

Tópicos Especiais em Estatística Aplicada – 2020-2 EPC9

Sejam as 8000 amostras de x e y fornecidos no arquivo epc9dat.txt, $Z=[y \ x]$;

1.1 Plotar x,y, e o gráfico de dispersão (x,y) e comentar os dados.

1.2 Medir a correlação entre x e y e o valor p relacionado ao teste de hipótese de que a correlação é significativamente diferente de zero. Selecionar aleatoriamente 500 pares (x,y) de amostras e refazer o cálculo da correlação, comentando eventuais diferenças e seus efeitos na análise de regressão.

2.1 Calcular a correlação cruzada entre x e y e plotar o gráfico de correlação versus atrasos

2.2 Selecionar aleatoriamente 500 pares (x,y) de amostras e refazer o cálculo da correlação cruzada, comparando com 2.1. Como estes resultados afetam a metodologia para obter um modelo de regressão?

3.1 Selecionar aleatoriamente 100 pares (x,y) de amostras e estimar 100 vezes β_1 e β_0 , fazendo seus histogramas com frequência relativa.

3.2 Calcule a média e o desvio padrão de β_1 e β_0 usando as expressões fornecidas e plote suas pdf com distribuição normal sobre os histogramas, comparando.

4.1 Faça uma regressão linear para 100 pares (x,y) de amostras selecionadas aleatoriamente, calcule β_1 e β_0 e seus intervalos de confiança, comparando com a pdf do item 3.2

4.2 Que parâmetros pode utilizar para concluir que tem um bom modelo de regressão? Quais seus valores?

Gere variáveis aleatórias independentes x e y com 50 amostras, com distribuição normal, média 0 e variância 1. (Não usar os dados fornecidos nesta atividade)

5.1 Mostre quais indicadores quantitativos sobre x e y permitem concluir que não se pode obter um bom modelo de regressão linear.

5.2 Calcule um modelo de regressão linear e mostre que indicadores permitem concluir que o modelo não é bom.