

## **PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS**

**Disciplina:** Unsupervised Machine Learning: Análise Fatorial e PCA I

**Data:** 22/06/2021

**Wilians Pereira Dutra**

professor, é dessa maneira que podemos então descartar algumas variáveis? mantemos apenas as variáveis que geraram, como no exemplo, os dois primeiros fatores?

William, é uma forma de melhorar o modelo sim. Mas depende muito dos seus dados e da sua pergunta de pesquisa.

**Thiago Escudero Augusto**

como interpretar as barras acima e abaixo de zero no gráfico dos eigenvectors? Elas significam as variâncias capturadas pelos fatores em relação à cada variável.

**Bruna Gennari**

No comando `sum(eigenvalues_rho$values)`, a sequência de valores ficou invertida

Oi Bruna. Durante a aula não verifiquei o problema. Inclusive o professor Rafael fez o cálculo e foi correto.

**Luiz Novi**

`plot_gg(ggobj = plot3d_rho, multicore = TRUE, width = 6, height = 6, scale = 300, background = "white", shadowcolor = "dodgerblue4")`

Não entendi bem a dúvida. Mas Tente rodar todo script novamente. Todas as funções foram testadas e estão corretas. Qualquer dúvida pode entrar em contato.

**Teresa Arlinda De Souza Campos**

quando executo a linha 549 - Erro: Cannot add ggproto objects together. Did you forget to add this object to a ggplot object?

Oi Teresa. Tente rodar todo script novamente. Todas as funções foram testadas e estão corretas. Qualquer dúvida pode entrar em contato.

**Larissa Ribeiro Rios**

e qual a conclusão final? não ficou claro

Na próxima aula iremos abordar novamente o assunto. De toda forma, tente assistir a aula novamente para se acostumar com o assunto. É normal ter certa dificuldade com esses novos conceitos.

**Eliane Chinaglia**

complementando a pergunta sobre o ajuste linear, então a função `smooth` não fornece os coef. linear e angular, só ajusta a reta. No `chart.correlation`, a curva no gráf de correlação é uma curva média?

Respondida pelo Prof. Rafael na durante a aula.

**Dilson Godoi Espenchitt**

apareceu a msg Loading required namespace: mgcv  
Tente instalar: install.packages("spatstat")

#### **Ricardo Rocha Pavan da Silva**

Por que não usou a função prcomp() para performar o PCA? Qual seria a diferença entre o pacote psych?

Estamos fazendo um passo a passo inicial. Na próxima aula veremos o modelo de uma forma mais dinâmico e resumida.

#### **Ricardo Santana Feitosa**

Professor, Na aula da professora Adriana ela comentou que a padronização deveria ocorrer quando você tem variáveis com características diferentes. Se tratando de notas porque precisamos padronizar?

Na análise fatorial não é necessário padronizar. Fizemos apenas um exemplo didático.

#### **Sintia Silva De Almeida**

Prof. poderia trazer um exemplo na área biológica , de expressão genica por exemplo?

Oi Sintia. Você pode seguir os mesmos passos só mudando a base de dados para alguma com variáveis de expressão genética que você pretende estudar.

#### **Cassio Fernando Ramos Caritá**

favor fornecer exemplos praticos em mercado financeiro nos materiais complementares

Oi Cassio. Você pode seguir os mesmos passos só mudando a base de dados para alguma com variáveis do mercado financeiro que você pretende estudar.

#### **Alessandro Nogueira Corrêa**

Não está claro como houve o agrupamento de variáveis, ou seja, não está claro a redução estrutural.

Na próxima aula iremos abordar novamente o assunto. De toda forma, tente assistir a aula novamente para se acostumar com o assunto. É normal ter certa dificuldade com esses novos conceitos.

#### **Heloísa Hilário**

Professor pode contar uma historinha novamente do passo a passo que precisamos fazer usando o exemplo prático pra ficar mais claro?

Oi Heloísa. Na próxima aula iremos abordar novamente o assunto. De toda forma, tente assistir a aula novamente para se acostumar com o assunto. É normal ter certa dificuldade com esses novos conceitos.

#### **Vitor Hugo De Marrafon**

Rafael, trabalhando com PCA já me deparei com o termo Varimax rotation, isso é um processo aparte da análise ?

O critério VARIMAX se concentra na simplificação das colunas da matriz fatorial. Ou seja, o método VARIMAX maximiza a soma de variâncias de cargas exigidas da matriz fatorial. Destaca-se que existem vários métodos de rotação dos fatores para conseguir a melhor distribuição da carga fatorial.

### Marina Bendit Szacher

Prof se ao final as correlações não forem 0, quer dizer que a hipótese nula não deveria ter sido rejeitada quando a matriz rho foi feita?

Provavelmente o teste de Bartlett não teria refutado a hipótese nula e o modelo fatorial estaria comprometido.

### Guilherme Santos Vasconcelos

A que conclusão chegamos com as correlações que tivemos e as variáveis desse exemplo?

Você precisa ver quais as variáveis formaram cada fator e ter seus insights.

### Rafael Cavalcante De Oliveira

Como posso interpretar o output final dos fatores?

Você precisa ver quais as variáveis formaram cada fator e ter seus insights.

### William Castro Bezerra

Como posso interpretar o output final dos fatores?

Você precisa ver quais as variáveis formaram cada fator e ter seus insights.

### Julia Ungaratti Nunes

Qual o principal outcome desse script? Se o professor puder resumir de forma prática e sumarizada quais as principais informações que conseguimos através dessas fórmulas e scripts seria ótimo!

Podemos sumarizar/reduzir as variáveis e englobar elas em fatores.

### Guilherme Santos Vasconcelos

E quem [e esse fator que explica os 63% do comportamento da nossa amostra

Os dois fatores agrupados explicar cerca de 63% da variabilidade da amostra.

### Gabriel Rodrigues Coutinho Pereira

Professor, como que uso o valor de pchisq para chegar na significância do teste? A partir de quanto esse pchisq indica que eu posso realizar PCA com minhas variáveis?

Para análise fatorial utiliza-se a correlação das variáveis para formar os fatores. O teste qui-quadrado tem função de testar hipóteses e para ser utilizado no teste de Bartlett que vai informar se seus dados suportam uma análise fatorial.

### AMANDA de Andrade Guimarães Beuthner

Professor, mas a correlação não tem que ser o? Não são 2 fatores?

A correlação entre os fatores sim, mas não há em tese como ter essa realidade nas variáveis.

### Samanta Lekecinskas Rodrigues

Alguém teve esse erro? `> notasfatorial %>% + kable() %>% + kable_styling(bootstrap_options = "striped", + full_width = T, + font_size = 12)`  
Error in notasfatoria

Tente instalar diretamente o pacote magrittr e depois chamar ele no library

### Thiago Silva Neves de Barros

Seria possível explicar isso desenhando o quadrado, cubo e etc?

Oi Thiago. Acho que não entendi bem a pergunta. Tente ver a aula novamente, caso não entenda envie um email pra nós. Estamos à postos para tentar ajudar.

**David Pohl**

Professor, é possível "superar" as ponderações arbitrárias com conhecimento teórico em variáveis qualitativas? Já vi algo semelhante em psicometria e equações estruturais.

O prof. Rafael respondeu na aula. Existindo uma teoria subjacente pode ser viável.

**Cainã Max Couto Da Silva**

Prof, vc vai ver a função `pca()` do R base?

Neste primeiro momento fazemos o passo a passo da técnica, para depois mostrar uma forma mais fácil de modelar.

**Dilson Godoi Espenchitt**

Não esta aparecendo em 3 d

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

**Felipe De Oliveira Ribeiro**

os graficos em 3d não funcionam aqui, será que é por causa de algum outro pacote? (Obs.: O car já está instalado)

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

**Felipe De Oliveira Ribeiro**

os graficos em 3d não funcionam aqui, será que é por causa de algum outro pacote? (Obs.: O car já está instalado)

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

**Israel Oliveira Rocha**

Qual a diferença disso que estamos vendo na aula de hoje para a análise de correlação múltipla? Em termos de resultados práticos, ou seja, analisar os dados e suas relações?

Os fatores são criados na análise fatorial a partir da correlação das variáveis.

**Augusto Ponso Alves**

Professor, a análise fatorial que estamos aprendendo é a mesma daquela usada em DoE (Design of Experiments), por exemplo análise fatorial 2<sup>k</sup>?

Oi Augusto. São métodos diferentes. Esse link explica o que é um experimento fatorial: <https://rogeriofvieira.com/wp-content/uploads/2016/05/Experimentos-fatoriais.pdf>

**Francisco Felipe Gomes De Sousa**

Na análise fatorial é possível usar variável categórica e na PCA não? Estou tentando entender a diferença entre as duas.

Em análise fatorial não é possível usar variável categórica. A PCA é uma das técnicas mais conhecidas de análise fatorial

#### Higino Tempski de Almeida Leite

... Referenced from:  
/Library/Frameworks/R.framework/Versions/4.1/Resources/library/rgl/libs/rgl.so  
Reason: image not found ...

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

Registered S3 methods overwritten by 'rgl': method from knit\_print.rglld  
knit\_print.rglOpen3d Error in dyn.load(dynlib - getDynlib(dir)) : ...

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

Agora ao final dessa sessão, o Professor comandou algo para aparecer a barra na parte inferior que permitiu visualizar o gráfico 3D e "girar" para observar. Para mim, não apareceu essa barra.

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

...  
dlopen(/Library/Frameworks/R.framework/Versions/4.1/Resources/library/rgl/lib  
s/rgl.so, 6): Library not loaded: /opt/X11/lib/libGLU.1.dylib ...

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

...erro: OpenGL is not available in this build Além disso: Warning messages: 1:  
Loading rgl's DLL failed. This build of rgl depends on XQuartz, which failed to  
load. ...

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>. Oi Higino. Tem alguns erros que são bem específicos de um tipo de computador. Assim que verificamos o erro vamos atrás imediatamente do conserto, mas às vezes pode demorar um pouco.

#### Fabrizio Rodrigues Passos

Existe algum limite para as aplicações das oportunidades no exterior?

Olá Fabrício. Tente entrar em contato com a Secretaria do Pecege. Eles podem te orientar sobre essa questão. Tente o email do prof: joao.guerreiro@ibs-americas.com

#### Dilson Godoi Espenchitt

Rodei a linha 35 algumas vezes e não apareceu o grafico em 3D

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

```
... impossível carregar objeto compartilhado  
'/Library/Frameworks/R.framework/Versions/4.1/Resources/library/rgl/libs/rgl.so'  
: ...
```

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Larissa Ribeiro Rios

Pq nesse comando, tem o ~ e + ? `scatter3d(notas_financas ~ notas_custos + notas_marketing,`

O "~" significa "em função de". No caso é notas\_financas em função de notas\_custos e notas\_marketing.

#### Higino Tempski de Almeida Leite

Amanda, help. Não apareceu a barra embaixo no R. No MAC.

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Ronaldo Lucinda de Oliveira

O meu RStudio não disponibilizou o gráfico após a linha 35.

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Higino Tempski de Almeida Leite

É o pop up para abrir o grafico

Isso mesmo. Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

#### Reinaldo Roberto Aranha

Boa noite. Porque no comando `scatter3d` tem a diferença neste trecho `notas_financas ~ notas_custos + notas_marketing` onde uma variavel utiliza o ~ e a outra o +?

O "~" significa "em função de". No caso é notas\_financas em função de notas\_custos e notas\_marketing.



### Mauricio Matos da Silva Leal

Poderia voltar e demonstrar como acessar o gráfico depois de comandar a linha 35?

Se você estiver usando MAC pode acontecer um erro bem específico do pacote "rgl". Tente seguir os passos deste link: <https://www.docow.com/53213/erro-ao-carregar-o-pacote-rgl-com-o-mac-os-x.html>

### Felipe De Oliveira Ribeiro

No Scatter3d: Porque após a variável "notas\_financas" temos um "~"?  
O "~" significa "em função de"

### Vanessa Nunes Sanchez Cavallini

omicro prof ficou baixinho do nada

Oi Vanessa, não entendi muito bem a sua dúvida. Se persistir pode mandar um email depois de assistir a aula novamente.

### Vanessa Firmino Barbosa

Se nao tiver uma nota, ficaria nulo?

Apareceria NA = no available na base de dados

### Alex De Lima Bassi

Boa noite !!! a minha instalação parou no SASS com o erro make[1]: \*\*\*  
[Makefile:228: src/extender.o] Interrupt make: \*\*\* [Makevars:7:  
libsass/lib/libsass.a] Interrupt ERROR: compilation failed for pac

Tenta instalar diretamente:

```
install.packages("sass")
```

```
library(sass)
```

### Alex De Lima Bassi

A minha instalação deu erro no SASS

Tenta instalar diretamente:

```
install.packages("sass")
```

```
library(sass)
```

### Alessandro Nogueira Corrêa

Favor explicar melhor a questão da ortogonalidade

Ortogonal significa capacidade de formar um ângulo reto de 90°. Neste sentido, ortogonalidade é a independência matemática (sem correlação) de eixos fatoriais, um em relação ao outro (ou seja, em ângulos retos ou de 90 graus), para evitar a correlação entre os fatores formados.

### Fabrcio de Oliveira Alves

Qual a diferença entre a padronização utilizando "z-score" e "scale"?

Na verdade, "scale" é a função que retorna uma padronização z-escore na linguagem R

### Roberto Mesquita Vono Silva

Professor, o que me leva a decidir entre análise de clusters vs. análise fatorial?  
Quais as diferenças, vantagens e desvantagens?

Análise de cluster busca classificar as observações com fundamento nas suas características. Já Análise Fatorial busca reduzir a quantidade de variáveis em fatores semelhantes.

#### Gustavo Murad

Prof não entendi bem a sintaxe do scatter3d em que ele agrega duas variáveis com sinal "~" e a terceira com sinal "+"? `scatter3d(notas_financas ~ notas_custos + notas_marketing`

O "~" significa "em função de". No caso é `notas_financas` em função de `notas_custos` e `notas_marketing`.

#### Roberto Mesquita Vono Silva

Professor, o que me leva a decidir entre análise de clusters vs. análise fatorial? Quais as diferenças, vantagens e desvantagens?

Análise de cluster busca classificar as observações com fundamento nas suas características. Já Análise Fatorial busca reduzir a quantidade de variáveis em fatores semelhantes.

#### Carla Porto Veiga

professor, desculpa se minha pergunta não é pertinente, mas tenho muita dificuldade de entender o que é "estatisticamente" igual a alguma coisa. poderia explicar brevemente? muito obrigada

Quer dizer que dois conjuntos de dados são estatisticamente semelhantes, ou seja, guardam uma relação de significância em termos de semelhança. Por exemplos, duas amostras com milhares de dados podem ser semelhantes a um nível de significância estatística de 95%, com erro de 5%

#### João Batista Duarte

Prof.: Como fazer um teste multivariado de ajuste das variáveis à distribuição normal? (pois o teste de Bartlett está apoiado em distribuição "qui-quadrado" - soma de normais ao quadrado).

Neste site ele explica o passo a passo para fazer um teste de normalidade: <https://rpubs.com/paternogbc/46768#:~:text=%20Como%20realizar%20teste%20de%20normalidade%20no%20R,entendermos%20melhor%20a%20distribui%C3%A7%C3%A3o%20de%20nossos...%20More%20>

#### Bruno Correa

A definição de mudança que demanda a atualização do algoritmo se faz quando a base de dados apresenta novos dados ( dados que extrapolam a amostra anterior) ou seria uma mudança de negocio / direção?

Sempre que novos dados são adicionados é necessário rodar o modelo novamente para verificar quais foram as mudanças nos fatores.