**Assignment 4**

By Harsh Sharma

P2. (Multivariate outlier detection) We will generate 2-dimensioanl data for multivariate outlier test.

Outliers can be detected separately from each attribute, and then take outliers as those data which are outers for both attributes.

(1) Generate 200 normally distributed random numbers for one attribute, x, and generate 200 normally distributed random numbers for another attribute, y, and then combine them for 2-dimensioanl data. For the vilification, display first 5 rows present in the data set.

(2) List the values of any data points which lie beyond the extremes of the whiskers in the boxplot of x

(3) List the values of any data points which lie beyond the extremes of the whiskers in the boxplot of y

(4) Report the values of outliers in both x and y

(5) Draw a 2-dimensional boxplot with the input data and mark the outlier data points with "orange" color.

**Solution:**

The solution is in the R file called ‘P2 Solution.R’. Here are the answers.

**(1)**

x y

1 1.3709584 -2.0009292

2 -0.5646982 0.3337772

3 0.3631284 1.1713251

4 0.6328626 2.0595392

5 0.4042683 -1.3768616

6 -0.1061245 -1.1508556

(2)

[1] "Outliers in x:"

> print(outliers\_x)

[1] -2.656455 -2.993090 2.701891

(3)

> print("Outliers in y:")

[1] "Outliers in y:"

> print(outliers\_y)

[1] -2.69993

(4)

[1] "Outliers in both x and y:"

> print(outliers\_both)

[1] x y

<0 rows> (or 0-length row.names)

There is no common elements between both.

**(5)** For the (5) there was no common outlier data points between x and y, and **hence no point is marked as orange.**

**A diagram of a box plot

Description automatically generated**