**A picture containing shape, arrow

Description automatically generatedInferential Statistics**

Instructions:

Please share your answers filled inline in the word document. Submit 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, etc.

Please ensure you update all the details:

**Name: \_**Uppara Naveen Kumar**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Batch Id: \_\_\_\_\_**07092022**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Topic: Basic Statistics**

**Problem Statements:**

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

Ans: (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT) = no.of possible outcomes = 8

Two heads and one tail = 3/8

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

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

1. Equal to 1 Ans: Minimum value of dice is 1. So, probability of the sum equal to 1 will never happen.
2. Less than or equal to 4 Ans: {(1,1), (1,2), (1,3), (2,1), (3,1), (2,2)} = 6/36 = 1/6
3. Sum is divisible by 2 and 3 Ans: {(5,1), (1,5), (4,2), (2,4), (3,3), (6,6)} = 6/36 = 1/6

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?

Ans: sample space =7c2 = 21. Let event E be none of the balls is blue = all balls are either red or green or both. n(E) = 5c2 =10 p(E) =10/21

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 Ans: 1\*0.015 = 0.015

ii. Child B – probability of having 4 candies is 0.2 = 4\*0.2 = 0.8

Ans: 1\*0.015+4\*0.2+3\*0.65+5\*0.005+6\*0.01+2\*0.12 = 3.09

|  |  |  |
| --- | --- | --- |
| 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 |

**A picture containing shape, arrow

Description automatically generated**

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.

Mean of points = Sum(3.9 ,3.9 , 3.85 ,3.08 ,3.15 ,2.76 ,3.21,3.69,3.92,3.92,3.92,3.07,3.07,3.07,2.93,3,3.23,4.08,4.93,4.22,3.7,2.76,3.15,3.73,3.08)/25 = 3.4928

Median = 3.23

Mode = 3.92,3.07

Range = 2.17

Variance = 0.279

Standard = 0.583

For score:

Mean = 2.62, 2.875,2.32,3.215,3.44,3.46,3.57,3.19,3.15,3.44,3.44,4.07,3.73,3.78,5.25,5.242,5.345,2.2,1.615,1.835,2.465,3.52,3.435,3.84,3.845 = 3.39568

Median = 3.44

Mode = 3.44

Standard deviation = 0.93564

Variance = 0.87542

For Weigh:

16.46,17.02,18.61,19.44,17.02,20.22,15.84,20,22.9,18.3,18.9,17.4,17.6,18,17.98,17.82,17.42,19.47,18.52,19.9,20.01,16.87,17.3,15.41,17.05

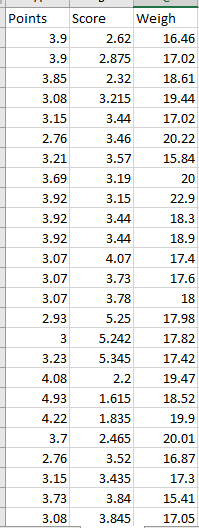
Mean = 18.218400000000003

Median = 17.98

Mode = 17.02

Standard deviation = 1.6049639995962521

Variance = 2.5759094399999984



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

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?

Ans: 1/9(108+110+123+134+135+145+167+187+199) = 145.333

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]

Ans : 24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00

Outliers are present

Mean = 33.27133333333333

Standard deviation = 16.945400921222028

Variance = 287.1466123809524

**A picture containing shape, arrow

Description automatically generated**

|  |  |
| --- | --- |
| **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% |

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]

Ans : IF 1 in 200 long-distance telephone calls are getting misdirected.  
probability of call misdirecting = 1/200 Probability of call not Misdirecting = 1-1/200 = 199/200 The probability for at least one in five attempted telephone calls reaches the wrong number Number of Calls = 5 n = 5 p = 1/200 q = 199/200 P(x) = at least one in five attempted telephone calls reaches the wrong number P(x) = ⁿCₓ pˣ qⁿ⁻ˣ P(x) = (nCx) (p^x) (q^n-x) # nCr = n! / r! \* (n - r)! P(1) = (5C1) (1/200)^1 (199/200)^5-1 P(1) = 0.0245037

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 |

**A picture containing shape, arrow

Description automatically generated**

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 monetary outcome of the business venture is 2000$ As for 2000$ the probability is 0.3 which is maximum as compared to others

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

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

**Ans:** Yes, the probability that the venture will make more than 0 or a profit p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit

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]

Ans: The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 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]

**Ans:** The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk Var (X) = E(X^2) –(E(X))^2 = 2800000 – 800^2 = 2160000

**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