Modelos de Dados Longitudinais

Ficha Prática Nº 2

1. Considere a base de dados da aula da pressão arterial.

   (a) Considere o modelo com erros independentes com o máximo de variáveis explicativas possíveis, mas sem interações. Compare os erros padrão **naive** e **robustos** dos parâmetros da média associados a este modelo. Que conclusões pode retirar?

   (b) Determine os valor de prova (p-value) associados aos erros padrão **robustos**.

   ~~(c)~~ Considere agora o modelo com erros correlacionados de um mesmo sujeito, mas com a mesma estrutura para a média (valor esperado) do modelo. Usando o método de *weighted least squares*, teste diferentes estruturas de correlação (correlação constante, não estruturado, AR1).

   ~~(d)~~ Compare os resultados dos diferentes modelos em (c), indicando qual o modelo que escolheria como o 'melhor' modelo.

2. Considere o caso do modelo linear geral com erros correlacionados (dados longitudinais)

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

onde $Var(\epsilon) = \boldsymbol{V}$. Em particular, consideremos que os erros são homocedásticos com variância constante igual a σ^2 e por isso podemos escrver a matriz $\boldsymbol{V} = \sigma^2\boldsymbol{G}$ onde a matriz $\boldsymbol{G}$ representa uma qualquer matriz geral de pesos. Neste caso o estimador de máxima verosimilhança (*weighted least squares*) é dado por (como provado na aula)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{G}^{-1}\boldsymbol{Y}.$$

Prove que, para qualquer matrix $\boldsymbol{G}$ este é um estimador não enviesado.