Curso de pós-graduação em Estatística Aplicada

Felipe Neres Silva Bezerra

Estatística Multivariada – Análise Fatorial Exploratória

1. Discuta cada um dos itens abaixo:

1. Cite as diferenças conceituais entre a análise de componentes principais e a análise fatorial. Dê um exemplo de aplicação de cada uma delas.

A análise de componentes principais preocupa-se mais em reduzir o número de variáveis mantendo a maior parte da variância explicada, enquanto que a análise fatorial tem o intuito de agrupar variáveis em busca de variáveis latentes que possam ser interpretadas.

1. Após uma rotação de fatores em uma Análise Fatorial, os autovalores e as comunalidades podem se alterar? Explique.

Como as cargas se alteram com a rotação, os autovalores de cada fator também se alteram, embora não haja modificação na variância acumulada. O mesmo ocorre com as comunalidades quando se utiliza rotação oblíqua.

Com a rotação ortogonal, por outro lado, as comunalidades são preservadas, já que não se alteram as posições das variáveis no espaço dos fatores.

2. Considere um modelo com dois fatores comuns de 7 características de vinhos representadas pelas seguintes combinações após rotação Varimax:

*Hedônico = − 0,71∗𝐹1 + 0,01∗𝐹2*

*Adequado para carnes = − 0,72∗𝐹1 − 0,21∗𝐹2*

*Adequado para sobremesas = − 0,11∗𝐹1 − 0,73∗𝐹2*

*Preço = 0,80∗𝐹1 − 0,20∗𝐹2*

*Açúcar = 0,13∗𝐹1 − 0,91∗𝐹2*

*Álcool = − 0,75∗𝐹1 − 0,05∗𝐹2*

*Acidez = − 0,78∗𝐹1 − 0,03∗𝐹2*

1. a) Quais são os autovalores dos fatores? Comente.

Os autovalores dos fatores 1 e 2 são, respectivamente, 2,8624 e 1,4486. O fator 1 se mostra mais relevante explicando as variáveis observadas, mas, seguindo o critério de Kaiser, ambos fatores devem ser considerados.

1. b) Qual o percentual da variabilidade é explicada pelos fatores individualmente e conjuntamente?

Os fatores 1 e 2 explicam, respectivamente, 40,89% e 20,69% e explicam conjuntamente 61,59 de toda variância observada.

1. c) Calcule as comunalidades. Qual característica dos vinhos foi menos explicada pelo modelo? E qual foi a mais explicada?

A característica menos explicada foi “Hedônico”, que teve 50,42% de sua variância explicada. A característica mais explicada foi “Açúcar”, com 84,5% da variância explicada pelos dois fatores.

1. d) Interprete os fatores.

O fator 1 está correlacionado negativamente com as variáveis “Hedônico”, “Adequado para carnes”, “Álcool” e “Acidez”, e positivamente correlacionado com a variável “Preço”. Se vinhos mais caros forem caracterizados pela ausência das demais variáveis supracitadas, pode-se assumir o fator 1 como uma medida de qualidade. Como o fator 2 está correlacionado negativamente com “Adequado para sobremesa” e “Açúcar”, e supondo que vinhos doces possam ser mais adequados para sobremesas, este fator estaria associado aos vinhos “não-doces” (ou qualquer outra nomenclatura antônima a “doce” que seria melhor definida por um sommelier).

3. No ano de 2013, 1010 jovens eslovacos de 15 a 30 anos participaram de uma pesquisa acerca de hobbies, personalidades, opiniões, hábitos de consumo e preferências quanto a música e filmes. Um questionário com 90 questões foi aplicado, em que as respostas eram escores de 1 a 5, com 1 correspondendo a “Discordo totalmente”, “Não gosto” ou “Não tenho interesse” e 5 correspondendo a “Concordo totalmente”, “Gosto muito” ou “Muito interessado”.

Na planilha “variáveis\_jovens.xlsx” encontram-se as questões aplicadas na pesquisa e seus respectivos nomes correspondentes na planilha “dados\_jovens.csv”.

1. Analisando o KMO e o teste de esfericidade, podemos descrever essa grande quantidade de interesses humanos em um número menor de conceitos latentes?

Segundo medida de adequacidade do critério de Kaiser-Meyer-Olkin, 81% da variabilidade vem de fator comum e, de acordo com o teste de esfericidade de Bartlett, rejeita-se a hipótese nula de que as variáveis não são correlacionadas. Ambos os testes indicam que é possível descrever as variáveis observadas em variáveis latentes.

1. Quantos fatores podem ser utilizados no modelo?

Seguindo o critério de Kaiser, os 24 primeiros fatores são relevantes explicando as variáveis observadas. O mesmo número de fatores são necessários para garantir ao mesmo 60% de variância explicada. Embora o screeplot demonstre queda abrupta do autovalor quando se chega a quatro fatores.

Visando manter a capacidade explicativa do modelo e seguindo métodos menos subjetivos, é aconselhável seguir com a análise de 24 fatores. Entretanto, avaliando as cargas (como segue nos itens posteriores), alguns fatores não foram relevantes para nenhuma variável original, sugerindo reduzir o número de fatores. Exigindo 19 fatores da análise fatorial com rotação Varimax, embora se perca variância explicada a níveis não aconselháveis, todos os fatores puderam ser interpretados.

1. Estime as cargas dos fatores por Eixos Principais e escolha um método de rotação (Varimax ou Oblimin) e explique por que é mais adequado para a análise dos dados.

Utilizando a rotação oblíqua, apenas 3 dentre os 24 fatores apresentaram correlação igual ou superior a 0,3, sendo que nenhuma destas correlações esteve demasiadamente longe dessa faixa (PA1 e PA22 com correlação de 0,32; PA2 e PA15 com correlação de 0,30, e PA1 e PA18 com correlação de 0,30). Tendo em vista que a magnitude das correlações entre os fatores ultrapassaram o limite recomendado por pouco e que os casos em que isso ocorreu não são tão expressivos visto o total de fatores, a análise segue com rotação Varimax com redução do número de fatores para 19 (como citado no item anterior), considerando que os fatores são independentes, o que favorece à interpretação dos fatores.

1. Interprete os fatores encontrados.

O fator *Classicismo* (PA1) está mais correlacionado diretamente à preferência por músicas dos gêneros folclórico, clássico, ópera, swing/jazz; aos interesses em psicologia, recitação e criação de poesia, exposição de artes, instrumentos musicais e teatro, e a doação para a caridade.

O fator *Dinamismo* (PA2) está mais correlacionado diretamente ao gosto pelos gêneros cinematográficos de ficção científica, faroeste e ação; aos interesses em matemática, física, internet, softwares de computador, carros e ciências e tecnologia, e ao dispêndio com aparatos tecnológicos.

O fator *Vaidade* (PA11) está mais correlacionado diretamente aos interesses no estilo de vida das celebridades e em fazer comprar; à preocupação com a saúde, com a aparência, com fazer contatos com as pessoas certas, e à preferência por grandes shopping centers, usar roupas de marca e com o gasto com a aparência.

O fator *Desgosto* (PA5) está mais correlacionado diretamente à autocrítica, à sensação de solidão, ao arrependimento, à alteração de humor e à raiva, e inversamente às variáveis relacionadas a ter sonhos bons, à felicidade e ao quanto se está satisfeito com a própria personalidade.

O fator *Receptividade* (PA14) está mais correlacionado diretamente a ter muitos amigos, à adaptação a novos ambientes, à socialização, ao nível de energia e ao interesse em hobbies diferentes, e inversamente às variáveis relacionadas ao medo em falar em público, ao uso do tempo para tomar decisões e à necessidade de preparo para falar em público.

O fator *Ciências Naturais* (PA7) está mais correlacionado diretamente aos interesses em biologia, química e medicina.

O fator *Música Agitada* (PA8) está mais correlacionado diretamente à preferência por músicas dos gêneros Rock, Metal Rock, Hardrock, Punk, Rock ‘n’ Roll e música alternativa.

O fator *Visão de Longo Prazo* (PA4) está mais correlacionado diretamente a um estilo de vida saudável, à presteza, à organização de informações, obsessão por trabalho/estudo e formação de poupança, e inversamente ao absenteísmo escolar e à dificuldade de acordar cedo.

O fator *Geopolítica* (PA6) está mais correlacionado diretamente aos interesses em história, política, geografia e eventos próximos.

O fator *Música Dançante* (PA16) está mais correlacionado diretamente à preferência por músicas dos gêneros Dance, Disco, Funk, Pop e música latina.

O fator *Religiosidade* (PA9) está mais correlacionado diretamente ao interesse em religião, à crença num “acerto de contas” *post mortem* e à crença em Deus.

O fator *Socialização* (PA3) está mais correlacionado diretamente ao interesse em socializar-se e ao gasto com festas e socialização.

O fator *Sensibilidade* (PA18) está mais correlacionado diretamente ao gosto pelos gêneros cinematográficos de comédia, romance e fantasia, à empatia e a suscetibilidade ao choro.

O fator *Esportes* (PA15) está mais correlacionado diretamente aos interesses em dança, esportes em nível competitivo e esportes radicais.

O fator *Confiabilidade* (PA19) está mais correlacionado diretamente à confiabilidade no trabalho, ao cumprimento de promessas e ao cuidado com o que se toma emprestado.

O fator *Filmografia Embasada em Fatos* (PA12) está mais correlacionado diretamente ao gosto por documentário e por filmes ambientados em período de guerra.

O fator *Filmografia que Instiga Medo* (PA10) está mais correlacionado diretamente ao gosto pelos gêneros cinematográficos de terror e suspense.

Os fatores *Línguas Estrangeiras* (PA13) e *Atenção aos Conselhos dos Pais* (PA17) estão mais correlacionado diretamente à apenas as próprias variáveis, respectivamente.

As variáveis correspondentes a interesse em dança, hábito de faltar na escola, preocupação com a saúde, doação para a caridade, número de amigos, conhecer as pessoas certas, suscetibilidade ao choro, nível de energia, atenção ao conselho dos pais e formação de poupança estão correlacionadas de forma muito próxima a mais de um fator, classificando-as como variáveis complexas.