



UNIVERSIDADE FEDERAL DO PARÁ
FACULDADE DE ESTATÍSTICA
Disciplina: Análise Multivariada II
Profa.: Marinalva Cardoso Maciel

Lista 3 – Manova a dois fatores e ACP

1. A Tabela a seguir apresenta dados sobre feijão-vagem mostrando os resultados de quatro variáveis: y_1 = precocidade do rendimento, y_2 = precocidade da área foliar específica (SLA), y_3 = rendimento total e y_4 = SLA médio. Os fatores são data de semeadura (S) e variedade (V).

S	V	Rep	y_1	y_2	y_3	y_4	S	V	Rep	y_1	y_2	y_3	y_4
1	1	1	59.3	4.5	38.4	295	3	1	1	68.1	3.4	42.2	280
1	1	2	60.3	4.5	38.6	302	3	1	2	68.0	2.9	42.4	284
1	1	3	60.9	5.3	37.2	318	3	1	3	68.5	3.3	41.5	286
1	1	4	60.6	5.8	38.1	345	3	1	4	68.6	3.1	41.9	284
1	1	5	60.4	6.0	38.8	325	3	1	5	68.6	3.3	42.1	268
1	2	1	59.3	6.7	37.9	275	3	2	1	64.0	3.6	40.9	233
1	2	2	59.4	4.8	36.6	290	3	2	2	63.4	3.9	41.4	248
1	2	3	60.0	5.1	38.7	295	3	2	3	63.5	3.7	41.6	244
1	2	4	58.9	5.8	37.5	296	3	2	4	63.4	3.7	41.4	266
1	2	5	59.5	4.8	37.0	330	3	2	5	63.5	4.1	41.1	244
1	3	1	59.4	5.1	38.7	299	3	3	1	68.0	3.7	42.3	293
1	3	2	60.2	5.3	37.0	315	3	3	2	68.7	3.5	41.6	284
1	3	3	60.7	6.4	37.4	304	3	3	3	68.7	3.8	40.7	277
1	3	4	60.5	7.1	37.0	302	3	3	4	68.4	3.5	42.0	299
1	3	5	60.1	7.8	36.9	308	3	3	5	68.6	3.4	42.4	285
2	1	1	63.7	5.4	39.5	271	4	1	1	69.8	1.4	48.4	265
2	1	2	64.1	5.4	39.2	284	4	1	2	69.5	1.3	47.8	247
2	1	3	63.4	5.4	39.0	281	4	1	3	69.5	1.3	46.9	231
2	1	4	63.2	5.3	39.0	291	4	1	4	69.9	1.3	47.5	268
2	1	5	63.2	5.0	39.0	270	4	1	5	70.3	1.1	47.1	247
2	2	1	60.6	6.8	38.1	248	4	2	1	66.6	1.8	45.7	205
2	2	2	61.0	6.5	38.6	264	4	2	2	66.5	1.7	46.8	239
2	2	3	60.7	6.8	38.8	257	4	2	3	67.1	1.7	46.3	230
2	2	4	60.6	7.1	38.6	260	4	2	4	65.8	1.8	46.3	235
2	2	5	60.3	6.0	38.5	261	4	2	5	65.6	1.9	46.1	220
2	3	1	63.8	5.7	40.5	282	4	3	1	70.1	1.7	48.1	253
2	3	2	63.2	6.1	40.2	284	4	3	2	72.3	0.7	47.8	249
2	3	3	63.3	6.0	40.0	291	4	3	3	69.7	1.5	46.7	226
2	3	4	63.2	5.9	40.0	299	4	3	4	69.9	1.3	47.1	248
2	3	5	63.1	5.4	39.7	295	4	3	5	69.8	1.4	46.7	236

- Faça a análise exploratória dos dados;
 - Organize o banco de dados
 - Teste a interação e efeitos principais usando as quatro estatísticas de teste da MANOVA a dois fatores.
 - Faça a análise descrevendo a metodologia, procedimentos e conclusão.
2. O pentatlo para mulheres foi realizado pela primeira vez na Alemanha, em 1928. Inicialmente a modalidade consistia no arremesso de peso, salto em distância, corrida de 100m, salto em altura e eventos de lançamento de dardo realizados durante dois dias. O pentatlo foi introduzido pela



UNIVERSIDADE FEDERAL DO PARÁ
FACULDADE DE ESTATÍSTICA
Disciplina: Análise Multivariada II
Profa.: Marinalva Cardoso Maciel

primeira vez em Jogos Olímpicos em 1964, consistindo dos 80 m com barreiras, tiros, salto em altura, salto em comprimento e corrida de 200 m. Em 1977, a corrida de 200 m foi substituída pela de 800 m e em 1981 a IAAF trouxe o heptatlo no lugar do pentatlo, com um dia que contém os eventos- corrida de 100 m com barreiras, tiro, salto em altura, corrida de 200 m e no segundo dia, o salto em comprimento, lançamento de dardo e corrida de 800 m. Um sistema de pontuação é utilizado para atribuir pontos aos resultados de cada evento, e o vencedor é a mulher que acumula mais pontos durante os dois dias. O evento fez sua primeira aparição olímpica em 1984.

Nos Jogos Olímpicos de 1988, em Seul, o heptatlo foi vencido por uma das estrelas do atletismo feminino, nos EUA, Jackie Joyner-Kersey. O pacote "HSAUR" contém os dados de 25 competidoras do heptatlo com 8 variáveis (heptathlon).

Variáveis:

hurdles: resultados de 100 m com barreiras

highjump: resultados de salto em altura

shot: resultados de arremesso de peso

run200m: resultados de 200 m rasos

longjump: resultados de salto em distância

javelin: resultados de lançamento de dardos

run800m: resultados de 800 m rasos

score: pontuação total

Realize a Análise de Componentes Principais visando a exploração da estrutura dos dados e avaliar como os escores das componentes principais se relacionam com os escores do sistema. Faça comentários sobre a matriz de correlação. Você observa alguma estrutura? Grupos se formando? Variáveis isoladas? Realize análise gráfica para uma melhor visualização.

Obs.

- Algumas variáveis são medidas em segundos (quanto menor o número, melhor a performance) e outras são medidas em metros (quanto maior o número, melhor a performance). Então, para que os escores apontem na mesma direção é necessário transformar os tempos para que representem a diferença em relação ao tempo da corredora mais lenta (que é o maior valor).

Ex: $\text{hurdles2} = \max(\text{heptathlon\$hurdles}) - \text{heptathlon\$hurdles}$

Desse jeito os maiores números são melhores em todas as variáveis.

- Não esqueça de padronizar os dados.

3. Para os dados de progênies de *Eucalyptus* sp., em que foram avaliadas 10 características (X1, X2, X3, X4, X5, X6, X7, X8, X9 e X10) em 10 progênies apresentados na apostila da Unidade 3, refaça a análise da apostila (até a modelagem), incluindo gráficos e comparando os seus resultados aos apresentados na apostila como conclusão.