

## TRABALHO 1

As questões a seguir referem-se ao estudo de Gross et al (1999) sobre a relação entre financiamento pelo National Institutes of Health e carga de 29 doenças. Os dados da Tabela 1 desde estudo estão no arquivo 3.ex.Funding.sav. Os nomes das variáveis e definições neste arquivo são:

Nome	Definição
disease	Condição ou Doença
id	Número de identificação da doença
dollars	Milhares de dólares de financiamento do NIH por ano
incid	Taxa de incidência da doença por 1000
preval	Taxa de prevalência da doença por 1000
hospdays	Dias de hospitalização em milhares
mort	Taxa de mortalidade pela doença por 1000
yrslost	Anos de vida perdidos em milhares
disabil	Anos de vida perdidos ajustados por incapacidade em milhares

1. Explore a relação entre *dollars* e as outras variáveis independentes (exceto *disease* e *id*). Construa manualmente um modelo que você julga que melhor explica o desfecho. Faça o gráfico dos valores preditos versus os resíduos *studentizados*. Interprete os resultados.
2. Transforme todas as variáveis utilizadas na questão 1 em logaritmo natural. Utilize o método *stepwise* para selecionar os preditores. Faça um resumo dos passos que o algoritmo tomou. Faça o gráfico dos valores preditos versus os resíduos *studentizados* para o modelo final. Compare com a resposta da questão 1.
3. Para o modelo final da questão 2, faça o diagnóstico completo avaliando as suposições sobre os resíduos, multicolinearidade e observações influentes. Interprete os resultados.
4. Faça a regressão entre  $\log(dollars)$  versus  $\log(disabil)$  e  $\log(hospdays)$ . Interprete os resultados.
5. Utilizando o modelo da questão 4, qual é a estimativa da quantidade média de fundos de pesquisa orçados para doenças que causam 1 milhão de dias de hospitalização em um ano e a perda de 1 milhão de anos de vida ajustados por incapacidade? Calcule o intervalo de 95% de confiança para este valor. Também calcule o IC 95% para um financiamento que seria fornecido a uma nova doença que causa 1 milhão de dias de hospitalização em um ano e a perda de 1 milhão de anos de vida ajustados por incapacidade.
6. Crie uma variável *dummy* que assume valor 1 se  $hospdays \geq 1000$  e 0 caso contrário. Faça a regressão entre  $\log(dollars)$  versus  $\log(disabil)$  e a variável *dummy*. Interprete os resultados, escrevendo as equações para cada categoria da variável *dummy*. Teste se é necessário incluir uma interação no modelo.