



Como fazer uma análise de regressão linear simples no R

A análise de regressão é normalmente utilizada quando se deseja obter um modelo preditivo do efeito de uma variável sobre outra sendo que essas variáveis devem ser contínuas. Na Regressão Linear Simples têm-se apenas duas variáveis, uma variável preditora (eixo X) e outra resposta (eixo y).



Existem quatro pressupostos básicos para análises de regressão:

✓ **LINEARIDADE** (o modelo linear descreve corretamente a relação funcional entre X e Y)

Se esse pressuposto for violado a estimativa do erro aumentará, já que os valores observados não se aproximam dos valores preditos (local onde passará a reta). Pressuposto fundamental já que essa regressão é um modelo linear.

✓ **VARIÁVEL PREDITORA MEDIDA SEM ERROS**

Garante que a variação observada na análise é devido apenas à sua associação com a variável resposta. Esse pressuposto tem sido ignorado devido à eficiência da solução dos mínimos quadrados.

✓ **NORMALIDADE**

Normalidade dos resíduos é esperada para que não existam tendências e que a estatística F funcione de forma correta.

✓ **VARIÂNCIAS HOMOGÊNEAS**

As variâncias dentro de cada grupo é igual (ou pelo menos aproximadamente) àquela dentro de todos os grupos. Desta forma, cada tratamento contribui de forma igual para a soma dos quadrados.

Se os pressupostos forem atendidos fica mais fácil afirmar que os resultados da análise são devido aos efeitos testados. Além disso, a confiabilidade do teste aumenta, já que se terá certeza que não há tendências nos resultados.

Segue um scrip para a realização da regressão linear simples no R (em verde são os comentários):

```
#Primeiramente, vamos criar dois conjuntos de dados ("x" e "y") apenas para realizar o exemplo
x<-c(1,2,3,4,5,6,7,8,9,10)
y<-c(2,5,8,5,3,9,10, 12, 9, 10)

#A função para regressão é "lm" e não requer pacote estatístico (variavel resposta ~ variavel preditora)
resmodelo<-lm(x~y)

#Sumário dos resultados do modelo
summary(resmodelo)

#Teste para NORMALIDADE (valores de p > 0,05 indicam dados normais)
shapiro.test(rstudent(resmodelo))    ##teste de shapiro wilk (normalidade)

# Análise visual para homogeneidade dos resíduos (visualmente eles devem se distribuir igualmente)
plot(rstudent(resmodelo) ~ fitted(resmodelo), pch = 19)
abline(h = 0, lty = 2)

#Visualização gráfica lty é o tipo da linha 1: linha contínua; 2: linha descontínua
plot(x~y)
abline(resmodelo,lty=2)
```

Ao digitar todas essas linhas de comando, o resultado final deverá ser um R^2 ajustado = 0,55. Para montar a equação é necessário pegar o valor do intercepto e da inclinação: $y = 0,3586 + 0,7043x$. Além disso, por meio do teste de Shapiro Wilk (W) é possível atestar a normalidade dos dados ($p = 0,33$) e através da inspeção visual dos resíduos verifica-se que os mesmos distribuem-se de forma igual.

Um dos principais resultados da regressão é o R^2 , que é o coeficiente de determinação. Esse valor varia de -1 a 1 (repostas positivas e negativas) e indica o grau de associação entre as duas variáveis testadas. Já o parâmetro da inclinação demonstra o quanto a variável resposta varia em função da preditora. Desta forma, uma inclinação igual a zero indicaria a falta de associação entre elas.

Este tutorial foi elaborado por [Herick Soares de Santana](#), aluno de doutorado do programa de pós-graduação em Ecologia de Ambientes Aquáticos Continentais. Gosto bastante de trabalhar com números e estatística

univariada e multivariada (principalmente no ambiente R) e, atualmente, tenho focado minhas pesquisas em dinâmica populacional e crescimento de peixes migradores, ecologia aquática e efeito da construção de hidroelétricas sobre as comunidades aquáticas. Espero que o tutorial possa ajudar. Qualquer dúvida, sugestão ou erro no script é só enviar um email para herick.bio@gmail.com que resolveremos da melhor forma possível.

Download Cleaner for Mac


7 years of professional care for Macs worldwide mackeeper.com



By [Colaboradores](#) | 01-05-2014 | [guia](#) | [15 Comments](#)

Compartilhe esse conteúdo com seus amigos no:



 Anúncios Google

[Regressão](#)

[Linear](#)

[Test](#)

About the Author: [Colaboradores](#)



O posgraduando.com é um espaço na internet para troca de experiências, opiniões, depoimentos, dicas, tutoriais, humor e debates saudáveis sobre o meio acadêmico. Para participar também, leia nossas [instruções aos colaboradores](#).

Postagens Relacionadas



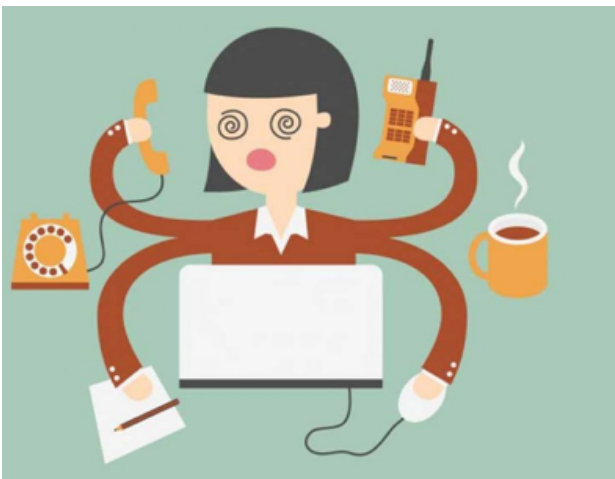
8 dicas para quem está começando (ou não) a pós-graduação

04-05-2018 | [1 Comentário](#)



4 técnicas de oratória para ir bem em concursos e apresentações

24-04-2018 | [1 Comentário](#)



3 técnicas da ciência da felicidade para combater a ansiedade e o estresse

02-04-2018 | [2 Comments](#)



7 motivos que fazem seu artigo ser recusado pelo periódico

19-03-2018 | [2 Comments](#)

15 Comentários



Luciana Barros (RR) 06.05.14 em 10:07- [Responder](#)

Bom dia Heric,

Gostei de seu texto. Tudo que venha ajudar a compreender melhor estatística é válido. Sou aluna do



Herick 06.05.14 em 15:15- [Responder](#)

Olá Luciana,

que bom que ajudou, espero que funcione bem!!! Acredito que seja uma boa forma para desmistificar um pouco daquilo que as pessoas pensam sobre estatística. Não é tão simples, mas acredito que não é tão difícil quanto imaginam.

Divulgue para pessoas que porventura precisarem!!!

Muito obrigado pelo elogio!



Mário Alberto 12.07.14 em 16:44- [Responder](#)

Olá Herick.

Gostei do texto, pois é bastante esclarecedor. Sou estudante do doutoramento em Estatística, mas sempre trabalhei com o SPSS. Só agora é que comecei a familiarizar-me com o R. Gostei de ver o seu trabalho sobre a regressão linear.



Herick 08.12.14 em 19:05- [Responder](#)

Olá Mário!

Muito obrigado!! Não sou “nato” da estatística, mas tento ajudar da forma que posso!!!

Nunca trabalhei com SPSS, mas já li a respeito! Qualquer coisa estamos aí!!

Abraços



Vinicius Giglio 13.02.15 em 22:04- [Responder](#)

As buscas na net me trouxeram até aqui e consegui esclarecer minha dúvida. Muito obrigado!!



Lucas 29.03.15 em 18:18- [Responder](#)

Ola Herick,

Muito legal sua iniciativa ajudando a galera fazer a regressao linear no R. So uma pequena correcao, o coeficiente de determinação varia de 0 a 1, pois ele e o valor ao quadrado do R (coeficiente de correlacao), esse sim varia de -1 a 1.

Continue fazendo esses tutoriais, parabens!

Abraco



Herick 13.04.15 em 21:13- [Responder](#)

Fala Lucas, muito obrigado, tanto pelo elogio, quanto pela correção!!

Falha minha no valor do coeficiente, hehee!!

Abraços!!



Carla 29.07.15 em 16:09- [Responder](#)

Muito boooom! Me ajudou bastante.



Dejair 25.08.15 em 13:50- [Responder](#)

Herick,

muito boa a dica. valeu.



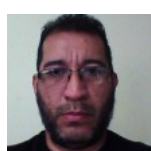
Mauricio 29.09.15 em 21:14- [Responder](#)

Herick, obrigado pelas dicas.

Mas ainda me resta uma dúvida.... nesse caso tu trabalhou apenas com as médias da variável resposta.

Mas caso eu tenha repetições para cada nível do fator “x”, como deveria proceder a análise?

Obrigado



Wilson Cohen 21.11.15 em 20:04- [Responder](#)

Fala Herick, valeu pela dica, bem simples e clara.

Você mostrou que visualmente os resíduos apresentam distribuição igual, mas não fez nenhum teste de homocedasticidade. Eles não são necessários para a Regressão Linear Simples?

Lembro de você na Oficina de Multivariada no R em Olinda no XXI EBI, confere??



Herick 01.02.16 em 15:30- [Responder](#)

Olá pessoal, me desculpe por não responder as mensagens, mas como o blog não é meu, eu não recebo nenhuma notificação sobre o recebimento das mesmas.

Qualquer dúvida e se eu puder ajudar (desde que eu saiba), escrevam para meu email que se encontra ali em cima na descrição.

Valeu



Eduardo 29.11.16 em 15:18- [Responder](#)

Olá Herick.

Muito bom seu texto, enxuto e esclarecedor.

Grande abraço.



Matheus Ferreira 16.03.17 em 15:42- [Responder](#)

Pra que serve esse comando (rstudent)? Eu sempre rodava o shapiro sem esse comando e esta dando resultado direfente, nao achei nada no help falando sobre



Willian Caetano 04.05.17 em 21:33- [Responder](#)

Muito bom o seu texto, mas acredito que vc está equivocado quanto à interpretação do R^2 .

O R^2 não determina o grau de associação das variáveis. Para isso é necessário um teste de Correlação de Pearson (teste paramétrico quando há suposição de normalidade das variáveis X e Y) ou o teste de Correlação de Sperman ou de Kendall (teste não paramétrico quando não há suposição de normalidade das variáveis X e Y). O teste de correlação só deve ser usado quando vc quer testar a correlação entre duas variáveis aleatórias (X e Y) e nem sempre, em um modelo de regressão linear simples a variável explicativa X precisa ser aleatória.

O R^2 representa a probabilidade de que Y é explicada pelo modelo proposto, no exemplo citado seria mais ou menos assim: A probabilidade do valor de Y ser explicada pelo modelo poposto ($y = 0,3586 + 0,7043x$) é de 55%, logo este não é um bom modelo para explicar a associação dos dados.

Deixar Um Comentário

Comentário...

Nome (obrigatório)

E-mail (obrigatório)

Site

POSTAR COMENTÁRIO

- ☐ Notifique-me sobre novos comentários por e-mail.
- ☐ Notifique-me sobre novas publicações por e-mail.

> Blog	> Guia da Pós	> Envie seu tutorial
> Humor	> Dúvidas	> Primeiros Passos
> Tutoriais	> Ciência	> Sábio Orientador
> Notícias	> Docência	> Mapa do site
> Debates	> Colaborações	> Ponto Didática
> Ponto Biologia	> Instituto Agro	