

ME 731 - Métodos em Análise Multivariada
Segundo semestre de 2020
Lista de Exercícios V

OBS: Para os exercícios relativos à análise de dados via MAFO, você deve obter (quando possível) a solução (via componentes principais e MV) bem como as soluções de MV rotacionadas via varimax e promax (ou seja, quatro soluções) e interpretar os fatores. Com base nos critérios discutidos em sala de aula você deve escolher uma entre as três últimas (MV, MV varimax e MV promax) e obter os escores fatoriais (via MQP ou via método da regressão) escolhidos de acordo com critérios que você julgar pertinentes. Não se esqueça de fazer o biplot usando a solução escolhida.

1. Resolva todos os exercícios deixados em sala.
2. No MAFO visto em sala de aula, considere que $Cov(\mathbf{F}) = \mathbf{\Omega}$. Refaça as contas (do modo mais simples possível), vistas em sala, que dependam de tal resultado.
3. Repita a questão anterior considerando $Cov(\mathbf{F}) = \mathbf{I}$ e $Cov(\mathbf{F}, \boldsymbol{\xi}) = \mathbf{\Sigma}$.
4. Repita a Questão 2 considerando $Cov(\mathbf{F}) = \mathbf{\Omega}$ e $Cov(\mathbf{F}, \boldsymbol{\xi}) = \mathbf{\Sigma}$.
5. Obtenha a distribuição conjunta de $(\mathbf{F}_i^{*MQP'}, \mathbf{F}_i^{*R'})'$.
6. Resolva todas as questões das listas anteriores, que foram resolvidas via ACP, através do MAFO.
7. No pacote “psych”, do programa R, há um conjunto de dados disponível sob o nome “sat.act”. As variáveis são: sexo, educação informada, idade, escores compostos do teste ACT, escores verbais do teste SAT, escores quantitativos do teste SAT. Veja maiores detalhes no “help” do conjunto de dados. Considere que a primeira variável é explicativa e as demais, resposta. Faça uma análise descritiva geral e por sexo. Analise os dados via MAFO, também comparando os grupos. Elimine os indivíduos com observações faltantes.
8. Considere os dados “diamonds” no pacote “ggplot2”. As observações consistem em preços e atributos de quase 54 mil diamantes, dados pelas variáveis quilate, profundidade, tabela, preço, x, y, z e qualidade do corte. Faça uma análise descritiva dos dados. Em seguida, faça uma análise dos dados via MAFO e compare os seus resultados com os grupos fornecidos pela variável qualidade do corte.

9. Neuroblastoma é um tipo de câncer cujas ocorrências são mais frequentes em crianças. Os dados disponíveis em ([aqui](#)), foram obtidos de 246 pacientes com neuroblastoma e consistem em expressões genéticas medidas em 10707 locais e as respectivas informações de sobrevivência dos pacientes (0: positivo, 1: negativo). Considerando os pacientes como observações e as medidas genéticas como variáveis, tente realizar uma análise via MAFO dos dados. Que desafios o conjunto conjunto de dados impõe à análise via MAFO? Que alternativas poderiam ser tomadas para contornar o problema da alta dimensionalidade dos dados utilizando ainda todos os dados disponíveis?