$$X \sim N_p(\mu, \Sigma)$$

$$f(\underbrace{x_1, x_2 \ldots x_p}_{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right\}$$

<u>Caso particular $p=2$</u>

<u>Distribuição Normal Bivariada</u>

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}'$$

$$\rho_{12} = \mathrm{Corr}(X_1, X_2) \Rightarrow \sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$$

$$(x-\mu)' \Sigma^{-1} (x-\mu) = \begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix} \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \times$$

$$\times \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$f(x) = f(x_1, x_2) = \frac{1}{2\pi\sqrt{\underbrace{\sigma_{11}\sigma_{22} - \rho_{12}^2\sigma_{11}\sigma_{22}}_{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

$$x' = (x_1, x_2, \dots x_p)$$

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

$f(x)$ é constante para todo x tal que

$(x-\mu)'\Sigma^{-1}(x-\mu)$ é constante.

Σ positiva definida $\Rightarrow$ Σ^{-1} é positiva definida

$\Rightarrow (x-\mu)'\Sigma^{-1}(x-\mu) = c^2$

Equação de um elipsoide com centro em μ.
O comprimento do maior eixo desse elipsoide é $2\sqrt{\lambda_1}\,c$, onde λ_1 é a maior raiz característica de Σ.
A direção do maior eixo é a do vetor característico associado.

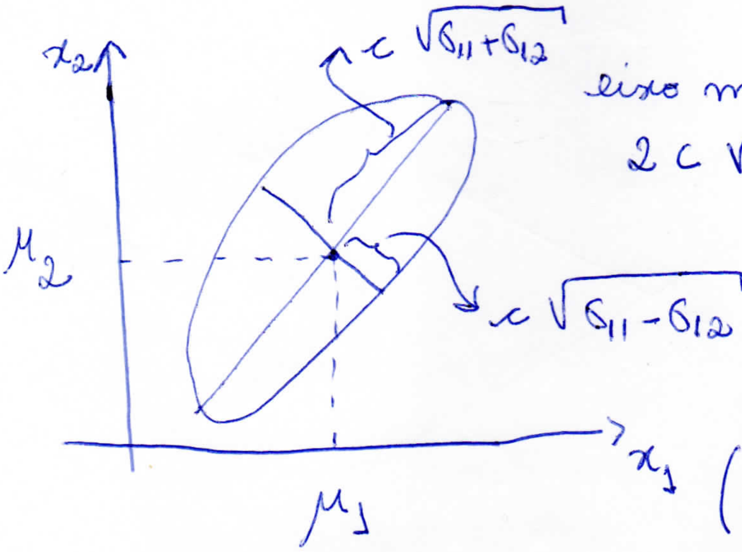
$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix}$$ Exemplo $p=2$ $\sigma_{11}=\sigma_{22}$

$$|\Sigma - \lambda I| = 0 \Rightarrow \begin{vmatrix} \sigma_{11}-\lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11}-\lambda \end{vmatrix} = 0 \Longrightarrow \underbrace{\begin{matrix} \lambda_1 = \sigma_{11}+\sigma_{12} \\ \lambda_2 = \sigma_{11}-\sigma_{12} \end{matrix}}_{\text{raízes características}}$$

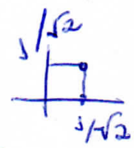
$$\begin{bmatrix} \sigma_{11}-\lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11}-\lambda \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ para } \lambda_1$$

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \text{ para } \lambda_2$$

Se $\sigma_{12} > 0$ ($\Rightarrow \rho_{12} > 0$), a maior raiz característica é $\lambda_1 = \sigma_{11} + \sigma_{12}$ e o vetor caract. associado é $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.

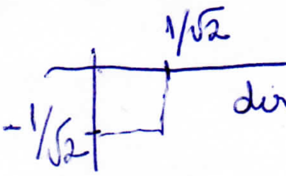


eixo maior, comprimento $2c\sqrt{\sigma_{11}+\sigma_{12}}$, e direção definida pelo vetor $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ (45°)



($\rho > 0$, relação linear crescente entre x_1 e x_2)

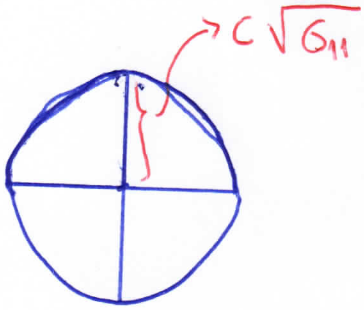
Se $\sigma_{12} < 0$ ($\Rightarrow \rho_{12} < 0$), maior r. c. $\lambda_2 = \sigma_{11} - \sigma_{12}$ e vetor característico $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$.



direção -45^0

Se $\rho_{12} = 0$ $(\sigma_{12} = 0)$, a elipse é um círculo

$\lambda_1 = \sigma_{11}$ e $\lambda_2 = \sigma_{11}$



$(x-\mu)' \Sigma^{-1} (x-\mu) = c^2$ nesse caso fica

$$\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix} \begin{bmatrix} 1/\sigma_{11} & 0 \\ 0 & 1/\sigma_{11} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} = c^2.$$

$$\frac{(x_1-\mu_1)^2}{\sigma_{11}} + \frac{(x_2-\mu_2)^2}{\sigma_{11}} = c^2$$

equação de uma circunferência de centro (μ_1, μ_2) e raio $\sqrt{\sigma_{11}}\, c$.

Para p qualquer, $X \sim N_p(\mu, \Sigma)$, Σ positiva definida.

<u>Resultado 1</u>

$X \sim N_p(\mu, \Sigma)$. Se A é uma matriz $m \times p$ de constantes, posto $m \leq p$, o vetor aleatório m dimensional

$$\underset{m \times 1}{Y} = \underset{m \times p}{A} \; \underset{p \times 1}{X}$$

tem distribuição normal multivariada com média $E(Y) = A\mu$ e matriz de covariância $A \Sigma A'$.

$$Y \sim N_m(A\mu, A\Sigma A').$$

Obs:

- Se $m = 1$ $A = a' = [a_1 \; a_2 \cdots a_p]$ $AX = a_1 X_1 + \cdots + a_p X_p$ é unidimensional, tem distribuição normal com média $a'\mu$ e matriz de covariância $a' \Sigma a$.

- Tomando $A = a' = [0 \dots 0 \; 1 \; 0 \dots 0]$, $AX = X_j$

$$\begin{array}{c} \\ \\ \end{array}$$

tem distribuição normal com média $a'\mu = \mu_j$ e variância

$$A \Sigma A' = [0 \dots 0 \; 1 \; 0 \dots 0] \, \Sigma \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{jj}.$$

$$\boxed{\therefore X_j \sim N(\mu_j, \sigma_{jj}).}$$

As marginais são normais univariadas.

Marginais normais $\not\Rightarrow$ conjunta normal multivariada

Marginais normais independentes $\Rightarrow$ Conjunta normal multivariada

- Qualquer subconjunto de q das variáveis $X_1, X_2, \ldots X_p$ tem distribuição normal q-variada.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \qquad A = \begin{bmatrix} I_q & 0_{q \times p-q} \end{bmatrix}$$

$$\underset{q \times p}{A} \underset{p \times 1}{X} = \begin{bmatrix} I_q & 0_{q \times p-q} \end{bmatrix} \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} = \underset{q \times 1}{X^{(1)}}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$A\mu = \begin{bmatrix} I_q & 0_{q \times p-q} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_1$$

$$A \Sigma A' = \begin{bmatrix} I_q & 0 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_q \\ 0_{p-q \times q} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \end{bmatrix} \begin{bmatrix} I_q \\ 0_{p-q \times q} \end{bmatrix}$$

$$= \Sigma_{11}$$

Pelo resultado $AX = X^{(1)} \sim N_q(\mu_1, \Sigma_{11})$

Variância Generalizada

X - Matriz de dados

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix}$$

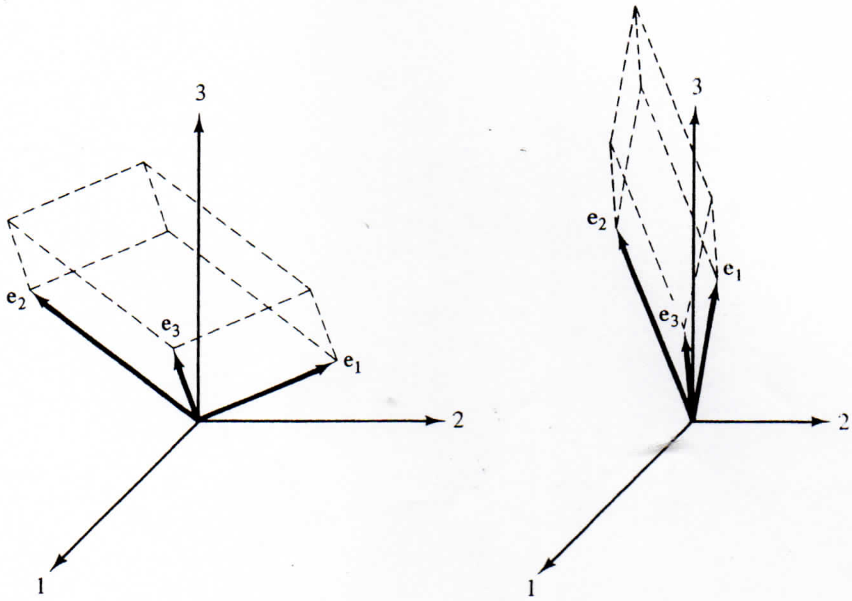
~~n~~ p pontos no espaço ~~p~~ n-dimensional

$$S^* = (s^*_{jk}) \qquad s^*_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

$$j, k = 1, 2, \dots p$$

Para resumir a variabilidade expressa em S^* através de um único número utiliza-se

Variância Amostral Generalizada $= |S^*|$



(a) (b)

Figure 3.6 (a) "Large" generalized sample variance for $p = 3$. (b) "Small" generalized sample variance for $p = 3$.

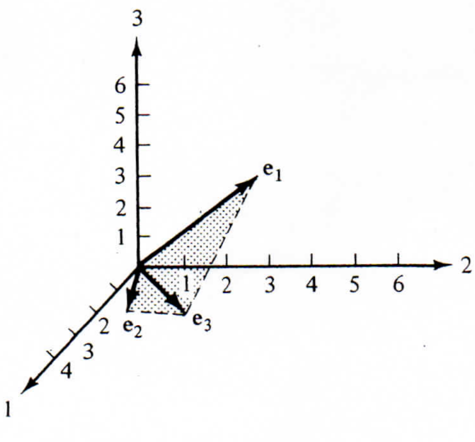


Figure 3.7 A case where the three-dimensional volume is zero ($|\mathbf{S}| = 0$).

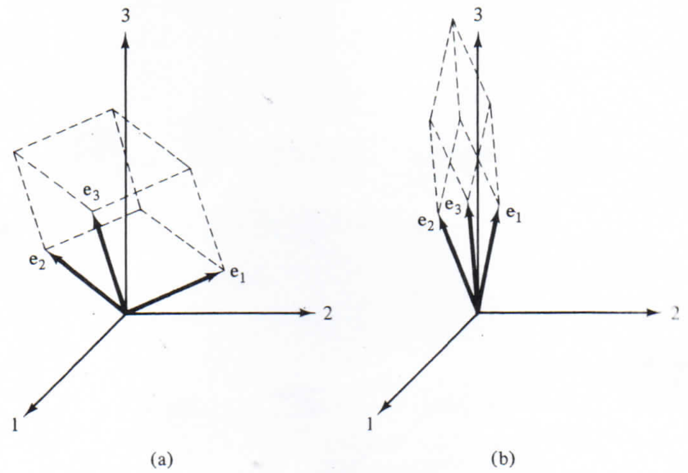


Figure 3.8 The volume generated by equal-length deviation vectors of the standardized variables.

Como gerar um vetor aleatório com distribuição normal p-variada

Gerar $Z_1, Z_2 \ldots Z_p$ independentes $\sim N(0,1)$

$Z' = [Z_1 \ Z_2 \ \ldots \ Z_p] \qquad Z \sim N_p(0, I)$

Desejamos gerar $X \sim N_p(\mu, \Sigma)$.

$\Sigma = L L'$ (decomposição de Cholesky)

Determina-se L e obtém-se

$X = \mu + L Z \Rightarrow X \sim N_p$

$E(X) = \mu + L E(Z) = \mu + L 0 = \mu$

$Var(X) = Var(\mu + L Z) = Var(L Z) =$

$= L \, Var(Z) \, L' = L I L' = \Sigma$

$\therefore X \sim N_p(\mu, \Sigma)$

Verificação da Normalidade Multivariada

$\underset{\sim}{X}_i \quad i = 1, 2, \ldots n \qquad \underset{\sim}{X}_i \sim N_p \, ?$

a) $X \sim N_p \Rightarrow$ As distribuições marginais de X são normais univariadas.

b) Combinações lineares das componentes são normais univariadas.

c) Diagramas de dispersão apresentam um formato "elíptico".

$X \sim N_p \Rightarrow$ a) b) c)

a) b) c) $\not\Rightarrow N_p$

não a) ou não b) ou não c) $\Rightarrow$ não N_p

Definição: Distância de Mahalanobis (ou Distância Estatística)

$$d^2 = (X-\mu)' \Sigma^{-1} (X-\mu)$$

Se $\mu = 0$ $\Sigma = I$ então $d^2 = X'X = \sum_{j=1}^{p} X_j^2$

Para verificar a suposição de normalidade p-variada, tomada uma amostra de n vetores aleatórios p-dimensionais $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots \underset{\sim}{x}_n$ calcula-se

$$d_i^2 = (\underset{\sim}{x}_i - \bar{x})' S^{-1} (\underset{\sim}{x}_i - \bar{x}) \qquad i = 1, 2 \ldots n$$

Se $X \sim N_p$ e n e $n-p$ são maiores que 30, $d_i = 1, 2 \ldots n$ se aproximam da distribuição qui-quadrado [d_i's não são independentes nem têm distribuição exata χ^2].

Gráfico

1 - $d^2_{(1)} \leq d^2_{(2)} \leq \cdots \leq d^2_{(n)}$

2 - Colocar no gráfico os pontos

$$\left(d^2_{(i)} , q_p\left((i - 0{,}5)/n \right) \right)$$

$q_p\left((i - 0{,}5)/n \right)$ quantil de ordem

$\dfrac{i - 0{,}5}{n}$ da distribuição χ^2_p.

Se $X \sim N_p$, o gráfico deve ser próximo da reta $y = x$ (reta pela origem, inclinação 1).

Detecção de "Outliers" Multivariados

13

1. Dot plot de cada X_j, $j = 1, 2 \dots p$

2. Diagrama de dispersão de cada par de variáveis X_j, X_k $j, k = 1, 2 \dots p$
$j \neq k$

3. Cálculo de $z_{jk} = \dfrac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}}$ $j = 1, 2 \dots n$
$k = 1, 2 \dots p$

1 e 3 não detectam outliers Multivariados (ver figura)

4. Cálculo de $d_i^2 = (\underset{\sim}{x}_i - \bar{x})' S^{*-1} (\underset{\sim}{x}_i - \bar{x})$

Altos valores de d_i^2 devem ser analisados

5. Cálculo de $r_i^2 = \dfrac{|S^*_{(i)}|}{|S^*|}$

(é equivalente a 4. pois $r_i^2 = 1 - \dfrac{n d_i^2}{(n-1)}$)

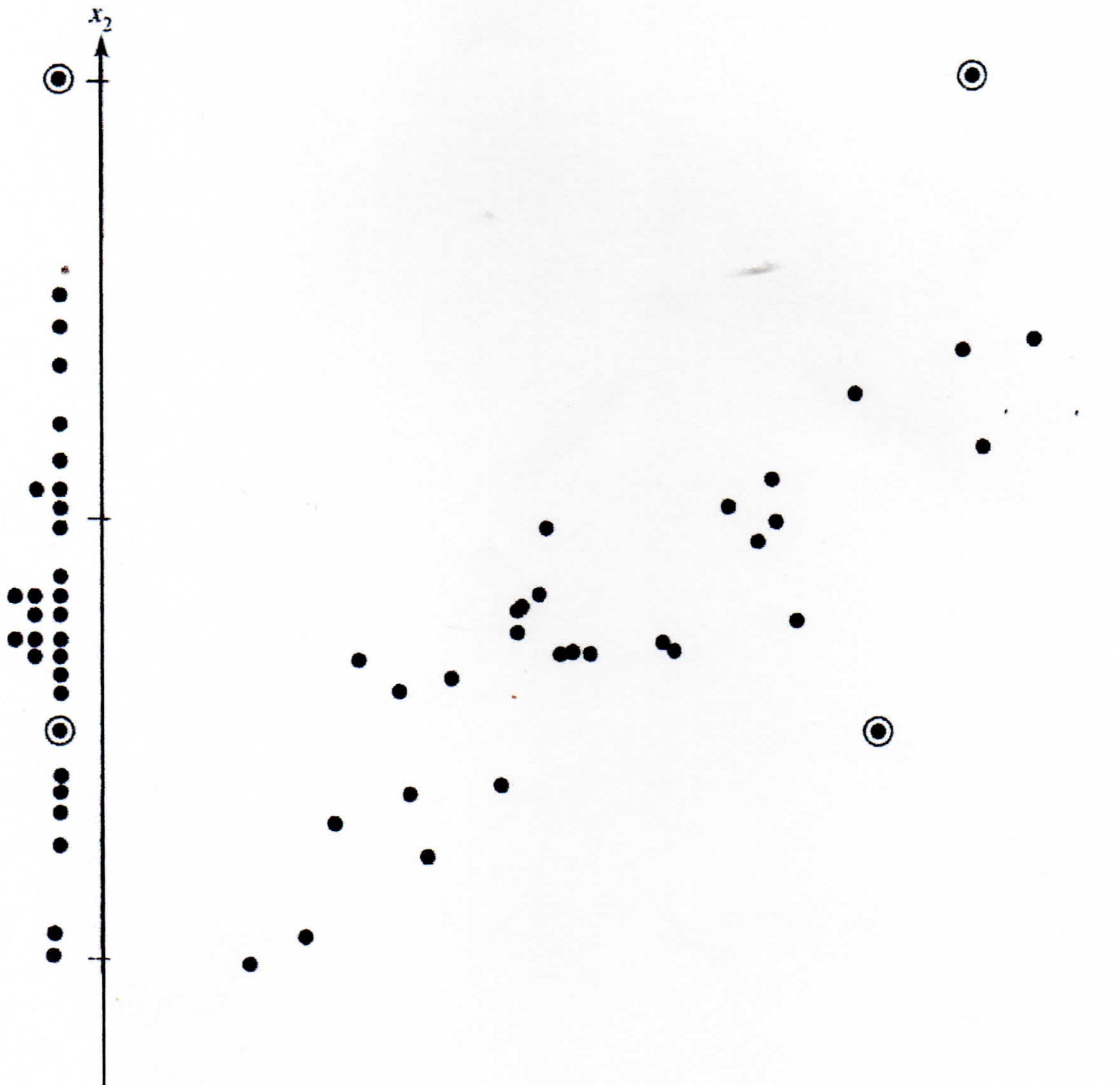


Figure 4.10 Two outliers: one univariate and one bivariate.

Função Geradora de Momentos da Distribuição Normal Multivariada

$$m_X(\underset{p\times 1}{t}) = E(e^{t'X}) = \exp\left(t'\mu + \frac{1}{2}t'\Sigma t\right)$$

$$E(X) = \mu \qquad Var(X) = \Sigma$$

$\nearrow t' = z'A$

$$Y = AX \quad m_Y(z) = E(e^{z'Y}) = E(e^{z'AX}) = m_X(A'z)$$

$$= \exp\left[z'A\mu + 1/2\, z'A\Sigma A' z\right] \quad \text{FGM de } N_p(A\mu, A\Sigma A')$$

$A_{m\times p}$

Definição "Sofisticada"

O vetor aleatório p-dimensional X tem distribuição normal p-variada se e somente se toda combinação linear dos elementos de X tem distribuição normal univariada.

$$X \sim N_p \iff t'X \sim N_1 \quad \forall t \in \mathbb{R}^p.$$