Questões de Concurso
Análise de Componentes Principais e Análise Fatorial

1 - Análise de Componentes Principais

(ADAF-AM - Estatístico (2018)) Uma das técnicas de Análise Multivariada é a análise por componentes principais. Dada a matriz de covariâncias do vetor aleatório $\boldsymbol{X'} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3})$, os resultados da análise de componentes principais foram os seguintes:

Componente	Autovalor	Percentagem da Variação	Percentagem Acumulada
1	5,813	69,095	69,095
2	2,350	27,933	97,028
3	0,250	2,971	100,000
Variável	**Autovetor 1**	**Autovetor 2**	**Autovetor 3**
X_1	-0,39	0,00	0,89
X_2	0,95	0,00	0,40
X_3	0,00	1,00	0,00

Considerando o exposto, assinale a alternativa que apresenta a primeira componente principal.

(A) $Y_1 = -0,39X_1 + 0,95X_2$
(B) $Y_1 = -0,39X_1 + 0,89X_2$
(C) $Y_1 = 5,813X_1 + 2,35X_2 + 0,25X_3$
(D) $Y_1 = -0,39X_1$
(E) $Y_1 = -2,267X_1 + 2,232X_2$

(ADAF-AM - Estatístico (2018)) Considerando as informações contidas na questão anterior, referente à análise de componentes principais, qual é a matriz de covariâncias do vetor de componentes principais?

(A) $V(Y) = \begin{bmatrix} 69,095 & 0 & 0 \\ 0 & 27,933 & 0 \\ 0 & 0 & 2,971 \end{bmatrix}$

(B) $V(Y) = \begin{bmatrix} 5,813 & 0 & 0 \\ 0 & 2,35 & 0 \\ 0 & 0 & 0,25 \end{bmatrix}$

(C) $V(Y) = \begin{bmatrix} 69,095 & 0 & 0 \\ 0 & 97,028 & 0 \\ 0 & 0 & 100 \end{bmatrix}$

(D) $V(Y) = \begin{bmatrix} 5,813 & 69,095 & 69,095 \\ 2,35 & 27,933 & 97,028 \\ 0,25 & 2,975 & 100 \end{bmatrix}$

(E) $V(Y) = \begin{bmatrix} 0,39 & 0 & 0,89 \\ 0,95 & 0 & 0,4 \\ 0 & 1 & 0 \end{bmatrix}$

(DATAPREV - 2012) A análise dos componentes principais é um método de se expressarem os dados multivariados. Ela permite que o pesquisador reoriente os dados para que algu-

mas poucas primeiras dimensões expliquem tantas informações quanto possível. A análise de componentes principais é também útil na identificação e compreensão dos padrões de associação entre as variáveis. Considere as cinco afirmações seguintes, sobre a análise dos componentes principais:

I. O primeiro componente principal, Z_1 é dado pela combinação linear das variáveis originais $X = [X_1, X_2, \ldots, X_p]$ com maior variância possível.

II. Todos os componentes principais subsequentes são escolhidos para que não sejam correlacionados a todos os componentes principais anteriores.

III. Em razão de a análise de componentes principais buscar maximizar a variância, ela pode ser altamente sensível às diferenças de escala entre variáveis. Assim, é uma boa ideia padronizar os dados e representá-los por X_S.

IV. A solução para o problema dos componentes principais é obtida realizando-se uma decomposição de autovalor da matriz de correlação. Cada autovetor, indicado por u_i, representa a direção de um desses eixos principais. O vetor u controla os pesos usados para formar a combinação linear de X_S, que resulta em $z_i = X_S u_i$.

V. A matriz das cargas, representada por F, é calculada por $F = UD$, em que U é a matriz dos autovetores, e D é a matriz de covariância dos componentes principais.

VI. No caso mais geral, só faz sentido utilizar a análise dos componentes principais quando os dados não são independentes. Barlett fornece um teste de qui-quadrado para determinar a esfericidade dos dados, representado por $\chi^2 = -\big[n - 1 + (2p + 6)/5\big] ln|R|$, com $(p^2 - p)/2$ graus de liberdade, onde p é o número de variáveis, n é o tamanho da amostra, e R é a matriz de correlação.

Dentre as seis afirmações dadas, quantas são falsas?

(A) Quatro
(B) Três
(C) Duas
(D) Cinco
(E) Seis

(TJ-RO - Analista Judiciário - Estatística (2012)) Uma análise de componentes principais considerou 20 variáveis. Com base na matriz de covariância entre essas variáveis, observou-se que os cinco maiores autovalores foram iguais a 6, 4, 3, 2 e 1. Considerando esses resultados, assinale a opção correspondente ao percentual de variação explicada por esses cinco maiores autovalores.

(A) 80%
(B) 40%
(C) 50%
(D) 60%
(E) 70%

(DATAPREV - 2009) O objetivo principal da Análise de Componentes Principais é:

(A) obtenção de um pequeno número de combinações lineares, de um conjunto de variáveis, que retenham o máximo possível da informação contida nas variáveis originais;
(B) calcular, para cada indivíduo da amostra, a probabilidade associada a um determinado fenômeno;
(C) testar, por meio de sub-amostras, a distribuição de probabilidades do conjunto de dados estudados;
(D) explicar a correlação ou covariância, entre um conjunto de variáveis, em termos de um número limitado de variáveis não-observáveis;
(E) identificar se há multicolinearidade nos resíduos, após a estimação de uma curva de regressão.

2 - Análise Fatorial

(DPE-PR - Estatístico (2017)) O governo do Estado do Paraná deseja avaliar as condições de vida dos seus habitantes, e para isso realizou um levantamento do perfil socioeconômico da população, utilizando informações do Censo Demográfico de 2010, disponibilizadas pelo Instituto de Geografia e Estatística (IBGE), com intuito de criar um indicador que represente as variáveis originais, obtendo assim uma medida geral e sintética da condição de vida da população paranaense. Nesse sentido, empregou-se a técnica de Análise Multivariada de Dados - Análise Fatorial. Diante do exposto e com base na tabela abaixo que apresenta os resultados da Análise Fatorial, é correto afirmar:

Tabela: Seleção de fatores pelo método de componentes principais, com rotação varimax.

Fator	Variável	KMO	Bartlett (p)	% Variância	Carga Total	MAA	Comum.
	% Domicílio sem banheiro.				0,87	0,757	0,77
	% Domicílio sem coleta de lixo.				0,84	0,706	0,74
1	% Pessoas responsáveis pelo domicílio com renda mensal de até 1 SM ou s/ rendimento.	0,79	<0,001	80,14	0,97	0,859	0,88
	% Chefes de família analfabetos e/ou com até um ano de estudo				0,68	0,694	0,69
2	% Domicílios sem abastecimento de água			19,8	0,95	0,938	0,93

Legenda: KMO – Kaiser - Meyer – Olklin / MAA – Medida de Adequação da Amostra / Comum – Comunalidade

(A) Para verificar a adequação do modelo da Análise Fatorial, somente o teste de esfericidade de Bartlett é suficiente, tendo em vista que este teste tem por hipótese nula o fato da matriz de correlação ser uma matriz identidade. Se esta hipótese for aceita, significa que as variáveis estão correlacionadas, sendo adequada a aplicação da Análise Fatorial.
(B) Pode-se afirmar que o valor do KMO igual a 0,79 indica a inadequação dos dados à técnica e o teste de esfericidade de Bartlett ($p < 0,001$) conduz a rejeição da hipótese sobre a matriz de correlação ser a matriz identidade, evidenciando que não há correlação entre as variáveis do estudo.
(C) As Medidas de Adequação das Amostras apresentadas na tabela são todas superiores a 0,50, o que indica que essas variáveis não se ajustaram à estrutura definida pelas outras variáveis, logo, devem ser retiradas da Análise Fatorial uma por vez.
(D) Com base na regra de retenção dos fatores pode-se concluir que foram retidos dois fatores que explicam <60% da variância total dos dados originais.
(E) O teste de esfericidade de Bartlett ($p < 0,001$) rejeita a hipótese nula H_0: a matriz de correlação é a identidade, o que evidencia que existe correlação entre as variáveis, mostrando que é indicada a utilização da Análise Fatorial.

(EBSERH - 2017) Na Análise Fatorial Exploratória, a comunalidade é um dos principais indicadores de ajuste de aplicação da técnica. A definição mais adequada para essa medida é:

(A) Representa a quantia total de variância no qual a solução fatorial é baseada.
(B) Correlação entre as variáveis originais e os fatores.
(C) Quantia total da variância que uma variável original compartilha com todas as outras variáveis incluídas na análise.
(D) Medida de confiabilidade que varia entre 0 e 1.
(E) Nível de dependência entre duas variáveis do modelo.

(TRE-MG - Analista Judiciário - Estatística (2013)) O modelo de análise fatorial representa a estrutura de covariância entre muitas variáveis aleatórias $\boldsymbol{X' = [X_1, X_2, \ldots, X_p]}$, através de poucas variáveis não observáveis $\boldsymbol{F' = [F_1, F_2, \ldots, F_m]}$ também conhecidas como fatores, construtos ou fatores comuns. Sendo $\boldsymbol{E(X) = \mu}$ e $\boldsymbol{V(X) = \Sigma}$, o modelo fatorial é expresso por $\boldsymbol{X - \mu = LF + \varepsilon}$. A matriz $\boldsymbol{L_{p\times m}}$ é conhecida como matriz das cargas fatoriais e seus elementos, $\boldsymbol{l_{ij}}$, carga da variável $\boldsymbol{i}$ no fator $\boldsymbol{j}$ e as variáveis aleatórias $\boldsymbol{F}$ e $\boldsymbol{\varepsilon m + p}$ são não observáveis. Analise as afirmativas, marque <u>V</u> para as verdadeiras e <u>F</u> para as falsas.

( ) No modelo fatorial ortogonal, as variáveis não observáveis F e ε são independentes, $E(F) = 0$, $V(F) = E(F'F) = I$, $E(\varepsilon) = 0$, $V(\varepsilon) = E(\varepsilon'\varepsilon) = \Psi$. A matriz Ψ é não diagonal, $V(X) = \Sigma = L'L + \Psi$ e $Cov(X, F) = L$.

( ) Um método de estimação para as cargas do modelo fatorial ortogonal é através de componentes principais, onde se utiliza a decomposição espectral da matriz Σ.

( ) Para se utilizar o método de máxima verossimilhança para estimar as cargas, é acrescida a suposição de que F e ε têm distribuição normal multivariada. As comunalidades (elementos da diagonal LL') têm como estimadores a proporção da variância total estimada pelo particular fator.

( ) Para melhorar a explicação do modelo fatorial, sem alterar a ortogonalidade dos fatores, muitas vezes, usa-se uma transformação ortogonal das cargas fatoriais, que, consequentemente, transforma os fatores. Esse procedimento é conhecido como rotação fatorial.

( ) Dependendo da natureza dos dados, os fatores não precisam ser ortogonais. Assim, para melhorar a explicação do modelo fatorial, pode-se utilizar a rotação oblíqua, onde cada variável é expressa em termos de um número máximo de fatores.

A sequência está correta em

(A) V, V, V, F, F
(B) F, V, V, F, V
(C) F, V, V, V, F
(D) V, F, V, V, F
(E) V, V, V, V, V

(TRF-2ª REGIÃO - 2017) Sobre a análise fatorial e suas propriedades, assinale a afirmativa INCORRETA.

(A) A rotação ajuda a tornar o resultado mais interpretável.
(B) Os fatores estimados não mudam com o acréscimo de novos fatores.
(C) Uma das suposições necessárias é que todos os fatores tenham média igual a zero.

(D) A estimação do número de fatores a ser utilizada pode ser feita com o auxílio dos autovalores da matriz de correlação amostral.

(PC-MG - Analista da Polícia Civil - Estatística (2013)) São métodos de rotação ortogonal dos fatores de uma análise fatorial, EXCETO:

(A) Promax.
(B) Varimax.
(C) Equamax.
(D) Quartimax.

(TJ-RS - 2012) A análise fatorial é uma técnica multivariada que busca identificar um número pequeno de fatores capazes de representar relações entre um conjunto de variáveis inter-relacionadas. Com base na teoria de análise fatorial, assinale a alternativa que apresenta a afirmação correta.

(A) O teste de Bartletts pode ser usado para testar a hipótese de que a matriz de correlação é uma matriz identidade; quando o valor da estatística do teste de Bartletts é grande e o nível de significância é pequeno, é provável que a matriz de correlação da população possa ser considerada a matriz identidade.
(B) Se as correlações entre as variáveis forem altas, é improvável que essas variáveis compartilhem fatores comuns.
(C) O coeficiente de correlação múltipla entre uma variável e todas as outras variáveis é também um indicador da força da associação linear entre as variáveis e os fatores, denominado de autovalor na matriz contraimagem.
(D) Se os fatores obedecem ou não à suposição de ortogonalidade, as cargas fatoriais são coeficientes de regressão padronizados em uma equação de regressão múltipla na qual a variável original de interesse no estudo é a variável dependente e os fatores são variáveis independentes.
(E) Em uma análise fatorial, para selecionar os fatores que permanecerão na análise, o autovalor associado a um determinado fator deve ser pequeno para garantir significância.