

Proposta de trabalho final

[veja aqui info sobre o fluxograma](#)

Plano A

Função “reordenar”

1. A função que recebe os dados de uma tabela tipo *dados.csv* que normalmente está em tabelas de dados obtidos da internet, em formato excel, ou em uma ordem lida por programas como ArcGIS entre outros.
2. A forma da tabela tem os fatores como cabeçalhos de colunas e linhas, e os dados no centro. Um usuário normal de excel vai cortar os dados e colá-los em uma coluna um abaixo do outro e repetir os nomes ao lado. Agora imagine fazer isso para uma tabela de 1000 linhas por 20 ou mais colunas!.
3. Como podemos observar no exemplo (Fig 1), a função vai reordenar os dados em um data.frame do tipo que é aceita para fazer as análises em R (Fig 2).
4. No Fluxograma é possível observar um resumo do input, os processos e o output da função (não é a estrutura completa!), que finalmente terá um data.frame re-estruturado e salvo em um arquivo na pasta do diretório de trabalho.
5. os argumentos da função também vão ter opções para o usuário entrar os cabeçalhos que deseja para as novas colunas, se deseja transpor os dados antes de ser reestruturados e finalmente se deseja o usuário também terá um resumo dos valores totais por fatores e máximos.

Nota (Esta é a função que realmente desejo fazer pois é muito útil não só para meus dados que vem nesse formato senão para qualquer usuário que tenha a mesma situação e prefira fazer isso facilmente em uma função que em um script que sempre tenha que ser modificado para ser adaptado a cada tabela)



Fluxograma da função



Olá,

Sua proposta está muito clara! As figuras ajudam muito a entender sua ideia. A função parece ser bem geral. Contudo, achei a função muito simples. Ela pode ser executada em algumas poucas linhas. Seria interessante se você pudesse acrescentar outras tarefas à sua função. Outra opção seria implementar partes da sua função sem utilizar as funções do R; e.g., fazer o resumo dos dados na

unha sem usar a função “summary”.

—[Mauro Sugawara](#)

Olá, Mauro

Obrigada pela resposta, então eu não falei que ia usar a função summary... e vou fazer esses valores do resumo de outro jeito, pois o summary não dá aquilo que eu quero que a função faça, ela também interage com o usuário pois permite colocar o nome e acusa de erro no caso de estiver errado o argumento. Na aula eu falei com o professor Alexandre para saber se ele achava completamente ruim ou muito simples minha ideia e ele achou boa ideia. Eu lhe perguntei isso precisamente porque eu precisava saber se podia ir trabalhando nela com antecedência devido ao pouco tempo que tenho..., e também perguntei para ele se seria bom que fizesse outras coisas mais complexas achou que não (pelo menos isso entendi tomara que não tenha mudado de opinião). Mas mesmo assim, a função não é um simples colar, pegar e salvar o arquivo, ela tem, além dos argumentos, tem if, apply, stop alias eu preciso calcular uma coisa com um for para esse resumo, mais ia perguntar isso mais para frente. O esquema não tem todos os detalhes. Tomara que você reconsidere por favor sua opinião, pois ela é útil e se bem não fez cálculos e simulações complexas.

Olá Nerida, Tive a mesma dúvida do Mauro e sua resposta já foi esclarecedora 😊. Uma coisa para você pensar em implementar é lidar com valores na sua planilha de entrada que não sejam apenas numéricos (poderiam ser caracteres lógicos, fatores, ou caracteres). Seria muito interessante incluir isto na função e, por exemplo, fazer análises exploratórias diferentes para cada caso. — [Sara Mortara](#)

Olá Sara

Já tinha falado pessoalmente com o Mauro depois que ele fez os comentários e ele aceitou a proposta, obrigada pelos comentários recém vi hoje (sexta 6), então ela tem fatores numéricos pois os cabeçalhos são caracteres e isso o usuário vai pegar na internet nas bases de dados públicas, que já vem por default nesse formato ex: ibge, centro vigilância de saúde etc., é um arquivo externo que você não tem como obter ele de outra forma, mas no formato original você não vai conseguir analisar bem esses dados, senão for modificada a ordem. a ideia é essa você tem um arquivo chato para analisar no R que pegou de um X base de dados nesse formato que alguém X fez. O que você propõe segundo entendo é como se a tabela fosse feita por você mesmo ou talvez por alguém mas não no formato “padrão” que me levou a idear em como reorganizar tudo isso, e só

precise de fazer análises exploratórios , o que não é a base da função 😞 .., pois já seria outra função tipo função de análises exploratórios que aceite outras classes de valores.

Agora os poucos cálculos no resumo que ela tem é opcional para o usuário, ela fez cálculos simples, totais por fator (cabeçalhos originalmente), os 3 principais máximos com os fatores associados(já não vai ter FOR nessa parte não faz sentido pois é resolvido com um sort() e um head()..), estava pensando também por mínimos e numero de 0 e talvez de NA (associados com seus valores). Já modifiquei também os argumentos que permitem ao usuário introduzir a tabela com separador diferente ao que ele vai salvar depois. agora talvez que faça um histograma optativo... para ver a distribuição dos dados pode ser, e poderia ser optativo para o usuário se quiser, mas não seria encher a função de coisas extras para fazer o código mais longo não é?

Estou ficando com medo, e muito medo a verdade! que finalmente minha função PLANO A (o B não vou fazer)que já tenho quase pronta (ela roda só estou aumentando ou restando alguma coisa de aquele resumo opcional) vocês achem ruim (e ponham pelo mesmo fato uma avaliação ruim ou regular) já que não faz cálculos estatísticos, eu trabalhei muito nela ainda estou mexendo nos detalhes para que fique melhor, se vocês tiveram sugestões de mais detalhes , obrigadão !

~~===== Plano B =====~~ **Função “multimodelo”**

~~1. A função inicialmente pretende ser um gráfico que plote um eixo Y vs vários eixos X. 2. Mas ela foi adaptada para que plote linhas de modelos de regressão linear, em um mesmo gráfico 3 linhas de diferentes cores com 3 modelos de regressão linear (Fig 1). 3. O usuário vai introduzir as variáveis preditoras X e a variável resposta Y que ele deseja plotar. 4. a função vai adicionando cada variável ao modelo mas não calcula interações entre elas. O usuário escolhe a ordem que ele deseja para cada variável. No fluxograma é possível observar um resumo do input, os processos e o output da função. 5. compara os modelos com anova() e também salva o valor R-squared de cada modelo. e os plota com um mtext() no gráfico em uma legenda.~~



Fluxograma da função



Olá,

Fiquei um pouco confuso com o que você quer dizer com “vários eixos X”. Pela sua figura você tem 3 retas, as quais (imagino eu) foram estimadas a partir de 3 conjuntos de dados diferentes, plotadas nos mesmos eixos, certo?

Assim como a proposta anterior, também achei esta função muito simples. Você pode implementá-la facilmente utilizando as funções “lm” e “plot”. Seria interessante acrescentar mais tarefas à função. Talvez fosse interessante juntar as propostas A e B em uma única função. Caso decida fazer isso você vai ter que pensar em uma nova proposta B.

—*Mauro Sugawara*

Olá, sim bom faz duas semanas que eu pedi ajuda precisamente porque não conseguia enfocar bem esta proposta e ela não ficou clara, então eu pensava primeiro em gráficos como aqueles que tem o pessoal que trabalha representando ambientes aquáticos (temp, oxigenio e pH em relação a profundidade)mas depois pensei em fazer alguma coisa com modelos de regressão. Por outro lado pessoalmente acho que juntar as propostas A e B, seria ter uma função que não faz muito sentido só pelo fato de dar mais tarefas a elas e dizer que virou mais complicada, no final nem sempre a pessoa que quer reorganizar a tabela vai querer misturar com outras variáveis que vai ter que agregar de fora ou vai querer fazer modelos lineares, não acha?.

Olá Nerida (e Mauro), Eu não vejo sentido em juntar as duas propostas. Mas concordo com o Mauro de que esta proposta é muito simples. De qualquer modo pra mim também não está claro se você teria apenas novos valores de x. Imagino que uma situação um pouco mais

geral seria ter novos valores de x e y para a mesma variável. E sua função poderia ser capaz de lidar com todas as situações e retornar também os coeficientes do modelo ajustado (só o R^2 não diz muito). De qualquer maneira, recomendo seguir com a primeira proposta! Se precisar executar o plano B, tente pensar em mais desafios e tarefas a implementar na função! — [Sara Mortara](#)

From:
<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2016:alunos:trabalho_final:nerina:propostas



Last update: **2020/08/12 06:04**