Universidade de Fortaleza

PPGIA – Introdução a estatística aplicada a ciência dos dados

Elaborado por Prof. Elizabeth S. Furtado em colaboração prof. Furlan Duarte

Exercício de participação

Objetivos:

Analisar a distribuição da amostra onde o fenômeno Tempo agora diz respeito aos valores de 1 fator e 3 níveis

Fazer os testes para análise da variância do fenômeno.

Considere o caso em que o tempo do usuário é a variável independente (calculada) e representando o tempo que o usuário passou em uma determinada conferência virtual, quando fez uso de um dos Meet virtuais, se usando Zoom, Hangout ou Skype.

A hipótese é saber se existe diferença significativa entre os três Meet.

O arquivo segue os seguintes princípios para a realização deste trabalho:

- independência dos dados, quem usou um meet não usou o outro.

- A variável tempo é mais próxima de uma log normalidade, porque a medida que o usuário usa um sistema, ele se torna mais especialista e o tempo, no eixo X tende a diminuir com o tempo; ou ainda tem poucas atividades que levam muito tempo e muitas que levam pouco tempo, afetadas pela experiência do usuário.

Pede-se:

Fazer os testes para análise da variância do fenômeno, realizando as três técnicas dadas a seguir e considerando que:

1. a técnica Shapiro-wilk permite testar a normalidade, para uma amostra pequena

2. a técnica Kolmogorov-Smirnov permite testar a lognormalidade da amostra.

3. a técnica Levene

4. Visualize os dados usando Boxplot, histograma e qqplot.

Senão houver normalidade da amostra, então transforme os dados em uma log normal, depois, verifique como ficaram os dados, e repita os testes dos passos.

Obs:

*In a test statistic: the result expresses in a single number how much my data differ from my*[*null hypothesis*](https://www.spss-tutorials.com/null-hypothesis/)*. So it indicates to what extent the observed scores deviate from a normal distribution.  
Now, if my null hypothesis is true, then this deviation percentage should probably be quite small. That is, a small deviation has a high probability value or p-value.  
Reversely, a huge deviation percentage is very unlikely and suggests that my reaction times don't follow a normal distribution in the entire population. So a****large deviation has a*low*p-value****. As a rule of thumb, we* ***reject the null hypothesis if p < 0.05.***

*So if p < 0.05, we*don't*believe that our variable follows a normal distribution in our population.*