

Disciplina: Estatística Aplicada
Professor: Gilberto S. Matos

Período: 2022.1

Aluno(a): _____ Nota: _____

Aluno(a): _____ Nota: _____

Aluno(a): _____ Nota: _____

3^a Avaliação

Análise de Correlação e Regressão Linear Simples

O Conjunto de Dados de Saúde e Desenvolvimento da Criança foi projetado para avaliar fatores relacionados ao baixo peso ao nascer em crianças. O conjunto de dados do software Stata *chdsmetric.dta*¹ contém as seguintes variáveis:

1. *bwtkg* (birth weight, in kg) / (peso ao nascer, em kg)
2. *blengthcm* (birth length, in cm) / (comprimento ao nascer, em cm)
3. *bheadcircm* (birth head circumference, in cm) / (perímetro cefálico de nascimento, em cm)
4. *gestwks* (gestational age, in weeks) / (idade gestacional, em semanas)
5. *mageyrs* (mother's age, in years) / (idade da mãe, em anos)
6. *mheightcm* (mother's height, in cm) / (altura da mãe, em cm)
7. *mweightkg* (mother's prepregnancy weight, in kg) / (peso pré-gestacional da mãe, em kg)
8. *mcig* (mother's number of cigarettes smoked per day) (número de cigarros fumados pela mãe por dia)
9. *fageyrs* (father's age, in years) / (idade do pai, em anos)
10. *fheightcm* (father's height, in cm) / (altura do pai, em cm)
11. *fedyrs* (father's education, in years) / (educação do pai, em anos)
12. *fcig* (father's number of cigarettes smoked per day) / (número de cigarros fumados por dia do pai)
13. *lowbwt* (binary outcome, = 1 for ≤ 2.75 kg, = 0 for > 2.75 kg) / (desfecho binário, = 1 para $\leq 2,75$ kg, = 0 para $> 2,75$ kg)

¹ *Pesquise como ler dados deste formato no R.*

- 1) Neste trabalho, solicito que comecem se concentrando nas associações de três variáveis (*gestwks*, *mageyrs*, *mheightcm*) com o resultado contínuo *bwtkg* (portanto, ignore as variáveis restantes por enquanto). Em trabalhos posteriores pode-se utilizar mais variáveis e assim desenvolver algum modelo multivariável; por exemplo regressão linear múltipla; com o mesmo conjunto de dados.
 - a) Leia os dados e realize algumas análises descritivas apropriadas dessas variáveis para começar a entender as relações entre elas. Por exemplo, use códigos para obter resumos e gráficos de maneira uni e bi-variada com o intuito de explorar a distribuição dos dados bem como a correlação entre as variáveis.
 - b) Tente algumas outras análises descritivas e gráficas que você acha que podem ser interessantes. Com base em seus resultados, quais são seus pensamentos preliminares sobre quais desses três fatores estão associados ao peso ao nascer? Inclua um conjunto limitado de resultados e gráficos relevantes em sua solução (você não precisa incluir tudo!). (Não use o comando *lm* ainda. Além disso, muitas dessas estatísticas descritivas e gráficos podem ser obtidos sob diversos códigos no *R*. Finalmente, veja se existe um comando do tipo *lowess* que cria um gráfico de dispersão junto com uma "curva suave de lowess" que se ajusta aproximadamente aos dados. Discutiremos mais sobre *lowess* em momento oportuno mas, por enquanto, veja-o como uma ferramenta útil à medida que você procura relacionamentos possivelmente não lineares em gráficos de dispersão.)
- 2) Desenvolva análises de Regressão Linear Simples completas considerando separadamente as variáveis independentes *gestwks*, *mageyrs*, e *mheightcm* com o intuito de verificar qual é o fator mais relevante/associado ou que possa melhor prever a variável dependente *bwtkg*.

Procure apresentar interpretações práticas dos resultados obtidos, verifique se as suposições do modelo de regressão linear simples são satisfeitas e comente tudo; realize algumas previsões com as devidas interpretações.

Bom Trabalho!!!