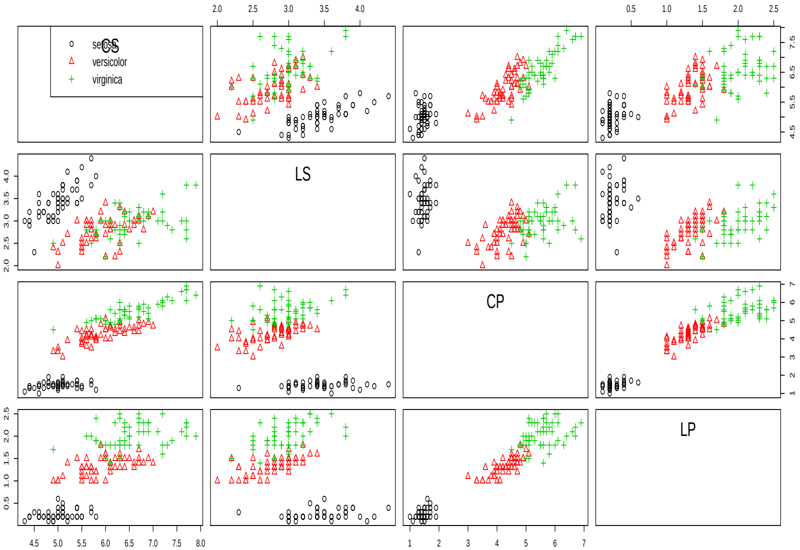
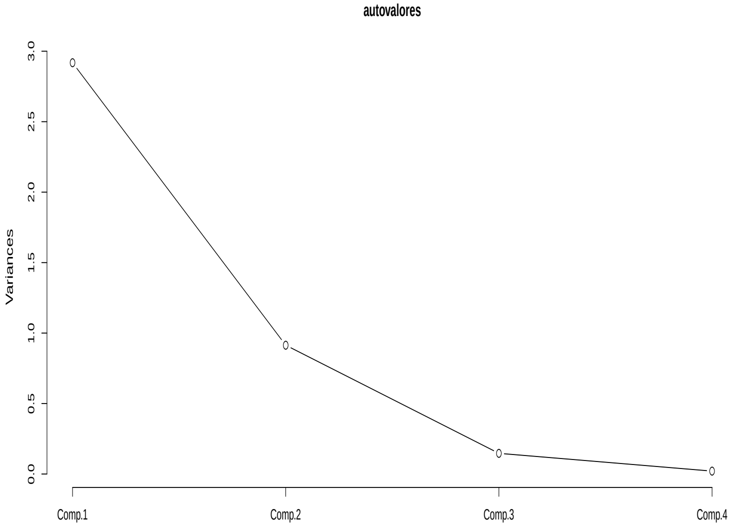
Análise de componentes principais: parte 2

Prof. Caio Azevedo

Exemplo 1: dados da íris “de Fisher”

- Recapitulando: quatro variáveis, três grupos, 50 observações por grupo.
- Objetivos: caracterizar os grupos em relação à essas quatro variáveis e compará-los.
- Utilizaremos a análise de componentes principais (usando a matriz de correlações) para esse fim.


setosa
versicolor
virginica
CS
LS
CP
LP


autovalores
Variances
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Comp.1
Comp.2
Comp.3
Comp.4

- Variâncias de cada componente:

$\widetilde{\lambda}_1 = 2,91, \widetilde{\lambda}_2 = 0,91, \widetilde{\lambda}_3 = 0,15, \widetilde{\lambda}_4 = 0,02.$

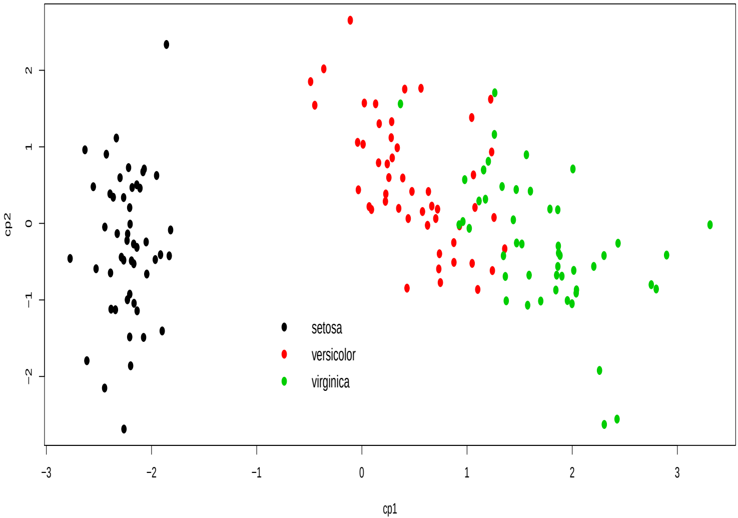
- Variância explicada

	Comp. 1	**Comp. 2**	**Comp. 3**	**Comp. 4**
PVE (%)	72,96	22,85	3,67	0,51
PVEA (%)	72,96	95,81	99,48	100,00

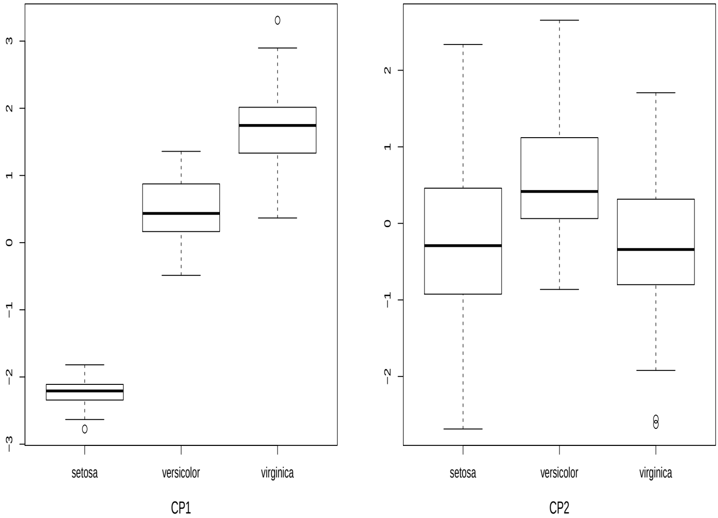
- Componentes principais

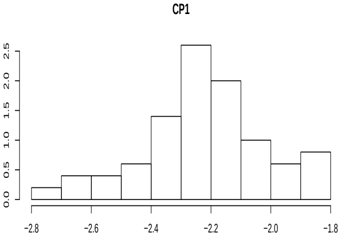
	Comp. 1	**Comp. 2**
CS	0,52 (0,89)	-0,37 (0,37)
LS	-0,27 (-0,46)	-0,92 (0,88)
CP	0,58 (0,99)	-0,02 (0,02)
LP	0,56 (0,96)	-0,07 (0,07)

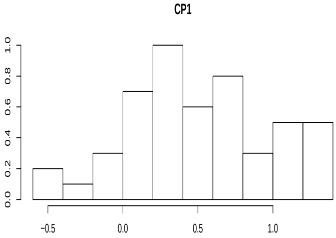
- Primeira componente: constrate entre as variáveis CS, CP e LP e a variável LS.
- Segunda componente: é uma média ponderada entre as variáveis CS e LS.

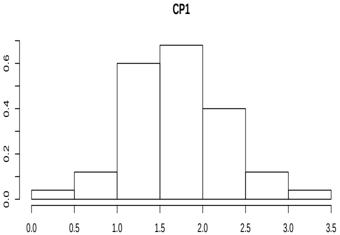


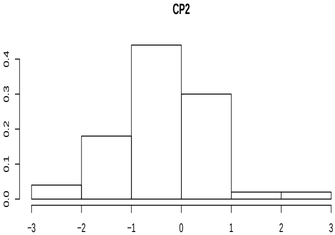
setosa
versicolor
virginica
cp1
cp2
-3
-2
-1
0
1
2
3

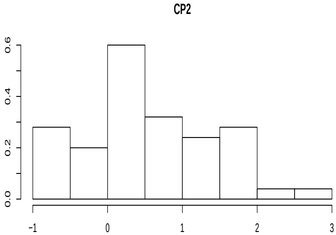

3
2
1
0
−1
−2
−3
setosa
versicolor
virginica
CP1
2
1
0
−1
−2
setosa
versicolor
virginica
CP2

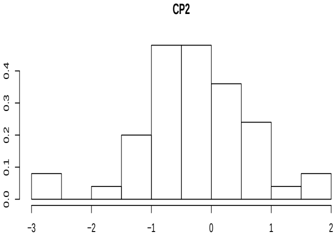

CP1
0.0
0.5
1.0
1.5
2.0
2.5
−2.8
−2.6
−2.4
−2.2
−2.0
−1.8

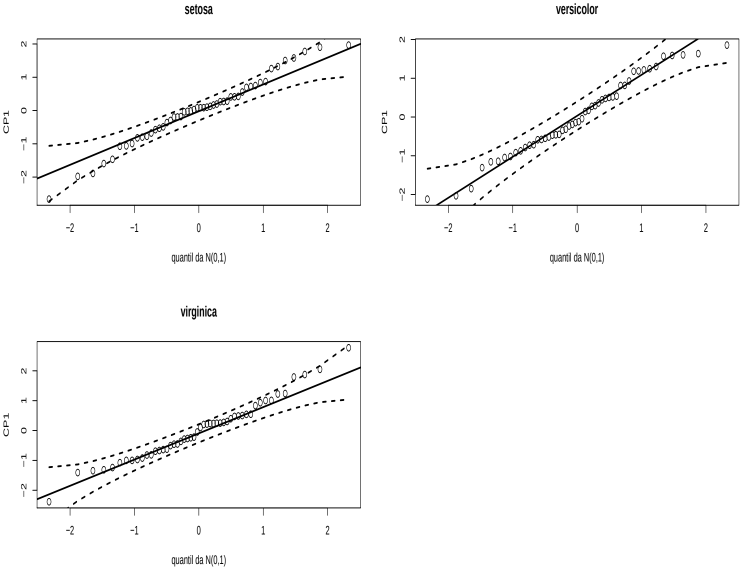

CP1
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.0
0.5
1.0


CP1
0.0
0.2
0.4
0.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

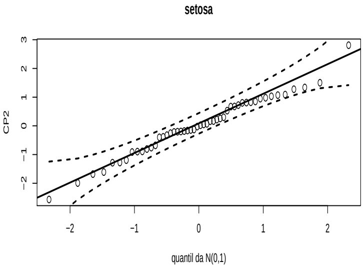

CP2
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3

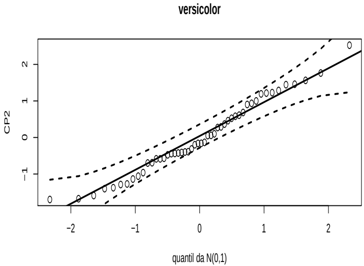

CP2
0.0
0.2
0.4
0.6
-1
0
1
2
3

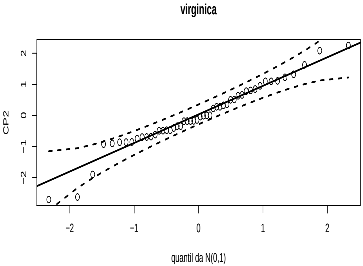

CP2
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2

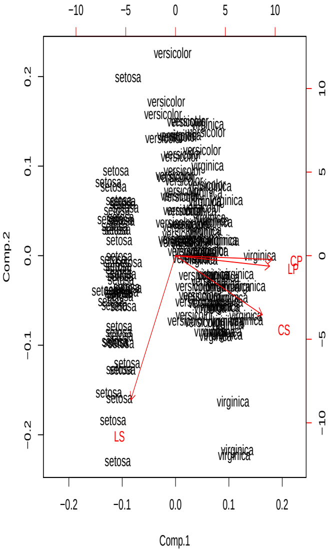


setosa
versicolor
virginica
CP1
quantil da N(0,1)


setosa
CP2
quantil da N(0,1)


versicolor
CP2
quantil da N(0,1)


virginica
CP2
quantil da N(0,1)


Comp.1
Comp.2
LS
CS
CP
LP

Comparação de grupos via modelos lineares

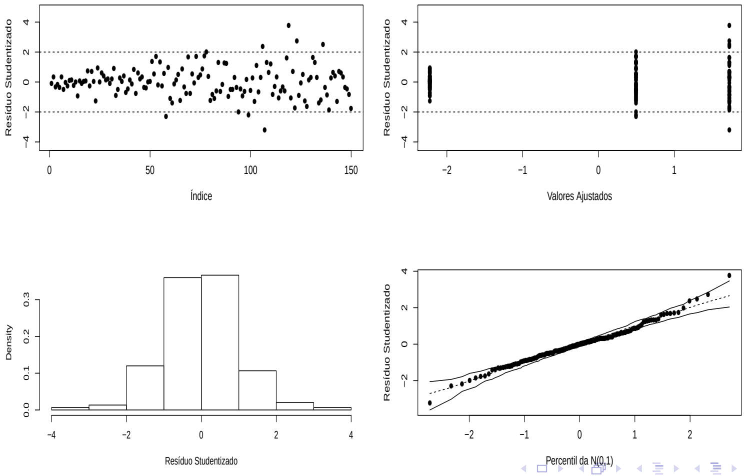
$Y_{ijk} = \mu_k + \alpha_{ik} + \xi_{ijk}$,

$i = 1, 2, 3$(tipo de iris, setosa, veriscolor, virginica), $j = 1, ..., n_i$,

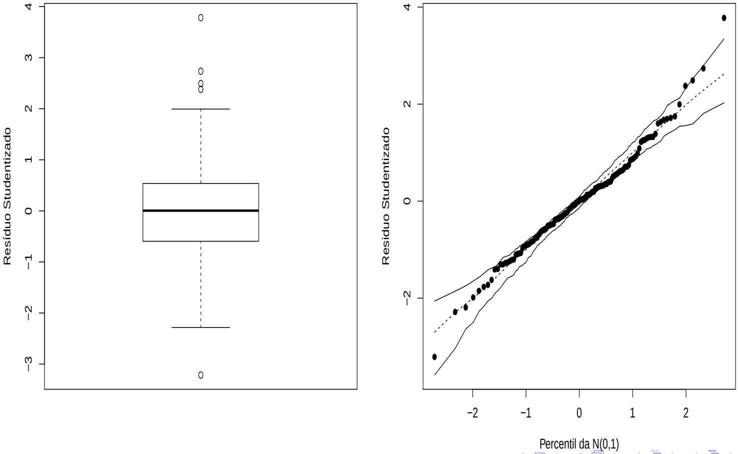
$n_i = 50, \forall i, k = 1, 2$(componente principal), $\xi_{ijk} \overset{i.i.d}{\sim} N(0, \sigma_k^2)$

- Y_{ijk} : valor da componente principal k, da planta j, do tipo de íris i.
- μ_k : média da componente principal k do grupo de referência (setosa).
- α_{ik} : incremento na média da componente principal k, do grupo i, em relação ao grupo de referência.
- Utilizou-se o resíduo "studentizado" (veja `http://www.ime.unicamp.br/~cnaber/aula_Diag_REG_2S_2014.pdf`) (para verificar a qualidade de ajuste do modelo).

Componente 1



Componente 1



Componente 1

Parâmetro	Estimativa	EP	Estat.t	p-valor
μ_1	-2,22	0,06	-35,65	$< 0{,}0001$
α_{21}	2,72	0,09	30,83	$< 0{,}0001$
α_{31}	3,95	0,09	44,79	$< 0{,}0001$

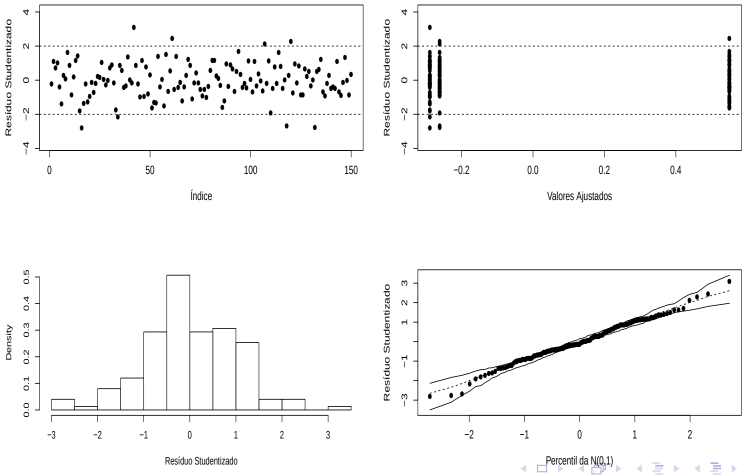
Há diferença entre o primeiro grupo e os outros dois. Vamos agora testar a igualdade entre as médias dos dois outros grupos através de testes do tipo $\mathbf{C}\boldsymbol{\beta} = \mathbf{M}$ (veja mais detalhes em `http://www.ime.unicamp.br/~cnaber/aula_testes_Cbeta_analise_REG_2S_2014.pdf`).

Componente 1

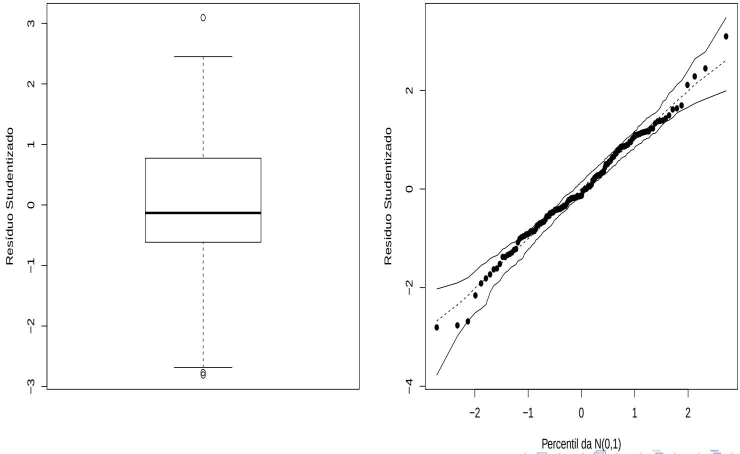
- Teste: $\alpha_{21} = \alpha_{31}$ vs $\alpha_{21} \neq \alpha_{31}$, $f_c = 194,83 (< 0,0001)$.
- Médias preditas pelo modelo.

grupo	Estimativa	EP	IC(95%)
setosa	-2,22	0,06	[-2,35 ; -2,10]
versicolor	0,50	0,06	[0,37 ; 0,62]
virginica	1,73	0,06	[1,60 ; 1,85]

Componente 2



Componente 2

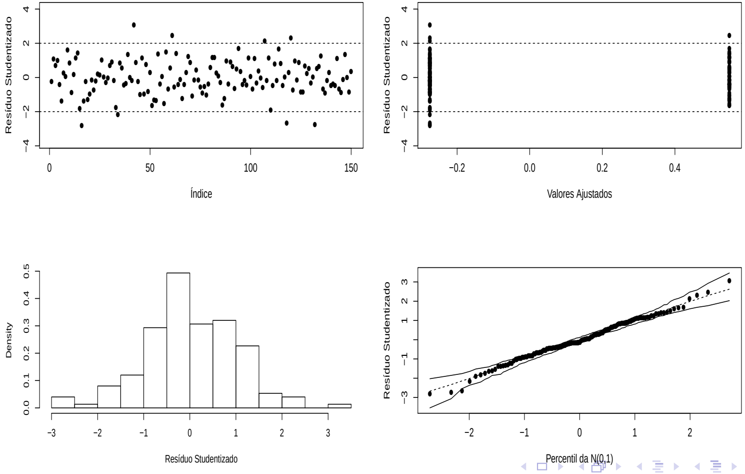


Componente 2

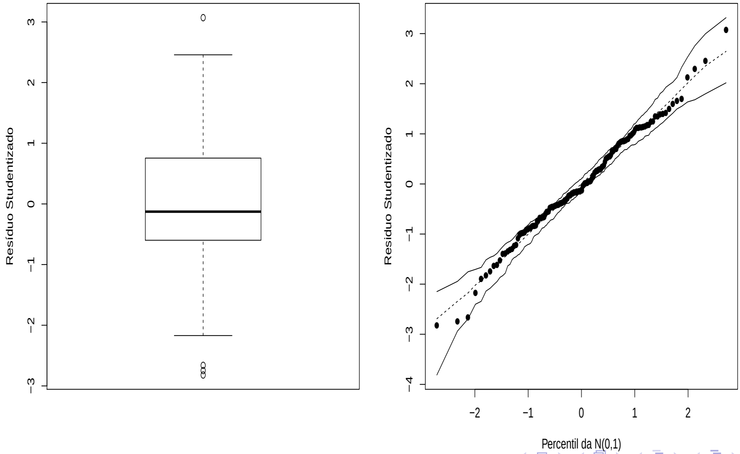
Parâmetro	Estimativa	EP	Estat.t	p-valor
μ_2	-0,2889	0,1247	-2,32	0,0219
α_{22}	0,8391	0,1764	4,76	$<$0,0001
α_{32}	0,0277	0,1764	0,16	0,8755

Há diferença entre o primeiro grupo e o segundo e uma equivalência entre aquele e o terceiro. Vamos ajustar um modelo reduzido ($\alpha_{32} = 0$).

Componente 2



Componente 2



Componente 2

- Estimativa dos parâmetros.

Parâmetro	Estimativa	EP	Estat.t	p-valor
μ_2	-0,2751	0,0879	-3,13	0,0021
α_{22}	0,8253	0,1523	5,42	$<$0,0001

- Médias preditas pelo modelo.

grupo	Estimativa	EP	IC(95%)
setosa/virginica	-0,28	0,09	[-0,45 ; -0,10]
versicolor	0,55	0,12	[0,30 ; 0,80]