**Inferential Statistics**

Instructions:

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Insights should be drawn from the plots about the data such as, is data normally distributed/not, outliers, measures like mean, median, mode, variance, std. deviation

Please ensure you update all the details:

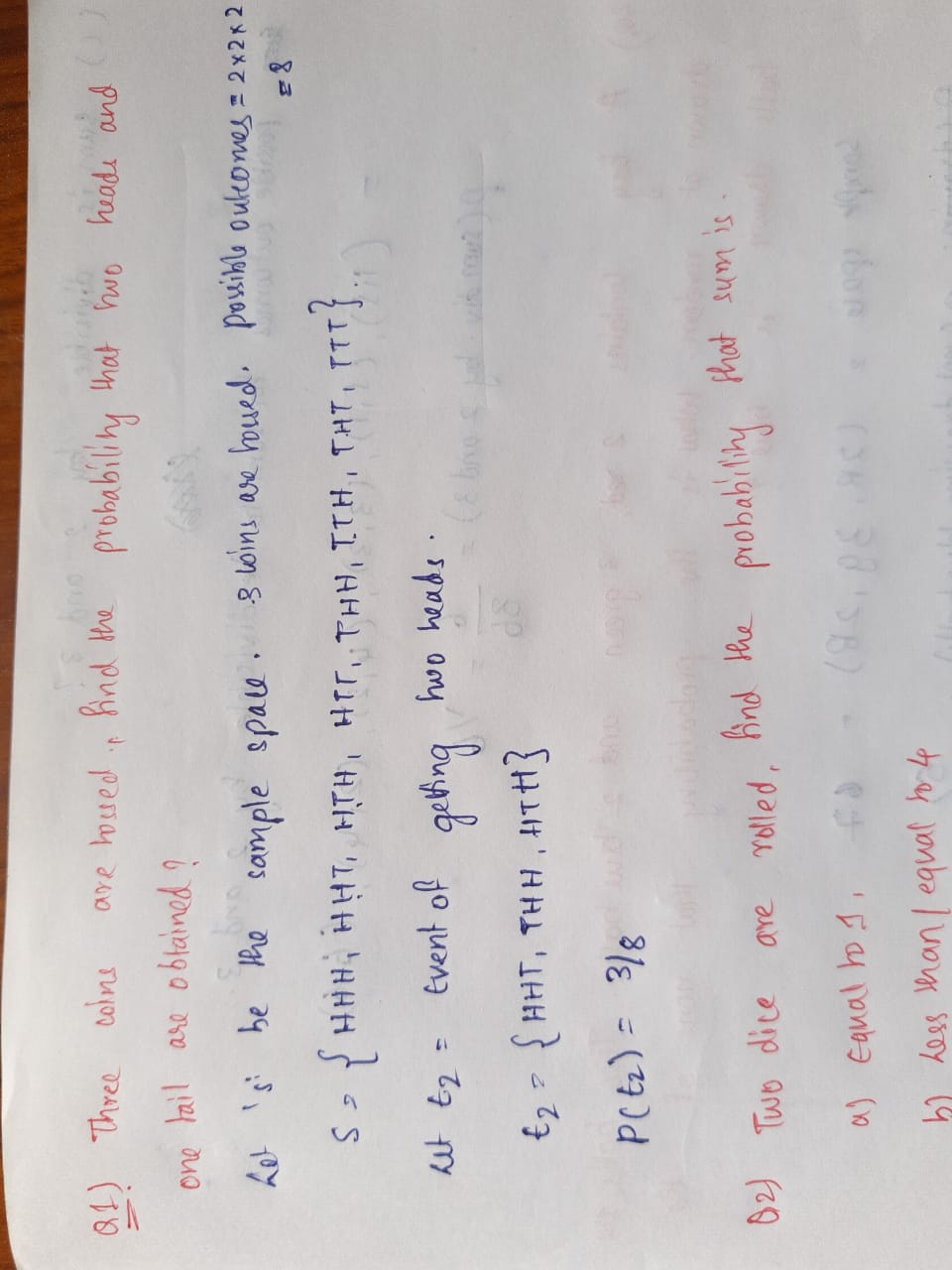
**Name: Navya.P**

**Batch Id: WDEO171220**

**Topic: Basic Statistics**

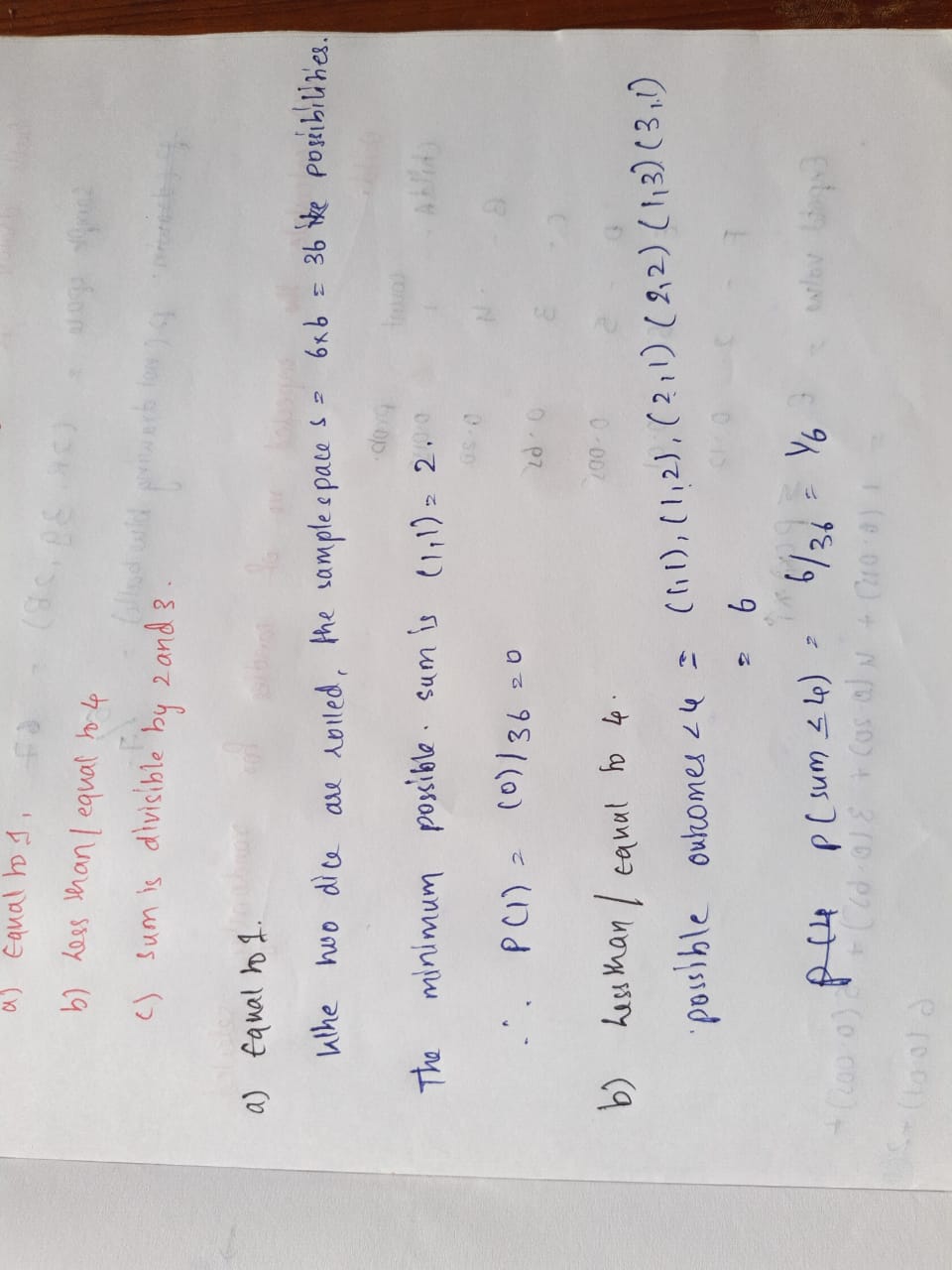
**Problem Statements:**

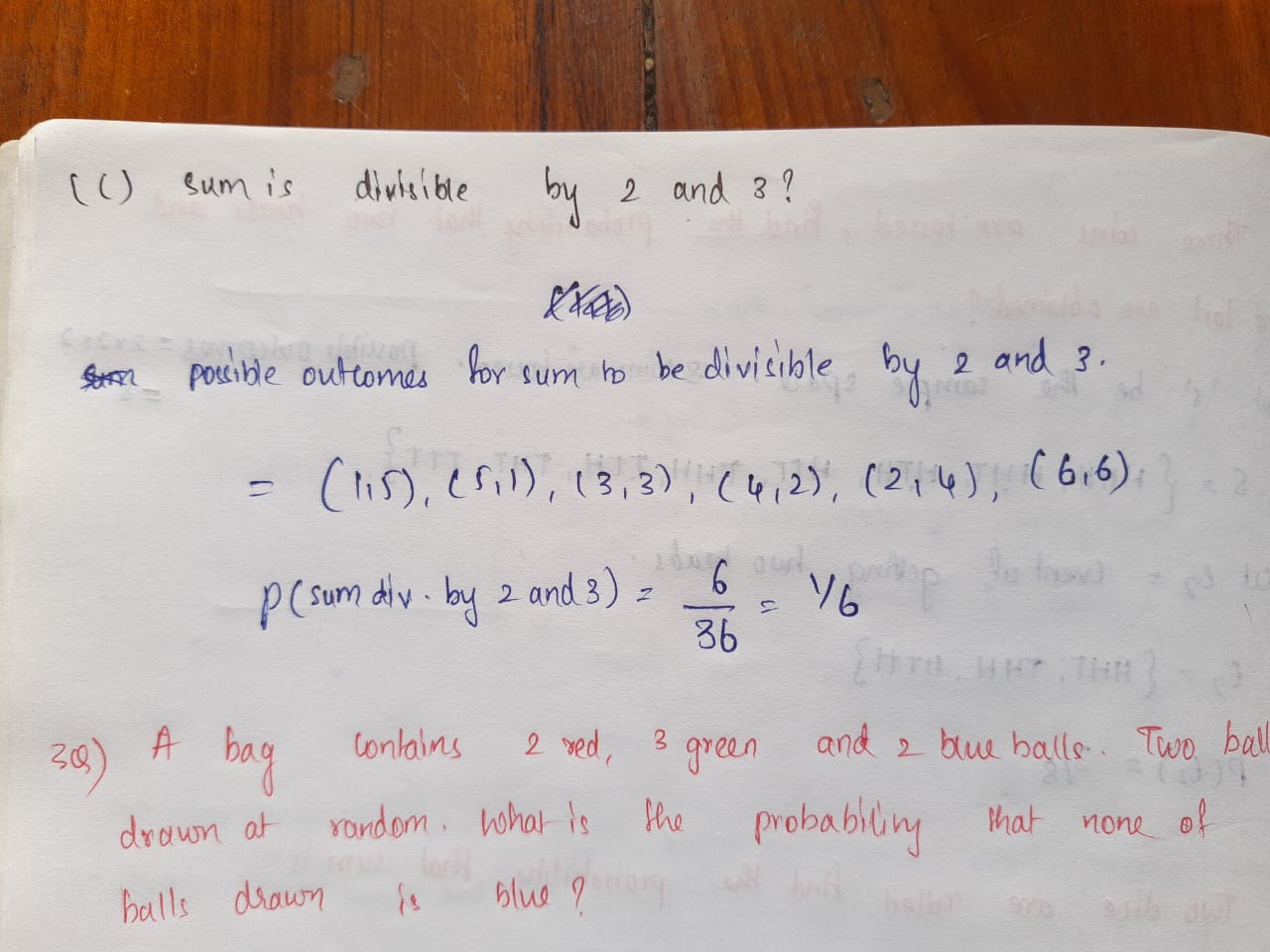
Q1) Three Coins are tossed, find the probability that two heads and one tail are obtained?



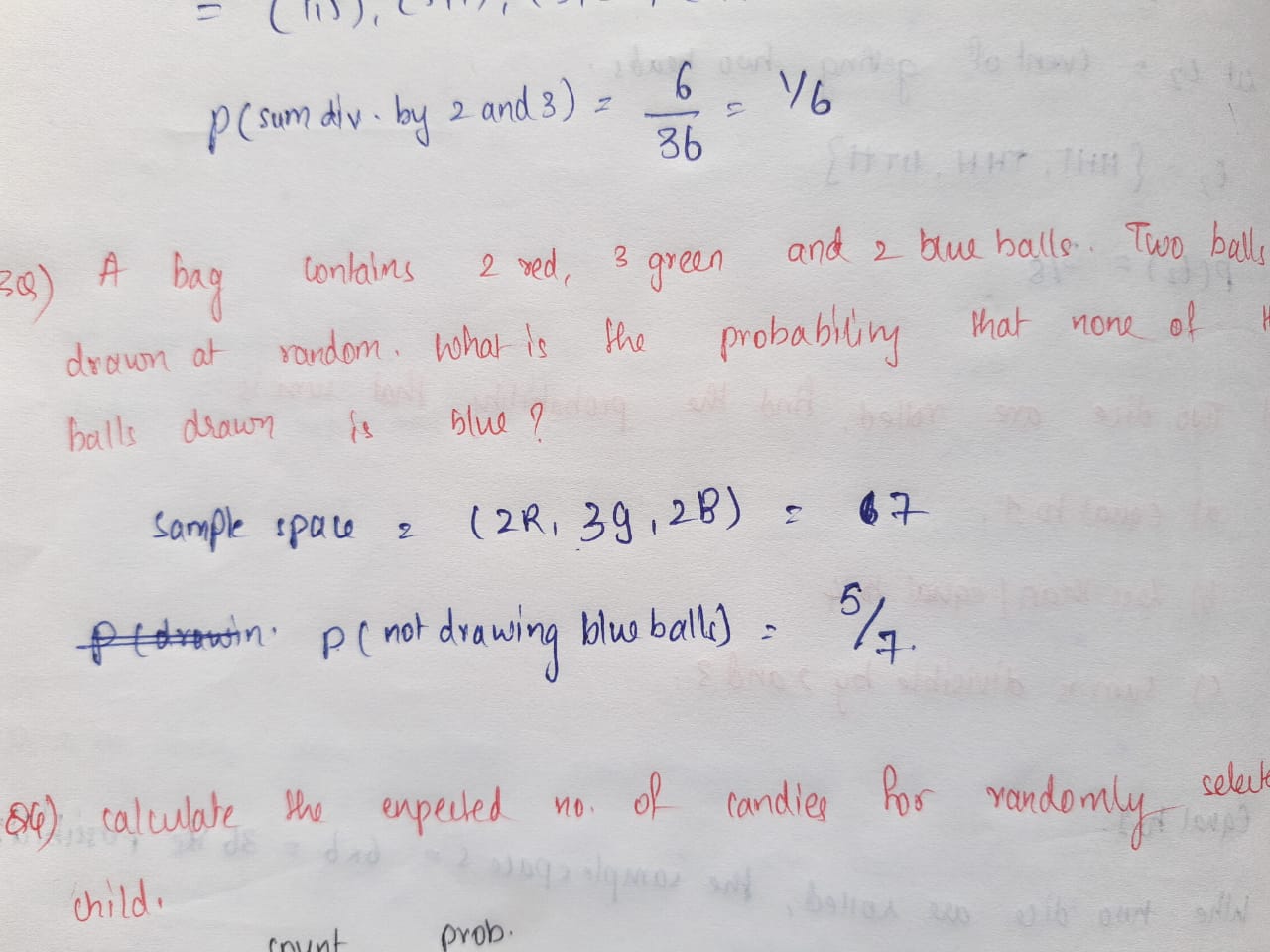
Q2) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3





Q3) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?



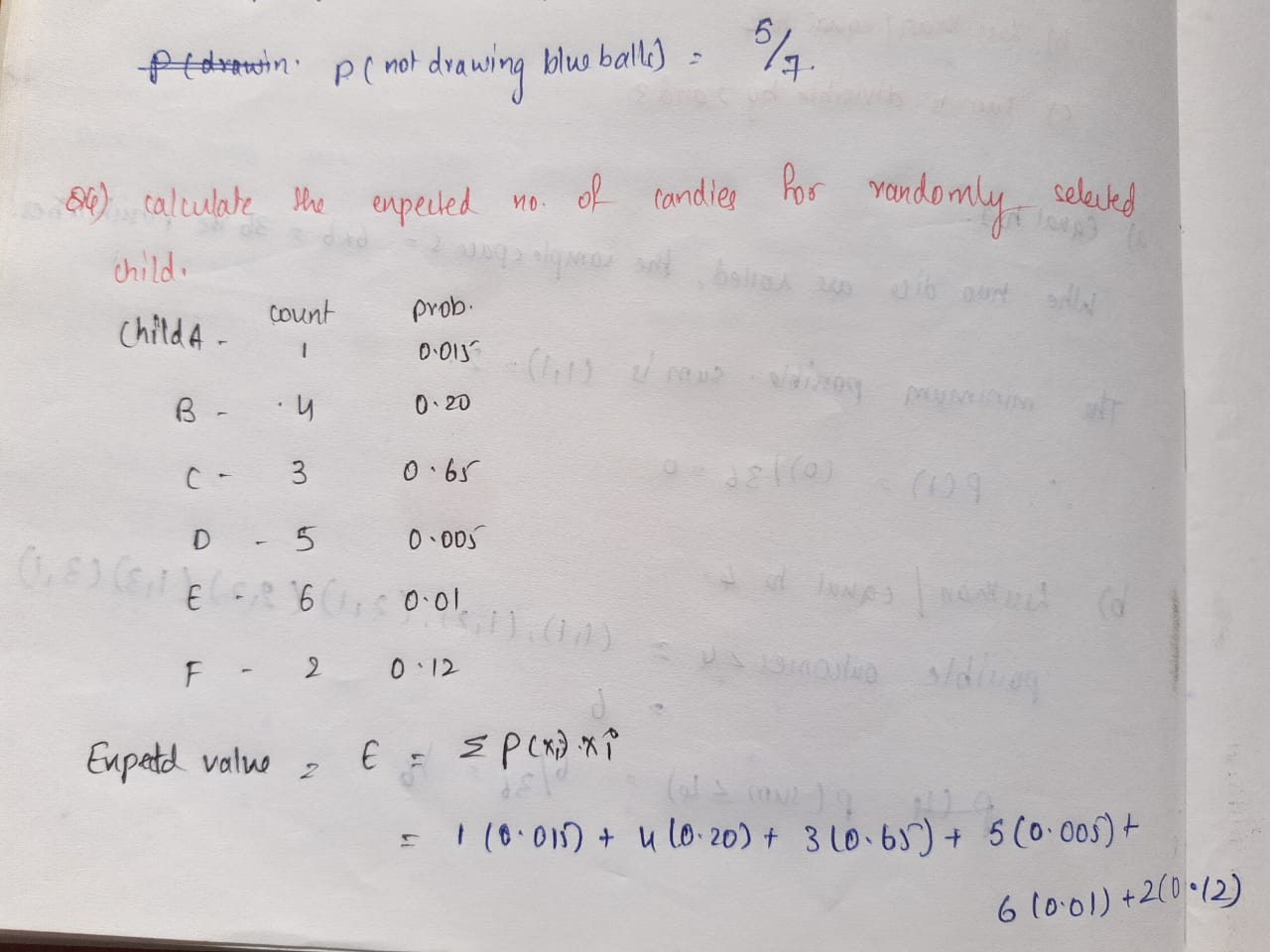
Q4) Calculate the Expected number of candies for a randomly selected child:

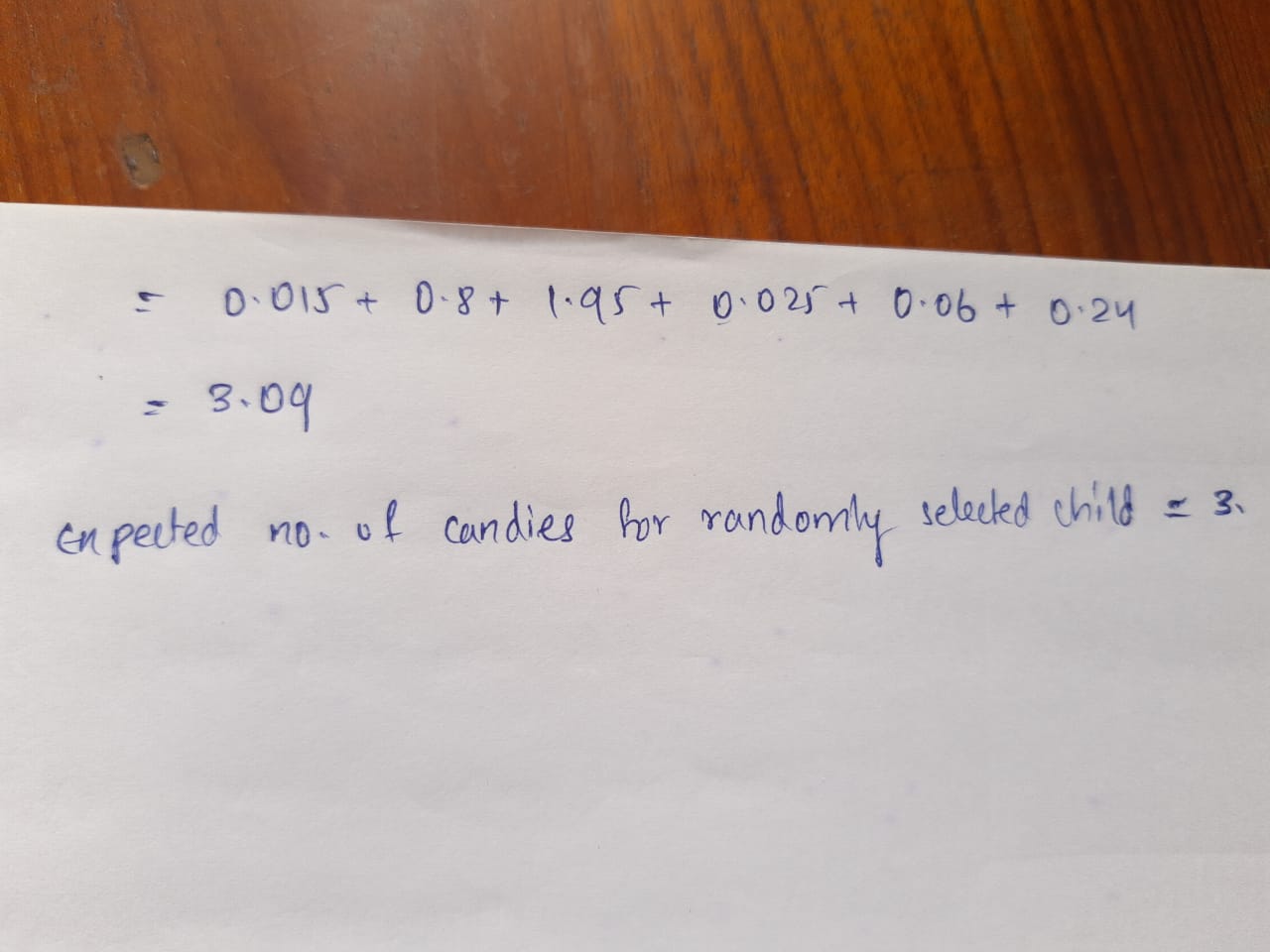
Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

i. Child A – probability of having 1 candy is 0.015

ii. Child B – probability of having 4 candies is 0.2

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.12 |

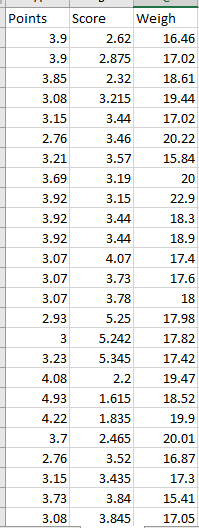




Q5) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and comment about the values/ Draw some inferences.



Dataset: Refer to Hands-on Material In LMS - Data Types EDA assignment snap shot of dataset is given above.

The Mean, Median, variance and standard deviation, range are measured for the dataset and the results are as following.

|  |  |
| --- | --- |
| mean of points | 3.596563 |
| median | 3.695 |
| variance | 0.285881 |
| SD | 0.534679 |
| range | 2.17 |

|  |  |
| --- | --- |
| Mean of score | 3.21725 |
| median | 3.325 |
| Variance | 0.957379 |
| SD | 0.978457 |
| range | 3.911 |

|  |  |
| --- | --- |
| Mean of weight | 17.84875 |
| median | 17.71 |
| variance | 3.193166 |
| SD | 1.786943 |
| range | 8.4 |

Observations :

Mean and Median of all the three data sets are almost equal, implying the chance of outliers is very less.

Range of weight is more implying the wide spread of the data.

Also, SD of points is less and nearby zero implying the data is near to the mean. SD of score is slightly more indicating widespread than points.

SD of weight is more which indicates the data is spread far away from the mean.

Q6) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

The Expected value is Mean of the data given incase of list

Mean = (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.3

Q7) Look at the data given below. Plot the data, find the outliers and find out

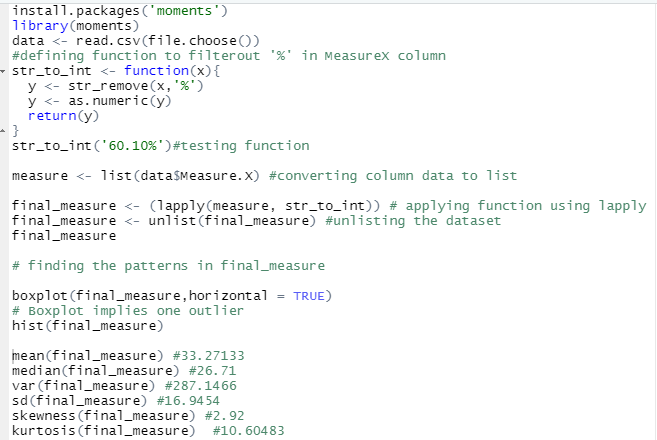
**Hint:** [Use a plot which shows the data distribution, skewness along with the outliers; also use R/Python code to evaluate measures of centrality and spread]

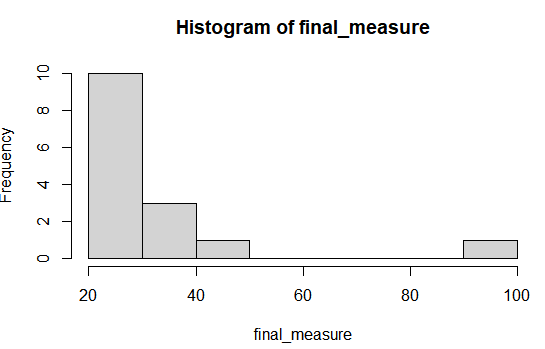
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

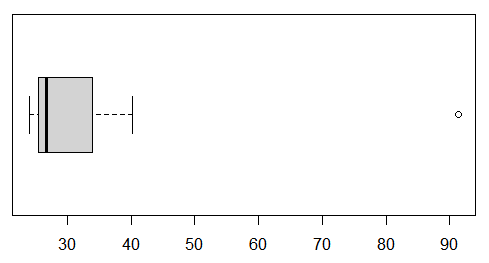
The above data is saved to excel file and worked in R to get the relevant information.

The column Measure X is preprocessed to get the information.

The code is as follows.



The obtained boxplot and histogram are as follows :



The data has positive skewness and kurtosis is also positive. We have a he difference between mean and mode owing to the outlier present.

Q8) AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected.

What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

**Hint:** [Using Probability formula evaluate the probability of one call being wrong out of five attempted calls]

one in 200 long-distance telephone calls is misdirected.  probability of call misdirecting

 p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one in five attempted telephone calls reaches the wrong number

= 1  -  none of the call reaches the wrong number

= 1  - P(0)

= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1  -  (199/200)⁵

= 0.02475

Q9) Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

|  |  |
| --- | --- |
| X | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

1. What is the most likely monetary outcome of the business venture?

**Hint:** [The outcome is most likely the expected returns of the venture]

**Ans: The most likely outcome is X = 2000 with P(x) = 0.3**

1. Is the venture likely to be successful? Explain.

**Hint:** [Probability of % of venture being a successful one]

**The probability of the venture to be successful is probability to profit/total probability**

**= 0.6/1**

1. What is the long-term average earning of business ventures of this kind? Explain.

**Hint:** [Here, the expected returns to the venture is considered as the

the required average]

**The longterm average is nothing but the expected value of the business, which is 800.**

**Expected value = ∑X\*P(x)**

**= -2000\*0.1+(-1000\*0.1)+0+1000\*0.2+2000\*0.3+3000\*0.1**

**= 800**

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure.

**Hint:** [Risk here stems from the possible variability in the expected returns, therefore, name the risk measure for this venture]

**Standard deviation (sd) = 0.08164**

**Hints:**

For each assignment, the solution should be submitted in the below format

1. Research and Perform all possible steps for obtaining solution.

2. For Statistics calculations, explanation of the solutions should be documented detail along with codes. Use the same word document to fill in your explanation

Must follow these guidelines:

2.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the

calculation stepwise

2.2. For True/False Questions, or short answer type questions explanation is must.

2.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) the data

distribution to be attached

3. All the codes (executable programs) should execute without errors

4. Code modularization should be followed

5. Each line of code should have comments explaining the logic and why you are using that function