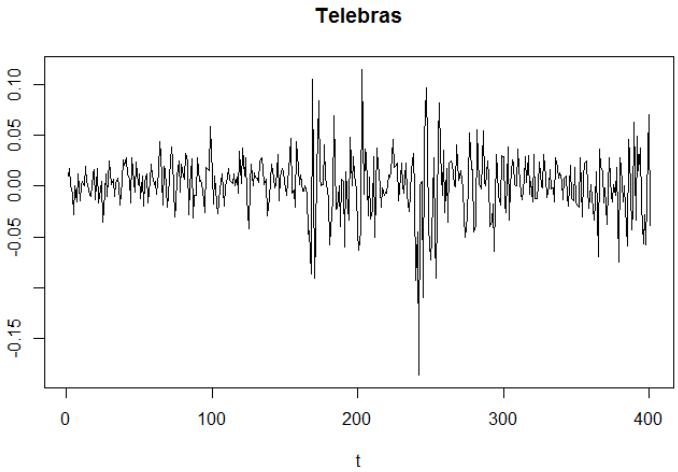
$$Z(t+1) = \frac{S(t+1) - S(t)}{S(t)} = 0.1 \Rightarrow S(t+1) = (1+0.1)S(t)$$

- Ou seja, um aumento de 10% no preço.
- Se $Z(t+1) = -0.1$, temos então $S(t+1) = (1-0.1)S(t)$, uma diminuição de 10% no preço.
- Uma suposição muito comum é que os retornos diários de uma ação segue uma normal.
- E que os retornos de várias ações num mesmo dia seguem uma normal multivariada

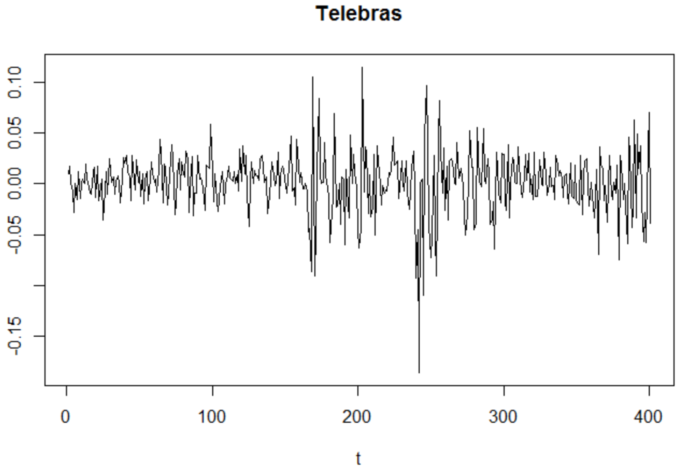
Quatro ações

- Vamos olhar os dados dos retornos diários de algumas das principais ações da Bolsa do Estado de São Paulo (BOVESPA).
- Dados diários do período de 4/Nov/1996 a 18/Junho/1998
- Vamos olhar apenas 4 ações, como ilustração:
    - Eletrobrás,
    - Vale do Rio Doce,
    - Petrobrás,
    - Suzano (empresa de papel e celulose)

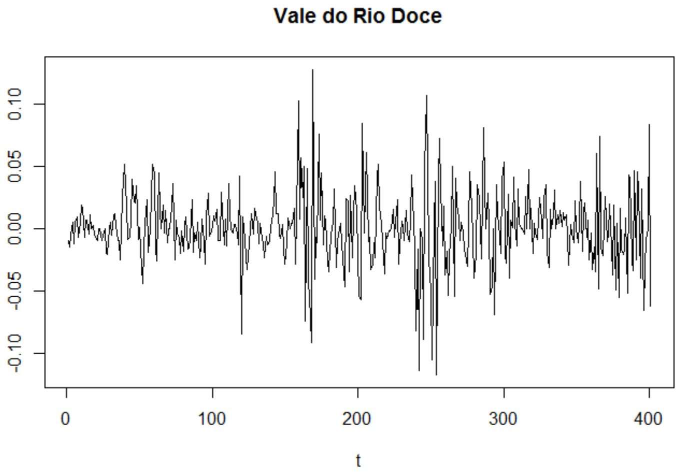
Telebrás no tempo



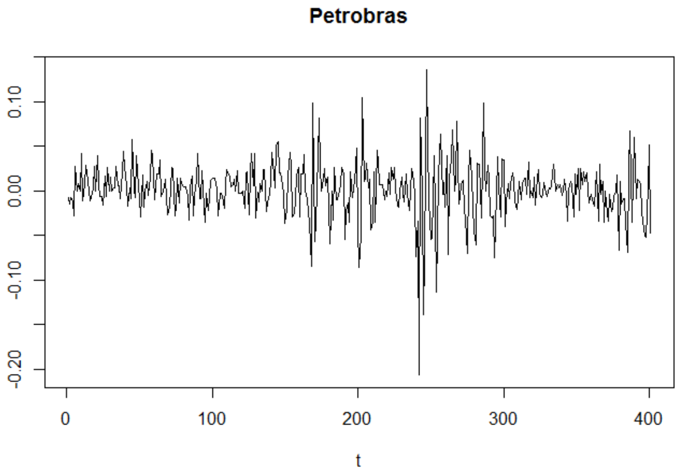
Telebrás no tempo



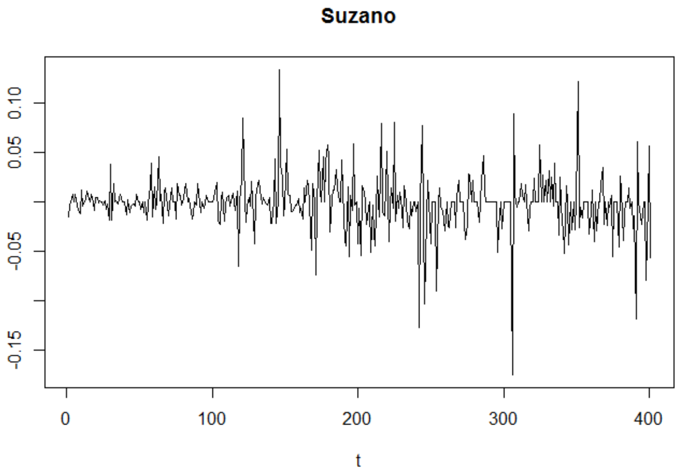
Vale no tempo



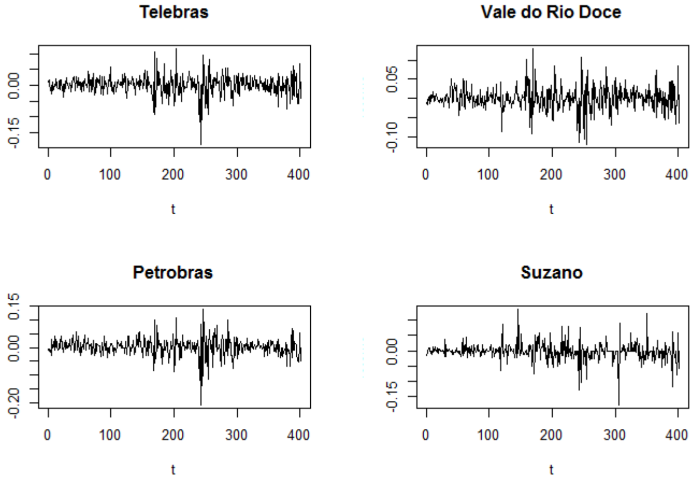
Petrorás no tempo



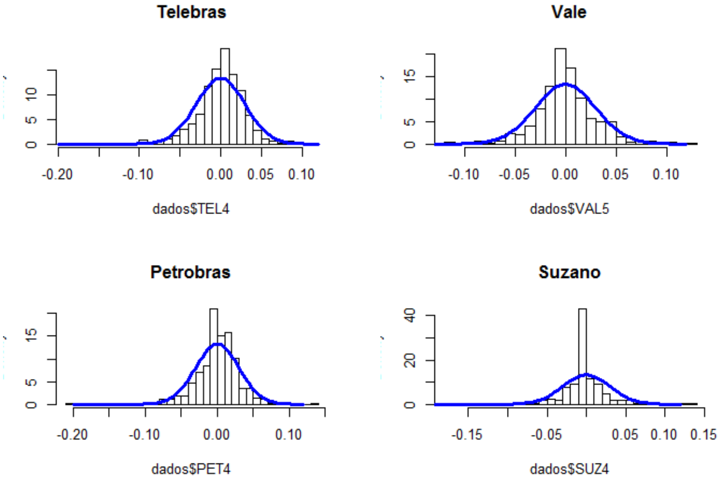
Suzano no tempo



As quatro no tempo



Histogramas e ajustes gaussianos



Scatterplot dos pares de ações

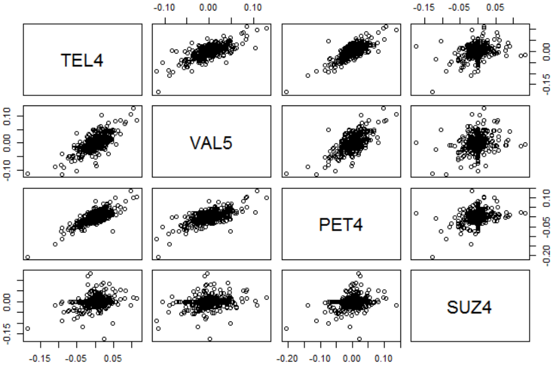


Figura: Tele, Vale e Pet são bem correlacionadas. Quando uma sobe muito, as outras duas também sobem. Suzano não parece ser muito correlacionada com estas outras.

Correlação no tempo

- E a correlaçao de UMA MESMA açao em dias sucessivos?
- Por exemplo, se a ação da Vale subir bastante hoje, o que podemos esperar para seu movimento amanhã?
- Surpreendente: quase não existe correlação.
- Não uma tendência detectável na VARIAÇÃO dos preços das ações em dias sucessivos.
- Não acredita? Veja os próximos gráficos.

Correlação no tempo

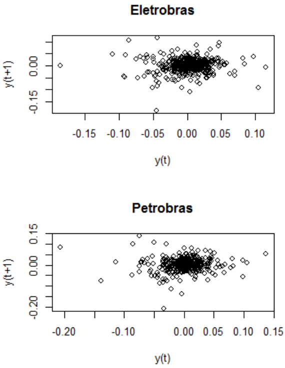


Figura: Gráfico de y_{t+1} versus y_t para as quatro ações. Não existe correlação entre os retornos de uma mesma aąão em dias sucessivos.

Matriz de correlação

- Seja $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})$ os retornos das quatro ações no dia t.
- Vamos calcular empiricamente o coeficiente de correlação de Pearson para os pares de ações.
- Temos 401 instâncias de $\mathbf{Y}_t$ correspondentes a 401 dias.
- O vetor esperado pode ser estimado a partir dos dados

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4) \approx (\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4) = \bar{\mathbf{y}} = (1.094, 0.276, 1.038, -1.459) \times 10^{-3}$$

onde $\bar{x}_j$ é a média aritmética

$$\bar{x}_j = \frac{1}{401} \sum_t Y_{tj}$$

- Assim, o retorno diário no período é pequeno, ligeiramente positivo pra as três primeiras, e negativo para SUZ.
- Quanto ao desvio padrão para cada uma delas, estimando dos dados encontramos

$$(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \approx (s_1, s_2, s_3, s_4) = (0.030, 0.030, 0.032, 0.028)$$

Matriz de correlação

- Seja $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})$ os retornos da quatro ações no dia t.
- Vamos calcular empiricamente o coeficiente de correlação de Pearson para os pares de ações.
- Temos 401 instâncias de $\mathbf{Y}_t$ correspondentes a 401 dias.
- Matriz de correlação:

$$\boldsymbol{\rho} = \left(\begin{array}{c|cccc} & \text{Tel4} & \text{Val5} & \text{Pet4} & \text{Suz4} \\ \hline \text{Tel4} & 1.00 & 0.72 & 0.79 & 0.30 \\ \text{Val5} & 0.72 & 1.00 & 0.68 & 0.24 \\ \text{Pet4} & 0.79 & 0.68 & 1.00 & 0.33 \\ \text{Suz4} & 0.30 & 0.24 & 0.33 & 1.00 \end{array} \right)$$

- De fato, TEL, PET e VALE são bem correlacionadas (positivamente) enquanto SUZ mostra pouca correlação com estas outras três ações.

Matriz de covariância

- Matriz de covariância:

$$\sum \approx \mathbf{S} = 10^{-4} \begin{array}{c|cccc} & \text{Tel4} & \text{Val5} & \text{Pet4} & \text{Suz4} \\ \hline \text{Tel4} & 9.08 & 6.58 & 7.49 & 2.53 \\ \text{Val5} & 6.58 & 9.27 & 6.53 & 2.00 \\ \text{Pet4} & 7.49 & 6.53 & 10.01 & 2.86 \\ \text{Suz4} & 2.53 & 2.00 & 2.86 & 7.69 \end{array}$$

- Veja que, ao contrário da matriz de correlação, os números dessa matriz de covariância são difíceis de interpretar.
- Como uma primeira aproximação, podemos dizer que os retornos das quatro ações num dado dia seguem $\mathbf{Y}_t \sim N_4(\boldsymbol{\mu}, \sum)$
- com $\boldsymbol{\mu} \approx \bar{\mathbf{y}}$ e$\sum \approx \mathbf{S}$.

Marginais da normal multivariada

- Suponha que $\mathbf{Y} = (Y_1, \ldots, Y_k) \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ seja um vetor aleatório com distribuição gaussiana multivariada.
- A distribuição de cada uma das entradas Y_i do vetor $\mathbf{Y}$ é uma gaussiana.
- Além disso, a esperança e variância da distribuição s ao extraídos diretamente dos parâmetros de $\mathbf{Y}$ sem necessidade de nenhum cálculo.
- $Y_i \sim N(\mu_i, \Sigma_{ii})$ onde μ_i é a i-ésima entrada do vetor $\boldsymbol{\mu}$ e $\sum_{ii}$ a i-ésima entrada da diagonal de $\sum$.

Exemplo: marginais de $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ tem vetor esperado $\boldsymbol{\mu} = (4, 3, -2, 2)$ e matriz de covariância

$$\sum = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}$$

- Então a distribuição marginal é diretamente obtida desta conjunta:

$$Y_3 \sim N(\mu_3, \Sigma_{33}) = N(-2, 9)$$

Marginais de $N_k(\boldsymbol{\mu}, \sum)$

- Obtemos não apenas a distribuição de cada entrada individual de $\mathbf{Y}$ mas a distribuição marginal de qualquer sub-vetor de $\mathbf{Y}$.
- Por exemplo, se $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ tem vetor esperado $\boldsymbol{\mu} = (4, 3, -2, 2)$ e matriz de covariância

$$\sum = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}$$

- Então a distribuição marginal do sub-vetor (Y_1, Y_3) é dada por

$$(Y_1, Y_3) \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}\right) = N_2\left(\begin{pmatrix} 4 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 2 & 9 \end{pmatrix}\right)$$

A importância desta propriedade

- Aparentemente, esta propriedade é boba.
- Da conjunta que é normal multivariada chegamos a marginais que também são normais.
- Isto não é óbvio?
- Não, não é.
- Acontece que é muito difícil e raro que a gente consiga saber quais são as marginais apenas mirando a fórmula da conjunta.
- Na maioria dos casos, a única maneira de obter as marginais é integrando ou somando sobre os valores das demais variáveis.

A importância desta propriedade

- Por exemplo, suponha que

$$f_{XY}(x,y) = \text{cte}\left(yx^2 + \sqrt{y}\exp(-xy + x^2)\right)$$

  com suporte em $[0,1]^2$.
- Não é possível saber quais são as marginais $f_X(x)$ e $f_Y(y)$ diretamente a partir da expressão da conjunta.
- A única maneira de obter $f_X(x)$ é integrando $f_{XY}(x,y)$ com respeito a y:

$$f_X(x) = \int_0^1 \text{cte}\left(yx^2 + \sqrt{y}\exp(-xy + x^2)\right)\ dy$$

- Este não é caso da normal multivariada.
- Para escrever a densidade conjunta precisamos de $\boldsymbol{\mu}$ e de $\sum$.
- Com estes dois elementos temos também todas as marginais.

Combinação linear de normais

- Suponha que $\mathbf{Y} = (Y_1, \ldots, Y_k) \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ seja um vetor aleatório com distribuição gaussiana multivariada.
- Queremos criar um indicador baseado nestas variáveis, uma nova variável aleatória:
$$X = c_1 Y_1 + \ldots + c_k Y_k$$
- Por exemplo, podemos criar um índice para o movimento no mercado de ações.
- Usando apenas os retornos das quatro ações que vimos antes, poderíamos estabelecer o índice
$$X_t = 0.2\text{PET}_t + 0.2\text{VALE}_t + 0.2\text{TEL}_t + 0.4\text{SUZ}_t$$
- Os coeficientes c_i náo precisam somar 1 ou serem positivos.
- Por exemplo, por alguma razão, poderíamos querer
$$X_t = 1.2\text{PET}_t + 2.0\text{VALE}_t + 4.3\text{TEL}_t - 3.5\text{SUZ}_t$$

Combinação linear de normais

- Se $\mathbf{Y} = (Y_1, \ldots, Y_k) \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ qual é a distribuição de probabilidade do indicador X?
- Note que este tipo de indicador pode ser escrito em forma vetorial:

$$X = c_1 Y_1 + \ldots + c_k Y_k = \mathbf{c}'\mathbf{Y} = (c_1, \ldots, c_k) \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$$

- Adivinhe: X continua gaussiano.
- Isto é, temos que $X \sim N(??, ??) = N(\mu_X, \sigma_X^2)$
- Como X é univariado, basta acharmos o seu valor esperado μ_X e sua variância σ_X^2.
- Estes momentos são facilmente obtidos a partir dos momentos $\boldsymbol{\mu}$ e $\sum$ da normal multivariada, como veremos a seguir.

Momentos de $\mathbf{c}'\mathbf{Y}$

- O cálculo do valor esperado $\mathbb{E}(X) = \mathbb{E}(\mathbf{c}'\mathbf{Y})$ e da variância $\mathbb{V}(X) = \mathbb{V}(\mathbf{c}'\mathbf{Y})$ *não depende da distribuição do vetor* $\mathbf{Y}$.
- Qualquer que seja a distribuição do vetor $\mathbf{Y}$, gaussiana ou não, contínua ou discreta, podemos obter $\mathbb{E}(X) = \mathbb{E}(\mathbf{c}'\mathbf{Y})$ e $\mathbb{V}(X) = \mathbb{V}(\mathbf{c}'\mathbf{Y})$ facilmente.
- Isto tem relevância pois é muito comum criar indicadores que são combinações lineares de algumas variáveis.
- Portanto, o cálculo a seguir usa apenas as propriedades de esperança e variância, sem recorrer à especificação de uma distribuição conjunta para o vetor $\mathbf{Y}$.

$\mathbb{E}(\mathbf{c}'\mathbf{Y})$

- Começando com o valor esperado:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(c_1 Y_1 + c_2 Y_2 + \ldots + c_k Y_k) \\
&= c_1\mathbb{E}(Y_1) + c_2\mathbb{E}(Y_2) + \ldots + c_k\mathbb{E}(Y_k)) \\
&= c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k \\
&= \mathbf{c}'\mu
\end{aligned}$$

- Isto é,

$$\mathbb{E}\left(\mathbf{c}'\mathbf{Y}\right) = \mathbf{c}'\mathbb{E}\left(\mathbf{Y}\right)$$

- O vetor de constantes **c** vai para fora do símbolo de esperança.

$\mathbb{V}(\mathbf{c}'\mathbf{Y})$

- Agora a variância.

$$\mathbb{V}(X) = \mathbb{V}(c_1 Y_1 + c_2 Y_2 + \ldots + c_k Y_k)$$

- Basta usar a definição de variância (e um pouco de paciência):

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left(X - \mathbb{E}(X)\right)^2 \\
&= \mathbb{E}\left((c_1 Y_1 + c_2 Y_2 + \ldots + c_k Y_k) - (c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k)\right)^2 \\
&= \mathbb{E}\left((c_1(Y_1 - \mu_1) + c_2(Y_2 - \mu_2) + \ldots + c_k(Y_k - \mu_k)\right)^2 \\
&= \mathbb{E}\left(\sum_i c_i^2 (Y_i - \mu_i)^2 + \sum_{i \neq j} c_i c_j (Y_i - \mu_i)(Y_j - \mu_j)\right) \quad \text{(abrindo o quadrado)} \\
&= \sum_i \mathbb{E}\left(c_i^2 (Y_i - \mu_i)^2\right) + \sum_{i \neq j} \mathbb{E}\left(c_i c_j (Y_i - \mu_i)(Y_j - \mu_j)\right) \quad \text{(linearidade da esperanç} \\
&= \sum_i c_i^2 \mathbb{E}\left(Y_i - \mu_i\right)^2 + \sum_{i \neq j} c_i c_j \mathbb{E}\left((Y_i - \mu_i)(Y_j - \mu_j)\right) \quad \text{(linearidade de novo)} \\
&= \sum_i c_i^2 \mathbb{V}(Y_i) + \sum_{i \neq j} c_i c_j \mathrm{Cov}(Y_i, Y_j) \quad \text{(pela def de Var e Cov)}
\end{aligned}$$

$\mathbb{V}(\mathbf{c}'\mathbf{Y})$

- Assim, temos

$$\begin{aligned}
\mathbb{V}(\mathbf{c}'\mathbf{Y}) &= \mathbb{V}(c_1 Y_1 + c_2 Y_2 + \ldots + c_k Y_k) \\
&= (c_1, \ldots, c_k) \sum \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \\
&= \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}
\end{aligned}$$

  onde $\sum$ é matriz de covariância do vetor $\mathbf{Y}$.
- Veja que no caso univariado tínhamos a fórmula

$$\mathbb{V}(cY) = c^2\ \mathbb{V}(Y) = c\ \mathbb{V}(Y)\ c\ ,$$

  que é versão univariada de $\mathbb{V}(\mathbf{c}'\mathbf{Y}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$

Enfatizando

- Mais uma vez, repetimos:
- Seja $\mathbf{Y} = (Y_1, \ldots, Y_k)$ um vetor aleatório com QUALQUER distribuição.
- Seja $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}$ com matriz de covariância $\mathbb{V}(\mathbf{Y}) = \sum$.
- Então
$$\mathbb{E}\left(\mathbf{c}'\mathbf{Y}\right) = \mathbf{c}'\boldsymbol{\mu}$$
e
$$\mathbb{V}(\mathbf{c}'\mathbf{Y}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$$
- Estes resultados são válidos mesmo no caso em que $\mathbf{Y}$ não é nomal multivariado.
- Caso $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ então $\mathbf{c}'\mathbf{Y} \sim N(\mathbf{c}'\boldsymbol{\mu}, \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c})$

Retorno de portfólio de ações

- Seja $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})$ os retornos das quatro ações no dia t.
- Suponha que $\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com

$$\boldsymbol{\mu} = 10^{-3} \times (1.10.31.0 - 1.5)$$

e

$$\sum = 10^{-4} \begin{pmatrix} & \text{Tel4} & \text{Tel5} & \text{Pet4} & \text{Suz4} \\ \text{Tel4} & 9.08 & 6.58 & 7.49 & 2.53 \\ \text{Tel5} & 6.58 & 9.27 & 6.53 & 2.00 \\ \text{Pet4} & 7.49 & 6.53 & 10.01 & 2.86 \\ \text{Suz4} & 2.53 & 2.00 & 2.86 & 7.69 \end{pmatrix}$$

- Vamos criar um portfólio com 30% de ações da Telebrás, 20% da Vale, 30% da Petrobrás e 20% da Suzano.
- O retorno deste mix de ações será o mix dos retornos das ações.
- Isto é, o retorno do portfólio é a v.a.

$$X = 0.3Y_1 + 0.2Y_2 + 0.3Y_3 + 0.2Y_4$$

- Qual a distribuição do retorno deste portfólio?

Retorno de portfólio de ações

- Temos

$$X = 0.3Y_1 + 0.2Y_2 + 0.3Y_3 + 0.2Y_4 \sim N(\mu_x, \sigma_x^2) = N(0.00039, (0.0247)^2)$$

onde

$$\mu_x = (0.3, 0.2, 0.3, 0.2)\ 10^{-3} \begin{pmatrix} 1.1 \\ 0.3 \\ 1.0 \\ -1.5 \end{pmatrix} = 0.00039$$

e

$$\sigma_x^2 = (0.3, 0.2, 0.3, 0.2) \sum \begin{pmatrix} 0.3 \\ 0.2 \\ 0.3 \\ 0.2 \end{pmatrix} = 6.1247 \times 10^{-4} = (0.0247)^2$$

Mais propriedades

- Podemos encontrar a distribuição conjunta de vários índices simulatâneos
- Seja $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \sum)$
- Seja $\mathbf{A}$ um matriz $q \times k$ de constantes e usada para gerar q índices:

$$\underset{q\times k}{\mathbf{A}}\ \underset{k\times 1}{\mathbf{Y}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ & & \vdots & \\ a_{q1} & a_{q2} & \dots & a_{qk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1k}Y_k \\ a_{21}Y_1 + a_{22}Y_2 + \ldots + a_{2k}Y_k \\ \vdots \\ a_{q1}Y_1 + a_{q2}Y_2 + \ldots + a_{qk}Y_k \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{bmatrix}$$

- Qual a distribuição CONJUNTA dos q índices no vetor $\mathbf{X}$?
- Uma maneira intuitiva de ver isto é pensar que, se duas linhas de $\mathbf{A}$ forem muito parecidas, esperamos que os dois índices associados sejam muito correlacionados.

Distribuição e combinações lineares

- RESULTADO: Se $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ então

$$\underset{q\times k}{\mathbf{A}} \underset{k\times 1}{\mathbf{Y}} \sim N_q\left(\mathbf{A}\mu, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\right)$$

- Além disso, $\mathbf{Y} + \mathbf{c}$, onde $\mathbf{c}$ é um vetor $k \times 1$ de constantes, é distribuído como

$$\mathbf{Y} + \mathbf{c} \sim N_k(\boldsymbol{\mu} + \mathbf{c}, \boldsymbol{\Sigma})$$

Independência e covariância

Result 4.5.

(a) If $\underset{(q_1\times 1)}{\mathbf{X}_1}$ and $\underset{(q_2\times 1)}{\mathbf{X}_2}$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hline \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

Distribuição condicional

- Outra propriedade fantástica da normal multivariada é a facilidade de obter a distribuição condicional de um sub-vetor dados os valores dos outros elementos do vetor aleatório $\mathbf{Y}$.
- Em geral, para uma distribuição conjunta arbitrária, isto não é possível.
- Dada a conjunta $f(y_1, \ldots, y_k)$ dificlmente conseguimos saber qual é a distribuição de Y_1 dados os valores das demais variáveis.
- Temos a fórmula para obter esta distribuição condicional,

$$f_{Y_1|Y_2,\ldots,Y_k}(y_1|y_2, \ldots, y_k) = \frac{f_{Y_1,Y_2,\ldots,Y_k}(y_1, y_2, \ldots, y_k)}{f_{Y_2,\ldots,Y_k}(y_2, \ldots, y_k)}$$

  mas não é óbvio de antemão qual será o resultado desta fórmula.
- Este não é o caso da normal multivariada.
- Podemos obter imediatamente e sem muitas contas a distribuiç ao condicional.

Distribuição condicional

Result 4.6. Let $\mathbf{X} = \left[\begin{array}{c}\mathbf{X}_1 \\ \hdashline \mathbf{X}_2\end{array}\right]$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \left[\begin{array}{c}\boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2\end{array}\right]$, $\boldsymbol{\Sigma} = \left[\begin{array}{c:c}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{array}\right]$, and $|\boldsymbol{\Sigma}_{22}| > 0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

$$\text{Mean} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$\text{Covariance} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

Condicional: exemplo

- Seja $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})$ os retornos das quatro ações no dia t.
- Suponha que $\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com

$$\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = N_4\left(\begin{pmatrix} 0.001 \\ 0.003 \\ \hline 0.001 \\ -0.001 \end{pmatrix}, \; 10^{-4} \begin{array}{c|cc|cc} & \text{Tel4} & \text{Vale5} & \text{Pet4} & \text{Suz4} \\ \hline \text{Tel4} & 9.08 & 6.58 & 7.49 & 2.53 \\ \text{Vale5} & 6.58 & 9.27 & 6.53 & 2.00 \\ \hline \text{Pet4} & 7.49 & 6.53 & 10.01 & 2.86 \\ \text{Suz4} & 2.53 & 2.00 & 2.86 & 7.69 \end{array} \right)$$

- Suponha que, de alguma forma,antecipamos o retorno das duas últimas ações, PET4 e SUZ4, no dia seguinte.
- Estima-se que PET4 e SUZ4 terão ambas um aumento de 5% (isto é, $Y_{3t} = Y_{4t} = 0.05$).
- O que podemos dizer sobre os valores mais prováveis para TEL4 e VALE5?

Condicional: exemplo

- Distribuição de (Y_{1t}, Y_{2t}) dado que $Y_{3t} = 0.05$ e $Y_{4t} = 0.05$: Como

$$\boldsymbol{\mu} = \begin{pmatrix} 0.001 \\ 0.003 \\ \hline 0.001 \\ -0.001 \end{pmatrix} \quad \text{e } \boldsymbol{\Sigma} = 10^{-4} \begin{pmatrix} & \text{Tel4} & \text{Vale5} & \text{Pet4} & \text{Suz4} \\ \text{Tel4} & 9.08 & 6.58 & 7.49 & 2.53 \\ \text{Vale5} & 6.58 & 9.27 & 6.53 & 2.00 \\ \text{Pet4} & 7.49 & 6.53 & 10.01 & 2.86 \\ \text{Suz4} & 2.53 & 2.00 & 2.86 & 7.69 \end{pmatrix}$$

- sabemos que (Y_{1t}, Y_{2t}) será uma normal bivariada $N_2(\mathbf{m}, \boldsymbol{\Phi})$ com os seguintes parâmetros:

$$\begin{aligned} \boldsymbol{m} &= \boldsymbol{\mu}_{12} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \begin{pmatrix} y_3 - \mu_3 \\ y_4 - \mu_4 \end{pmatrix} \\ &= \begin{pmatrix} 0.001 \\ 0.003 \end{pmatrix} + \begin{pmatrix} 7.49 & 2.53 \\ 6.53 & 2.00 \end{pmatrix} \begin{pmatrix} 10.01 & 2.86 \\ 2.86 & 7.69 \end{pmatrix}^{-1} \begin{pmatrix} 0.05 - 0.001 \\ 0.05 + 0.001 \end{pmatrix} \\ &= \begin{pmatrix} 0.040 \\ 0.036 \end{pmatrix} \end{aligned}$$

Condicional: exemplo

- A matriz de covariância de (Y_{1t}, Y_{2t}) dado que $Y_{3t} = 0.05$ e $Y_{4t} = 0.05$:

$$\begin{aligned}
\mathbf{\Phi} &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\,\mathbf{\Sigma}_{22}^{-1}\,\mathbf{\Sigma}_{21} \\
&= 10^{-4}\left(\begin{pmatrix} 9.08 & 6.58 \\ 6.58 & 9.27 \end{pmatrix} - \begin{pmatrix} 7.49 & 2.53 \\ 6.53 & 2.00 \end{pmatrix}\begin{pmatrix} 10.01 & 2.86 \\ 2.86 & 7.69 \end{pmatrix}^{-1}\begin{pmatrix} 7.49 & 6.53 \\ 2.53 & 2.00 \end{pmatrix}\right) \\
&= 10^{-4}\begin{pmatrix} 3.45 & 1.69 \\ 1.69 & 5.01 \end{pmatrix}
\end{aligned}$$

- Assim, olhando diretamente as duas marginais CONDICIONADAS ao evento $Y_{3t} = 0.05$ e $Y_{4t} = 0.05$ e a correlação condicional:

$$(Y_{1t} | Y_{3t} = 0.05, Y_{4t} = 0.05) \sim N(0.040, (0.019)^2)$$

$$(Y_{2t} | Y_{3t} = 0.05, Y_{4t} = 0.05) \sim N(0.036, (0.022)^2)$$

Condicional: exemplo

- Assim, antes de saber o valor de Y_{3t} e Y_{4t}, sabíamos que

$$Y_{1t} \sim N(0.001, (0.0301)^2) \quad texte \quad Y_{2t} \sim N(0.003, (0.0304)^2)$$

- Isto é, espera-se uma valorização de 0.1% ao dia para PET4 e de 0.3% para VALE5.
- Agora que somos informados de que as ações PET4 e SUZ4 tiveram uma grande valorização de 5% de um dia para o outro, podemos revisar o que esperamos para as outras duas ações, PET4 e VALE5:

$$(Y_{1t} | Y_{3t} = 0.05, Y_{4t} = 0.05) \sim N(0.040, (0.019)^2)$$

$$(Y_{2t} | Y_{3t} = 0.05, Y_{4t} = 0.05) \sim N(0.036, (0.022)^2)$$

- Note que agora esperamos uma valorização de 4% para PET4 (40 vezes maior que antes) e de 3.6% para VALE5 (ou 12 vezes maior que antes).
- Note também como o DP diminuiu nos dois casos.

Condicional: exemplo

- Marginalmente, Y_{1t}ce Y_{3t} são correlacionadas e $\rho_{12} = 0.72$.
- A correlação condicional Cor$(Y_{1t}, Y_{2t}|Y_{3t} = 0.05, Y_{4t} = 0.05)$ entre estas duas ações PET4 e VALE5 é dada por

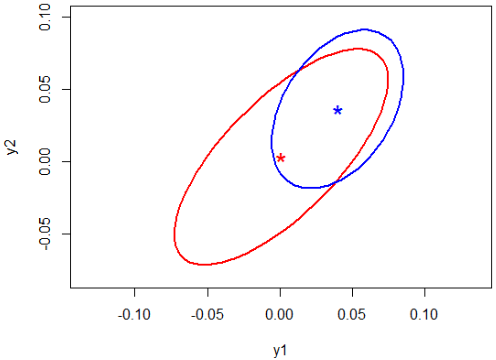
$$\frac{\text{Cov}(Y_{1t}, Y_{2t}|Y_{3t} = 0.05, Y_{4t} = 0.05)}{\sqrt{\mathbb{V}(Y_{1t}|Y_{3t} = 0.05, Y_{4t} = 0.05)\mathbb{V}(Y_{2t}|Y_{3t} = 0.05, Y_{4t} = 0.05)}}$$

$$\frac{1.69 \times 10^{-4}}{0.019 \times 0.022} = 0.40$$

- Assim, conhecer o evento $Y_{3t} = 0.05$ e $Y_{4t} = 0.05)$ DIMINUI a correlação entre as variáveis Y_{1t} e Y_{4t} de 0.70 para 0.40.
- Em resumo: as distribuições de Y_{1t} e Y_{4t} bem como sua correlação são bastante impactadas pelo evento $Y_{3t} = 0.05$ e $Y_{4t} = 0.05)$.

Condicional: exemplo

- Elipse vermelha: região em que (Y_1, Y_2) cai com probabilidade 95%.
- Elipse azul: dado que $Y_3 = 0.05$ e $Y_4 = 0.05$, esta é a região em que (Y_1, Y_2) cai com probabilidade 95%.



Propriedades: resumo

The following are true for a random vector **X** having a multivariate normal distribution:

1. Linear combinations of the components of **X** are normally distributed.
2. All subsets of the components of **X** have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.

Distância estatística e anomalias

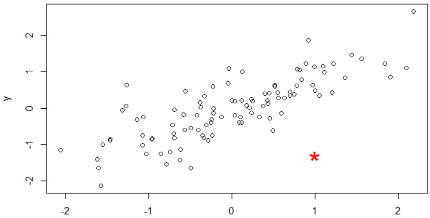
- Seja $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Considere a distância estatística de um ponto aleatório $\mathbf{Y}$ até seu vetor médio $\boldsymbol{\mu}$:
$$D^2 = (\mathbf{Y} - \boldsymbol{\mu})' \, \Sigma^{-1} \, (\mathbf{Y} - \boldsymbol{\mu})$$
- Se $\mathbf{Y}$ é um ponto aleatório, a distância D^2 também é aleatória.
- Algumas vezes esta distância é grande, algumas vezes ela é pequena.
- Qual a distribuição de probabilidade desta distância?
- Quando um ponto aleatório $\mathbf{Y}$ poderá ser considerado anômalo?
- Quando sua distância poderá ser considerada excessiva?

Uma coordenada por vez?

- Basta olhar cada coordenada, não?
- Sabemos que cada coordenada Y_i segue uma normal.
- Neste caso, a chance de cada coordenada Y_i se afastar de sua média μ_i por mais de 2DPs σ_i é aproximadamente 5%.
- E de se afastar mais de 3 DPs σ_i é 0.3% enquanto se afastar mais de 4 DPs é apenas 0.006%.
- Assim, temos um critério para determinar se um ponto é moderadamente ou muito anômalo olhando uma coordenada de cada vez.
- Acontece que podemos fazer melhor.
- Podemos ter um ponto muito anômalo n espaço k-dimensional mas NENHUMA de suas coordenadas é anômala!!

Anomalias em k dimensões

- Para ilustrar este fenômeno, vamos considerar apenas o caso bi-dimensional.
- Veja o gráfico abaixo de uma amostra de uma normal bi-variada $N_2(\mathbf{0}, \mathbf{\Sigma})$ onde o valor esperado de cada variável é zero, a variância é 1 e a correlação é $\rho = 0.85$.
- O ponto vermelho é claramente uma anomalia mas olhando-se os valores de *cada uma de suas coordenadas isoladamente* não nenhuma evidência de que seja uma anomalia.



Distribuição de D^2

- Como usar o resultado abaixo?

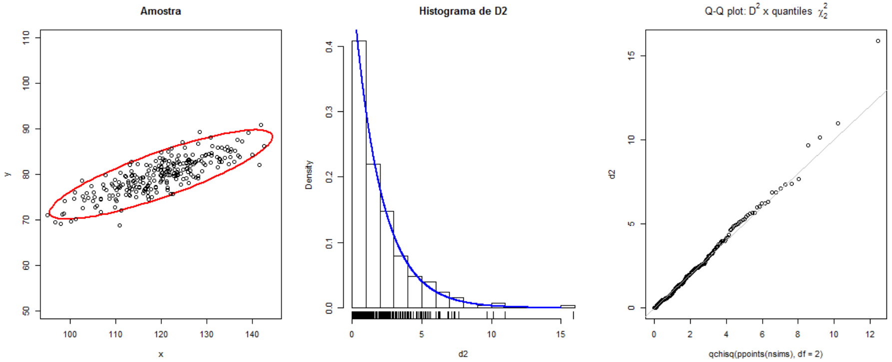
Result 4.7. Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. Then

(a) $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

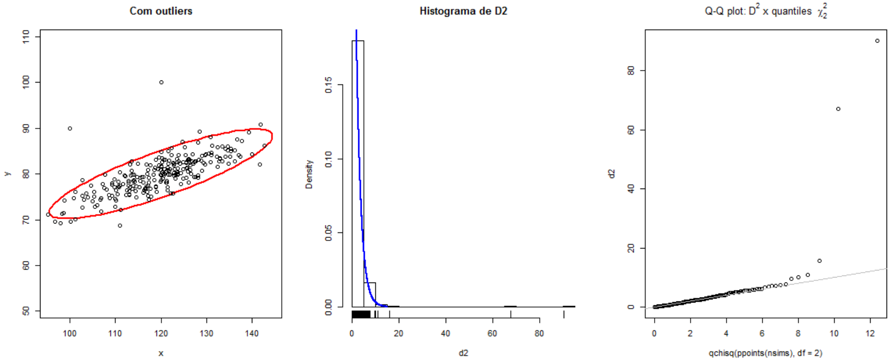
(b) The $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution assigns probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x}: (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\}$, where $\chi^2_p(\alpha)$ denotes the upper (100α)th percentile of the χ^2_p distribution.

- Calcule a distâcia estatística d^2 para cada ponto $\mathbf{y}$.
- Plote estes distância versus um certo valor esperado sob a hipótese de normalidade.
- O plot deveria se parecer com uma linha reta com inclinação 1 e passando pela origem.
- Uma curva sistematicamente desviando-seda reta indica que a distribuição não é normal multivariada.
- Apenas um ou dois pontos desviando-se muito acima da reta indicam grandes distâncias ou anomalias.

Exemplo: N_2 sem outliers



Exemplo: N_2 com dois outliers



2-dim ou k-dim

- O mais relevante é o QQ-plot,o terceiro plot.
- Se os vetores são 2-dim, um simples scatterplot das duas variáveis mostra quais são as anomalias: QQ-plot não ajuda muito.
- A principal utilidade dos QQ-plots é quando a dimensão do vetor for maior que 2.
- Não conseguimos visualizar todos as dimensões ao mesmo tempo.
- Em R:

```
d2 = mahalanobis(x, center = mu, cov=S)
qqplot(qchisq(ppoints(length(d2)), df = ncol(x)), d2)
```

Distribuição qui-quadrado

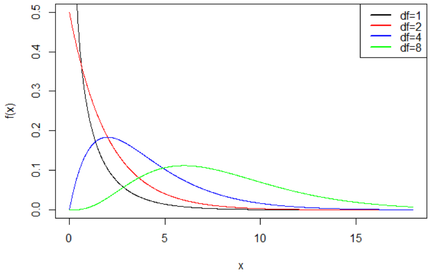
- Vimos que D^2 segue uma distribuição qui-quadrado com k graus de liberdade quando o vetor $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Já encontramos esta distribuição no teste qui-quadrado.
- Esta distribuição só depende da dimensão do vetor.
- Não interessa os valores típicos nem a escala dos elementos do vetor $\mathbf{Y}$.
- Ela é uma métrica universal para desvios em k dimensões quano os dados são gaussianos.
- Como é a cara desta distribuição qui-quadrado?
- Ela é uma distribuição contínua com uma densidade

$$f(x) = \text{ cte } x^{k/2-1} e^{-x/2}$$

Distribuição qui-quadrado

- Por exemplo:

$$f(x) = \text{ cte } e^{-x/2} \quad \text{com } k = 2 \text{ degrees of freedom}$$
$$f(x) = \text{ cte } xe^{-x/2} \quad \text{com } k = 4 \text{ d.f.}$$
$$f(x) = \text{ cte } x^9 e^{-x/2} \quad \text{com } k = 20 \text{ d.f.}$$



Quantil de qui-quadrado

- Como $D^2 \sim \chi^2(k)$, achamos o valor de x que deixa uma proporção p da área da $\chi^2(k)$ abaixo dele.
- Este valor é chamado de quantil e denotado por $q(p)$.

$$\text{quantil } p = q(p) \text{ tal que } \mathbb{P}(\chi^2(k) \leq q(p)) = p$$

- Por exemplo, para a proporção $p = 0.65$, queremos

$$\text{quantil } 0.65 = q(0.65) \text{ tal que } \mathbb{P}(\chi^2(k) \leq q(0.65)) = 0.65$$

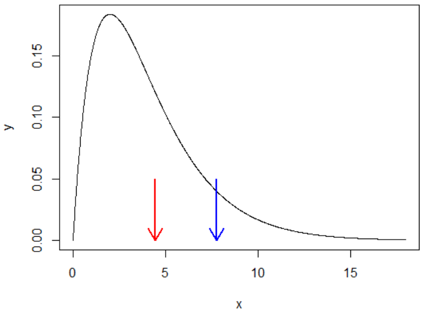
- Sendo mais específicos, suponha que a dimensão do vetor é $k = 4$.
- Então queremos

$$\text{quantil } 0.65 = q(0.65) \text{ tal que } \mathbb{P}(\chi^2(4) \leq q(0.65)) = 0.65$$

- Esta valor é $q(0.65) = 4.438$ e facilmente encontrado em R com `qchisq(0.65, 4)`.

Quantil de qui-quadrado

- Dois quantis de uma qui-quadrado com 4 graus de liberdades.
- Seta vermelha: com $p = 0.65$, o quantil $q(0.65)$
- Seta azul: com $p = 0.90$, o quantil $q(0.90)$
- Estes quantis são valores TEÓRICOS, não precisam de dados para serem obtidos, apenas da densidade de probabilidade.



Os quantis $q(p)$

- Para fazer o QQ-plot com uma amostra de tamanho n, os quantis $q(p)$ de interesse são os valores p no conjunto

$$p \in \left\{ \frac{1-0.5}{n}, \frac{2-0.5}{n}, \frac{3-0.5}{n}, \ldots, \frac{n-0.5}{n} \right\}$$

- Por exemplo, se $n = 137$ então calculamos os quantis $q(p)$ para p no conjunto

$$p \in \left\{ \frac{1-0.5}{137}, \frac{2-0.5}{137}, \frac{3-0.5}{137}, \ldots, \frac{137-0.5}{137} \right\} = \{0.0036, 0.0109, 0.0182, \ldots, 0.9963\}$$

- A subtração de 0.5 no numerador épara evitar problemas nos extremos: não queremos o quantil associado com a proporção $p = n/n = 1$. Para manter a consistância, 0.5 é subtraído de todos os elementos.
- Se a dimensão do vetor é $k = 4$, estes quantis são obtidos em R com os comandos

```
prop = ppoints(nsims) # conjunto de pontos p
qp = qchisq(prop, df = 4)
```

Os quantis $q(p)$ e a amostra

- Sabemos que o valor teórico $q(p)$ para a distância estatística é aquele valor que deixaria aproximadamente $p100\%$ das distâncias de uma amostra abaixo dele.
- Isto é, como
$$\mathbb{P}(D^2 \leq q(p)) = p$$
então esperamos $p100\%$ dos pontos com uma distância estatística menor que $q(p)$.
- Por exemplo, com um vetor de dimensão $k = 4$, o quantil $q(0.65) = 4.438$ pois
$$\mathbb{P}(\chi^2(4) \leq 4.438) = 0.65$$
- Assim, ao calcular a distância estatística D^2 para cada ponto da amostra, esperamos que 65% deles tenham distância menor que 4.438.

Quantil teórico e empírico

- Podemos obter uma boa estimativa do valor TEÓRICO $q(p)$ a partir da amostra.
- Calcular a distância estatística D^2 para cada ponto da amostra e *ordene os valores*:
$$D^2_{(1)} \leq D^2_{(2)} \leq \ldots \leq D^2_{(n)}$$
- O símbolo (i) no sub-índice indica a i-ésima *estatística de ordem*.
- Considere uma proporção $p = i/n$.
- Por exemplo, suponha que $k = 4$, $n = 100$ e queremos $p = 76/100 = 0.76$.
- Temos $q(0.76) = 5.4969 =$ `qchisq(0.76, 4)`:
$$\mathbb{P}(D^2 \leq 5.4969)) = 0.76$$

Quantil teórico e empírico

- A 76-ésima estatística de ordem D^2_{76} deixa 76% dos pontos abaixo ou igual a ela.
- Assim, se quiseremos uma estimativa *EMPÍRICA* (baseada nos dados) do valor teórico $q(0.76) = 5.4969$ (obtido da densidade da χ^2), o valor D^2_{76} é um bom candidato.
- Esperamos $D^2_{(76)} \approx q(0.76) = 5.4969$ pois como

$$\mathbb{P}(D^2 \leq 5.4969) = 0.76 \, ,$$

esperamos 76% da amostra abaixo do valor teórico $q(0.76) = 5.4969$.

- Como temos exatamente 76% abaixo do valor EMPÍRICO $D^2_{(76)}$,

$$\text{proporção das variáveis } leqD^2_{(76)} = 76/100 \, ,$$

podemos esperar $D^2_{(76)} \approx q(0.76) = 5.4969$.

Então, fazendo o QQ-plot

- Tome a proporção $p = (i - 0.5)/n$ e ache o quantil TEÓRICO

$$\text{quantil } \frac{i - 0.5}{n} = q_i \text{ tal que } \mathbb{P}(\chi^2(k) \leq q_i) = \frac{i - 0.5}{n}$$

- Variamos i calculamos os quantis q_i para $i = 1, 2, \ldots, n$.
- Para cada i, pareamos q_i e a i-ésima distância ordenada $D^2_{(i)}$.
- Devemos ter $q_i \approx D^2_{(i)}$ para todo i.
- Portanto, os pares $(q_i, D^2_{(i)})$ devem cair ao longo da reta $y = x$ se os pontos são uma amostra aleatória de uma nomal multivariada.

Checando em R

- Vamos simular alguns dados de uma normal multivariada e verficar que seu QQ-plot segue o padrão esperado.
- Seja $\mathbf{Y} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ onde $\boldsymbol{\mu} = (4, 3, 2, 1)$ e matriz de covariância

$$\boldsymbol{\Sigma} = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}$$

- Vamos gerar uma amostra de `nsims`=200 vetores i.i.d. desta distribuiçãao e fazer o QQ-plot com as distâncias estatísticas.
- NOTE QUE: Na prática, temos APENAS a amostra, sem os valores teóricos $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, que devem ser estimados a partir dos dados.

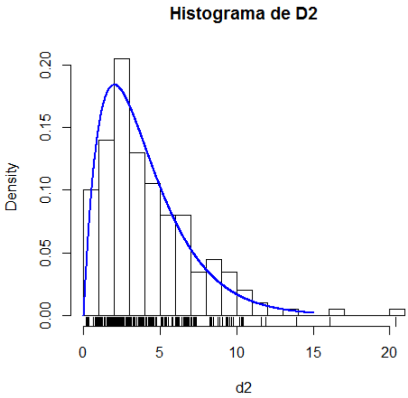
Script R

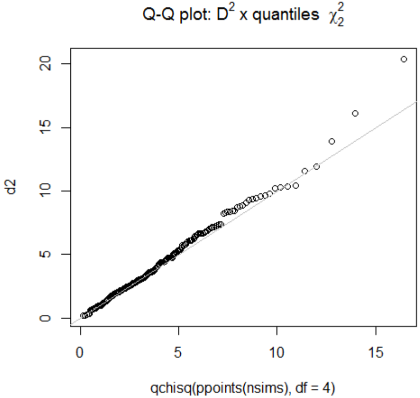
```
require(MASS); nsims=200; set.seed(1)
Sigma = matrix(c(3,0,2,2,0,1,1,0,2,1,9,-2,2,0,-2,4),4,4)
pts = mvrnorm(nsims, c(4,3,2,1), Sigma)
m = apply(pts, 2, mean)
round(m, 3) # [1] 3.863 2.997 1.909 0.998
S = cov(pts); round(S, 2)
#       [,1]  [,2]  [,3]  [,4]
#[1,]  2.97 -0.09  1.50  2.17
#[2,] -0.09  1.16  0.98 -0.02
#[3,]  1.50  0.98  7.87 -1.66
#[4,]  2.17 -0.02 -1.66  3.96
d2 = mahalanobis(pts, m, S)
# Passamos m e S, as ESTIMATIVAS de mu e Sigma
```

Script R

```
par(mfrow=c(1,2))
hist(d2, n=15, prob=T, main = "Histograma de D2")
rug(d2)
aux = seq(0, 15, by=0.01)
yaux = dchisq(xaux, df=ncol(pts))
lines(xaux, yaux, , lwd=2, col="blue")
qqplot(qchisq(ppoints(nrow(pts)), df = ncol(pts)), d2,
  main = expression("Q-Q plot:" * ~D^2 *
         " x quantiles " * ~ chi[2]^2))
abline(0, 1, col = 'gray')
```

Amostra de $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$





Amostra com outliers

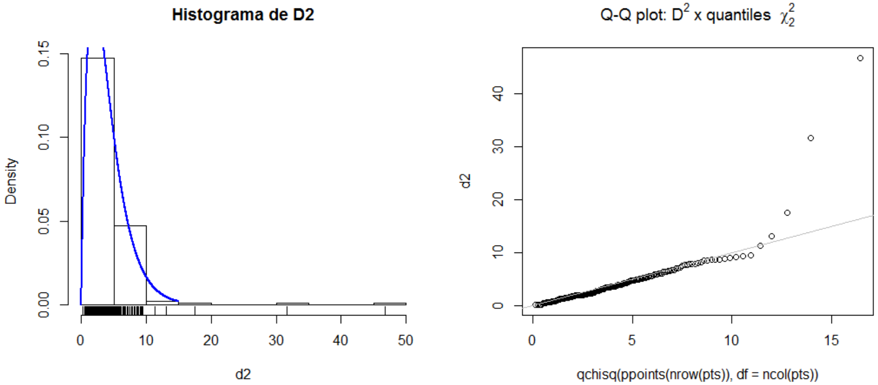


Figura: Adicionando dois outliers

Script: mais dois outliers

```
set.seed(1)
pts = mvrnorm(nsims, c(4,3,2,1), Sigma)
pts = rbind(pts, c(4,3,-3,-5), c(10,6,2,4))
m = apply(pts, 2, mean); S = cov(pts)
d2 = mahalanobis(pts, m, S)
par(mfrow=c(1,2))
hist(d2, n=15, prob=T, main = "Histograma de D2"); rug(d2)
aux = seq(0, 15, by=0.01); yaux = dchisq(xaux, df=ncol(pts))
lines(xaux, yaux, , lwd=2, col="blue")
qqplot(qchisq(ppoints(nrow(pts)), df = ncol(pts)), d2,
       main = expression("Q-Q plot:" * ~D^2 *
       " x quantiles " * ~ chi[2]^2))
abline(0, 1, col = 'gray')
```