

Análise de Dados I - Fundamentos de estatística

Lab 2 - Tópicos de probabilidade para análise de dados

A base de dados utilizada nesse Lab é “tweets_by_hour.csv”, é um log com quantidade de tweets por dia do Twitter do ano de 2006 à 2010. A tabela é composta pelos atributos:

timestamp - data da observação, no formato aaaa-mm-dd-HH-MM.

tweets - número estimado de tweets observados por hora.

Utilizando os conceitos vistos em sala plote as funções FDA, FDP ou FDM para o atributo **tweets**. Em seguida analise os gráficos gerados em termos de dispersão e concentração dos dados.

Após realizar a análise proposta acima, fiquei interessado em estudar o comportamento dos dados de acordo com o alguns horários do dia. Organize os dados de acordo com a **média** de tweets por hora e trace novamente a FDA e um gráfico de linhas para os novos dados. Foi observado algum comportamento interessante? Como os dados se comportam em termos de dispersão e concentração?

Foi visto em sala que muitos fenômenos na natureza tendem a ter uma distribuição normal, com o auxílio do R gere uma população normal com o mesmo tamanho dos dados da tabela “tweets_by_hour.csv”. Gere gráficos da FDA e histograma para os dois conjuntos de dados e responda com base nos gráficos se os dados utilizados no lab são normais, caso contrário, quais as diferenças nos gráficos que levaram você a essa conclusão.

Detalhe: pode ser interessante quebrar a coluna timestamp em outras várias. Você decide.

Siga as recomendações de entrega que figuram no Piazza!