

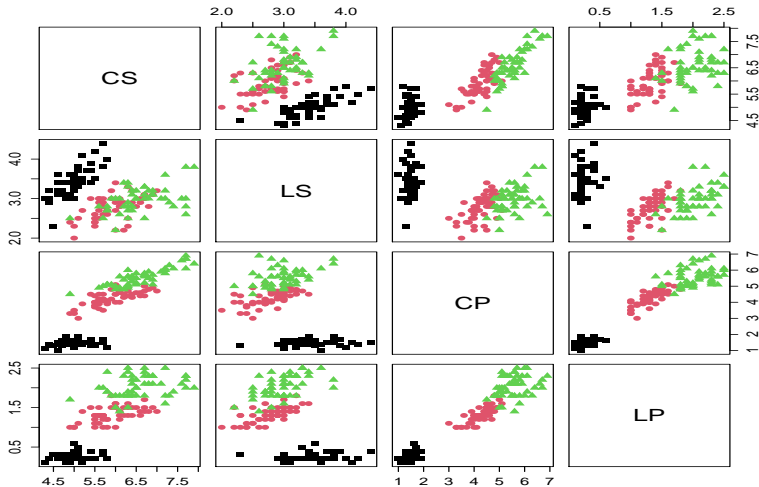
Análise de componentes principais: parte 2

Prof. Caio Azevedo

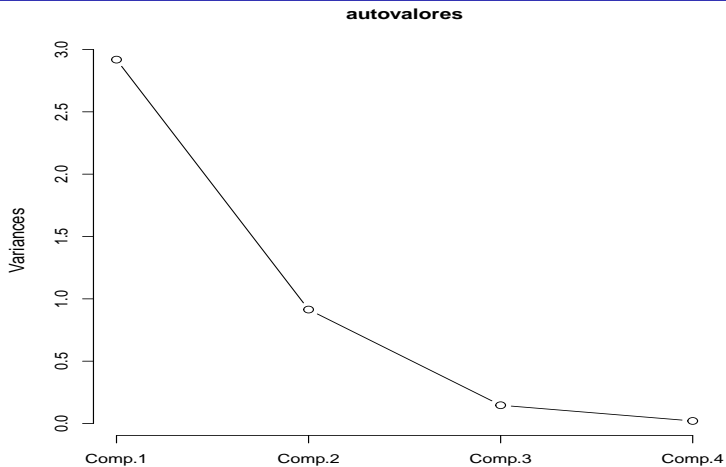
Exemplo 1: dados da íris “de Fisher”

- 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.
- Denotaremos por “grupo” os tipo de iris (setosa, versicolor e virginica).

Matriz de dispersão: ■ - S, ● - Ver, ▲ - Vir



Scree-plot



Resultados da ACP

- Variância de cada componente:

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

- 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

Resultados da ACP

■ Componentes principais

	Comp. 1 (Y_1)	Comp. 2 (Y_2)
CS (Z_1)	0,52 (0,89)	-0,37 (0,37)
LS (Z_2)	-0,27 (-0,46)	-0,92 (0,88)
CP (Z_3)	0,58 (0,99)	-0,02 (0,02)
LP (Z_4)	0,56 (0,96)	-0,07 (0,07)

- Primeira componente (CP1): constraste entre as variáveis CS, CP e LP e a variável LS.
- Segunda componente (CP2): é uma média ponderada entre as variáveis CS e LS.

Equações das componentes

- Para $i = 1, 2, \dots, 150$ (flores)

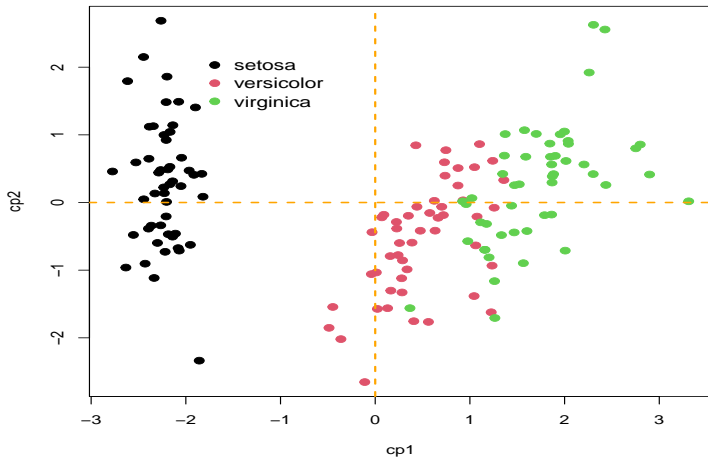
$$Y_{1i} = 0,52Z_{i1} - 0,27Z_{i2} + 0,58Z_{i3} + 0,56Z_{i4}$$

$$Y_{2i} = -0,37Z_{i1} - 0,92Z_{i2}$$

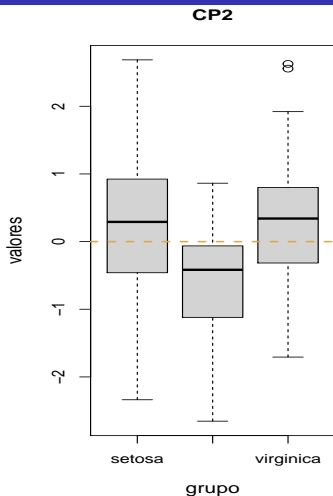
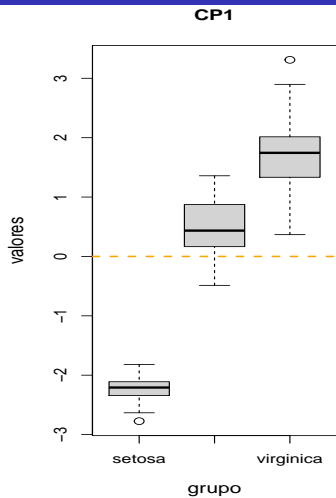
Interpretações

- Valores positivos para CP1 e CP2:
 - Valores acima da média (CS, CP, LP) e abaixo da média (LS).
- Valores negativos para CP1 e CP2:
 - Valores abaixo da média (CS, CP, LP) e acima da média (LS).
- Valores positivos para CP1 e negativos CP2:
 - Valores acima da média (CS, CP, LP,LS) .
- Valores negativos para CP1 e positivos para CP2:
 - Valores abaixo da média (CS, CP, LP, LS).

Gráfico anterior, com divisão de quadrantes

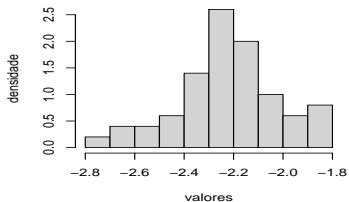


Box plots das duas CP's

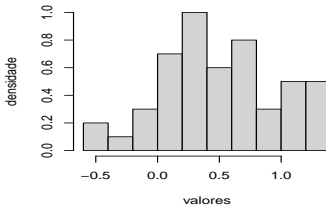


Histogramas da CP1

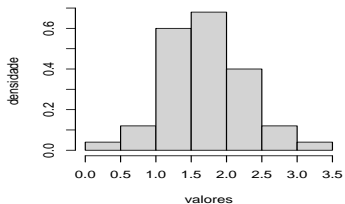
CP1 – setosa



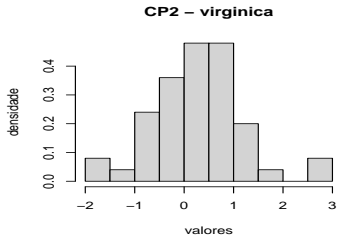
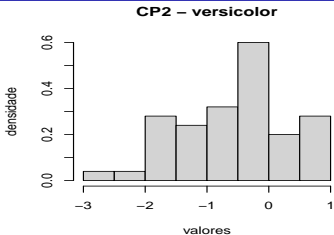
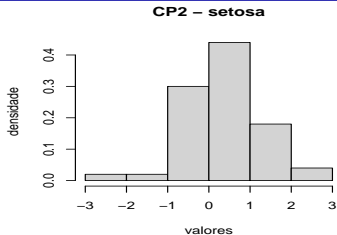
CP1 – versicolor



CP1 – virginica

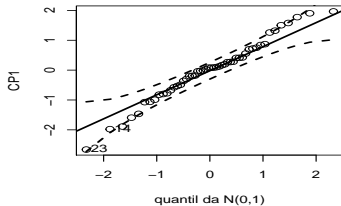


Histogramas da CP2

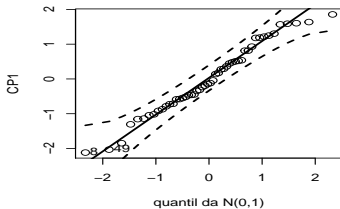


QQplot para CP1

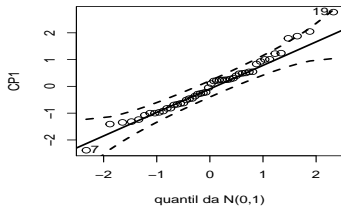
setosa, KS = 0.8394



versicolor, KS = 0.8713

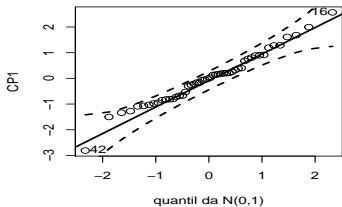


virginica, KS = 0.7891

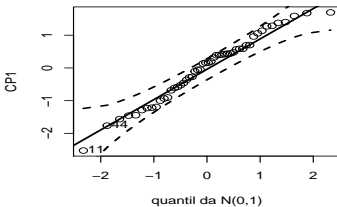


QQplot para CP2

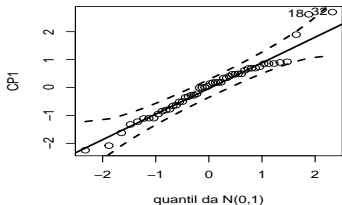
setosa, KS = 0.8326



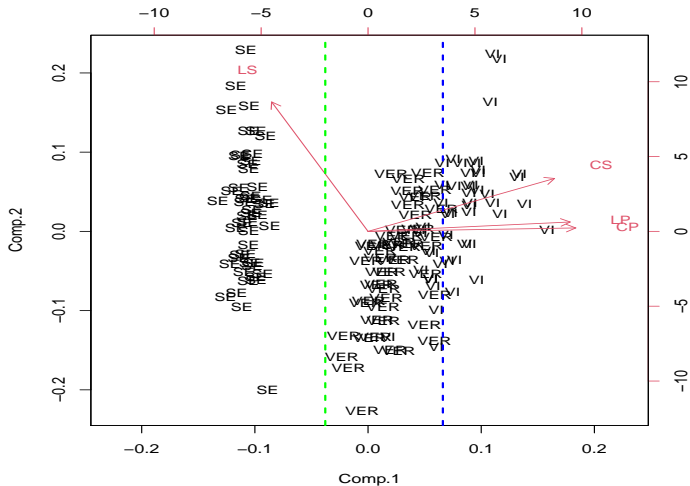
versicolor, KS = 0.7441



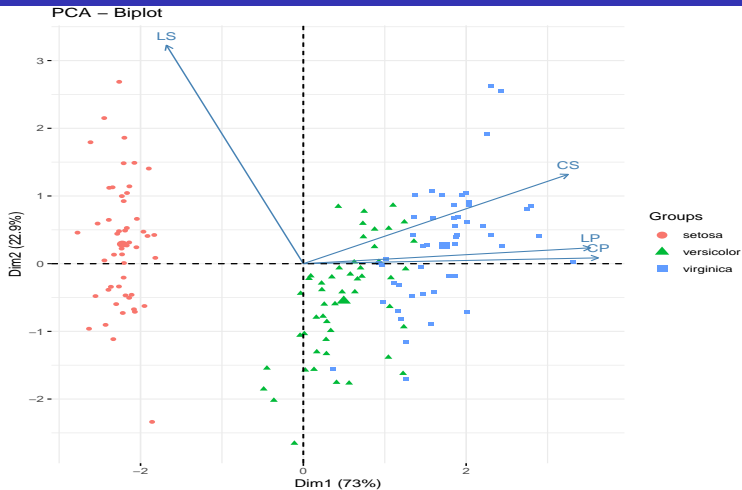
virginica, KS = 0.4855



Biplot: duas componentes principais



Biplot: duas componentes principais



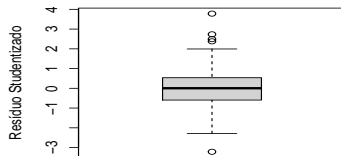
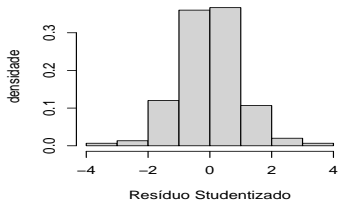
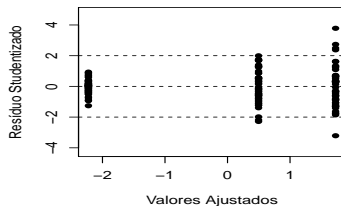
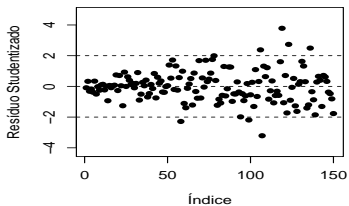
Comparação de grupos via **modelos lineares**

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

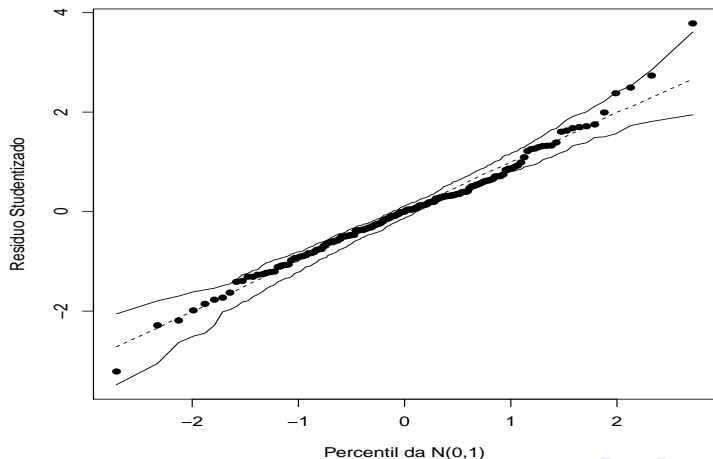
$i = 1, 2, 3$ (tipo de iris, setosa, versicolor, virginica), $j = 1, \dots, n_i$,
 $n_i = 50, \forall i, k = 1, 2$ (componente principal), $\xi_{ijk} \stackrel{i.i.d}{\sim} N(0, \sigma_k^2)$

- 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 [aqui](#)), para verificar a qualidade de ajuste do modelo.

Componente 1: gráficos de diagnóstico



Componente 1: QQ-plot com envelopes



Componente 1

Parâmetro	Estimativa	EP	IC(95%)	Estat.t	p-valor
μ_1	-2,22	0,06	[-2,35 ; -2,10]	-35,65	< 0,0001
α_{21}	2,72	0,09	[2,55 ; 2,89]	30,83	< 0,0001
α_{31}	3,95	0,09	[3,78 ; 4,13]	44,79	< 0,0001

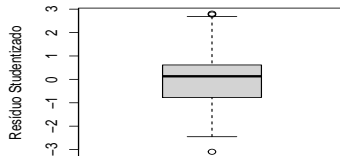
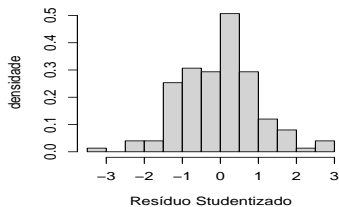
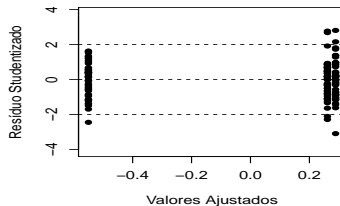
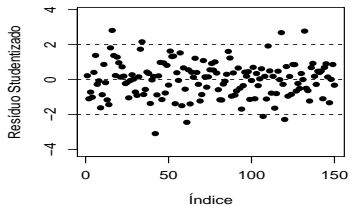
Há diferença entre o primeiro grupo e os outros dois. Vamos agora testar a igualdade entre as médias dos dois outros grupos através de testes do tipo $C\beta = M$ (veja mais [aqui](#)).

Componente 1

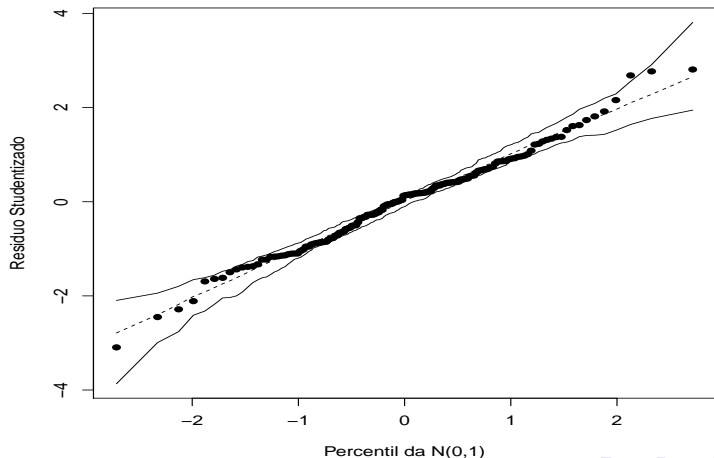
- Teste: $\alpha_{21} = \alpha_{31}$ vs $\alpha_{21} \neq \alpha_{31}$, $f_c = 194,83 (< 0,0001)$.
- Médias previstas pelo modelo.

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: gráficos de diagnóstico



Componente 2: QQ-plot com envelopes

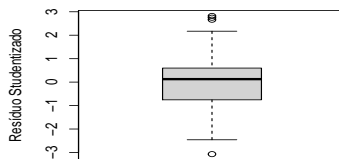
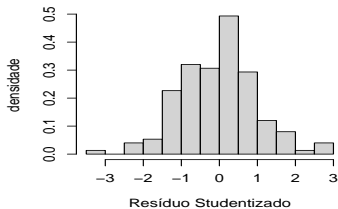
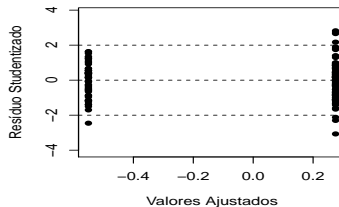
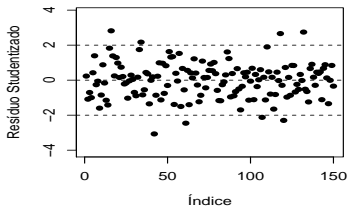


Componente 2

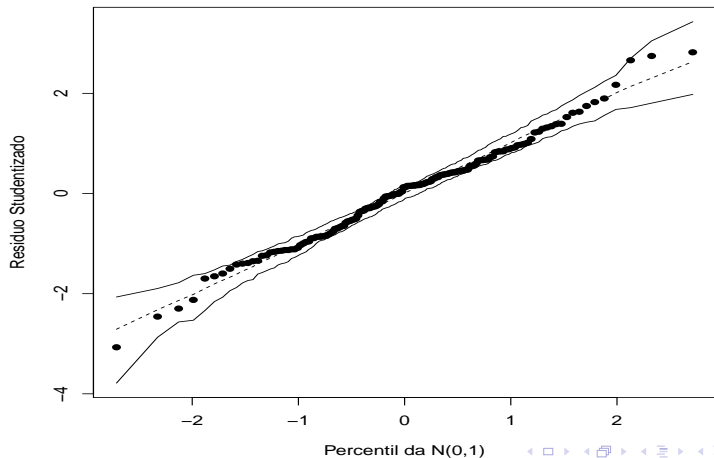
Parâmetro	Estimativa	EP	IC(95%)	Estat.t	p-valor
μ_2	0,29	0,12	[0,04 ; 0,53]	-2,32	0,0219
α_{22}	-0,84	0,18	[-1,18 ; -0,49]	4,76	<0,0001
α_{32}	-0,03	0,18	[-0,37 ; 0,32]	0,16	0,8755

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: gráficos de diagnóstico



Componente 2: QQ-plot com envelopes



Componente 2

- Estimativa dos parâmetros.

Parâmetro	Estimativa	EP	IC(95%)	Estat.t	p-valor
μ_2	0,28	0,09	[0,10 ; 0,45]	-3,13	0,0021
α_{22}	-0,83	0,15	[-1,12 ; -0,53]	5,42	<0,0001

- Médias previstas 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]