

UNICAMP - Universidade Estadual de Campinas
IMECC - Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística
ME731 - Métodos em Análise Multivariada

Trabalho

Carlos Caminho Oco RA 168624
marfloy Chaves eduardo RA 173733
Caio Henrique de Sousa Lima RA 214144
Piupiu Mendes Panela RA 219494
Nathan Blusa Amaarela Soltinho RA 222854

Campinas
2020

1 Questão 1

hist apenas no relatório, qq e boxplot apendice

1.1 Introdução

O banco de dados é composto pelo perfil de tamanho e formato de 48 tartarugas, sendo 24 machos e 24 fêmeas. As informações obtidas foram largura, comprimento e altura. O estudo tinha como objetivo observar e comparar o sexo ao qual o animal pertencia e sua relação com essas medidas físicas. Para realizar a análise desse banco, foi utilizado o software estatístico R e aplicado algumas metodologias como a Análise de Variância Multivariada (MANOVA).

1.2 Análise descritiva

Tabela 1: Medidas resumo (F = Fem, M = Masc)

| | Média | Var. | DP | CV(%) | Min. | Med. | Max. | CA | Cur. |
|------------|--------|--------|-------|-------|-------|--------|--------|-------|------|
| length (F) | 136.00 | 451.39 | 21.25 | 15.62 | 98.00 | 136.50 | 177.00 | -0.23 | 2.26 |
| width (F) | 102.58 | 171.73 | 13.10 | 12.77 | 81.00 | 102.00 | 132.00 | 0.31 | 2.55 |
| height (F) | 51.96 | 66.65 | 8.16 | 15.71 | 38.00 | 51.00 | 67.00 | -0.03 | 2.19 |
| length (M) | 113.38 | 138.77 | 11.78 | 10.39 | 93.00 | 115.00 | 135.00 | -0.08 | 2.10 |
| width (M) | 88.29 | 50.04 | 7.07 | 8.01 | 74.00 | 89.00 | 106.00 | 0.20 | 3.15 |
| height (M) | 40.71 | 11.26 | 3.36 | 8.24 | 35.00 | 40.00 | 47.00 | 0.18 | 2.20 |

Inicialmente, visualizando as medidas resumo para cada sexo, podemos observar que aparentemente o sexo feminino tem, em média, medidas maiores que o sexo masculino.

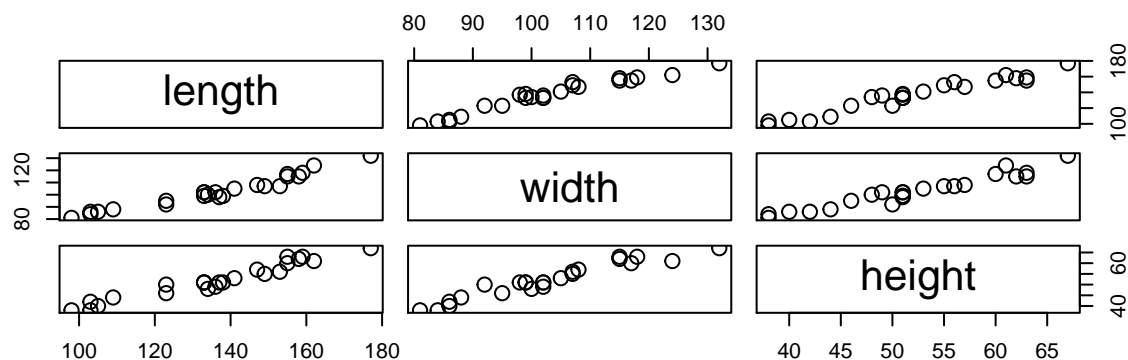


Figura 1: Matriz de gráfico de dispersões sexo feminino.

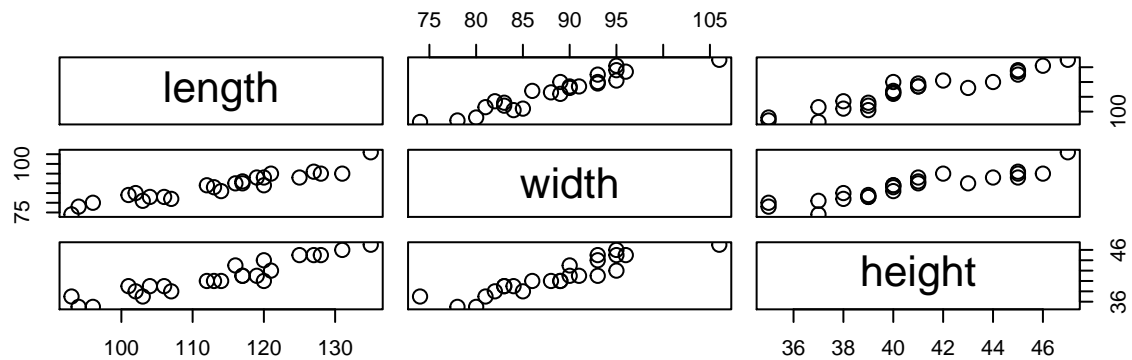


Figura 2: Matriz de gráfico de dispersões sexo masculino.

Para ambos os sexos, as medidas de altura, largura e comprimento são bastante correlacionadas.

Tabela 2: Matriz de covariâncias (F = Fem, M = Masc)

| | length (F) | width (F) | height (F) | length (M) | width (M) | height (M) |
|--------|------------|-----------|------------|------------|-----------|------------|
| length | 451.39 | 271.17 | 168.70 | 138.77 | 79.15 | 37.38 |
| width | 271.17 | 171.73 | 103.29 | 79.15 | 50.04 | 21.65 |
| height | 168.70 | 103.29 | 66.65 | 37.38 | 21.65 | 11.26 |

A matriz de covariância para ambos os sexos parecem ser diferentes, na qual as medidas do sexo masculino aparentam ter variância bem menor se comparado ao sexo feminino.

4) Matriz de correlações

Feminino

```
##           length      width      height
## length 1.0000000 0.9739706 0.9725816
## width  0.9739706 1.0000000 0.9654187
## height 0.9725816 0.9654187 1.0000000
```

Masculino

```
##           length      width      height
## length 1.0000000 0.9497846 0.9455580
## width  0.9497846 1.0000000 0.9122648
## height 0.9455580 0.9122648 1.0000000
```

Como dito anteriormente, as correlações parecem ser bem altas entre as medidas (já era esperado).

#####

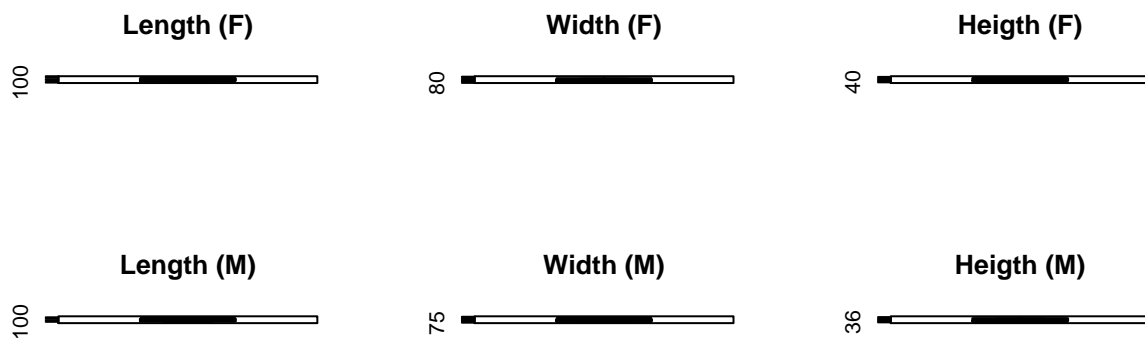


Figura 3: Boxplots das variáveis por sexo.

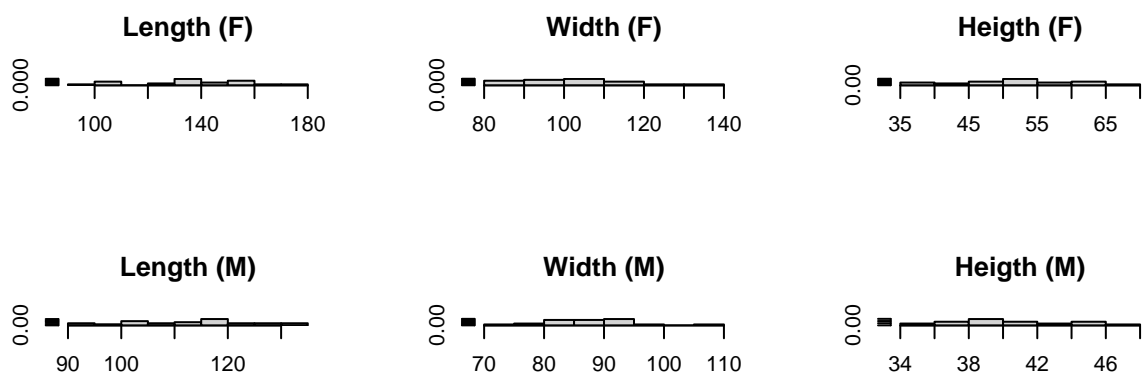


Figura 4: Histogramas das variáveis por sexo.

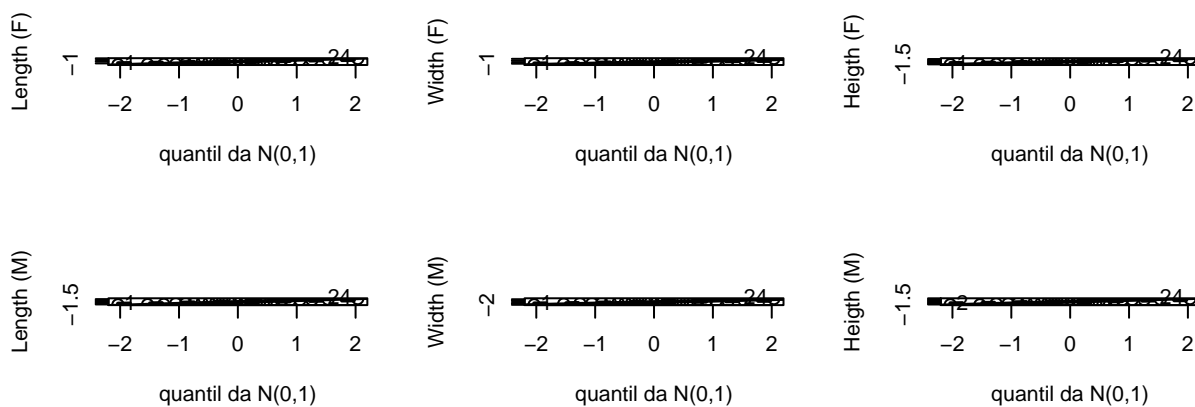


Figura 5: QQplots das variáveis por sexo.

Olhando para boxplot, histogramas e qqplots, as variáveis não parecem seguir uma distribuição nor-

mal.

1.3 Análise inferencial

Tabela 3: Teste de Box-Cox para homocedasticidade

| | Estatística do teste | p-valor |
|---------|----------------------|---------|
| Box-Cox | 24.04 | 0.0005 |

Tabela 4: Resultados da MANOVA

| | Valor | Aproximação pela distribuição F | p-valor |
|------------------|-------|---------------------------------|----------|
| Wilks | 0.41 | 21.28 | < 0.0001 |
| Pillai | 0.59 | 21.28 | < 0.0001 |
| Hotelling-Lawley | 1.45 | 21.28 | < 0.0001 |
| Roy | 1.45 | 21.28 | < 0.0001 |

Para observar inferencialmente a diferença entre as matrizes de covariância de cada grupo, foi realizado um teste de Box-cox, e tivemos como resultado que as matrizes realmente são diferentes. Após isso, foi realizado uma MANOVA com o objetivo de visualizar se existe diferença entre os vetores de médias para ambos os sexos, e também foi rejeitada a hipótese de que elas são iguais. Dessa forma, queremos ver em qual medidas essas diferenças são mais acentuadas, portanto aplicamos um teste CBU.

Tabela 5: Estimativas dos parâmetros do modelo

| | Estimativa | EP | Estatística t | p-valor |
|---------------|------------|------|---------------|----------|
| μ_1 | 136.00 | 3.51 | 38.79 | < 0.0001 |
| μ_2 | 102.58 | 2.15 | 47.72 | < 0.0001 |
| μ_3 | 51.96 | 1.27 | 40.78 | < 0.0001 |
| α_{21} | -22.63 | 4.96 | -4.56 | < 0.0001 |
| α_{22} | -14.29 | 3.04 | -4.70 | < 0.0001 |
| α_{23} | -11.25 | 1.80 | -6.24 | < 0.0001 |

Observando o resultado dos teste CBU, podemos observar que as diferenças estão em todas as medidas (largura, comprimento e altura).

Aqui eu não sei o que falar

OBS: DEPOIS TEM QUE VER A LISTA FEITA PRA VER SE FALTA ALGO

2 Questão 2:

Questão 6 Lista 4 ACP

Dados

Tabela 6: Testes CBU ($\alpha_{2i} = 0$)

| Parâmetro | Estatística Qui-quadrado | p-valor |
|------------------------|--------------------------|----------|
| α_{21} (Length) | 20.82 | < 0.0001 |
| α_{22} (Width) | 22.10 | < 0.0001 |
| α_{23} (Height) | 38.99 | < 0.0001 |

removido score pelo fato de ser combinação das demais variáveis (resposta) se fico em primeiro nos 200m ganho 500 pontos

2.1 Análise descritiva

Tabela 7: Medidas resumo dos resultados

| | Média | Var. | DP | CV(%) | Min. | Med. | Max. | CA | Cur. |
|----------|--------|-------|------|-------|--------|--------|--------|-------|------|
| hurdles | 13.84 | 0.54 | 0.74 | 5.32 | 12.69 | 13.75 | 16.42 | 1.65 | 7.22 |
| highjump | 1.78 | 0.01 | 0.08 | 4.37 | 1.50 | 1.80 | 1.86 | -1.99 | 7.87 |
| shot | 13.12 | 2.23 | 1.49 | 11.37 | 10.00 | 12.88 | 16.23 | 0.18 | 2.78 |
| run200m | 24.65 | 0.94 | 0.97 | 3.93 | 22.56 | 24.83 | 26.61 | -0.17 | 2.62 |
| longjump | 6.15 | 0.22 | 0.47 | 7.71 | 4.88 | 6.25 | 7.27 | -0.48 | 4.28 |
| javelin | 41.48 | 12.57 | 3.55 | 8.55 | 35.68 | 40.28 | 47.50 | 0.16 | 1.89 |
| run800m | 136.05 | 68.74 | 8.29 | 6.09 | 124.20 | 134.74 | 163.43 | 1.40 | 5.89 |

2) Diagrama de dispersão

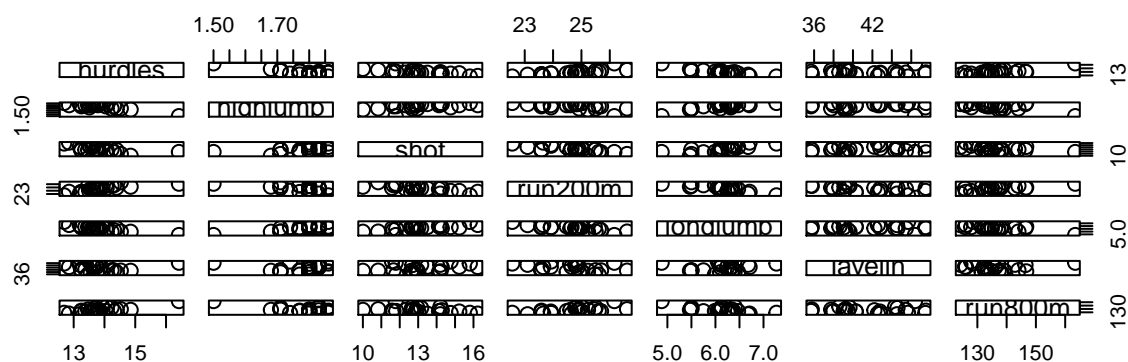


Figura 6: Matriz de gráfico de dispersões.

5) Boxplot

Tabela 8: Matriz de covariâncias

| | hurdles | highjump | shot | run200m | longjump | javelin | run800m |
|----------|---------|----------|-------|---------|----------|---------|---------|
| hurdles | 0.54 | -0.05 | -0.72 | 0.55 | -0.32 | -0.02 | 4.76 |
| highjump | -0.05 | 0.01 | 0.05 | -0.04 | 0.03 | 0.00 | -0.38 |
| shot | -0.72 | 0.05 | 2.23 | -0.99 | 0.53 | 1.42 | -5.19 |
| run200m | 0.55 | -0.04 | -0.99 | 0.94 | -0.38 | -1.14 | 4.96 |
| longjump | -0.32 | 0.03 | 0.53 | -0.38 | 0.22 | 0.11 | -2.75 |
| javelin | -0.02 | 0.00 | 1.42 | -1.14 | 0.11 | 12.57 | 0.59 |
| run800m | 4.76 | -0.38 | -5.19 | 4.96 | -2.75 | 0.59 | 68.74 |

Tabela 9: Matriz de correlações

| | hurdles | highjump | shot | run200m | longjump | javelin | run800m |
|----------|---------|----------|-------|---------|----------|---------|---------|
| hurdles | 1.00 | -0.81 | -0.65 | 0.77 | -0.91 | -0.01 | 0.78 |
| highjump | -0.81 | 1.00 | 0.44 | -0.49 | 0.78 | 0.00 | -0.59 |
| shot | -0.65 | 0.44 | 1.00 | -0.68 | 0.74 | 0.27 | -0.42 |
| run200m | 0.77 | -0.49 | -0.68 | 1.00 | -0.82 | -0.33 | 0.62 |
| longjump | -0.91 | 0.78 | 0.74 | -0.82 | 1.00 | 0.07 | -0.70 |
| javelin | -0.01 | 0.00 | 0.27 | -0.33 | 0.07 | 1.00 | 0.02 |
| run800m | 0.78 | -0.59 | -0.42 | 0.62 | -0.70 | 0.02 | 1.00 |

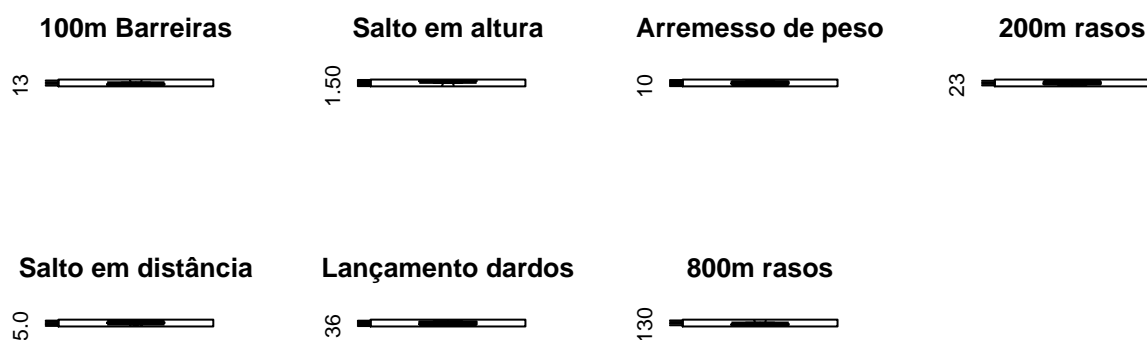


Figura 7: Boxplots das variáveis.

#histograms

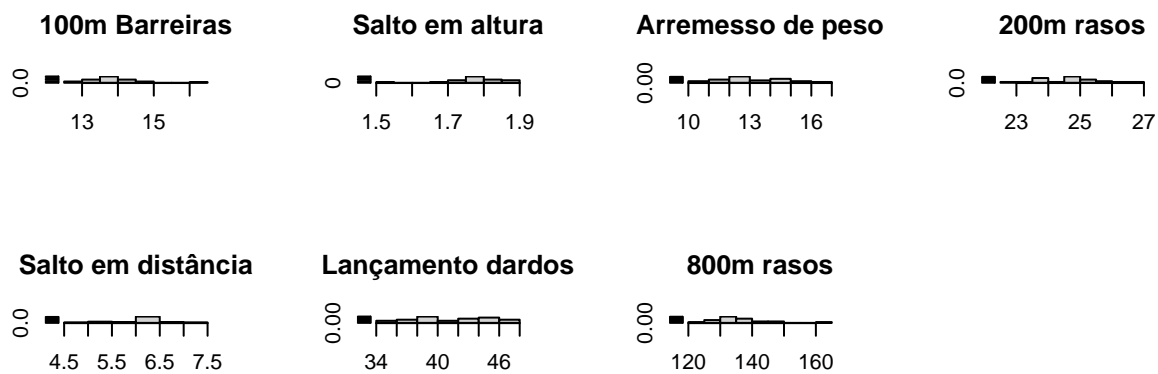


Figura 8: Histogramas das variáveis.

3 qqplots

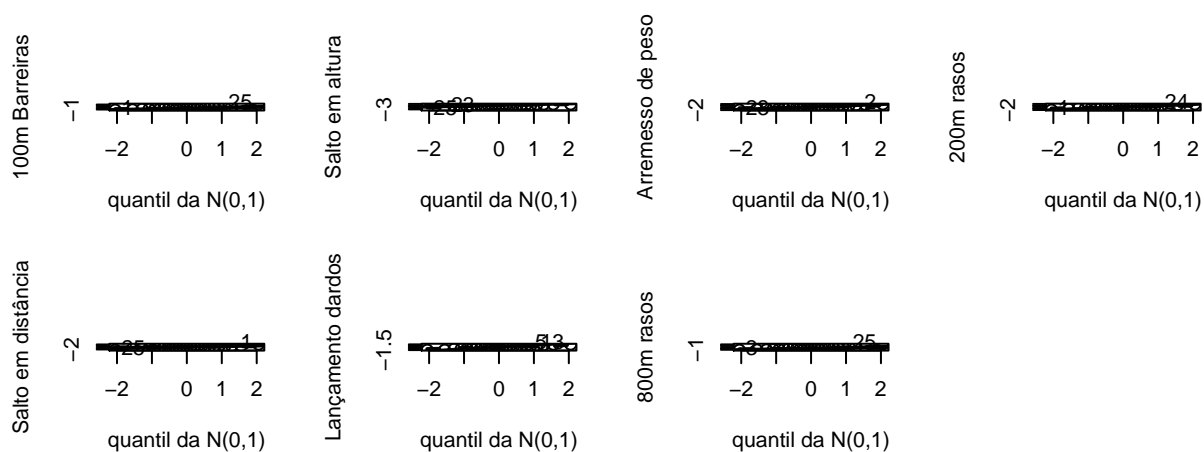


Figura 9: QQplots das variáveis.

Análise inferencial

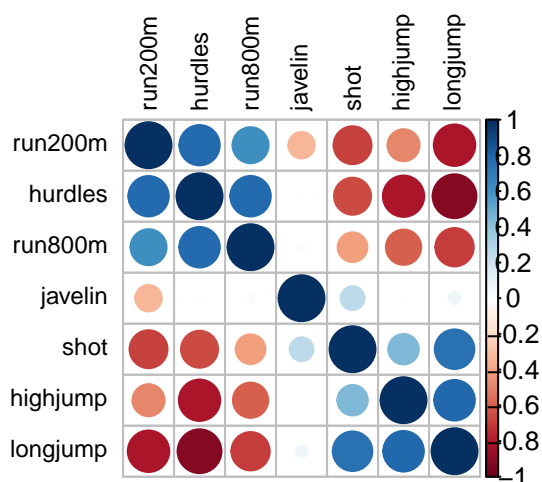


Figura 10: Correlograma das variáveis.

3.1 ACP

0 - Descritiva 1 - Suposições, dois grupos (variáveis tempo e distancia) 2 - ACP 2 comp 80% 3 - CP1 para o rank devido ao fato de fazer sentido com a direção das variáveis. 4 - Abordar novas classificações bem diferentes (12º->6º)

autovalores

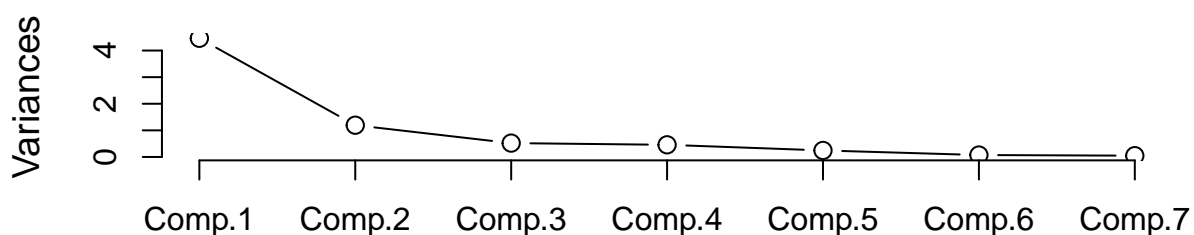


Tabela 10: Variância Explicada

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
|----------|--------|--------|--------|--------|--------|--------|--------|
| PVE (%) | 63.72 | 17.06 | 7.44 | 6.53 | 3.50 | 1.04 | 0.70 |
| PVEA (%) | 63.72 | 80.78 | 88.22 | 94.75 | 98.26 | 99.30 | 100.00 |

#Classificação #Score da primeira componente principal. acho q isso ta certo

Tabela 11: Componentes principais

| | Comp.1 | Comp.2 |
|----------|--------|--------|
| hurdles | 0.45 | 0.16 |
| highjump | -0.38 | -0.25 |
| shot | -0.36 | 0.29 |
| run200m | 0.41 | -0.26 |
| longjump | -0.46 | -0.06 |
| javelin | -0.08 | 0.84 |
| run800m | 0.37 | 0.22 |

Tabela 12: Rankings e escores dos atletas

| Atleta | Rank | Escore PCA | Novo rank | Atleta | Rank | Escore PCA | Novo rank |
|---------------------|------|---------------|--------------|--------------------|------|---------------|--------------|
| Joyner-Kersee (USA) | 1 | 4.12 | 1 | Braun (FRG) | 14 | -0.00 | 14 |
| John (GDR) | 2 | 2.88 | 2 | Ruotsalainen (FIN) | 13 | -0.02 | 15 |
| Behmer (GDR) | 3 | 2.65 | 3 | Yuping (CHN) | 15 | -0.09 | 16 |
| Sablovskaitė (URS) | 5 | 1.36 | 4 | Hagger (GB) | 17 | -0.17 | 17 |
| Choubenkova (URS) | 4 | 1.34 | 5 | Brown (USA) | 18 | -0.52 | 18 |
| Schulz (GDR) | 12 | 1.19 | 6 | Mulliner (GB) | 20 | -1.09 | 19 |
| Fleming (AUS) | 7 | 1.10 | 7 | Hautenauve (BEL) | 19 | -1.13 | 20 |
| Greiner (USA) | 6 | 1.04 | 8 | Kytola (FIN) | 21 | -1.45 | 21 |
| Lajbnerova (CZE) | 8 | 0.92 | 9 | Geremias (BRA) | 22 | -2.01 | 22 |
| Bouraga (URS) | 10 | 0.76 | 10 | Hui-Ing (TAI) | 23 | -2.88 | 23 |
| Wijnsma (HOL) | 11 | 0.56 | 11 | Jeong-Mi (KOR) | 24 | -2.97 | 24 |
| Dimitrova (BUL) | 9 | 0.53 | 12 | Launa (PNG) | 25 | -6.27 | 25 |
| Scheider (SWI) | 16 | 0.14 | 13 | | | | |