

PERGUNTAS E RESPOSTAS – MBA EM DATA SCIENCE E ANALYTICS

Disciplina: Supervised Machine Learning: Análise de Regressão Simples e Múltipla IV

Data: 17/08/2021

Ricardo Rocha Pavan da Silva

a função `lm` do R já transforma em dummies as variáveis categóricas. Mesmo assim, é mais seguro transformar em dummies antes de rodar essa função?

Resposta: Há vários caminhos para se chegar em determinado objetivo em softwares de programação orientados a objetos, como o R. Acreditamos no efeito didático da função `dummy_columns` do pacote `fastDummiesSim`, já que permite que o analista não tenha dúvidas do que está sendo feito e considerado, inclusive no que diz respeito à definição das categorias de referência.

Vanessa Magnani

Prof. qual a diferença entre o teste Breusch-Godfrey e o Breusch-Pagan?

Resposta: Breusch-Pagan é para diagnóstico de heterocedasticidade, como vimos nas aulas. Já Breusch-Godfrey é para diagnóstico de autocorrelação dos resíduos em séries temporais.

Felipe de Souza Mendes e Silva

Prof. como faríamos para utilizar para uma análise os métodos PCA e Regressão Linear? Gostaria de alguns exemplos práticos. obrigado

Resposta: Podemos gerar os fatores (F1 e F2), por exemplo, e colocá-los como variáveis preditoras em determinado modelo OLS. Você irá notar que a multicolinearidade será nula entre estas duas variáveis, já que a correlação entre elas, pela construção da PCA, será zero. O problema de se fazer isso é a perda de capacidade inferencial para outras observações, já que a técnica de PCA é não supervisionada e, portanto, deverá ser novamente elaborada quando da inclusão de novas observações no dataset, para que sejam gerados novos fatores antes da estimação de outro modelo de regressão.

Felipe Francisco Nusda

no diagnóstico de multicolinearidade, essa matriz parece Newton-Raphson, é isso mesmo?

Resposta: Sim, a lógica é a mesma, já que o objetivo é a estimação de raízes de uma função.

Thiago Colette Vegi

Nas linguagens de programação data normalmente é representado como a quantidade de milissegundo depois de 1/1/1970. Não seria melhor usar esta referência numérica ao invés de transformar em dummies?

Resposta: Depende do objetivo: entender o período como evolução temporal, ou tentar apenas capturar, por exemplo, sazonalidade para determinado período específico (neste caso, o período pode ser entendido como categórico).

Diego Barbosa Batista

Neste gráfico criado, conseguimos inserir o percentual de correlação?

Resposta: No modo como está estruturado, não. Mas os outros gráficos análogos, como aquele gerado pelo pacote PerformanceAnalytics traz esta informação.

Anna Rita Moraes de Souza Aguirre Mazo

e se um ponto no zero fosse incluído nas observações? em vez de passar o modelo pelo zero?

Resposta: Aí, sim, poderia ser elaborada uma interpolação a partir da estimação dos novos parâmetros alfa e beta.

Carlos Henrique de Oliveira

Quando o R quadrado do box-cox fica menor que antes da transformada, devo considerar que o modelo linear é mais apropriado? Ou devemos utilizar o maior p-valor do Shapiro-Francia?

Resposta: Devemos olhar o modelo com o maior p-valor do teste de Shapiro-Francia.

Fairus Manfro

box-cox e cox-score são a mesma coisa?

Resposta: Não. Cox score são os scores gerados para cada observação a partir da estimação de um modelo de análise de sobrevivência de riscos proporcionais de Cox.

Heberson Barbosa Nicoli

Professor, quando executou o dummy, vi que o p-value de algumas UFs não passaram. Não é para fazer stepwise de dummy? E se tiver outras variáveis além da dummy?

Resposta: Sim, deve-se elaborar o procedimento Stepwise após a geração de todas as dummies, para se chegar ao modelo final.

Pedro Henrique Esteves Trindade

Como correr comparações múltiplas no pós-teste com as variáveis Dummy?

Resposta: Deve-se elaborar o procedimento Stepwise após a geração de todas as dummies, para se chegar ao modelo final. A escolha do melhor modelo preditivo estimado por OLS, verificado o diagnóstico de normalidade dos resíduos, pode ser feita a partir da comparação do R^2 ajustado.

Juliano Sartorelli Dias

se funcionar, não precisaria usar o predict.

Resposta: Acredito que sim.

Pedro Henrique Esteves Trindade

Quando uma das dependentes for 'sexo' eu preciso transformá-las em Dummy? No exemplo do pacote lme4 na função lmer (modelos mistos) basta apenas a variável ser entendida como factor. Fiquei confuso!

Resposta: No pacote lme4, conforme veremos nas aulas de modelagem multinível, as variáveis qualitativas podem ser entendidas como contexto e, portanto, não precisam diretamente ser transformadas em dummy. Iremos estudar este efeito em profundidade em mais algumas aulas.

Karina P. P. Marques

Poderia comentar sobre interação e confusão? Como interpretar os betas nesse caso?

Resposta: Se uma dummy é interativa (multiplicativa) com outras variável, o beta só terá influência sobre o comportamento de Y nos casos em que a categoria 1 da dummy ocorrer. Caso contrário, não haverá influência da interação sobre a Y.

Samya de Lara Lins de Araujo Pinheiro

No gancho da pergunta da maravilhosa Prof. Ivette, como fica VIF_TOL para um modelo com mais de duas variáveis? testei aqui tem mais valores, como são calculados daí?

Resposta: Mostramos este ponto com códigos ao final da aula. Para três ou mais variáveis X, os VIFs e Tolerances são diferentes para as variáveis preditoras. Isso ocorre pela própria definição de Tolerance ($1 - R^2$) resultante de um modelo auxiliar entre cada variável e as demais (lembrando que $VIF = 1 / \text{Tolerance}$).

Hernandes Matias Junior 124.799.11540