

PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

Disciplina: Unsupervised Machine Learning: Análise Fatorial e PCA II

Data: 29/06/2021

Vinicius Ribeiro

Hey pessoal, alguém tem para disponibilizar o link pra acesso à base de dados completa de indicadores socioeconômicos por região? De onde o professor fez esse recorte que tá sendo explorado no exemplo

Olá, Vinicius! Tudo bem?

Dados secundários governamentais, geralmente, são disponibilizados nos sites de cada instância. Prefeitura, governo estadual ou federal. O IBGE disponibiliza, por exemplo, pesquisa de orçamentos familiares. Você pode buscar bases de dados no portal de dados abertos do governo federal <https://dados.gov.br/>

Alexandre Gonçalves da Rocha

Recebi um erro assim: Fill argument neither colors nor valid variable name(s). O que eu esqueci de instalar ou carregar?

Olá, Alexandre! Tudo bem?

O erro aponta que você está utilizando alguma variável válida como argumento de algum parâmetro. Verifique isso! Porém, verifique também se os pacotes estão sendo chamados pelo R

Pedro Henrique Esteves Trindade

Não seria apropriado deletar as variáveis que pouco contribuem no PCA e refazer a análise? Isso poderia retornar um escore melhor, não?!

Olá, Pedro! Tudo bem?

Neste caso, vai depender dos seus objetivos para análise e no conhecimento dados

Paulo Renato Leite

Professor, pergunta prática, se chegar na minha mão uma solicitação de um ranking para uma base de dados que possua variáveis quantitativas, a melhor solução será o PCA?

Olá, Paulo, tudo bem?

Vai depender dos objetivos de pesquisa.

Vanessa Hoffmann de Quadros

Professor, por que é usada a idade mediana em vez da idade média nessa base de dados?

Olá, Vanessa! Tudo bem?

Talvez tenham utilizado a mediana para compor a base pois essa mostra o valor que está exatamente no meio de um conjunto de dados. No exemplo do professor, no distrito de Água Rasa, a idade de 32 anos está no meio do total de idades encontradas. E para o objetivo de pesquisa foi a melhor forma que encontraram.

Gustavo Murad

Prof, ao analisar a tabela com as notas de todas as matérias para cada aluno, me parece que somando as 4 materiais e tirando a média, chegaríamos no tal ranking que veio via PCA...que prob. resolvemos?

No caso utilizado na aula, utilizou-se para a técnica para verificar como se comportam as notas dos alunos para depois propor o ranking.

RAFAEL GARCIA

Para estabelecer o Ranking, nesse caso, não basta somar as colunas?

No caso utilizado na aula, utilizou-se para a técnica para verificar como se comportam as notas dos alunos para depois propor o ranking.

Luiz Fernando De Menezes Calábria

Professor, podemos então trabalhar com o módulo de f1 e f2?

Olá, Luiz! Tudo bem?

Para criar o ranking sim.

Victor Antonio Pituco Junior

Professor, se eu precisar usar 3 fatores ou mais, precisarei espelhar todos além do F1?

Sim, para poder visualizar seus dados da melhor forma.

Rogério Assiz

library("tmap") Erro: package or namespace load failed for 'tmap': .onLoad falhou em loadNamespace() para 'units', detalhes: chamada: udunits_init(path) erro: no database found!

Olá, Rogério! Tudo bem?

Poderia atualizar os pacotes rodando esse comando no console? Ele irá atualizar suas bibliotecas.

update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)

Adriana Melges Quintanilha Weingart

Onde podemos obter esses Shape Files para outros municípios, Estados ou regiões? Há alguma biblioteca pública?

Olá, Adriana! Tudo bem?

Dados secundários governamentais, geralmente, são disponibilizados nos sites de cada instância. Prefeitura, governo estadual ou federal. Você pode buscar bases de dados no portal de dados abertos do governo federal <https://dados.gov.br/>

Eliane Chinaglia

Rafael, qual é a fonte do arquivo mapa_sp.Rdata? Se eu quisesse fazer do estado de SP por cidade, como encontro o mapa?

Olá, Eliane! Tudo bem?

Dados secundários governamentais, geralmente, são disponibilizados nos sites de cada instância. Prefeitura, governo estadual ou federal. Você pode buscar bases de dados no portal de dados abertos do governo federal <https://dados.gov.br/>

Daniela Maciel Pinto

Outro erro (uso o mac): Erro: Fill argument neither colors nor valid variable name(s). O que seria e o que posso fazer?

O erro aponta que você está utilizando alguma variável válida como argumento de algum parâmetro. Por gentileza, verifique isso!

Matheus de Lima

Poderia fazer o mapa a partir do QGIS? Exportar a base atlas e importar ao QGIS e trabalhar o mapa lá? Acredito que é até melhor e abre mais possibilidades de explorar

Olá, Matheus tudo bem?

Há possibilidade de integração do QGIS com o R aqui há um tutorial <https://www.youtube.com/watch?v=Gvlp1EhhAo>

Isabella Montanhal de Araujo

Problema no tmap, não funcionou mesmo instalando todos os pacotes

Olá, Isabella! Tudo bem?

Verifique se a biblioteca foi chamada se caso sim, tente atualizar as bibliotecas com o comando abaixo:

```
update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)
```

Vitor Bruno da Silveira Guimarães

Consigo usar esse mapaSP.RData para manipular dados da minha cidade?

Olá, Eliane! Tudo bem?

Dados secundários governamentais, geralmente, são disponibilizados nos sites de cada instância. Prefeitura, governo estadual ou federal. Você pode buscar bases de dados no portal de dados abertos do governo federal <https://dados.gov.br/>. No caso do mapaSP.RData, são dados da cidade de São Paulo.

Isabella Montanhal de Araujo

Carregando pacotes exigidos: tmap Error: package or namespace load failed for 'tmap': .onLoad falhou em loadNamespace() para 'units', detalhes: chamada: udunits_init(path) erro: no database found

Olá, Isabella! Tudo bem?

Verifique se a biblioteca foi chamada se caso sim, tente atualizar as bibliotecas com o comando abaixo:

```
update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)
```

Rogério Assiz

Ocorreu o erro ao executar o tm_shape: Error in tm_shape(mapa_sp) : não foi possível encontrar a função "tm_shape"

Olá, Rogério! Tudo bem?

Verifique se as bibliotecas estão instaladas ou se foram chamadas.

Gabriel Cinati Sundfeld

Quando tento abrir o zoom para ver esse mapa, abre a aba "não é possível acessar esse site". Pq?

Olá, Gabriel!

Por gentileza, tente reiniciar o R para ver se o erro persiste.

Renato Calhau Codá

No último comando tive o erro "Erro: Fill argument neither colors nor valid variable name(s)"

Olá, Renato! Tudo bem?

O erro aponta que você está utilizando alguma variável válida como argumento de algum parâmetro. Verifique isso! Porém, verifique também se os pacotes estão sendo chamados pelo R

Isabella Montanhal de Araujo

No meu deu erro nessa parte do mapa de SP Error in tm_shape(mapa_sp) : could not find function "tm_shape"

Olá, Isabella!

Verifique se a biblioteca foi chamada se caso sim, tente atualizar as bibliotecas com o comando abaixo:

```
update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)
```

Leandro Albernaz Alves

Problema no tmap:> library(tmap) Erro: package or namespace load failed for 'tmap': .onLoad falhou em loadNamespace() para 'units', detalhes: chamada: udunits_init(path) erro: no database found!

Olá, Leandro!

Tente, por gentileza, atualizar as bibliotecas com o comando abaixo:

update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)

Ludmilla Manera Conti

Onde encontra esses dados geográficos em RData pra outras regiões?

Olá, Ludmilla!

Dados secundários governamentais, geralmente, são disponibilizados nos sites de cada instância. Prefeitura, governo estadual ou federal. Você pode buscar bases de dados no portal de dados abertos do governo federal <https://dados.gov.br/>.

Rafael Tomio Castro Okabayashi

Professor, como concluir o uso da pontuação na prática? No exercício das notas quanto maior a pontuação, maiores são as notas, menor pontuação então menor as notas. Como ler o resultado dos distritos?

Olá, Rafael tudo bem?

Isso vai depender dos seus objetivos de pesquisa. Porém, conforme falado em aula, esse ranking dos distritos pode ser feito para apontar a qualidade de vida dos distritos. Você pode ver do menor para o pior em qualidade de vida.

Luciana da Costa Amorim

Estou com este problema: Error in tm_shape(mapa_sp) : could not find function "tm_shape"

Olá, Luciana!

Verifique se a biblioteca foi chamada se caso sim, tente atualizar as bibliotecas com o comando abaixo:

update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)

Maurício Luz Mafra e Silva

Olá, eu não consigo carregar o TMAP, que posso fazer para carregar?

Olá, Maurício!

Verifique se a biblioteca foi chamada.

Daniela Maciel Pinto

Alguém sabe o que é o erro: Erro: arrange() failed at implicit mutate() step. * Problem with `mutate()` column `..1`. [i] `..1` = pontuacao. x objeto 'pontuacao' não encontrado

Olá, Daniela! Tudo bem?

Por gentileza, verifique se você rodou as linhas de código correspondentes ao objeto pontuação. O erro aponta que ele não foi encontrado.

Alessandro Nogueira Corrêa

Professor, pode explicar o porquê da rotação?

A rotação tem como objetivo rotacionar os eixos de referência para formar os fatores de forma simples para a interpretação das cargas fatoriais.

Carlos Roberto Nunes

Pergunta: Se F2 basicamente reflete apenas uma variável, não é mais lógico utilizar a própria variável nas análises ao invés de usar o F2 espelha apenas.

Olá, Carlos! Tudo bem?

O F2 é um fator gerado na análise. A análise fatorial é utilizada para redução das variáveis e no caso das variáveis notas de finanças, custos e atuária apresentam uma forte correlação com o F1 e o analista de dados pode interpretar que esse conjunto pode ter sido agrupado em F1 por ser disciplinas quantitativas e no F2 temos a de marketing com alta correlação e pode ter sido agrupada lá por ser uma disciplina qualitativa (não é obrigatório dar nomes aos fatores, pois não é o objetivo da técnica). Ou poderia ter agrupamentos diferentes. Você precisa observar novas variáveis (fatores) que você não poderia enxergar. Como o professor falou em aula, utilizar 4 fatores você capturaria toda a variância e dessa forma, perde o sentido e é melhor utilizar as variáveis.

Rafael Lopes De Souza

O que significa os valores negativos no gráfico dos eigen vectors?????

Olá, Rafael! Tudo bem?

Os eigenvectors são projeções da variância capturada, utilizados para o cálculo dos scores fatoriais e das cargas fatoriais. Por gentileza, verifique a página 26 da apresentação, lá consta a explicação completa.

Mariana Guimarães Castelo Borges Maié

Podemos relacionar o Scree Plot com o Elbow da aula de Cluster?

Olá, Mariana Tudo bem?

No sentido da retenção sim! O primeiro é utilizado para a determinação dos fatores em uma análise fatorial exploratória ou para manter os componentes em uma PCA

Isadora Salvador Rocco

Eu posso entender os eigenvalues como o quanto as observações da matriz de correlação desviam da matriz identidade elevado ao quadrado?

Os eigenvalues são vetores independentes, advindos de uma matriz quadrada, que capturam o máximo da variabilidade dos dados. Pode-se entender, portanto, os autovalores como a parcelas da variância total capturada das variáveis estudadas.

Kauê Bonato De Araujo

No caso do exercício 1, não deveria ser usado o $n=96$ para o teste de Bartlett, uma vez que a base tem 96 observações? Como nada foi definido, por padrão ele usou $n=100$.

Olá, Kauê! Tudo bem?

Sim, você está correto a função automaticamente vai setar para $n=100$. Porém, caso você sete para um $n=96$ na função o p-valor ainda vai ser $< 0,05$.

Wilson Palhao

Professor, como tratar base variaveis maior melhor, e quanto menor melhor (ex. Renda +, Mortalidade -)

Olá, Wilson! Vai depender da técnica escolhida para a análise de dados. A análise fatorial tem como amostragem ideal 100 ou mais observações ou um mínimo de 5 observações por variável porém é recomendado acima de 10.

Bruno Ruffo

Estou com problemas de encoding usando o RStudio no Mac: "Error in make.names(col.names, unique = TRUE) : invalid multibyte string 1" . Como posso resolver esse problema?

Olá, Bruno! Tudo bem?

Por gentileza, tente desinstalar e reinstalar o R novamente. Pode ocorrer algum erro de encoding do R com sua máquina. Caso persista, por gentileza, nos comunique por e-mail.

Higino Tempski de Almeida Leite

Msg erro: Warning messages: 1: Loading rgl's DLL failed. This build of rgl depends on XQuartz, which failed to load. See the discussion in <https://stackoverflow.com/a/66127391/2554330> 2: Trying w

Heginio, tudo bem?

Essas mensagens são de aviso. Acredito que não vá impactar no seu script ao rodá-lo. Mas caso aconteça de não rodar, por gentileza, tente atualizar suas bibliotecas através do console com esse código.

`update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)`

Vitor Bruno da Silveira Guimarães

O que é a quota mesmo?

Olá, Vitor. Poderia enviar um email sendo mais específico com essa dúvida.

Ellen Nogueira

Não ficou muito claro por que multiplica por -1 (linha 168=F2 - data.frame(F2) %>% mutate(fator2 = rowSums(.) * -1), será sempre visual?

Neste caso é uma especificidade do próprio R. Como detalhado pelo professor, na construção de rankings no R, devemos efetuar a multiplicação por -1, visto que os scores fatoriais das observações mais fortes são, por padrão, apresentados acompanhados do sinal de menos.

Ian Dal Grande Furlani

professor, se a gente fizesse um ranking apenas das medias dos alunos o resultado seria diferente?

Provavelmente, pois as médias diferem dos valores discretos e absolutos.

Roseli de Souza Maia

Na linha 156 não entendi muito bem como escolhemos qual fator multiplicamos por -1... se tivesse mais de uma nota perto da nota de marketing multiplicaríamos ela também?

Neste caso é uma especificidade do próprio R. Como detalhado pelo professor, na construção de rankings no R, devemos efetuar a multiplicação por -1, visto que os scores fatoriais das observações mais fortes são, por padrão, apresentados acompanhados do sinal de menos.

Marcos Henrique Da Silva

Professor, nesta base do atlas de SP, eu poderia selecionar apenas a variáveis que têm relação com minha necessidade? por exemplo, olhar apenas para faixa de renda, idade e escolaridade?

Com certeza, Marcos. Tente pesquisar a função `select()` para escolher a variáveis que você tem interesse e criar um banco de dados para fazer suas análises.

Sabrina Portela Costa Ferreira

O R não disponibiliza o log dos comandos rodados?

Oi Sabrina. Sabemos que existem pacotes específicos para fazer essa depuração, mas cuidam de programação bem avançada que foge do escopo do curso.

Yasmin Bezerra Matsunobu de Oliveira

Não entendo por que estamos calculando esses scores para multiplicar pela base padronizada. Poderia explicar melhor o objetivo dessa multiplicação?

Estabelecidos os fatores, podemos definir as cargas fatoriais, que correspondem aos coeficientes de correlação de Pearson entre as variáveis originais e cada um dos fatores. Os scores fatoriais correspondentes a cada fator são exatamente os parâmetros estimados de um modelo de regressão linear múltipla que apresenta, como variável dependente, o próprio fator, e como variáveis explicativas, as variáveis padronizadas.

Higino Tempski de Almeida Leite

rayshader e rgl não carregaram novamente

Oi Higino. Vamos tentar: sair e entrar no R, após tente atualizar os pacotes. Em packages do lado direito e abaixo tente ver se o pacote está marcado. Antes de começar, tente esse código: `update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)/`

Priscila Morgante Bohlsen

tmap precisa de algo diferente pra instalar no mac?

Não vimos esse erro, ainda. Saia e entre no R, após tente atualizar os pacotes. Em packages do lado direito e abaixo tente ver se o pacote está marcado. Antes de começar, tente esse código: `update.packages(repos='http://cran.rstudio.com/', ask=FALSE, checkBuilt=TRUE)/`

Guilherme Pimpão Pessoa de Oliveira

Como visualizar para qual sentido os dados estão sendo estressados quando estamos trabalhando com mais de 2 fatores?

A partir da visualização das cargas fatoriais, como foi feito na linha 76. A roteiro da técnica não muda, enfim.

Fernando Gonçalves

professor, não seria melhor calcular F1 só usando as 3 variáveis que tem forte correlação com o componente 1?

Oi Fernando. É exatamente isso que fizemos. No gráfico resultante da linha 104 fica bem visível.

Wagner Seizo Hokama

ainda não consegui enxergar onde estamos agrupando variáveis

Os fatores utilizam todas as variáveis para ser formados. Todavia, somente as que tiverem maior correlação serão agrupadas no mesmo fator. No gráfico resultante da linha 104 fica bem visível.

Daniel Da Silva

O professor pegou os Fatores F1 e F2 devido à marketing? Poderia pegar F1 e F2?

Os fatores foram escolhidos em função do critério de kaise. Como o critério mais adotado para a escolha da quantidade de fatores é o critério da raiz la- tente (critério de Kaiser), em que são levados em consideração apenas os fatores correspondentes a autovalores maiores que 1, o pesquisador pode optar por elaborar toda a análise subsequente apenas com os dois primeiros fatores, formados pelo compartilhamento de 87,985% (F1 e F2) da variância total das variáveis originais, ou seja, com perda total de variância de 12,015% (F3 e F4)

Hyago Marinho Reis

O meu não aparece o GRAFICO: aparece isso:
/session/viewhtml3758d648cd/index.html?viewer_pane=1&capabilities=1&host=http%3A%2F%2F127.0.0.1%3A45038 not found

Oi Hyago. Tente rodar todo o código novamente, inclusive com fechamento e reabertura do R, com instalação novamente os pacotes. Todos os códigos foram testados, pode ser algo específico do seu computador.

Hyago Marinho Reis

Quando rodo as funções de visualização aparece o erro:
/session/viewhtml3758d648cd/index.html?viewer_pane=1&capabilities=1&host=http%3A%2F%2F127.0.0.1%3A45038 not found

Oi Hyago. Tente rodar todo o código novamente, inclusive com fechamento e reabertura do R, com instalação novamente os pacotes. Todos os códigos foram testados, pode ser algo específico do seu computador.

Rafael Lopes De Souza

tenho dúvida em entende o sinal negativo pq em minha cabeça eu faço o seguinte raciocínio: esses valores representam quanto da variância está naquele fator... mas como isso se torna negativo? e pq a s

Os autovalores derivam da matriz de correlação que podem ter variáveis com correlação forte, mas negativa.

Sandra Oliveira Daniel Trindade

Professor, pode concluir a análise dos gráficos anteriores ao intervalo por favor ?

Apresenta como os fatores conseguem capturar as variâncias dos autovalores.

Thiago Colette Vegi

Os fatores tem uma relação direta com as variáveis? Ex: F1 com notas_financas?

Isso mesmo. Os 2 fatores conseguem capturar o máximo de variância do banco de dados inteiro. No caso, F1 tem alta correlação com a variável finanças.

Ellen Nogueira

Na tabela de cargas fatoriais, a variável marketing apresenta correlação com F2 de -0,99. E se hipoteticamente não encontrasse nenhum fator que guardasse correlação alta com essa variável? Oque fazer?

Provavelmente não existiria esse fator, pois não há qualquer variável correlacionada com esse fator.

Douglas Fernandes de Albuquerque

a função sum (linha 73) soma os autovalores > 1? ou conta o número de autovalores > 1?

Conta o número de autovalores maiores que 1, conforme o critério da raiz latente.

Rafael Lopes De Souza

o que significa valores negativos no eigenvectors?

Os eigenvector são extraídos a partir dos autovalores que, por sua vez, são oriundos da matriz de correlação que podem apresentar valores negativos.

Andres Salas

Como a correlação negativa das variáveis é interpretada na aplicação de um problema?

Quando uma variável aumenta seu valor, a outra variável tende a diminuir quando tem correlação negativa. Por exemplo: Quanto o preço de um produto aumenta, o volume de vendas tende a diminuir, em regra.

Alisson Tavares Corrêa

Boa noite Prof poderia explicar o comportamento no gráfico o PC3 e o PC4 ?

Eles capturam pouca variância total. O fator 3 captura apenas 7,44% e o fator 4 captura apenas 4,57% da variância total, ou seja, explicam muito pouco os dados. Tente olha o summary para verificar isso.

William Castro Bezerra

qual o critério para desprezar os fatores 3 e 4?

Eles capturam pouca variância total. O fator 3 captura apenas 7,44% e o fator 4 captura apenas 4,57% da variância total, ou seja, explicam muito pouco os dados. Tente olha o summary para verificar isso.

Vitor Macedo De Medeiros

Essa captura sempre contemplará todas as variáveis, ou pode ser que o processo ignore uma ou mais delas?

Sempre usará todas as variáveis do seu banco de dados, pois é o padrão do algoritmo. De toda forma, você pode montar outros bancos de dados excluindo variáveis para usar somente aquelas do seu interesse.

Valter Abranches

Tem como explicar de forma mais ilustrativa o que são esses eigenvalues e eigenvectors?

Os autovalores correspondem ao percentual de variância compartilhada pelas variáveis originais para a formação de cada fator. Você pode verificar, por exemplo, que $(1.5872^2 = \text{variância}) / 4(\text{fatores}) = 0.6298$.

Eduardo Ferreira Rego

Poderia por favor, explicar novamente de forma resumida eigenvalues e eigenvectors ?

Os autovalores correspondem ao percentual de variância compartilhada pelas variáveis originais para a formação de cada fator. Você pode verificar, por exemplo, que $(1.5872^2 = \text{variância}) / 4(\text{fatores}) = 0.6298$.

Alessandro Nogueira Corrêa

Professor, os eigenvalues, que são as raízes do polinômio característico, capturam a variância de toda a amostra, ou seja, de cada variável?

Isso mesmo, os autovalores correspondem ao percentual de variância compartilhada pelas variáveis originais para a formação de cada fator.

Isadora Salvador Rocco

Eu posso entender os eigenvalues como o quanto as observações da matriz de correlação desviam da matriz identidade elevado ao quadrado?

Os autovalores correspondem ao percentual de variância compartilhada pelas variáveis originais para a formação de cada fator.

Marcus Origenes Pereira Rodrigues

Dúvida área da educação, digamos que eu tenha colunas com a quantidade de inscritos, de aprovados, que pagaram a inscrição. Essas colunas são qualitativas, mesmo que numéricas?

Depende de como seu banco de dados está estruturado. Se as observações forem escolas ou cidades serão métricas, pois você estará medido quantos inscritos, aprovado e pagante por cidade ou escola, por exemplo.

Vitor Macedo De Medeiros

Essa captura sempre contemplará todas as variáveis, ou pode ser que o processo ignore uma mais delas?

Sempre usará todas as variáveis do seu banco de dados, pois é o padrão do algoritmo. De toda forma, você pode montar outros bancos de dados excluindo variáveis para usar somente aquelas do seu interesse.

Debora Duarte Pinheiro

PC1, PC2, PC3 e PC4 são os fatores?

Isso mesmo. Cada componente é um fator que captura parcela da variância das variáveis originais.

Anna Rita Moraes de Souza Aguirre Mazo

na primeira linha estão os desvios padrão, não seria a segunda linha, da variância?

O desvio-padrão é a raiz quadrada da variância. O valor da segunda linha é a variância compartilhada. Assim, como os autovalores correspondem ao percentual de variância compartilhada pelas variáveis originais para a formação de cada fator, podemos chegar a variância compartilhada na segunda linha (proportion of variance). No caso, você pode verificar que $(1.5872^2 = \text{variância}) / 4(\text{fatores}) = 0.6298$.

Alessandro Nogueira Corrêa

Por que no summary aparece a proporção da variância, uma vez que a matriz de correlação é baseada na razão da covariância pelo produto dos dps?

É um padrão do algoritmo. No caso, você pode verificar que $(1.5872^2 = \text{variância}) / 4(\text{fatores}) = 0.6298$.

José Nataniel Centeno Klug

Rafael, porque fazer a inserção do objeto no começo da linha se pode-se aplicar ao final ficando portanto alinhado com a intenção do código com pipes? Exemplo: `data.frame()` -> `notasfatorial_std`

Pode sim. A linguagem R aceita criação de objetos tanto partindo do final como começo. Todavia, o padrão mais conhecido é no sentido do código da aula. A forma de programar é muito pessoal para cada pessoa.

Erika Aparecida Gava

Se tivesse dois nomes iguais como ficaria quando eu transforme a coluna de nome como "column_to_rownames"?

Ele pega a variável selecionada e transforma em nome de linha. Caso você tenha dois nomes iguais seu banco de dados tem um problema, pois provavelmente dará erro. Assim, é melhor renomear uma das variáveis.

Marcos Batalha

Seria possível, nos scripts, sempre informar os pacotes para familiarizarmos?

Oi Marcos. Na aba Help você pode verificar a documentação dos pacotes. Tenta rodar: `library(tidyverse)` e depois: `tidyverse_packages(include_self = TRUE)`. Antes instale o pacote tidyverse, caso ainda não esteja instalado.

Vera Lucia Candiani

Digamos que temos uma base com 70 variáveis. Como escolho inicialmente as minhas variáveis para aplicar a Análise fatorial?

Tente rodar uma análise com todas as variáveis, primeiramente.

Thiago Colette Vegi

Pode explicar como analisar o resultado da linha 202 do script de teoria. `rho - cor(notasfatorial[,2:5])`

É uma matriz de correlação das notas dos alunos é criada quando você roda essa função. Verifique que neste momento é criado um objeto no “environment” (lado esquerdo direito). Após, com a função “`chart.Correlation`” do pacote é um atalho para criar um gráfico de correlação em R com histogramas, funções de densidade, linhas de regressão suavizadas e coeficientes de correlação com os níveis de significância correspondentes (se não houver estrelas, a variável não é estatisticamente significativa, enquanto uma, duas e três estrelas significam que a variável correspondente é significativa em 10%, 5% e 1% níveis , respectivamente).

Augusto Ponso Alves

Professor, não entendi o Scree Plot. Ele é usado como se fosse o método Elbow para escolha do número de fatores? Poderia explicar novamente?

Isso mesmo, Augusto. O professor fez essa comparação na aula.

Alex Virgilio

Prof, quais são as implicações de uma PC ser positiva ou negativa?

Os autovalores derivam da matriz de correlação que podem ter variáveis com correlação forte, mas negativa.