

Lista nº 5 – Relações entre Variáveis Quantitativas

- 1) Simplifique os dados do arquivo `USGDP.txt` e use sua tabela simplificada para comentar sobre a importância relativa dos vários setores econômicos que contribuem para o Produto Nacional Bruto (PNB) americano em 1994. Comente também sobre quaisquer tendências temporais observadas e quaisquer exceções a essas tendências. Parte do crescimento temporal do PNB é devido simplesmente à inflação.
 - a) O que você deveria observar para detectar mudanças na importância dos diferentes setores econômicos para a economia dos Estados Unidos.
 - b) Aplique os gráficos adequados para investigar as tendências ao longo do tempo nas contribuições dos vários setores econômicos ao PNB norte-americano.
- 2) Há alguns anos, um colecionador desejava comprar um modelo particular de caminhão GMC de algum ano no intervalo 1979-1981. Ele começou a olhar anúncios nos jornais e registrou ANO (year) e PREÇO (price). Os dados obtidos estão disponíveis no arquivo `gmc.txt`. De fato, ele manteve um gráfico de dispersão atualizado. Em outras palavras, cada vez que ele via o anúncio de outro caminhão, adicionava ANO e PREÇO no gráfico.
 - a) Construa um gráfico de dispersão das 23 primeiras observações. O que você pode dizer a respeito da relação entre ANO e PREÇO? *Resp.*: É aproximadamente linear.
 - b) A maioria das pessoas em busca de um modelo 1979-1981 acompanharia apenas os preços de caminhão para aqueles anos. Esse colecionador registrou tudo que viu desde 1975 até 1988. Analise o gráfico e comente porque foi uma boa ideia incluir os anos adicionais? *Resp.*: Devido à tendência, há informação útil nos dados dos outros anos sobre os preços de 1979-1981. Pareceria, a partir dessa tendência que os preços de 1981 são mais altos
 - c) Ajuste uma reta de regressão aos dados para modelar a tendência resumidora de seu gráfico. Qual a interpretação dos parâmetros estimados? Qual seria sua estimativa do preço médio do caminhão em 1980? *Resp.*: $PREÇO = -78.23 + 1.067 ANO$; intercepto: sem interpretação, 1.067 a mais no preço para cada ano mais novo; \$ 7.116,88.
 - d) Há obviamente muita variação nos preços de caminhão mesmo para o mesmo ano. Queremos levar isso em consideração. Desenhe a olho uma envoltória em torno dos pontos. Qual a amplitude de valores você esperaria para os preços de um caminhão 1981? *Resp.*: \$3.600 a \$8.900.
 - e) O vigésimo quarto caminhão foi de 1980 por \$2000. Adicione-o ao gráfico. O caminhão estava em boas condições? O que você faria? *Resp.*: Compre-o!
 - f) Que outras variáveis além de ANO poderiam afetar o preço de um carro ou caminhão? Como você mediria cada uma dessas quantidades? Você pode imaginar maneira pelas quais você poderia adicionar qualquer dessas informações adicionais ao gráfico? *Resp.*:
- 3) Os dados do arquivo `áudio.txt` consistem em resultado de estudo com 27 meninos com idades de 8 a 9 anos no grupo de controle de um estudo maior. As variáveis são escores de ASSOCIAÇÃO auditiva e de MEMÓRIA auditiva fornecidas por um teste denominado ITPA (Illinois Test of Psycholinguistic Ability). Deseja-se explorar a relação, se houver alguma, entre essas duas variáveis.

- a) Faça o gráfico de MEMÓRIA *versus* ASSOCIAÇÃO. Parece haver uma relação? *Resp.*: a partir do gráfico de dispersão parece haver uma tendência crescente que é aproximadamente linear.
- b) Há um valor atípico claro nesses dados. Ele é visível quando você olha a distribuição dos escores de ASSOCIAÇÃO sozinhos? Dos escores de MEMÓRIA sozinhos? *Resp.*: É visível apenas com os escores de MEMÓRIA.
- c) Calcule a correlação entre escore de ASSOCIAÇÃO e escore de MEMÓRIA, tanto com o valor atípico como sem ele. *Resp.*: $r = 0,457$, com o valor atípico e $r = 0,542$, sem o valor atípico.
- d) Quando as pessoas dizem que há uma “correlação significativa”, o que você acha que essa afirmação quer dizer? A que hipótese ela está relacionada? *Resp.*: Significa que a correlação é significativamente diferente de zero e refere-se a $H_0: \rho = 0$, a hipótese de nenhuma relação linear.
- e) Omitindo o valor atípico, verifique se a correlação em (c) é “significante”. *Resp.*: Sem o valor atípico, $p\text{-valor} = 0,004$, indicando a existência de uma relação linear.

[O arquivo áudio.txt é um subconjunto dos dados descritos em SEBER, G. A. F. *Multivariate observations*. New York: Wiley, 1984.]

Fonte: MOORE, D. S.; MCCABE, G. P; DUCKWORTH, W. M.; SCLOVE, S. L. *A Prática da estatística empresarial: como usar dados para tomar decisões*. Rio de Janeiro: LTC, 2006.