

Lista 10 – Componentes Principais – Estatística 2025

1. Responda as questões abaixo:

- a) A transformação por componentes principais baseia-se no cálculo de autovalores e autovetores. O que representam os autovalores e os autovetores? Qual a relação dos autovalores e as variâncias originais das variáveis?
- b) Em muitos casos, as componentes principais são obtidas a partir da matriz de correlação. Em que casos isso é recomendado?
- c) Por que se diz que a informação é menor nas últimas componentes principais? Nesse caso, sempre podemos descartar estas últimas componentes por conterem apenas ruídos?

2. A partir dos dados abaixo, calcule a matriz de variância/covariância e a matriz de correlação. Obtenha os autovalores e autovetores considerando apenas a matriz de variância/covariância. Responda:

- a) Qual a variável que apresenta a maior variância? E a menor?
- b) Qual o par de variáveis apresenta a maior correlação (em módulo)? E a menor?
- c) Calcule a matriz de variância-covariância das componentes principais. Por que nesse caso as covariâncias são nulas?
- d) Por que a variável X_2 só está representada efetivamente na última componente principal, aquela com menor autovalor?
- e) Represente, através de um gráfico de dispersão, a primeira e a segunda componentes principais. Qual a proporção da variância total que está representada nestas duas primeiras componentes?

X_1	X_2	X_3	X_4	X_5
34,78	-0,94	41,77	89,34	-63,17
16,56	-0,46	17,54	58,15	-47,84
82,34	-1,82	77,05	118,70	-75,06
30,47	-0,83	33,96	80,42	-38,09
4,48	-0,43	29,82	71,39	-33,48
16,49	-0,72	62,61	105,82	-51,25
69,25	-1,70	66,08	108,69	-67,18
69,68	-1,50	56,42	104,96	-58,34
70,60	-1,69	56,00	105,97	-72,49
42,82	-1,08	71,08	112,19	-51,73
65,49	-1,53	78,45	120,63	-68,31
16,47	-0,63	34,79	79,42	-39,05
54,40	-1,25	33,34	75,61	-41,60
98,68	-2,22	90,11	137,04	-75,66
58,58	-1,38	69,72	117,56	-56,78

3. A fim de retirar a forte relação entre duas variáveis X_1 e X_2 , uma pessoa pensou em aplicar esses dados a uma análise de componentes principais. No entanto, ela observou que, quando plotou o diagrama de dispersão entre as componentes principais resultantes, aparentemente a relação entre elas ainda existia. Ela esperava que a distribuição dos pontos fosse totalmente casual, indicando a independência entre as componentes. Qual a explicação para isso? Você imagina uma maneira de conseguir de fato produzir componentes que sejam independentes entre si? Em seguida são apresentados os dados utilizados e os diagramas de dispersão das variáveis originais e das componentes principais resultantes.

X_1	X_2
6,62	3,72
0,41	0,18
4,76	1,54
7,51	5,92
7,89	6,96
3,58	0,77
7,22	4,9
6,45	3,03
7,07	4,8
6,01	2,37
5,59	2,07
7,03	4,16
7,04	4,48
7,01	4,55
6,09	2,48

