

**FACULDADE DE ECONOMIA,
ADMINISTRAÇÃO E CONTABILIDADE
FEA/USP**

São Paulo - Nov/2017

Exercícios Revisão

Professora: Paula Pereda

EAE0324

Monitores: Bernardo Ostrovski e Igor Barreto

Nesta lista de exercícios iremos treinar todo o conteúdo visto no curso, desde os princípios de estimação por MQO simples, as primeiras hipóteses clássicas e suas violações. Além disso, lembraremos toda a abordagem matricial e, por fim, chegaremos aos Modelos de Probabilidade Linear. Alguns exercícios de Verdadeiro ou Falso são retirados de questões antigas de Exames da ANPEC. Recomenda-se fortemente que essa Lista seja feita tendo em vista a Prova 2 e a Prova de Reavaliação.

As questões de provas da ANPEC estão indicadas antes dos enunciados. Todas nesta Lista são do tipo Verdadeiro ou Falso. Lembre-se que apesar desta Lista ser extra, é importante sempre justificar o motivo da resposta ser V ou F a fim de elevar nosso aprendizado acerca do assunto tratado nas questões.

1 Causalidade versus Correlação

Questão 1 Uma discussão muito importante trata da diferença entre os conceitos de *correlação* e *causalidade*. O primeiro trata apenas de uma dependência linear, ou seja, uma ligação entre dois eventos ou variáveis. Já o segundo é mais forte: trata do efeito de um evento/variável em cima de outro, ou seja, porque X ocorreu, então Y necessariamente ocorrerá. Em Econometria buscamos sempre relações de causalidade entre variáveis e precisamos ter o cuidado de não tirar conclusões equivocadas. As vezes as variáveis são apenas correlacionadas mas não possuem nenhum tipo de causalidade, ou então uma terceira variável que causa as duas simultaneamente. Até mesmo é possível que as duas possuam causalidade uma na outra (simultaneamente). Analise os seguintes exemplos abaixo, corrigindo as afirmações feitas (se achar necessário) e dizendo qual o tipo de relação causal você é capaz de enxergar:

- (a) No dia 19 de Junho desse ano o jornal *Gazeta do Povo* de Curitiba publicou uma notícia com a seguinte manchete: “Saiba de onde ‘surgem’ os vendedores de guarda-chuva de Curitiba”. Sempre que vemos muitos vendedores de guarda-chuva na rua, é praticamente certa de que irá chover. Portanto pode-se concluir que mais vendedores de guarda-chuva na rua aumentam a probabilidade de chover.¹
- (b) Uma professora certa vez classificou os alunos de sua classe em duas categorias: alunos que estudaram ao longo do curso (A) e alunos que só estudaram na véspera das provas (B). Além disso ela perguntou aos alunos o quanto (em litros) de cerveja que eles haviam consumido na semana anterior a prova. Após apurar todas as informações ela constatou que dentro da categoria A, um maior consumo de cerveja na semana anterior a prova estava correlacionado a uma maior nota. Portanto, se conclui que um maior consumo de cerveja próximo a prova aumenta a nota do aluno.

¹fonte: <http://www.gazetadopovo.com.br/curitiba/saiba-de-onde-surgem-os-vendedores-de-guarda-chuva-de-curitiba-7sqruk7d3to1qlug1lmsxclv>

2 Métodos de Estimação

Questão 2 Suponha que você esteja interessado em estudar o efeito do aumento da velocidade máxima permitida (X) nas principais vias de sua cidade em cima do número de acidentes de trânsito (Y). Após uma leitura na bibliografia relacionada, você está confiante de que pode expressar tal relação através do seguinte modelo linear:

$$Y = a + bX + u$$

onde u capta o efeito de outras variáveis, independentes a velocidade máxima permitida, em cima do número de acidentes de trânsito, de tal forma que $\mathbb{E}[u] = 0$ e $\text{Var}[u] = \sigma^2$. Suponha que $Y|X$ possua distribuição Log-Normal (veja a dica ao final da questão).

A partir de uma base de dados com N observações, você passa a estimar os parâmetros de interesse (a e b) por diferentes métodos.

- (a) Estime a e b através do Método dos Mínimos Quadrados.
- (b) Estime a e b através do Método da Máxima Verossimilhança (lembre que você é capaz de encontrar $\mathbb{E}[Y|X]$ e $\text{Var}[Y|X]$).

Dica: Se W tem distribuição Log-Normal com média μ e variância σ^2 , então a função de densidade de probabilidade de W é dada por:

$$f(w; \mu, \sigma) = \begin{cases} \frac{1}{w\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log(w) - \mu)^2}{2\sigma^2}\right], & \text{se } w > 0 \\ 0 & \text{caso contrário} \end{cases}$$

- (c) Qual(is) momento(s) populacional(is) você é capaz de identificar a partir das informações do enunciado? Estime a e b a partir do Método dos Momentos.

3 Introdução a Regressão Simples

Questão 3 Considere o mesmo modelo da questão anterior. Suponha que sua base de dados tenha lhe fornecido a seguinte tabela:

Observação	Nº acidentes de transito	Velocidade máxima permitida (Km/h)
1	12	55
2	15	60
3	17	70
4	24	75

Nota: cada observação representa uma rodovia em um período de um mês

- (a) Estime a e b pelo Método dos Mínimos Quadrados Ordinários (MQO) (Você não precisa demonstrar a fórmula se já a obteve na questão anterior).

- (b) Agora que você estimou os coeficientes, calcule o valor dos resíduos.
- (c) Calcule a qualidade do ajuste, isto é, calcule o R^2 da regressão.
- (d) A partir do que você respondeu nos itens 3 e 4, é possível afirmar que $E[u] = 0$? E que $E[ux] = 0$?

Questão 4 (Anpec 2012) Considere o seguinte modelo de regressão:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

em que β_0 e β_1 são parâmetros estimados pelo método dos mínimos quadrados ordinários e u representa o erro do modelo. Julgue as alternativas:

- (a) A hipótese de que $\mathbb{E}[u|x] = 0$ assegura que a soma dos resíduos da regressão é igual a zero.
- (b) Nesse modelo, a soma dos quadrados totais é igual a soma dos quadrados explicada mais a soma dos quadrados residuais da regressão.
- (c) A covariância amostral entre a variável independente x e os resíduos da regressão é zero se a hipótese de que $\mathbb{E}[u|x] = 0$ for verdadeira.

Dica: A covariância amostral entre duas variáveis, W e Z digamos, a partir de uma amostra com N observações é dada por:

$$Cov(W, Z) = \frac{\sum_{i=1}^N (W_i - \bar{W})(Z_i - \bar{Z})}{N} = \frac{\sum_{i=1}^N (W_i - \bar{W})Z_i}{N} = \frac{\sum_{i=1}^N W_i(Z_i - \bar{Z})}{N}$$

- (d) Neste modelo, a covariância amostral entre os valores preditos pela regressão ($\hat{y}$), e os resíduos da regressão ($\hat{u}$) é sempre igual a zero.

4 Hipóteses Clássicas e Propriedades dos Estimadores: Regressão Simples

Questão 5 Considere o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

tal que u_i é um termo de erro não-observado e ortogonal as variáveis incluídas no modelo. Suponha ainda que $\mathbb{E}[u_i] = 0$ e que $Var[u_i|x_i] = \sigma^2$. Considere também que você estime os coeficientes pelo método de MQO.

- (a) Mostre que $\hat{\beta}_1$ é um estimador não-viesado para β_1 . Enuncie as hipóteses utilizadas para tal.
- (b) Mostre que $\hat{\beta}_0$ é um estimador não-viesado para β_0 . Enuncie as hipóteses utilizadas para tal.
- (c) Calcule $Var[\hat{\beta}_1|x]$ como função de σ^2 .

Questão 6 (Anpec 2011) Considere o seguinte modelo de regressão:

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Suponha que x_i não é estocástico e que $\mathbb{E}[u_i] = 0$, $\mathbb{E}[u_i^2] = \sigma^2$, $\mathbb{E}[u_i, u_j] = 0$ para todo $i \neq j$. Considere os dois estimadores alternativos de β_1 :

$$b_1 = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}; \hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

É correto afirmar que:

- (a) b_1 em geral é um estimador não viesado de β_1 .
- (b) $\hat{\beta}_1$ é um estimador não viesado de β_1 se e somente se $\beta_0 = 0$.
- (c) $\hat{\beta}_1$ é mais eficiente que b_1 se $\beta_0 = 0$.
- (d) b_1 é um estimador não viesado de β_1 se, para qualquer amostra de tamanho N , $\bar{x} = 0$.
- (e) b_1 é um estimador não viesado de β_1 se, para qualquer amostra de tamanho N , $\bar{y} = 0$.

Questão 7 Suponha um modelo de regressão simples, onde

$$y_i = \alpha + \beta x_i + u_i,$$

sendo u_i um termo de erro não observável. Explique por que cada um dos problemas abaixo causa ou não viés no estimador do coeficiente de inclinação por MQO:

- (a) Heterocedasticidade.
- (b) Omissão de Variáveis.
- (c) Erro de Amostragem.
- (d) Baixa correlação entre variável independente e variável explicativa.
- (e) Correlação serial entre os erros.

5 Hipóteses Clássicas e Propriedades dos Estimadores: Regressão Múltipla

Questão 8 Suponha que você esteja interessado no seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

onde ε_i é um termo de erro não correlacionado com x_{1i} e x_{2i} . Como você prestou atenção na última aula de econometria, achou melhor escrever no formato matricial:

$$Y = X\beta + \varepsilon,$$

onde Y é um vetor com os valores da variável explicada para cada aluno e X inclui os valores das variáveis independentes. O vetor ε contém os termos de erro para cada observação. Assim,

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} \end{bmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

- (a) Encontre $\hat{\beta}$ no formato matricial.
- (b) Suponha que você tenha as seguintes informações após usar sua amostra:

$$X'X = \begin{pmatrix} 33 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 60 \end{pmatrix}, X'y = \begin{pmatrix} 132 \\ 24 \\ 92 \end{pmatrix}$$

Obtenha as estimativas de MQO do vetor $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$.

- (c) Calcule $\bar{y}$

Questão 9 Suponha que você esteja interessado em pesquisar os maiores fatores que influenciam o aprendizado do aluno em uma universidade. Primeiramente, você imagina um modelo do tipo:

$$Nota_i = \beta_0 + \beta_1 Estudo_i + \beta_2 Dist_i + \beta_3 Monit_i + u_i, \quad (1)$$

onde $Nota_i$ é a nota (de 0 a 10) do aluno i em Econometria (medida em pontos); $Estudo_i$ é o quanto o aluno i estudou ao longo do semestre (em minutos); $Dist_i$ é a distância (em quilômetros) da residência do aluno i para a universidade em que estuda; $Monit_i$ é o quanto (em minutos) de monitorias o aluno i assistiu; e, por fim, u_i são outros fatores não observados que influenciam a nota do aluno, independentes as variáveis incluídas no modelo. Além disso, suponha que, condicional as variáveis independentes, o erro possui variância constante. Com sua base de 5 dados, você obteve estimativas para os coeficientes rodando um MQO no *Stata*.

- (a) Demonstre que o vetor de estimadores de β feito por MQO é não viesado para estimar β .
- (b) Obtenha uma expressão para $Var[\hat{\beta}_{MQO}|X]$, colocando $Var[U|X]$ como $\sigma^2 I_5$.
- (c) Demonstre que $\hat{\beta}_{MQO}$ é consistente.
- (d) Enuncie as hipóteses clássicas necessárias dentro do arcabouço matricial de regressão múltipla. Indique quais são necessárias para a estimação por MQO, quais são necessária para mostrar que $\hat{\beta}_{MQO}$ é não-viesado e quais são necessárias para mostrar o valor de $Var[\hat{\beta}_{MQO}|X]$ obtido por você no item b

(Note que nem toda(s) hipótese(s) utilizada(s) no Modelo de Regressão Linear Simples é(são) necessária(s), assim como alguma(s) a mais pode(m) ser necessária(s)).

Questão 10 (Anpec 2008) Considere a regressão múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

cujos parâmetros são estimados via MQO. Julgue as afirmativas:

- (a) Se $\mathbb{E}(u|x_1, x_2, x_3) = 0$ e o modelo não é perfeitamente colinear, então os estimadores não são viesados.
- (b) Se o $R^2 = 1$, então y é uma combinação linear de x_1, x_2 e x_3 .
- (c) Se o modelo satisfaz as hipóteses do teorema de Gauss-Markov, então $\hat{\beta}_1$ é o estimador linear não viesado de β_1 com menor variância possível.
- (d) Se omitirmos x_3 da regressão, os estimadores de β_0, β_1 e β_2 serão viesados.
- (e) Se omitirmos x_3 da regressão, os estimadores de β_0, β_1 e β_2 podem ser viesados.

Questão 11 (Anpec 2012) Suponha que um pesquisador esteja interessado em investigar os determinantes da delinquência juvenil e tenha acesso aos seguintes dados provenientes de 100 cidades de um dado país: A , o número de internações por 1000 adolescentes; P , o número de residências por 1000 domicílios na cidade com renda abaixo da linha da pobreza; S , o número de residências por 1000 domicílios na cidade com apenas um dos pais. Ele estima a seguinte regressão:

$$A = \beta_0 + \beta_1 P + \beta_2 S + u,$$

em que u é o termo de erro que satisfaz todas as hipóteses usuais do modelo de regressão. A correlação populacional entre P e S é 0,96, isto é, $\text{Corr}(P, S) = 0,96$. Julgue as seguintes afirmativas:

- (a) A alta correlação populacional entre P e S dará origem ao problema conhecido como multicolinearidade.
- (b) Multicolinearidade não torna viesados os estimadores de mínimos quadrados ordinários dos coeficientes, mas faz com que eles sejam inconsistentes.
- (c) As estimativas dos desvios padrões serão viesadas e provavelmente subestimarão os valores verdadeiros.
- (d) Na presença de multicolinearidade, os testes t e F não são mais válidos.

Questão 12 (Anpec 2017) Considere o modelo de regressão linear:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i,$$

$i = 1, 2, \dots, N$, em que $\mathbb{E}(u_i|x_{i1}, x_{i2}) = 0$. Com base nesse modelo, é correto afirmar:

- (a) A hipótese $\mathbb{E}(u_i|x_{i1}, x_{i2}) = 0$ não é necessária para que o estimador de Mínimos Quadrados Ordinários (MQO) de β_1 seja consistente.
- (b) Se $\text{Var}(u_i|x_{i1}, x_{i2}) = \sigma^2$, o estimador de MQO de β_1 tem distribuição normal.
- (c) Se $\text{Var}(u_i|x_{i1}, x_{i2}) = \sigma^2 x_{i1}$, o estimador de MQO de β_1 é tendencioso.
- (d) Se a correlação entre x_{i1} e x_{i2} é igual a 0,95, o estimador de MQO de β_1 não é eficiente.
- (e) Suponha que os parâmetros do modelo tenham sido estimados por MQO. Se $\text{Var}(u_i|x_{i1}, x_{i2}) = \sigma^2 x_{i1}$, a estatística t não é válida para testar a significância dos parâmetros do modelo.

6 Especificações: Formas Funcionais

Questão 13 Suponha que você esteja interessado em pesquisar os maiores fatores que influenciam o aprendizado do aluno em uma universidade. Primeiramente, você imagina um modelo do tipo:

$$\text{Nota}_i = \beta_0 + \beta_1 \text{Estudo}_i + u_i, \quad (2)$$

onde Nota_i é a nota (de 0 a 10) do aluno i em Econometria (medida em pontos), Estudo_i é o quanto o aluno i estudou ao longo do semestre (em minutos) e, por fim, u_i são outros fatores não observados que influenciam a nota do aluno. Com sua base de dados, você obteve o seguinte resultado rodando um MQO no Stata:

$$\widehat{\text{Nota}_i} = 2 + 0,008 \text{Estudo}_i \quad (3)$$

- (a) Interprete os coeficientes estimados.

Suponha agora que ocorra uma mudança na variável Nota_i e que esta passe a variar de 0 a 100 pontos. Portanto, o modelo a ser estimado agora é:

$$\text{Nota}^* = \beta_0^* + \beta_1^* \text{Estudo} + u^*,$$

onde $\text{Nota}^* = \text{Nota}_i \times 10$.

- (b) Quais seriam os estimadores para β_0^* e β_1^* ? Demonstre, colocando os resultados em função de $\hat{\beta}_0$ e $\hat{\beta}_1$.
- (c) Quanto vale $\hat{u}^*$? Demonstre, colocando o resultado em função de $\hat{u}$.

Suponha agora que a nota dos alunos continue medida de 0 a 10 mas que você também passe a considerar o tempo em que os alunos estavam na aula e nas monitorias. Como os alunos dessa faculdade assistem a todas as aulas e monitorias, você considerou um

acréscimo de 2000 minutos para cada aluno na variável $Estudo_i$. Assim, agora o modelo a ser estimado é:

$$Nota = \beta_0^* + \beta_1^* Estudo^* + u^*,$$

onde $Estudo^* = Estudo + 2000$.

- (d) Quais seriam os estimadores para β_0^* e β_1^* ? Demonstre, colocando os resultados em função de $\hat{\beta}_0$ e $\hat{\beta}_1$.
- (e) Quanto vale R^{2*} ? Demonstre, colocando o resultado em função de R^2 .

Depois de ler um pouco mais sobre o tema, você passa a acreditar que os ganhos percentuais na nota com os estudos aumentem de forma constante. Portanto, sua regressão estimada foi:

$$\widehat{\log(Nota_i)} = 7,29 + 0,002Estudo_i$$

- (f) Interprete o coeficiente de $Estudo$.
- (g) O fato de não termos mais uma relação linear entre y e x afeta RLM 1?

Por fim, você passa a ficar desconfiado de que a relação entre $Nota$ e $Estudo$ não seja linear. Sua hipótese é de que o modelo abaixo expresse melhor essa relação:

$$Nota_i = \delta_0 + \delta_1 Estudo_i + \delta_2 Estudo_i^2 + \epsilon_i.$$

Rodando um MQO pelo Stata, você obteve:

$$\widehat{Nota_i} = 2 + 0,01Estudo_i - 0,000005Estudo_i^2. \quad (4)$$

- (h) A partir do resultado do item anterior, o que você pode dizer da relação das variáveis $Nota$ e $Estudo$? Pense no resultado previsto pelo modelo em (3) para um aluno que estudou 100 minutos e outro que estudou 1000 minutos. Veja o que ocorre quando usamos (4) para prever a nota desses mesmos alunos.
- (i) Qual o nível ótimo de estudo para o aluno maximizar sua nota?
- (j) Esboce um gráfico de $\mathbb{E}[Nota|Estudo]$ e $Estudo$.

7 Especificações: Variáveis Binárias

Questão 14 Suponha que você ainda esteja interessado em pesquisar os maiores fatores que influenciam o aprendizado do aluno em uma universidade. Agora você deseja entender como a participação dos alunos nas monitorias é um fator importante para explicar a nota deles. Porém, desconfiado de que muitos deles apenas assinam a lista de presença e vão embora, você usa uma variável que difere se o aluno realmente levou a sério a monitoria, através do seguinte modelo:

$$Nota_i = \beta_0 + \beta_1 Estudo_i + \beta_2 S\acute{e}rio_i + u_i, \quad (5)$$

onde $Nota_i$ é a nota (de 0 a 10) do aluno i em Econometria (medida em pontos), $Estudo_i$ é o quanto o aluno i estudou ao longo do semestre (em minutos) e, por fim, u_i são outros fatores não observados que influenciam a nota do aluno. A variável $S\acute{e}rio_i$ é definida da seguinte forma:

$$S\acute{e}rio_i = \begin{cases} 1, & \text{se o aluno } i \text{ ficou em 90\% do tempo nas monitorias.} \\ 0, & \text{caso contrário.} \end{cases}$$

- (a) Interprete o parâmetro β_2 do modelo (5) .
- (b) Qual seria o problema em incluirmos no modelo (5) uma variável binária que assuma 1 quando o aluno ficou em menos de 90% do tempo nas monitorias e 0 caso contrário?

Agora você desconfia que o retorno médio de um minuto a mais de estudo é diferente para as duas categorias. Portanto, você passa a utilizar o seguinte modelo:

$$Nota_i = \delta_0 + \delta_1 Estudo_i + \delta_2 S\acute{e}rio_i + \delta_3 (S\acute{e}rio_i \times Estudo_i) + u_i$$

Após rodar sua regressão no *Stata*, você obteve as seguintes estimativas:

$$\widehat{Nota_i} = 1,5 + 0,004 Estudo_i + 1,5 S\acute{e}rio_i + 0,008 (S\acute{e}rio_i \times Estudo_i)$$

- (c) Qual o retorno médio de um minuto a mais de estudo para um aluno que levou a sério as monitorias? E para um aluno que não levou a sério?
- (d) O sinal de $\hat{\delta}_3$ faz sentido? Explique.
- (e) Faça um esboço gráfico do modelo estimado.

8 Distribuição de Probabilidades dos Estimadores

Questão 15 Considere o seguinte modelo de regressão populacional:

$$y = \beta_0 + \beta_1 x + u,$$

onde valem todas as hipóteses de Gauss-Markov. Considere agora que vale também RLM6. Chame de N o tamanho da amostra a ser usada. Além disso, você sabe que

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \sim N(0, 1)$$

e

$$\left(\frac{\hat{\sigma}^2}{\sigma^2} \right) (N - 2) = \frac{SQR}{\sigma^2} \sim \chi_{N-2}^2.$$

Lembre que podemos escrever uma v.a. com distribuição t-student como a razão entre uma normal-padrão e a raiz de uma chi-quadrada sobre seus graus de liberdade (ver questão 8 da Lista 1), isto é:

$$\frac{N(0, 1)}{\sqrt{\chi_L^2/L}} \sim t_L.$$

(a) Conclua agora que

$$\frac{\hat{\beta}_1 - \beta_1}{\text{ep}(\hat{\beta}_1)} \sim t_{N-2}, \tag{6}$$

partindo do fato de que:

$$\frac{\hat{\beta}_1 - \beta_1 / \sqrt{\text{Var}(\hat{\beta}_1)}}{\sqrt{(SQR/\sigma^2)/(N-2)}} = \frac{N(0, 1)}{\sqrt{\chi_{N-2}^2/(N-2)}} \sim t_{N-2}.$$

- (b) Supondo agora que $N \rightarrow \infty$, qual a distribuição de $\hat{\beta}_1$?
- (c) Responda se Verdadeiro ou Falso, justificando: Se a hipótese de normalidade nos erros for comprometida (RLM 6), então a hipótese de homocedasticidade não é mais válida.
- (d) Responda se Verdadeiro ou Falso, justificando: Ao fazermos a hipótese RLM 6, necessariamente estamos assumindo RLM 4 e RLM 5.
- (e) Responda se Verdadeiro ou Falso, justificando: Para chegarmos a distribuição exata de $\hat{\beta}_1$, a hipótese RLM 2 se torna essencial.

9 Inferência

Questão 16 Suponha que você esteja interessado em estudar os impactos de blitz de Lei Seca no número de acidentes de trânsito em algumas cidades do país, durante o período de uma semana. Seu modelo é dado por:

$$acid = \beta_0 + \beta_1 blitz + \beta_2 carro + \varepsilon,$$

onde *acid* é a taxa de acidentes de trânsito por 100.000 habitantes no município, *blitz* é o número de blitz realizadas no município durante a semana, *carro* é a média de carros por habitantes do município e ε é um termo de erro não observado ortogonal as variáveis incluídas no modelo e homocedástico, tal que $\varepsilon|X \sim N(0, 2)$. Sua amostra de 25 observações lhe forneceu:

$$(X'X)^{-1} = \begin{pmatrix} 0,737 & -0,074 & -0,452 \\ -0,074 & 0,010 & -0,002 \\ -0,452 & -0,002 & 1,302 \end{pmatrix} \text{ e } X'y = \begin{pmatrix} 493 \\ 3519 \\ 177 \end{pmatrix}.$$

- (a) Obtenha o vetor $\hat{\beta}$.
- (b) Teste se o efeito de serem feitas blitz no número de mortes por acidentes de trânsito é nulo. Use 7% de significância estatística.
- (c) Calcule o p-valor para a estimativa de β_0 .
- (d) Monte um intervalo de confiança para o coeficiente estimado de *carro*. Use 3% de significância estatística.
- (e) Suponha que você esteja interessado em testar a hipótese de que $\beta_1 + \beta_2 = 2$. Como você montaria esse teste?
- (f) Suponha que você esteja desconfiado que seu modelo está completamente errado, ou seja, o efeito de cada variável incluída (e da constante) é nulo. Como você procederia para testar essa hipótese?
- (g) Como seria possível você testar conjuntamente as hipóteses da letra (b) e da letra (e)?

Questão 17 Ainda no contexto da questão anterior, suponha que o valor de σ^2 não é mais conhecido. Porém, supõe-se que $\hat{\varepsilon}'\hat{\varepsilon} = 52$.

- (a) Teste se o impacto de *blitz* em *acid* é nulo, usando 10% de significância estatística.
- (b) Novamente, calcule o p-valor para a estimativa de β_0 .
- (c) Monte um intervalo de confiança para o coeficiente estimado de *carro*. Use 5% de significância estatística.

Questão 18 (Anpec 2011) Considere as seguintes estimativas obtidas pelo método de mínimos quadrados ordinários para o modelo de regressão abaixo (desvios-padrão entre parênteses):

$$\ln(\text{salário}) = 0,600 + 0,175\text{ Sind} + 0,090\text{ homem} + 0,080\text{ educ} + 0,030\text{ exp} - 0,003\text{ exp}^2 + \hat{u}$$

(0,201)
(0,10)
(0,050)
(0,032)
(0,009)
(0,001)

$$R^2 = 0,36$$

em que *educ* e *exp* denotam, respectivamente, o número de anos de estudo e o número de anos de experiência profissional, *Sind* é uma variável dummy que assume o valor 1 se o trabalhador for sindicalizado e 0 caso contrário e *homem* é uma variável dummy igual a 1 se o trabalhador for do sexo masculino e igual a 0 se for do sexo feminino. O resíduo da regressão é o termo $\hat{u}$. Todas as suposições usuais acerca do modelo de regressão linear clássico são satisfeitas. É correto afirmar que:

- (a) É possível rejeitar, ao nível de significância de 5%, a hipótese nula de que os salários de trabalhadores sindicalizados e não sindicalizados são iguais. A hipótese alternativa é que os trabalhadores sindicalizados ganham mais do que os não sindicalizados.
- (b) É possível rejeitar, ao nível de significância de 5%, a hipótese nula de que os salários de homens e mulheres são iguais. A hipótese alternativa é que os salários de homens e mulheres são diferentes.
- (c) Um ano adicional de experiência eleva em média o salário em 3,0%.
- (d) Se incluirmos um regressor adicional entre as variáveis explicativas, o R^2 não diminuirá.
- (e) Supondo que os erros tenham distribuição normal e que o tamanho da amostra seja $N = 206$, é possível rejeitar, ao nível de significância de 5%, a hipótese de que os coeficientes da regressão, com exceção do intercepto, são simultaneamente iguais a 0. Dica: $F_{(0,95;5,200)} = 2,2592$.

Questão 19 (Anpec 2014) Suponha que queremos estimar como a renda de um indivíduo varia ao longo do ciclo de vida. Queremos testar a teoria de que a renda do indivíduo cresce a partir do momento que ele entra no mercado de trabalho até uma idade média, e depois começa a decrescer até o final do ciclo de vida. Usando dados de uma pesquisa anual para 14.368 trabalhadores, estimamos o seguinte modelo:

$$\ln(\text{renda}) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{homem} + \beta_3 \text{escol} + \beta_4 \text{idade}^2 + \varepsilon,$$

em que $\ln(\text{renda})$ é o logaritmo da renda mensal, *idade* é a idade do indivíduo, *homem* é uma variável binária que é igual 1 se o indivíduo é homem e *escol* representa o número de anos de estudo do indivíduo. Estimando o modelo por Mínimos Quadrados Ordinários, obtemos o seguinte resultado, em que os valores em parênteses abaixo dos coeficientes representam os erros-padrão:

$$\widehat{\ln(\text{renda})} = 49,66 + 0,45\text{idade} + 9,55\text{homem} + 1,10\text{escol} - 0,06\text{idade}^2$$

(1,67)
(0,08)
(0,46)
(0,08)
(0,0009)

Responda se as afirmações abaixo são Verdadeiras ou Falsas (Dica: se uma variável aleatória Z tem distribuição normal padrão, então $P(|Z| > 1,645) = 0,10$ e $P(|Z| > 1,96) = 0,05$):

- (a) Se a teoria descrita acima é verdadeira, esperamos que o sinal de β_1 seja positivo e o sinal de β_4 negativo.
- (b) Neste modelo, o intercepto do modelo para homens é $\beta_0 + \beta_2$ e o do modelo para mulheres é somente β_0 .
- (c) O resultado indica que, mantendo tudo mais constante, o aumento de 1 ano da idade do indivíduo aumenta a sua renda em 45%.
- (d) Temos evidência de que a equação de salários dos homens apresenta um intercepto diferente do modelo para mulheres.
- (e) Com os resultados do modelo, podemos afirmar que idade e educação têm um efeito conjunto significativo no logaritmo do salário, isto é, temos evidência para rejeitar a hipótese nula $H_0 : \beta_2 = 0, \beta_3 = 0$.

10 Heterocedasticidade

Questão 20 Suponha que as variáveis S (índice de sucesso da pessoa i no carnaval), $bloco$ (número de blocos onde a pessoa i esteve presente no carnaval) e $cerveja$ (quantidade de cerveja que a pessoa i ingeriu, em latas, no carnaval) estão relacionadas em uma determinada população de acordo com o modelo linear abaixo:

$$S_i = \beta_0 + \beta_1 bloco_i + \beta_2 cerveja_i + u_i, \quad (7)$$

onde u é um termo de erro estocástico não observado, com média zero e variância dada por:

$$\mathbb{E}[u^2 | bloco, cerveja] = \sigma^2 \sqrt{cerveja}$$

- (a) Suponha que você não saiba qual a forma funcional da variância dos erros. Descreva em poucas palavras algum procedimento para identificá-la.
- (b) Sabendo a forma funcional de $Var[u | bloco, cerveja]$, escreva um novo modelo que contenha um termo de erro homocedástico (**chame-o de u^***), demonstrando a validade da hipótese RLM5.

Em formato matricial, temos esse modelo dado pela seguinte equação:

$$S = X\beta + U,$$

onde $\hat{\beta}$ será o estimador do vetor β por MQO.

- (c) Suponha que você não sabe qual o formato de $Var[u|X]$ e que o tamanho da amostra seja N . Chamando a matriz $E[uu'|X]$ de Ω , descreva alguns elementos dela, assim como sua dimensão. (você só precisa colocar alguns elementos da matriz para não perder muito tempo)
- (d) Demonstre que, nesse caso, $Var[\hat{\beta}|X] = (X'X)^{-1}X'\Omega X(X'X)^{-1}$.
- (e) Escreva qual deve ser a matriz $\hat{\Omega}$ proposta por White, robusta a heterocedasticidade.

Questão 21 (Anpec 2007) Julgue as afirmativas:

- (a) Heterocedasticidade ocorre quando o erro aleatório em um modelo de regressão é correlacionado com uma das variáveis explicativas.
- (b) Na presença de heterocedasticidade, estimadores dos coeficientes populacionais de inclinação por Mínimos Quadrados Ordinários são ineficientes.
- (c) Os testes t e F usuais não são válidos na presença de heterocedasticidade.
- (d) Na presença de heterocedasticidade, estimadores de Mínimos Quadrados Ordinários são não viesados, mas são inconsistentes.

Questão 22 Suponha que você tenha rodado um modelo tal como em (7), obtendo as estimativas $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{u}$ (e como consequência desta última, $\hat{u}^2$). Após, você roda a seguinte regressão a partir de 87 observações:

$$\hat{u}^2 = \gamma_0 + \gamma_1 bloco + \gamma_2 cerveja + \varepsilon,$$

obtendo um $R^2 = 0,1601$.

- (a) Diga qual o nome deste procedimento.
- (b) Diga qual a conclusão do teste (use 5% de significância estatística).

Por fim, suponha que você tenha rodado o seguinte modelo, com uma amostra de 144 observações:

$$\hat{u}^2 = \gamma_0 + \gamma_1 bloco + \gamma_2 cerveja + \gamma_3 bloco^2 + \gamma_4 cerveja^2 + \gamma_5(bloco \times cerveja) + \varepsilon,$$

obtendo $R^2 = 0,083$.

- (c) Diga qual o nome deste procedimento.
- (d) Diga qual a conclusão do teste (use 5% de significância estatística).

11 Autocorrelação

Questão 23 Suponha que você tenha sido contratado por uma empresa para trabalhar na área macroeconômica. Eles te pedem para estimar um modelo relacionando o produto (Y) e juros (r), tal como:

$$Y_t = \beta_0 + \beta_1 r_{t-1} + \varepsilon_t,$$

onde valem as 4 primeiras hipóteses de Gauss-Markov. Um dos analistas da área te diz que está desconfiado de que $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, sendo $|\rho| < 1$. Além disso, $u_t \sim N(0, \sigma^2)$.

- (a) Suponha que ρ seja diferente de zero e seja conhecido. Elabore uma estratégia de estimação eficiente do modelo e que leve em consideração essas informações.
- (b) Descreva como você poderia corrigir esse mesmo problema no caso de não conhecer o verdadeiro valor de ρ .
- (c) Vamos supor ainda que ρ seja desconhecido. Sugira e descreva em detalhes um procedimento que lhe ajude a testar a presença de autocorrelação no modelo acima.

Questão 24 Diga se as afirmações seguintes são verdadeiras ou falsas:

- (a) Quando a autocorrelação está presente, os estimadores de MQO são tendenciosos, bem como ineficientes.
- (b) O teste d de Durbin-Watson pressupõe que o termo de erro, u_t , é homocedástico.
- (c) Um d de Durbin-Watson significativo não implica necessariamente a existência de autocorrelação de primeira ordem.
- (d) Na presença de autocorrelação, a variância e os erros padrão dos valores previstos são ineficientes.