Week 6: Daily Morning Challenge

Day 1: Tuesday 28th January 2020

Question 1: What is the difference between the art and science of data analysis?

Answer: Art is an expression of human ideas, creativity, imagination and intuition. Art is also about beauty, but beauty lies in the eyes of the beholder, so art is very subjective and intangible. It really can't be truly measured, evaluated and compared.

Science, on the other hand, is objective by definition. It can be measured precisely, and its truth indisputably proven. Science is also methodical, predictive, and repeatable.

Like any other art, analysis does require the intangible and somewhat unexplainable factor of intuition to make it successful. The subjectivity of analysis comes through intuition, inquisitiveness, creativity, and the imagination of the people involved, not just through hard numbers and facts. Analysis requires thinking out of the box to come up with solutions that have not been thought of before – or to put it differently: thinking like an artist. The science in analytics is essentially just an enabler, not a solution. It provides a vehicle for translating intuition into results.

Question 2: Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(a)What is the mean of the data? What is the median?

Mean = 29.962963

Median = 25.0

(b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).

Mode = 25, 35

Modality = Multimodal

(c) What is the midrange of the data?

Midrange = 41.5

(d) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?

Q1 = 20.5

Q3 = 35

(e) Give the five-number summary of the data.

Minimum = 13.0

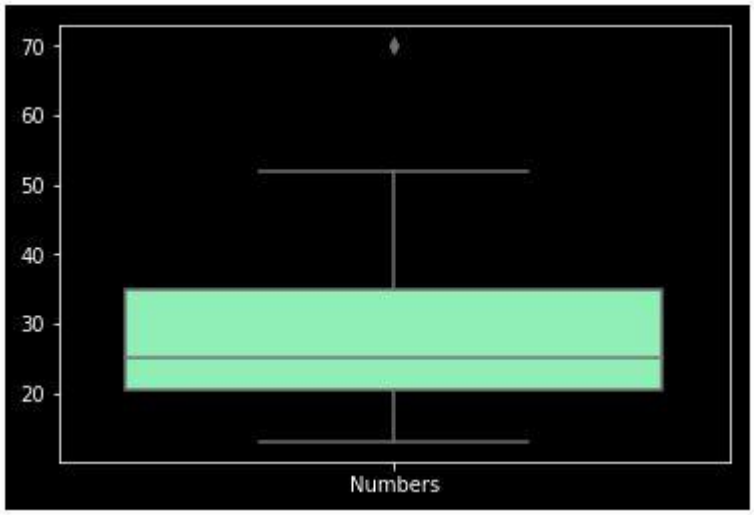
Q1 = 20.5

Median = 25.0

Q3 = 35.0

Maximum = 70.0

(f) Show a boxplot of the data.



(g) How is a quantile–quantile plot different from a quantile plot?

Quantile-Quantile plots are plots of two different quantiles against each other. It is used to assess the distribution of a data set.

Quantile plot is used to show peculiarities of the shape of a sample distribution, which might be symmetrical or skewed.

Question 3: Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

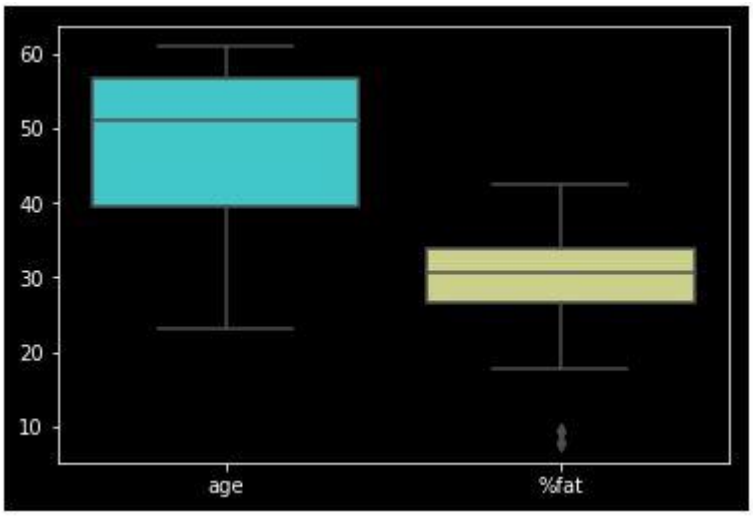
(a) Calculate the mean, median, and standard deviation of age and %fat.

Mean – 46.44444 and 28.783333

Median – 51.0 and 30.7

SD - 13.218624 and 9.254395

(b) Draw the boxplots for age and %fat.



(c) Draw a scatter plot and a q-q plot based on these two variables.