**Atividade Final Regressões**

**A atividade deverá ser feita em duplas e enviada para o email** [**danilosouto@gmail.com**](mailto:danilosouto@gmail.com) **com o assunto UP2020-PTI-REG até o dia 30/11**

**Aluno 1:**

**Aluno 2:**

**Informe o método utilizado e justifique. A formulação da resposta faz parte da avaliação. Exibir o código utilizado no R.**

1. **Faça a análise conforme descrito a seguir:** 
   1. defina A renda média per-capta do estado em relação a média de escolaridade do estado (y = renda, x = escolaridade) em outras palavras renda ~ escolaridade) dos dados públicos a seguir:

|  |
| --- |
| #dados para o exercício copie e cole no R  mec <- data.frame(  row.names = c("RR","AC","PA","TO","MA","SE","BA","AL","SP","ES", "SC","PR","GO","DF","AP","RO","AM","PB","RN","PI","PE","CE","RJ", "MG","RS","MT","MS"),    escolaridade = c(5.7, 4.5 ,4.7 ,4.5 ,3.6,4.3 , 4.1 ,3.7 ,6.8 ,5.7 ,6.3 ,6.0 ,5.5 ,8.2 ,6.0 ,4.9 ,5.5 ,3.9 ,4.5 ,3.5 ,4.6 ,4.0 ,7.1 ,5.4 ,6.4,5.4,5.7),    renda = c(685,526 ,536, 520 ,343,462 ,460 ,454 ,1076,722 ,814 ,782 ,689 ,1499,683 ,662 ,627 ,423 ,513 ,383 ,517 ,448 ,970 ,681,800,775,731)  ) |

* 1. veja os gráficos de dispersão

|  |
| --- |
| Código: |

* 1. Exiba as correlações

|  |
| --- |
| Código: |

* 1. plote os histogramas de renda e escolaridade

|  |
| --- |
| Código: |

* 1. teste de normalidade

|  |
| --- |
| Código: |

* 1. faça a regressão linear *lm()*

|  |
| --- |
| Código: |

* 1. quais são os pontos com maior alavancagem?

|  |
| --- |
| Código: |

* 1. qual o coeficiente de determinação (R-squared)?

|  |
| --- |
| Código: |

* 1. verifique os resíduos com a biblioteca *library(hnp)*

|  |
| --- |
| Código: |

* 1. A regressão linear parece ser uma boa escolha? Por que?

|  |
| --- |
| Resposta: |

* 1. Qual das distribuições que estudamos (Binomial, Normal, Poisson, Gama, Gaussiana Inversa) tem uma semelhança com os dados mostrados pelo histograma de renda?

|  |
| --- |
| Resposta: |

* 1. Faça uma regressão *glm()* com essa distribuição

|  |
| --- |
| Código: |

* 1. Agora estime (utilize a função *predict()* ) os valores de renda para os valores de escolaridade utilizando os dois modelos (*lm()* e *glm()*) e plote os gráficos com as curvas. Mostre no mesmo gráfico os valores **observados em preto**, preditos do modelo 1 em vermelho, e preditos no modelo 2 em verde.( utilize as funçoes *plot(), points() e points()*)

|  |
| --- |
| Código: |

* 1. Compare os modelos com a função AIC e informe qual modelo você escolhe e por que?

|  |
| --- |
| Código: |
| Resposta: |

1. Encontre uma base de dados de sua preferência, caso não possua alguma há várias disponíveis no [**https://www.kaggle.com/datasets**](https://www.kaggle.com/datasets) e [**http://dados.gov.br/dataset**](http://dados.gov.br/dataset), e faça uma análise de regressão ou *forecast* sobre alguma informação que lhe pareça importante. Atenção que todas as **análises dos resultados e gráficos** devem ser **exibidas, comentadas e descritas** abaixo.