

SME0822 - Análise Multivariada e Aprendizado Não Supervisionado (2023)

🏠 Início / Meus Ambientes / 2023 / ICMC / SME / SME0822-201-2023 / Questionários / Questionário Q4 - até 03/10/2023

Iniciado em	quarta, 4 out 2023, 19:19
Estado	Finalizada
Concluída em	quarta, 4 out 2023, 20:52
Tempo empregado	1 hora 32 minutos
Avaliar	7,33 de um máximo de 10,00(73,33%)

Questão 1

Correto

Atingiu 1,00 de 1,00

Marcar questão

Suponha um banco de dados com 10 variáveis quantitativas disponíveis, em que se deseja aplicar a análise de componentes principais. Para saber quantas componentes principais serão utilizadas, foi considerado o método de proporção da variância total explicada. Sabe-se que os autovalores foram ordenados de forma decrescente, que a proporção mínima da variância total que deve ser explicada pelas componentes está fixada em 95%, e que as 3 últimas componentes explicam 15% da variância total. Dadas as informações disponíveis, quantas componentes principais serão usadas, no mínimo, para esta análise de componentes principais?

Escolha uma opção:

- ☒ a. 8✔
- ☐ b. 9
- ☐ c. 10
- ☐ d. 7
- ☐ e. 6

A resposta correta é: 8

Questão 2

Correto

Atingiu 1,00 de 1,00

Marcar questão

Sobre a Análise de Componentes Principais (ACP ou PCA), assinale todas as alternativas corretas.

Escolha uma ou mais:

- ☐ a. A ACP transforma linearmente p variáveis correlacionadas em k variáveis correlacionadas, com $k < p$.
- ☐ b. O vetor de componentes principais tem distribuição normal multivariada, não importa qual a distribuição do vetor aleatório original de dados.
- ☒ c. Se o vetor aleatório original possui distribuição normal multivariada, o vetor de componentes principais também terá distribuição normal multivariada e e sua matriz variâncias e covariâncias será diagonal.✔
- ☒ d. As componentes principais podem ser obtidas a partir da matriz de variâncias e covariâncias ou de correlações populacionais do vetor original de dados \underline{X} , mas em geral são estimadas a partir da matriz de variâncias e covariâncias ou de correlações amostrais observadas.✔
- ☒ e. A ACP transforma linearmente variáveis correlacionadas em variáveis não-correlacionadas.✔
- ☐ f. A ACP sempre reduz a dimensionalidade do problema.

As respostas corretas são: A ACP transforma linearmente variáveis correlacionadas em variáveis não-correlacionadas,. Se o vetor aleatório original possui distribuição normal multivariada, o vetor de componentes principais também terá distribuição normal multivariada e e sua matriz variâncias e covariâncias será diagonal,. As componentes principais podem ser obtidas a partir da matriz de variâncias e covariâncias ou de correlações populacionais do vetor original de dados \underline{X} , mas em geral são estimadas a partir da matriz de variâncias e covariâncias ou de correlações amostrais observadas.

Questão 3

Correto

Atingiu 2,00 de 2,00

Marcar questão

Sobre a análise de componentes principais, assinale todos os itens corretos.

Escolha uma ou mais:

- ☐ a. Se as variáveis originais tiverem distribuição normal multivariada e forem correlacionadas, as componentes principais também terão distribuição normal multivariada e serão correlacionadas.
- ☒ b. As componentes principais, quando são extraídas via matriz de covariâncias, podem sofrer influência das diferentes magnitudes das variâncias de diferentes variáveis. Uma alternativa é utilizar a matriz de correlações para extrair as componentes principais.✔
- ☐ c. A análise de componentes principais pode reduzir fortemente a dimensionalidade dos dados quando as variáveis são independentes.
- ☒ d. Nem sempre há perda de informação quando se aplica o método das componentes principais, mas a interpretação das variáveis pode não ser tão direta quanto na análise com todas as variáveis originais.✔
- ☒ e. Os escores das componentes principais podem ser usados para ordenação dos dados, mas também como entradas para outras técnicas, como modelos de regressão.✔

As respostas corretas são: As componentes principais, quando são extraídas via matriz de covariâncias, podem sofrer influência das diferentes magnitudes das variâncias de diferentes variáveis. Uma alternativa é utilizar a matriz de correlações para extrair as componentes principais,. Os escores das componentes principais podem ser usados para ordenação dos dados, mas também como entradas para outras técnicas, como modelos de regressão,. Nem sempre há perda de informação quando se aplica o método das componentes principais, mas a interpretação das variáveis pode não ser tão direta quanto na análise com todas as variáveis originais.

Questão 4

Correto

Atingiu 1,00 de 1,00

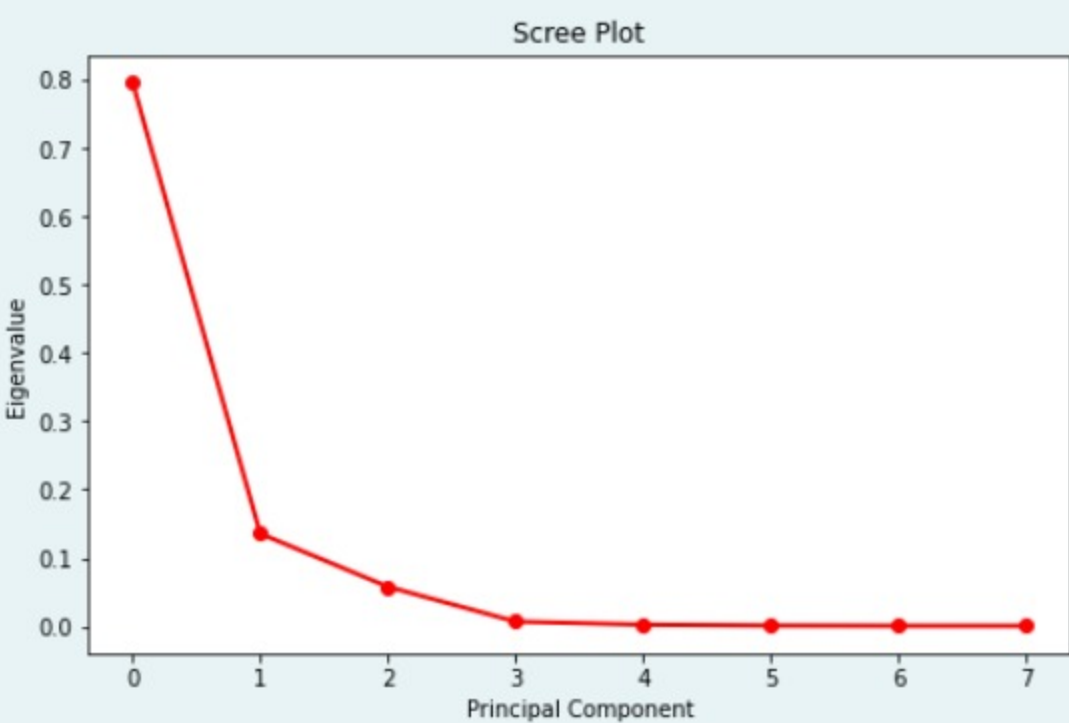
Marcar questão

Uma base de dados de decatlo foi observada com 41 linhas e 13 colunas, de forma que as dez primeiras colunas correspondem ao desempenho dos atletas para as 10 provas do decatlo. As colunas 11 e 12 correspondem respectivamente à classificação e aos pontos obtidos. A última coluna é uma variável categórica correspondente ao evento esportivo (Jogos Olímpicos de 2004 ou Decastar de 2004)

As variáveis são

- 100m (100 metros rasos),
- long.jump (salto em distância),
- shot.put (arremesso de peso),
- High.jump (salto em altura),
- 400m (400 metros rasos),
- 110m.hurdle (110 metros com barreiras),
- Discus (lançamento de disco),
- Pole.vault (salto com vara),
- Javeline (lançamento de dardo),
- 1500m (1500 metros).

Um analista desenvolveu uma análise de componentes principais e obteve o gráfico scree-plot a seguir



Complete cada sentença a seguir com a resposta correta:

Se ele quiser um número de componentes principais que expliquem ao menos 75% da variabilidade dos dados, ele deve escolher ✔ componente(s) principal(is).

Se ele quiser um número de componentes principais que expliquem ao menos 90% da variabilidade dos dados, ele deve escolher ✔ componente(s) principal(is).

Pelo critério do scree-plot, entretanto, ele poderá escolher até ✔ componentes principais.

Para obter o gráfico scree-plot acima, o analista considerou a transformação de 10 variáveis originais em

- ☐ 2
- ☐ 4
- ☐ 6
- ☒ 8✔
- ☐ 10
- componentes principais.

Questão 5

Correto

Atingiu 2,00 de 2,00

Marcar questão

Uma análise de componentes principais foi desenvolvida a partir de uma matriz de variâncias e covariâncias amostrais, a partir de uma amostra de tamanho n=50, da qual foi obtido o maior autovalor $\hat{\lambda}_1 = 3$. Selecione o intervalo de confiança com nível de confiança 95% mais adequado para λ_1 .

Escolha uma opção:

- ☐ (2.35, 4.15)
- ☐ (2.43, 3.91)
- ☐ (1.62, 2.61)
- ☐ (1.57, 2.77)
- ☒ (2.16, 4.93)✔
- ☐ (1.8, 4.11)
- ☐ (2.51, 5.76)

A resposta correta é: (2.16, 4.93)

Questão 6

Incorreto

Atingiu 0,00 de 1,00

Marcar questão

Dada a matriz de covariâncias amostrais igual a $\begin{pmatrix} 8 & 7 \\ 7 & 7 \end{pmatrix}$.

Qual o percentual aproximado da variância total que é explicado por cada componente principal estimada?

Escolha uma opção:

- ☐ 95% e 5%
- ☐ 70% e 30%
- ☐ 90% e 10%
- ☐ 94% e 6%
- ☐ 97% e 3%
- ☒ 99% e 1%✖
- ☐ 69% e 31%

A resposta correta é: 97% e 3%

Questão 7

Parcialmente correto

Atingiu 0,33 de 2,00

Marcar questão

Seja \underline{X} e um vetor aleatório 3×1 com $E(\underline{X}) = \underline{\mu}$ e $\text{Var}(\underline{X}) = \Sigma$, em que $\Sigma = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Uma análise de componentes principais foi feita a partir da decomposição em autovalores e autovetores de Σ , como discutido em aula. Assinale as alternativas corretas.

Escolha uma ou mais:

- ☒ As duas primeiras componentes principais explicam mais de 90% da variabilidade dos dados.✔
- ☐ As duas primeiras componentes principais são dadas por $Y_1 = 0.89X_1 + 0.45X_2$
 $Y_2 = X_3$
- ☐ As duas primeiras componentes principais são dadas por $Y_1 = -0.86X_1 - 0.36X_2 + 0.36X_3$
 $Y_2 = 0.51X_1 - 0.61X_2 + 0.61X_3$
- ☐ As componentes principais explicam 67%, 22% e 11% da variância total dos dados.
- ☐ As componentes principais explicam 77%, 12% e 11% da variância total dos dados.
- ☒ As componentes principais obtidas nesta aplicação são chamadas componentes principais exatas.✔

As respostas corretas são: As componentes principais explicam 67%, 22% e 11% da variância total dos dados,. As componentes principais obtidas nesta aplicação são chamadas componentes principais exatas.

Terminar revisão