**1)** Utilize os dados em WAGE2.wf1 para estimar uma regressão simples que explique o wage mensal (*wage*) em termos do escore do IQ (*IQ*).

(i) Ache o wage médio e o IQ médio da amostra. Qual é o desvio-padrão de *IQ*? (Os escores do IQ são padronizados, de modo que a média é 100, na população, com um desvio-padrão igual a 15.)



(ii) Estime um modelo de regressão simples em que um aumento de um ponto em *IQ* faça com que *wage* varie em uma quantidade constante em dólar. Use esse modelo para achar o aumento predito no wage a partir de um aumento de 15 pontos em *IQ*. O *IQ* explica muito da variação em *wage*?





Substituindo Beta\_1 por 15, beta\_0 por 117 e IQ por 8.30 obtém:



Entretanto, como a regressão obteve um de apenas 9.5%, pode-se afirmar que a variável IQ não explica muito a variação no salário do sujeito.



(iii) Agora, estime um modelo em que cada aumento de um ponto em *IQ* tenha o mesmo efeito percentual sobre *wage*. Se *IQ* aumenta em 15 pontos, qual é o aumento percentual aproximado no *wage* predito?

Não é a mesma coisa do de cima?

**2)** Um problema de interesse das autoridades da saúde (e outras) é determinar os efeitos que fumar durante a gravidez exerce sobre a saúde do recém-nascido. Uma medida da saúde do recém-nascido é o peso de nascimento; um peso de nascimento muito baixo pode atribuir à criança o risco de contrair várias doenças. Como outros fatores que afetam o peso de nascimento, além de fumar cigarros, estão provavelmente correlacionados com o fumo, devemos levar em consideração tais fatores. Por exemplo, uma inca maior geralmente permite acesso a pré-natais melhores, bem como a uma nutrição melhor da mulher. Uma equação que reconhece isso é:

i) Qual é o sinal mais provável de B2?

Positivo. Quanto maior a renda da população, maior será o acesso à exames, melhores cuidados e práticas durante a gravidez.

(ii) Você acha que *cigs* e *faminc* estão, provavelmente, correlacionados? Explique por que a correlação pode ser positiva ou negativa.

Sim. Quanto maior a renda da família, maior deve ser o grau de instrução e, quanto maior o grau de instrução, maior deve ser a chance de saberem que fumar durante a gravidez aumenta os riscos ao feto.

A correlação, desta forma, deve ser negativa, pois, conforme aumentar a renda, diminui a quantidade de cigarros consumidos durante a gravidez.

(iii) Agora, estime a equação com e sem *faminc*, usando os dados em BWGHT.wf1. Relate os resultados na forma de uma equação, incluindo o tamanho da amostra e o *R*quadrado. Discuta seus resultados, dando ênfase ao fato de acrescentar *faminc* mudar ou não substancialmente o efeito estimado de *cigs* sobre *bwghtlbs*.



O Rˆ2 desta equação é d 2.2%, não muito representativo para o modelo como um todo com 1386 observações. A equação fica da seguinte forma:





Controlando o mesmo modelo de regressão com a variável “renda familiar”, há um aumento da explicação do modelo quando comparando o Rˆ2 anterior com o Rˆ2 ajustado, passou de 2.2% para 2.8%, ainda pouco representativo para o modelo como um todo.

Devido à utilização de mais um grau de liberdade, é necessário reduzir uma observação, portanto, o número de pontos observáveis ficou em 1385 e a equação ficou da seguinte forma:



**3)** O arquivo CEOSAL2.wf1 contém dados de 177 diretores, os quais podem ser utilizados para examinar os efeitos do desempenho da firma sobre o wage do CEO.

(i) Estime um modelo que relacione o wage anual às vendas da firma e ao seu valor de mercado. Faça um modelo de elasticidade constante para ambas as variáveis independentes. Escreva os resultados na forma de uma equação.







(ii) Acrescente *profits* ao modelo da parte (i). Por que essa variável não pode ser incluída na forma logarítmica? Você diria que as variáveis de desempenho dessa firma explicam muito da variação nos wages do CEO?





A variável profits não pode ser incluída de forma logarítmica porque a empresa pode ter prejuízos em algum período e não existe logaritmo de número negativo.

Este modelo não explica muito a variação dos salários dos CEOs devido ao fato do Rˆ2 ajustado ser de apenas 29%.

(iii) Acrescente a variável *ceoten* ao modelo da parte (ii). Qual é o retorno percentual estimado para um ano a mais da permanência do CEO no emprego atual, mantendo fixos os outros fatores?



1.2% a mais por ano de permanência

(iv) Ache o coeficiente de correlação amostral entre as variáveis log(*lmktval*) e *profits*. Essas variáveis são altame nte correlacionadas? O que isso diz sobre os estimadores de MQO?



São altamente correlacionadas (78%), provavelmente porque uma empresa lucrativa vai ter maior valor de mercado. Utilizar as duas variáveis irá enfraquecer o modelo proposto pois são redundantes.

4) Ver depois

5) **\*5)** O modelo seguinte pode ser usado para estudar se os gastos de campanha afetam os resultados da eleição:

em que *voteA* é a porcentagem de votos recebidos pelo Candidato A, *lexpendA* e *lexpendB* são os gastos de campanha dos Candidatos A e B, e *prtystrA* é uma medida da força do partido do Candidato A (a porcentagem dos mais recentes votos presidenciais que foram para o partido de A).

1. Qual é a interpretação de Beta1?

Quanto o gasto de campanha do partido A irá influenciar na quantidade total de votos, quando controlados os gastos do partido B e a força do partido A

1. **Em termos dos parâmetros, formule a hipótese nula de que um aumento de 1% nos gastos de A é compensado por um aumento de 1% nos gastos de B. (VERIFICAR)**

****

1. Estime o modelo dado usando os dados em VOTE1.wf1 e relate os resultados na forma usual. Os gastos de A afetam o resultado? E os gastos de B? Você pode usar esses resultados para testar a hipótese da parte (ii)?



Tanto os gastos de A quanto os gastos de B afetam o resultado pois possuem um pvalue menor que 5%.

De acordo com a regressão, quando o partido A aumenta o gasto, o partido B diminui na mesma proporção e não aumenta.

**(iv) Estime um modelo que dê diretamente a estatística *t* para testar a hipótese da parte (ii).O que você conclui? (Use uma alternativa bilateral.)**

**VERIFICAR**

**6)** Consulte o exemplo usado na Seção 4.4. Você usará os dados em TWOYEAR.wf1.

(i) A variável *phsrank* é o percentil no ensino médio da pessoa. (Um número maior é melhor. Por exemplo, 90 significa que sua classificação é melhor que 90 por cento de sua turma de graduação.) Ache o *phsrank* menor, maior e médio da amostra.



(ii) Acrescente *phsrank* à equação (4.26) e informe as estimativas de MQO na forma usual. O *phsrank* é estatisticamente significativo? Quanto vale, em termos salariais, dez pontos percentuais na classificação do ensino médio?





phsrank não é estatisticamente significativo, o pvalue é maior que 5% o que aceita a hipótese nula. 10 classificações no colegial aumenta o salário em 0.03%.

1. Acrescentar *phsrank* a (4.26) mudou substancialmente as conclusões sobre os retornos das faculdades de dois e quatro anos? Explique.

Não. Tanto o R^2 quanto o R^2 ajustado possuem o valor de 22%.

(iv) O conjunto de dados contém uma variável chamada *id*. Explique por que, ao acrescentar *id* à equação (4.17) ou (4.26), você espera que ele seja insignificante. Verifique que ele é insignificante.



ID é insignificante porque esta variável apenas serve para identificar as ocorrências e não possui qualquer valor intrínseco significativo para qualquer regressão.

**7)** O conjunto de dados 401KSUBS.wf1 contém informações sobre a riqueza financeira líquida (*nettfa*), a age do respondente da pesquisa (*age*), a renda familiar anual (*inc*), o tamanho da família (*fsize*), e informações sobre a participação em certos planos de pensão para pessoas dos Estados Unidos. As variáveis de riqueza e inca são registradas em milhares de dólares. Para esta questão, use somente os dados para pessoas solteiras (portanto, *fsize* = 1).

(i) Quantas pessoas solteiras há no conjunto de dados?



(ii) Use MQO para estimar o modelo:

e relate os resultados usando o formato habitual. Esteja seguro de usar somente as pessoas solteiras da amostra. Interprete os coeficientes de inclinação. Há alguma surpresa nas estimativas de inclinação?



Se a pessoa não tiver renda ou idade, ela possui apenas passivos.

(iii) O intercepto da regressão da parte (ii) tem um significado interessante? Explique.

 A pessoa começa com passivo.

1. **Ache o *p*-valor para o teste 􏰟􏰉: 􏰈􏰎 􏰇 1 contra 􏰟􏰠: 􏰈􏰎 􏰡 1. Você rejeita H􏰉 ao nível de 5%?**

(v) Se você fizer uma regressão simples de *nettfa* sobre *inc*, o coeficiente estimado de *inc* é muito diferente do estimado na parte (ii)? Por quê?



O coeficiente não é muito diferente provavelmente porque a renda da pessoa é o maior indicador do que será o patrimônio total.

**\*8)** Use os dados em W AGE1.wf1 para este exercício.

(i) Estime a equação



Salve os resíduos e faça um histograma.



(ii) Repita a parte (i), mas com log(*wage*) como a variável dependente.



(iii) Você diria que a Hipótese RLM.6 está mais próxima de ser satisfeita para o modelo nível-nível ou para o modelo log-nível?

Para o modelo log-nível pois a distribuição está mais parecida com uma normal.

**9)** Na equação (4.42) do Capítulo 4, calcule a estatística *LM* para testar se *motheduc e fatheduc* são conjuntamente significativos. Ao obter os resíduos do modelo restrito, esteja seguro de que o modelo restrito seja estimado usando somente aquelas observações para as quais todas as variáveis do modelo irrestrito estejam disponíveis (veja Exemplo 4.9).