Q1: Sejam X e Y duas variáveis aleatórias, enquanto a, b, c e d são quatro constantes diferentes de zero. Julgue as proposições:

(a) $Var(aX + b) = a^2Var(X)$;
(b) $Var(aX - cY) = aVar(X) + cVar(Y) - 2Cov(X, Y)$;
(c) $Cov(aX+bY, cX+dY) = acVar(X)+bdVar(Y)+(ad+bc)Cov(X, Y)$;
(d) $Corr(aX + b, cX + d) = Corr(X, Y)$;

Q2: Considere o seguinte modelo de regressão:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \ i = 1, ..., n,$$

onde β_0 e β_1 são parâmetros estimados por MQO e u_i representa os resíduos do modelo.

Julgue as seguintes afirmativas:

(a) A hipótese de que $E(Y|X_1) = 0$ assegura que a soma dos resíduos da regressão é igual a zero.
(b) Nesse modelo, a soma dos quadrados total é igual a soma dos quadrados explicada mais a soma dos quadrados dos resíduos da regressão.
(c) A covariância amostral entre a variável independente X_1 e os resíduos da regressão é zero se a hipótese de que $E(Y|X_1) = 0$ for verdadeira.
(d) Neste modelo, a covariância amostral entre os valores preditos pela regressão, $\widehat{Y}_i$, e os resíduos da regressão é sempre igual a zero.
(e) Para verificar quão bom é o ajuste da regressão podemos usar o R^2, que é
igual ao quadrado do coeficiente de correlação entre o Y_i observado e o predito, $\widehat{Y}_i$.

Q3: Considere as seguintes afirmativas sobre os estimadores de Mínimos Quadrados Ordinários em um modelo de regressão múltipla:

(a) A presença de colinearidade imperfeita entre as variáveis explicativas gera estimadores viesados.
(b) Se a hipótese de homocedasticidade for violada, os estimadores de MQO serão viesados.

(c) Assuma que todas as suposições de Gauss Markov foram satisfeitas, então os estimadores de MQO serão os melhores estimadores na classe dos lineares.

(d) Se o valor esperado dos erros estimados do modelo for diferente de zero, então os estimadores de todos os parâmetros, inclusive o intercepto, não serão viesados.

(e) As estimativas de modelos cross-section com a presença de correlação serial geram estimadores viesados.

Q4: Julgue como verdadeiras ou falsas as afirmativas que se seguem:

(a) Na presença de heterocedasticidade dos erros de um modelo de regressão linear, os estimadores de mínimos quadrados ordinários são inconsistentes.

(b) Na presença de erros autocorrelacionados, os estimadores dos parâmetros de um modelo de regressão linear serão viesados.

(c) A condição de exogeneidade das variáveis explicativas é suficiente para que os

estimadores de mínimos quadrados sejam não viesados.

(d) A omissão de uma variável relevante implica que os estimadores dos parâmetros de um modelo de regressão linear serão viesados.

(e) Caso os estimadores dos parâmetros de um modelo de regressão linear sejam consistentes, sob a suposição de normalidade e homocedasticidade dos erros, então esses estimadores de mínimos quadrados ordinários serão idênticos aos obtidos via Máxima Verossimilhança.

Q5: O governo gostaria de estimar o efeito do Programa Saúde da Família sobre a taxa de internação por difteria das crianças entre 0 e 4 anos de idade. Para isso, ele gostaria de estimar o seguinte modelo de regressão:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, ..., n,$$

no qual Y_i é a taxa de internação no município i, X_i é uma variável binária que é igual a 1, se o município i participa do programa, e 0, caso contrário. Usando os dados para o Brasil em 2013, temos os seguintes resultados: $\overline{Y}_1 = 85$, $\overline{Y}_0 = 65$. Neste caso, $\overline{Y}_1$ é a média da taxa de internação para os municípios que participaram do programa e $\overline{Y}_0$ é a média da taxa de internação para os municípios que não participaram. Além disso, 70% dos municípios brasileiros participam do programa. Qual o valor obtido para o coeficiente associado a X_i se o modelo é estimado por MQO?

Q6: Considere o modelo de regressão linear simples

$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, ..., n,$ em que $E(u_i | X_i) = 0$ e $Var(u_i | X_i) = \sigma^2$.

Considere três estimadores de β_1 :

$$b_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2},\ b_1^* = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2},\ b_1^{**} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}$$

Sobre esses estimadores, é correto afirmar:

(a) b_1 é um estimador tendencioso para β_1.
(b) b_1 é um estimador consistente para β_1.
(c) b_1^* é um estimador tendencioso para β_1.
(d) b_1^* é um estimador consistente para β_1.
(e) b_1^{**} é um estimador tendencioso para β_1

Q7: Considere o modelo de regressão linear simples

$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, ..., n,$ em que $E(u_i | X_i) = 0$ e $Var(u_i | X_i) = \sigma^2$.

Considere os seguintes estimadores de β_1 :

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right) Y_i}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right) X_i} \text{ e } \widetilde{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

$\widetilde{\beta}_1$ e $\widehat{\beta}_1$ são estimadores tendenciosos de β_1? Se $\beta_0 = 0$, $\widetilde{\beta}_1$ e $\widehat{\beta}_1$ são estimadores consistentes de β_1? $\widehat{\beta}_1$ é mais eficiente do que $\widetilde{\beta}_1$ neste caso? Se $\beta_0 > 0$, $E\left(\widetilde{\beta}_1\right) > E\left(\widehat{\beta}_1\right)$? $\widetilde{\beta}_1$ é um estimador não viesado de β_1 se, para qualquer tamanho de n, $\overline{Y} = 0$? E se $\overline{X} = 0$?

Q8: Considere o modelo de regressão linear:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i,\ i = 1, ..., n, \text{em que } E\left(u_i \left| X_{1i}, X_{2i}\right.\right) = 0$$

(0) Mostre que a hipótese $E\left(u_i \left| X_{1i}, X_{2i}\right.\right) = 0$ é necessária para que o estimadores de MQO dos parâmetros sejam não viesados e consistentes. Se $E\left(u_i \left| X_{1i}\right.\right) = 0$, isto seria suficinte para que o EMQO de β_1 fosse não viesado e consistente?

(1) Se $Var\left(u_i \left| X_{1i}, X_{2i}\right.\right) = \sigma^2$ os EMQO têm distribuição normal? Eles serão tendenciosos? Explique.

(2) Os EMQO são eficientes se a correlação entre X_{1i} e X_{2i} é igual a 0,95?

(3) Se $R^2 = 0$ então Y é uma combinação linear de X_1 e X_2? Explique.

(4) Suponha que X_2 seja relevante e correlacionada com X_1. Se omitirmos X_2 da regressão, considerando que $E\left(u_i \left| X_{1i}\right.\right) = 0$, os EMQO de β_0 e β_1 serão não viesados?

(5) O R^2 ajustado aumenta ao se incluir uma variável adicional irrelevante?

(6) Podemos afirmar que $\sum_{i=1}^{n}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}\right) = 0$? Se $Z_i = a_0 + a_1 X_{1i} + a_2 X_{2i}$, em que a_0, a_1 e a_2 são constantes, então:

$$\sum_{i=1}^{n} Z_i \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}\right) = 0?$$

(7) Se $\sum_{i=1}^{n} X_{2i} > \sum_{i=1}^{n} X_{1i}$, então $\sum_{i=1}^{n} X_{2i}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}\right) > \sum_{i=1}^{n} X_{1i}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}\right)$?

(8) $\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{X}_1 + \widehat{\beta}_2 \overline{X}_2$?

(9) Sendo $\widehat{u}_i = Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i}$, temos $\sum_{i=1}^{n}\widehat{u}_i = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) - \widehat{\beta}_1 \sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right) - \widehat{\beta}_2 \sum_{i=1}^{n}\left(X_{2i} - \overline{X}_2\right)$

(10) Os estimadores de mínimos quadrados ordinários dos parâmetros são eficientes dentro da classe de estimadores lineares de β_0, β_1 e β_2, mesmo se os erros da regressão não forem normalmente distribuídos?

(11) Se a hipótese de homoscedasticidade for violada, os estimadores de mínimos quadrados ordinários de β_0, β_1 e β_2 serão viesados?

(12) Suponha que β_0, β_1 e β_2 sejam estimados por mínimos quadrados ordinários. Denote por $\widehat{Y}_i$ o valor previsto da regressão para i-ésima observação. Então $\sum_{i=1}^{n}\widehat{Y}_i = \sum_{i=1}^{n} Y_i$?

(13) Se omitirmos X_{2i} da regressão, o estimador de mínimos quadrados ordinários de β_1 será necessariamente inconsistente?

(14) Os estimadores de mínimos quadrados ordinários dos parâmetros não são eficientes se a hipótese de ausência de autocorrelação dos erros for violada.

Q9: (a) O estimador de Mínimos Quadrados do coeficiente angular em uma regressão simples com constante faz uma média ponderada da razão entre desvios da variável explicada de sua média e desvios da variável explicativa de sua média, tendo como ponderador a diferença da explicativa de sua média;

(b) O estimador de Mínimos Quadrados do coeficiente angular em uma regressão simples, com a constante erroneamente omitida, será consistente se a média da variável explicativa for zero e se o erro for independente da explicativa e possuir média zero;

(c) Considere o modelo $Y = a + b_1 X_1 + b_2 X_2 + e$, com exogeneidade estrita das explicativas e erro com média zero. O modelo é estimado em dois estágios: primeiro, uma regressão de Y sobre X_1, salvando-se o resíduo (e_1); segundo, uma regressão de e_1 sobre X_2. A estimativa do coeficiente angular da segunda regressão será igual à estimativa de b_2 na regressão múltipla;

(d) Em um modelo de regressão simples sem constante, em que o coeficiente angular é estimado por Mínimos Quadrados, os resíduos têm média amostral zero por construção;

(e) Se X e Y têm distribuição conjunta Normal, com correlação $r = 0,80$, v será independente de X, em que v é o desvio de Y com relação a sua média condicional, isto é, $v = Y - (\gamma_0 + \gamma_1 X)$, sendo γ_0 e γ_1 parâmetros.

Q10: Considere o seguinte modelo amostral:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + 2X_{2i} + \widehat{u}_i, \;\; i = 1, ..., n, \text{em que } E\left(u_i \left| X_{1i}, X_{2i}\right.\right) = 0.$$

Sabe-se que $Cov\left(X_{1i}, X_{2i}\right) = 40$, $Cov\left(Y_i, X_{1i}\right) = 60$, $Cov\left(Y_i, X_{2i}\right) = 50$ e que $Var\left(X_{1i}\right) = 20$ e $Var\left(X_{2i}\right) = 165$. Qual o valor de $\widehat{\beta}_1$?

Q11: Julgue as seguintes afirmativas:

(a) Colinearidade quase perfeita na matriz de variáveis explicativas causa um viés no
estimador de Mínimos Quadrados Ordinários?
(b) Colinearidade quase perfeita na matriz de variáveis explicativas causa um viés no
estimador da variância do estimador de Mínimos Quadrados Ordinários?
(c) Colinearidade quase perfeita na matriz de variáveis explicativas gera uma perda da
propriedade de eficiência do estimador de Mínimos Quadrados Ordinários?
(d) Colinearidade quase perfeita faz com que o erro-padrão de algumas estimativas dos coeficientes de Mínimos Quadrados Ordinários seja grande?
(e) Colinearidade quase perfeita faz com que o estimador de Mínimos Quadrados Ordinários deixe de ser linear?

Q12: O governo gostaria de estimar o efeito do Programa Saúde da Família sobre a taxa de internação por difteria das crianças entre 0 e 4 anos de idade. Para isso, ele gostaria de estimar o seguinte modelo de regressão:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \; i = 1, ..., n,$$

no qual Y_i é a taxa de internação no município i, X_i é uma variável binária que é igual a 1, se o município i participa do programa, e 0, caso contrário. Usando os dados para o Brasil em 2013, temos os seguintes resultados: $\overline{Y}_1 = 85$, $\overline{Y}_0 = 65$. Neste caso, $\overline{Y}_1$ é a média da taxa de internação para os municípios que participaram do programa e $\overline{Y}_0$ é a média da taxa de internação para os municípios que não participaram. Além disso, 70% dos municípios brasileiros participam do programa. Qual o valor obtido para o coeficiente associado a X_i se o modelo é estimado por MQO?

Q13: Considere o seguinte modelo de regressão:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, ..., n,$$

onde β_0 e β_1 são parâmetros estimados por MQO e u_i representa os resíduos do modelo.

Julgue as seguintes afirmativas:

(a) A hipótese de que $E(Y|X_1) = 0$ assegura que a soma dos resíduos da regressão é igual a zero.

(B) Nesse modelo, a soma dos quadrados total é igual a soma dos quadrados explicada mais a soma dos quadrados dos resíduos da regressão.

(c) A covariância amostral entre a variável independente X_1 e os resíduos da regressão é zero se a hipótese de que $E(Y|X_1) = 0$ for verdadeira.

(d) Neste modelo, a covariância amostral entre os valores preditos pela regressão, $\widehat{Y}_i$, e os resíduos da regressão é sempre igual a zero.

(e) Para verificar quão bom é o ajuste da regressão podemos usar o R^2, que é

igual ao quadrado do coeficiente de correlação entre o Y_i observado e o predito, $\widehat{Y}_i$.

Q14: Usando uma base de dados que têm informação de 65.535 trabalhadores, queremos verificar se existe desigualdade salarial entre os setores da economia. Consideremos que a economia está dividida em 4 setores: indústria, comércio, serviços e construção. Cada um dos trabalhadores está em um dos quatro setores e eles são mutuamente exclusivos. Seja Y_i o salário mensal do trabalhador i e definimos para cada setor uma variável binária que é igual a 1 se o trabalhador está em determinado setor e 0 caso contrário. Estimando um modelo linear de regressão, obtemos o seguinte resultado:

$$\begin{aligned} \widehat{Y}_i &= \underset{(0{,}02)}{4,00} + \underset{(0{,}008)}{0.12}\,\mathrm{educ}_i + \underset{(0{,}0001)}{0,03}\,\mathrm{idade}_i + \underset{(0{,}0005)}{0,40}\,\mathrm{Homem}_i + \underset{(0{,}001)}{0,05}\,\mathrm{DI}_i - \underset{(0{,}003)}{0,15}\,\mathrm{DC}_i \\ &\quad - \underset{(0{,}005)}{0,25}\,\mathrm{DCons}_i \\ R^2 &= 0,83. \end{aligned}$$

em que educ representa o número de anos de estudos de cada trabalhador, idade é medida em anos, Homem é uma variável binária que assume valor igual a 1 se i é homem e 0 caso contrário, DI representa a dummy para a indústria, DC para o comércio e DCons para o setor de construção. Entre parênteses encontra-se o erro padrão.

(a) Com base nos resultados acima, é possível rejeitar ao nível de 5% de significância a hipótese nula de que o salário do setor da indústria é igual ao salário do setor de serviços para trabalhadores com o mesmo nível educacional, a mesma idade e do mesmo sexo. A hipótese alternativa é que os salários nestes setores sejam diferentes.

(b) Com base nos resultados acima, é possível rejeitar ao nível de 5% de significância a hipótese nula de que o salário no setor de construção é igual ao salário no setor de comércio, mantendo educação, idade e sexo fixos. A hipótese alternativa é que os salários nestes setores sejam diferentes.

(c) Com base nos resultados acima, é possível rejeitar ao nível de 5% de significância a hipótese nula de que o salário nos 4 setores da economia são iguais, mantendo constante educação, idade e sexo.

(d) Os resultados do modelo acima permitem testar a hipótese de que o retorno salarial entre homem e mulher é diferente para cada nível educacional, ao nível de 5% de significância.

(e) Com base nos resultados acima, podemos testar a hipótese de que o intercepto do modelo linear de salário em função da educação, idade e setor para homem é diferente do intercepto do mesmo modelo linear de salário para mulher.

Q15: 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. O pesquisador estima a seguinte regressão:

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

em que u é um termo de erro que satisfaz todas as hipóteses usuais do modelo clássico de regressão. Assuma que a correlação populacional entre P e S é 0,96. Julgue as seguintes afirmativas abaixo:

(a) A alta correlação populacional entre P e S dará origem ao problema conhecido como multicolineariedade.

(b) Multicolineariedade 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 multicolineariedade, os testes t e F não são válidos.

(e) Se ao invés de uma alta correlação populacional entre P e S, houvesse uma alta correlação populacional entre A e P ou entre A e S, o problema de multicolineariedade seria ainda pior.

Q16: Um economista deseja avaliar o consumo de carne bovina em 2 estados brasileiros: Rio Grande do Sul (RS) e Rio Grande do Norte (RN). Para tanto, ele seleciona uma amostra de 50.000 unidades de consumo, 35.000 localizadas no Rio Grande do Sul (primeira sub-amostra) e 15.000 no Rio Grande do Norte (segunda sub-amostra). Inicialmente, o economista preferiu trabalhar com as sub-amostras em separado.

Para as duas sub-amostras ele estima a Curva de Engel para o consumo de carne bovina pelo método de Mínimos Quadrados Ordinários. Os resultados das regressões estão abaixo, em que os erros-padrão estão entre parênteses:

$$\begin{aligned} \ln \widehat{(\text{cons})} &= \underset{(0{,}25)}{0,30} + \underset{(0{,}04)}{1,15} \ln(\text{renda}) \text{ - RS } (1) \\ R^2 &= 0,45 \end{aligned}$$

$$\begin{aligned} \ln \widehat{(\text{cons})} &= \underset{(0{,}65)}{0,80} + \underset{(0{,}07)}{0,67} \ln(\text{renda}) \text{ - RN } (2) \\ R^2 &= 0,38. \end{aligned}$$

em que ln(cons) é o logaritmo natural do consumo de carne bovina, em quilogramas, e ln(renda) é o logaritmo natural da renda total do domicílio, em milhares de reais. Todas as suposições usuais acerca do modelo de regressão linear clássico são satisfeitas.

Com base nos resultados acima, e supondo que a amostra é suficientemente grande para que aproximações assintóticas sejam válidas, é correto afirmar que:

(a) Na equação (1), mantendo os preços constantes, com um aumento de 1% na renda das unidades de consumo, o consumo de carne bovina terá um aumento esperado de 1,15%?

(b) De acordo com os resultados das regressões, para um nível de renda igual a R$ 1,00, o consumo de carne no Rio Grande do Sul será maior do que no Rio Grande do Norte, mantendo todas as demais condições constantes?

(c) É possível afirmar, ao nível de significância de 10%, que no Rio Grande do Norte a carne bovina depende exclusivamente do nível de renda, portanto, não é um bem de primeira necessidade?

(d) É possível afirmar, com 1% de significância, que a demanda de carne bovina no estado do Rio Grande do Sul é superior a do Rio Grande do Norte em 67%, para um nível de renda média igual R$ 1.000,00?

Q18: Foram obtidos os seguintes resultados via análise de regressão linear:

$$\begin{aligned}\widehat{Y_i} &= \underset{(5{,}45)}{10,2} - \underset{(-9{,}06)}{125,4X_i} \\ R^2 &= 0,50\end{aligned}$$

Na pressa, o pesquisador se esqueceu de incluir a estatística F nos resultados. Este pesquisador precisa verificar se a regressão é significante. Ajude-o, calculando o valor da estatística F do teste a ser empregado.

Q19: Um pesquisador tem dados de 50 países das seguintes variáveis: N, número médio de jornais comprados durante um ano; Y, PIB per capita medido em dólares. Ele roda a seguinte regressão (desvios padrões entre parênteses, SQR = soma dos quadrados dos resíduos):

$$\begin{aligned}\widehat{N} &= = \underset{(10{,}00)}{25,0} + \underset{(-0{,}010)}{0,020\,Y}, \\ R^2 &= 0.06,\ F = 4,00, \\ SQR &= 4000\end{aligned}$$

Suponha que você rode a mesma regressão com Y medido em reais. Assuma, por simplicidade, que a taxa de câmbio seja quatro reais por dólar.

É correto afirmar que:

(a) A estimativa do coeficiente de Y permanecerá inalterada?
(b) A estimativa do intercepto permanecerá inalterada?
(c) SQR permanecerá inalterado?
(d) A estimativa do desvio padrão do coeficiente de Y permanecerá inalterada?
(e) R^2 permanecerá inalterado?

Q20: Usando uma base de dados que contém informações sobre 65.000 indivíduos, estimamos o retorno da educação usando educação da mãe do indivíduo i como instrumento para educação do indivíduo i, obtendo o seguinte resultado:

$$\begin{aligned} \widehat{Y_i} &= \underset{(220,75)}{-320,89} + \underset{(38,68)}{67,21}X_i + \underset{(1,60)}{5,49}W_i + u_i, \ i = 1, ..., n, \\ R^2 &= 0,46. \end{aligned}$$

no qual Y_i representa a renda mensal do indivíduo i, X_i o número de anos de estudo do indivíduo i, W_i a idade do indivíduo i e Z_i representa a educação da mãe. O termo em parênteses representa o desvio padrão respectivo.

Baseado nas informações acima, julgue as seguintes afirmativas:

(a) Para educação da mãe (Z_i) ser um bom instrumento para educação do filho (X_i), ele deve atender a duas condições: (1) $Cov\left(X_i, Z_i\right) \neq 0$ e (2) $Cov\left(X_i, u_i\right) = 0$?
(b) Com base nos resultados acima, podemos testar a condição (1), isto é, que educação da mãe é correlacionada com a educação do filho?
(c) Com base nos resultados acima, é possível rejeitar a hipótese de que educação da mãe tem um efeito parcial significativo na renda mensal do indivíduo ao nível de significância de 5%?
(d) Suponha que educação do pai seja correlacionada com educação da mãe e tenha uma correlação não-nula com a renda mensal do indivíduo. Neste caso, educação da mãe continua sendo um instrumento válido para a educação do indivíduo?
(e) Se houver uma correlação positiva entre idade e educação da mãe, educação da mãe deixa de ser um instrumento válido para educação do indivíduo.

Q21: Usando uma base de dados que contém informação sobre 437 firmas, estimamos uma função de produção Cobb-Douglas:

$$\begin{aligned} \widehat{Y}_i &= \underset{(0,003)}{0,99} + \underset{(0,035)}{0,64}\, L_i + \underset{(0,023)}{0,45}\, K_i, \;\; i = 1, ..., n, \\ R^2 &= 0,91 \end{aligned}$$

$\widehat{Y}_i$ denota o produto (em logaritmo), L_i representa o insumo trabalho (em logaritmo) e K_i , o insumo capital (em logaritmo). Os números entre parênteses representam o erro-padrão associado a cada coeficiente.

Baseado no resultado acima, julgue as afirmativas:

(a) Considerando que o tamanho da amostra é grande o suficiente para que aproximações assintóticas sejam válidas, é possível rejeitar a hipótese de que o retorno marginal do insumo capital, mantendo o insumo trabalho constante, é igual a zero ao nível de significância de 5%.

(b) Mantendo o capital em dado nível, um aumento de 10 para 11 unidades de trabalho causa um aumento no produto de $0,99 + 0,64 = 1,63$.

(c) Com base nas informações acima, podemos testar a hipótese de retornos constantes de escala, isto é, a hipótese nula de que $\beta_L + \beta_K = 1$.

(d) Com base nos dados acima, construímos um intervalo de 95% de confiança para β_K, [0,41, 0,495]. Supondo que o tamanho da amostra seja grande o suficiente para que aproximações assintóticas sejam válidas, com base neste intervalo, podemos rejeitar a hipótese nula de $\beta_k = \frac{2}{3}$ ao nível de significância de 5%.

(e) Suponha que estimamos uma nova função de produção que relaciona o produto com capital, trabalho e uma medida das condições climáticas enfrentadas por cada firma. Podemos afirmar que R_2 deste modelo será maior que 0,91.

Q22: Usando dados de uma amostra aleatória da população com 80.000 indivíduos, é estimada uma regressão pelo método de Mínimos Quadrados Ordinários. Os resultados dessa regressão são mostrados abaixo, em que os erros-padrão são mostrados entre parênteses:

$$\begin{aligned} ln(\widehat{salário}) &= \underset{(0{,}10)}{0,30} + \underset{(0{,}04)}{0,10}\text{escol} + \underset{(0{,}01)}{0,03}\text{idade} - \underset{(0{,}03)}{0,15}\text{mulher} - \underset{(0{,}05)}{0,05}(\text{mulher x escol}) \\ R^2 &= 0,45, \end{aligned}$$

em que escol representa o número de anos de estudo, idade é a idade do indivíduo em anos e mulher é uma variável dummy igual a 1 se o trabalhador for do sexo feminino e igual a 0 se for do sexo masculino. Todas as suposições usuais acerca do modelo de regressão linear clássico são satisfeitas.

Com base nos resultados acima, e supondo que a amostra é suficientemente grande para que aproximações assintóticas sejam válidas, é correto afirmar que:

(a) É possível rejeitar, ao nível de significância de 10%, a hipótese nula de que o coeficiente associado a variável escol é igual a zero. A hipótese alternativa é a de que o coeficiente associado a variável escol é diferente de zero;

(b) A média dos salários dos homens é maior do que a média dos salários das mulheres;

(c) Cada ano adicional de escolaridade deve elevar os salários em 10%;

(d) O coeficiente de interação (mulher x escol) é significante (hipótese alternativa de que é diferente de zero) ao nível de 10%;

(e) É possível rejeitar, ao nível de significância de 5%, a hipótese nula de que o coeficiente associado a variável idade é igual a zero. A hipótese alternativa é que o coeficiente associado a variável idade é maior do que zero.

Q23: Considere as seguintes estimativas obtidas pelo método de mínimos quadrados ordinários para o modelo de regressão abaixo (desvios-padrões entre parênteses):

$$
\begin{aligned}
\ln(\text{salário}) &= \underset{(0{,}201)}{0,600} + \underset{(0{,}100)}{0,175}\text{sindicato} + \underset{(0{,}050)}{0,090}\text{sexo} + \underset{(0{,}032)}{0,080}\text{educ} + \underset{(0{,}009)}{0,030}\text{exper} \\
&\quad \underset{(0{,}001)}{-0,003}\text{exper}^2 + \widehat{u}, \\
R^2 &= 0,36.
\end{aligned}
$$

em que educ e exper denotam, respectivamente, o número de anos de estudo e o número de anos de experiência profissional, sindicato é uma variável dummy que assume o valor 1 se o trabalhador for sindicalizado e 0 caso contrário e sexo 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 $\widehat{u}$. Todas as suposições usuais acerca do modelo de regressão linear clássico são satisfeitas.

É correto afirmar que:

(0) Supondo que o tamanho da amostra seja grande o suficiente para que aproximações assintóticas sejam válidas, 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.

(1) Supondo que o tamanho da amostra seja grande o suficiente para que aproximações assintóticas sejam válidas, 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.

(2) Um ano adicional de experiência eleva o salário em 3,00%.

(3) Se incluirmos um regressor adicional entre as variáveis explicativas, o R^2 não diminuirá.

(4) Supondo que os erros tenham distribuição normal e que o tamanho da amostra seja 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 zero ($F_{0,95;5,200} = 2.2592$).

Q24: 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:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{1i}^2 + \varepsilon_i,$$

em que Y_i é o logaritmo da renda mensal do indivíduo i, X_{1i} é a idade do indivíduo i, X_{2i} é uma variável binária que é igual 1 se o indivíduo é homem e X_{3i} representa o número de anos de estudo do indivíduo i.

Estimando o modelo por Mínimos Quadrados Ordinários, obtemos o seguinte resultado, em que os valores em parênteses abaixo dos coeficientes representam os erros-padrão:

$$\widehat{Y}_i = \underset{(1{,}67)}{49{,}66} + \underset{(0{,}08)}{0{,}45} X_{1i} + \underset{(0{,}46)}{9{,}55} X_{2i} + \underset{(0{,}08)}{1{,}10} X_{3i} - \underset{(0{,}0009)}{0{,}06} X_{1i}^2.$$

(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 $H_0 : \beta_2 = 0,\ \beta_3 = 0$.

Q25: Considere o modelo de regressão linear múltipla

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$$

no qual

$$\varepsilon_t \left| x_{1t}, x_{2t} \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right), \forall t,\ t = 1, ..., T\right.$$

Por simplicidade, assuma que as variáveis são expressas como desvios em relação às respectivas médias. É correto afirmar que:

(a) Se $\beta_2 \neq 0$ e excluirmos x_{2t} da regressão, o estimador de mínimos quadrados ordinários de β_1 será, em geral, inconsistente.

(b) Suponha que x_{2t} seja medido com erro, isto é, que $x^*_{2t} = x_{2t} + u_{2t,}$ e que $E\left(u_{2t}|x_{1t}, x_{2t}\right) = 0$, $E\left(u_{2t}\varepsilon_t|x_{1t}, x_{2t}\right) = 0$ e $E\left(u^2_{2t}|x_{1t}, x_{2t}\right) = \sigma^2_u$. Se substituirmos x_{2t} por x^*_{2t}, o estimador de mínimos quadrados ordinários de β_1 será inconsistente.

(c) Os estimadores de mínimos quadrados ordinários de β_1 e β_2 serão não viesados, porém não serão eficientes, se y_t for uma variável binária, assumindo apenas dois valores, 0 ou 1, e $\sigma^2 = 1$.

(d) Seja c uma constante diferente de zero. Defina $\widetilde{y}_t = cy_t$, $\widetilde{x}_{1t} = cx_{1t}$ e $\widetilde{x}_{2t} = cx_{2t}$. Os estimadores de mínimos quadrados ordinários (MQO) em uma regressão de $\widetilde{y}_t$ sobre $\widetilde{x}_{1t}$ e $\widetilde{x}_{2t}$ coincidem com os estimadores de MQO em uma regressão de y_t sobre x_{1t} e x_{2t}.

Q26: Considere o seguinte modelo de regressão linear clássico em que as variáveis são expressas como desvios em relação às respectivas médias:

$$y_i = \alpha x_i + u_i, i = 1, ..., n$$

e

$$E[ui] = 0,\ E[ui^2] = \sigma^2,\ E(u_i u_j) = 0 \text{ para todo } i \neq j$$

Suponha, por simplicidade, que x_i é um regressor escalar não estocástico. Propõe-se estimar α através da razão entre as médias amostrais de y_i e x_i:

Calcule a variância de $\overline{\alpha} = \frac{\overline{y}}{\overline{x}}$. Sabe-se que $\sigma^2 = 100$, $n = 100$ e $\overline{x} = 5$).

Q27: (i) Considere a seguinte regressão

$$\overrightarrow{Y} = \mathbf{X}\overrightarrow{\beta} + \overrightarrow{\varepsilon}$$

em que Y, X e são vetores de dimensão $n \times 1$ e β é um escalar. Adicionalmente, suponha que

$$E(\varepsilon|X) = 0$$

e que $\Omega = E\left(\varepsilon^T\varepsilon\,|X\right) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}$, $X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ e $Y = \begin{bmatrix} 0 \\ 7 \\ 5 \\ 0 \\ 0 \end{bmatrix}$

Compute $Var\left(\overrightarrow{\widehat{\beta}}\,|X\right)$.

(ii) Assuma agora que $\Omega = E\left(\varepsilon^T\varepsilon\,|X\right) = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 6 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$, $X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ e

$Y = \begin{bmatrix} 0 \\ 7 \\ 5 \\ 0 \end{bmatrix}$

Compute a estimativa eficiente de $\overrightarrow{\widehat{\beta}}$.

Q28: Considere o seguinte modelo de regressão linear simples: $Y = \beta_0 + \beta_1 X + u$.

Para uma amostra com 32 observações são observados os seguintes resultados:

$$\begin{aligned} \overline{Y} &= 30,\ \overline{X} = 30, \\ \sum_{i=1}^{32}\left(Y_i - \overline{Y}\right)^2 &= 90, \sum_{i=1}^{32}\left(X_i - \overline{X}\right)^2 = 60, \\ \sum_{i=1}^{32}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) &= 30. \end{aligned}$$

A partir dessas informações, obtenha a Soma dos Quadrados dos Resíduos (SQR) correspondente aos estimadores de MQO para esse modelo.

Q29: Considere um modelo de demanda por um produto de consumo, estimado com dados de séries de tempo mensais (t), para várias regiões (i):

$$\begin{aligned}
\widehat{ln(q_{i_t})} &= \underset{(0{,}02)}{-0,27} - \underset{(0{,}15)}{0,83}\ln(p_{i_t}) + \underset{(0{,}12)}{0,33}\left(\ln(p_{i_t}) \times ve_t\right) + \underset{(0{,}75)}{1,15} br_{i_t} + \\
&\quad \underset{(0{,}10)}{0,57}\left(br_{i_t}) \times ve_t\right) \underset{(0{,}88)}{+2,11}\ln(y_{i_t}) \\
R^2 &= 0,24, \ n = 870,
\end{aligned}$$

em que $ln\,(q)$ representa o logaritmo natural da quantidade consumida (em mil litros), $ln\,(p)$ o logaritmo natural do preço do produto por litro, ve uma variável que representa se o mês é de verão (0 em outros casos), br se no período havia uma promoção de compra com brinde gratuito (0 em outros casos) e $ln\,(y)$ o log da renda média dos consumidores. Os desvios padrões estão entre parênteses. O tamanho da amostra valida o uso de resultados assintóticos com pequeno erro.

(a) Se os preços forem convertidos para preços em mil litros, os coeficientes de $ln(p)$ e $ln(p) \times ve$ irão aumentar;

(b) No verão a demanda tende a ser menos preço-elástica, a 5% de significância;

(c) Os coeficientes estimados indicam que os preços são maiores no verão;

(d) O baixo valor do R^2 sugere que as estimativas dos coeficientes são inconsistentes por omissão de explicativas;

(e) Conforme o valor dos coeficientes, é possível concluir que, em média, as vendas são menores no verão.

Q30: Considere o modelo de regressão linear múltipla com regressores estocásticos

$$
\begin{aligned}
Y_t &= \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t,\ t = 1, ..., n, \text{em que } E\left(\varepsilon_t \left| x_{1t}, x_{2t}\right.\right) = 0, \\
Var\left(\varepsilon_t \left| x_{1t}, x_{2t}\right.\right) &= \sigma^2 \text{ e os resíduos são não autocorrelacionados.}
\end{aligned}
$$

Por simplicidade, suponha que as variáveis são expressas como desvios com relação às respectivas médias. É correto afirmar que:

(a) Se $\beta_2 = 0$ e incluirmos x_{2t} na regressão, o estimador de mínimos quadrados ordinários de β_1 será viesado;

(b) Se não conseguirmos observar x_{1t}, mas apenas $x_{1t*} = x_{1t} + u_t$, em que u_t é
um erro de medida, e se substituirmos x_{1t} por x_{1t}^* na regressão, o
estimador de mínimos quadrados ordinários de β_1 ainda assim será consistente;

(c) Se $x_{2t} = y_{t-1}$ e relaxarmos a hipótese de que os erros $\varepsilon_t' s$ não são autocorrelacionados, o estimador de mínimos quadrados ordinários de β_2 será consistente, porém não será eficiente;

(d) A variância do estimador de mínimos quadrados ordinários $\widehat{\beta}_1$diverge para infinito à medida que a correlação entre x_{1t} e x_{2t} aproxima-se de 1;