MI 406 - Regressão
Primeiro semestre de 2021
Lista de Exercícios V.

1. Resolva TODOS os exercícios deixados em sala.

2. Seja $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Encontre a distribuição (exata) de $Y_1 = \ln X$ e $Y_2 = \sqrt{X}$.

3. Para o item 2), usando a expansão em série de Taylor, em torno de λ, até a segunda ordem, calcule $E(Y_i)$ e $\mathcal{V}(Y_i)$, i=1,2.

4. Seja $X \sim \text{gama}(\mu, \phi)$ (veja aqui), $\mu, \phi > 0$. Encontre a distribuição (exata) de $Y_1 = \ln X$ e $Y_2 = \sqrt{X}$.

5. Para o item 4), usando a expansão em série de Taylor, em torno de μ, até a segunda ordem, calcule $E(Y_i)$ e $\mathcal{V}(Y_i)$, i=1,2.

6. Seja $X \sim \text{beta}(\mu, \phi)$ (veja aqui), $\mu \in (0, 1), \phi > 0$. Encontre a distribuição (exata) de $Y = \ln\left(\dfrac{X}{1-X}\right)$.

7. Para o item 6), usando a expansão em série de Taylor, em torno de μ, até a segunda ordem, calcule $E(Y)$ e $\mathcal{V}(Y)$.

8. Compare de forma apropriada (veja aqui o Teorema de Gauss Markov (link), slides de 30 a 38), os estimadores de MQO, MQP e MQG, considerando a matriz de variâncias e covariâncias dos erros, conhecida. Ou seja, as comparações tem de ser feitas elemento a elemento, para combinações lineares dos estimadores e para os vetores como um todo.

9. Reanalize todos os conjuntos de dados vistos em sala e analisados nas listas, quando pertinente, usando as metodologias de MQP e MQG, de forma apropriada. Neste caso, não é necessário fazer o diagnóstico do ajuste do modelo.

10. Reanalize todos os conjuntos de dados vistos em sala e analisados nas listas, quando pertinente, usando alguma transformação apropriada (pode ser apenas uma) para a resposta, comparando todos os resultados, inclusive a verificação da qualidade de ajuste do modelo, com aqueles obtidos nas análises que você já tinha feito.

11. Reanalize todos os conjuntos de dados vistos em sala e analisados nas listas, quando pertinente, (ou seja, essencilamente quando o problema envolver várias covariáveis quantitativas, independentemente de haver ou não covariáives qualitativas) em relação à multicolinariedade. Se algum outro modelo for ajustado, realize todas as análises (inferenciais e de verificação da qualidade do ajuste) com base nele, reduzindo-o, se for possível.

12. Suponha o MRNLH em sua forma matricial, ou seja, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\xi}$. Suponha que as colunas de $\boldsymbol{X}$ são mutuamente ortogonais. Simplique todas as espressões anteriormente desenvolvidas (estimadores, testes de hipótese, resíduos etc) quando possível.

13. Seja X uma variável aleatória contínua (vac) com função de distribuição acumulada (fda) F e Φ a fda de uma N(0,1). Encontre a distribuição de $Y = \Phi(F^{-1}(X))$. Indique como esse resultado pode ser útil para verificar a qualidade de ajuste de um dado modelo de regressão.

14. Seja $Y \sim LN(\alpha, \beta)$. Reparametrize essa distribuição em termos de sua média (μ) e parâmetro de dispersão (ϕ). Com base nisso, proponha um modelo de regressão linear simple apropriado à semelhança dos modelos lineares generalizados.

15. Mostre como estimar os parâmetros do modelo da Questão 14), por máxima verossimilhança (MV), através do algoritmo Escore de Fisher. Pesquise sobre a respectiva distribuição assintótica dos estimadores de MV.

16. No modelo de regressão linear simples com erro nas variáveis sob o enfoque estrutural, visto em sala, prove que:

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim N_2\left[\begin{pmatrix} \beta_0 + \beta_1\mu_W \\ \mu_W \end{pmatrix}; \begin{pmatrix} \beta_1^2\sigma_W^2 + \sigma^2 & \beta_1\sigma_W^2 \\ \beta_1\sigma_W^2 & \sigma_W^2 + \sigma_\epsilon^2 \end{pmatrix}\right].$$

17. Seja $(Y_i, X_i)' \overset{iid}{\sim} N_2\left[\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_X \\ \rho\sigma_Y\sigma_X & \sigma_X^2 \end{pmatrix}\right]$, i=1,2,...,n, $\boldsymbol{\theta} = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)'$e $\sigma_{(.)} = \sqrt{\sigma_{(.)}^2}$. Responda os itens abaixo:

    a) Obtenha a distribuição de $Y|X = x$.
    b) Proponha um modelo de regressão normal linear homocedástico simples ($Y_i = \beta_0 + \beta_1 x_i + \xi_i,\ \xi_i \overset{iid}{\sim} N(0, \sigma^2)$), através da distribuição condicional obtida no item a), relacionando (interpretando) $(\beta_0, \beta_1, \sigma^2)$ com $\boldsymbol{\theta}$.

18. Repita o exercício 16, considerando :

$$(Y_i, X_{1i}, X_{2i}, .., X_{pi})' \overset{iid}{\sim} N_{(p+1)}\left[\begin{pmatrix} \mu_Y \\ \mu_{X_1} \\ \vdots \\ \mu_{X_p} \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_X & \rho\sigma_Y\sigma_X & \dots & \rho\sigma_Y\sigma_X \\ \rho\sigma_Y\sigma_X & \sigma_X^2 & \rho\sigma_Y\sigma_X & \dots & \rho\sigma_Y\sigma_X \\ \rho\sigma_Y\sigma_X & \rho\sigma_Y\sigma_X & \sigma_X^2 & \dots & \rho\sigma_Y\sigma_X \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_Y\sigma_X & \rho\sigma_Y\sigma_X & \rho\sigma_Y\sigma_X & \dots & \sigma_X^2 \end{pmatrix}\right]$$

para o modelo de regressão normal linear homocedástico múltiplo.

19. Repita o exercício 16 mas, agora, determine a distribuição bivariada correspondente ao modelo de regressão normal linear heterocedástico simples: $Y_i = \beta_0 + \beta_1 x_i + \xi_i$, $\xi \overset{ind.}{\sim} N(0, \sigma_i^2)$, $\sigma_i^2 = \sigma^2 x_i$, i=1,2,...,n.

20. Repita o exercício 16 mas, agora, determine a distribuição (n+2)-variada correspondente ao modelo de regressão normal linear homocedástico simples. $Y_i = \beta_0 + \beta_1 x_i + \xi_i$, $\xi \overset{iid}{\sim} N(0, \sigma^2)$, em que $Cov(\xi_i, \xi_{i'}) = \rho$, $i \neq i'$.

21. Para todos os dados analisados (slides das aulas e listas de exercícios), caso o(s) modelo(s) utilizado(s) não tenham se ajustado bem, proponha um único modelo que você entenda ser adequado (eventualmente fora da classe estudada), justificando sua escolha e apresentando interepretações para os parâmetros e /ou funções deles.