**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: Vaibhav Chaudhari**

**Batch Id: DSWDEOS 280422**

**Topic: Basic Statistics**

**Problem Statements:**

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

**Sol:** For 3 coins combinations = 2^3 = 8

{ HHH, THH, HTH, HHT, TTH, TTT, THT, HTT }

Two Heads and one tail: THH, HTH, HHT ----> 3

Probability = 3 / 8

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

**Sol:** Two Dices are rolled: 6^2 = 36 outcomes

1. Probability of sum equal to 1 = **0**
2. Probability less than or equal to 4 = (1,1) (1,2) (1,3) (2,1) (2,2) (3,1) = 6

= 6 / 36 = **1/6**

1. Probability of sum divisible by 2 and 3 : (5,1) (1,5) (3,3) (2,4) (4,2) = 5

= **5/36**

**A picture containing shape, arrow

Description automatically generated**

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?

**Sol:**

1st pick: Probability of ball red or green = 5 / 7

2nd pick: Probability of ball red or green = 4 / 6

Probability that none of the balls are blue = (5/7) x (4/6) = **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

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 |

**Sol:**

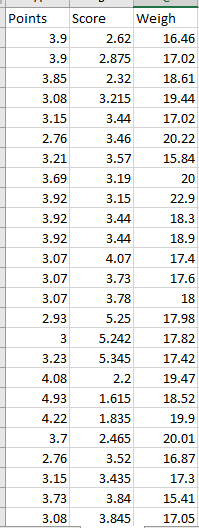
Expected No. of Candies E(X) = ∑ x P ( x )

E(X) = (1 x 0.015) + (4 x 0.20) + (3 x 0.65) + (5 x 0.005) + (6 x 0.01) + (2 x 0.12) = **3.09**

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.

**Sol:**

Mean = Sum of Observations / Total no. of observations.

Median = It is the middle value of the ordered observations.

Mode = Most repeated value in the observations.

Variance = ∑(xi−μ)2/n

Std Deviation = sqrt(Variance)

Range = (max – min)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Points** | **Score** | **Weigh** | | **Mean** | 3.49 | 3.40 | 18.22 | | **Median** | 3.23 | 3.44 | 17.98 | | **Mode** | 3.92 | 3.44 | 17.02 | | **Variance** | 0.28 | 0.88 | 2.58 | | **Std Dev** | 0.53 | 0.94 | 1.60 | | **Range** | 2.17 | 3.73 | 7.49 | |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

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?

**Sol:** Expected Value E(X) = ∑ (x \* P ( x ) )

Probability of each patient = 1/9

E(X) = (108\*1/9) + (110\*1/9) + (123\*1/9) + (134\*1/9) + (135\*1/9) + (145\*1/9) + (167\*1/9)

+ (187\*1/9) + (199\*1/9)

= **145.33**

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]

**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]

**Sol:**

Probability of getting atleast one mis-direction = prob of one misdirected call + prob of two misdirected call + prob of three misdirected call + prob of four misdirected call + prob of five misdirected call **(or)**

1 - prob of zero misdirected call

Probability of call = 199 / 200

Prob of misdirected call = 1/200

Probability of getting atleast one misdirected call = 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 |

**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]

**Sol:** The most likely outcome will be the outcome with the highest probability = **$ 2,000**

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

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

**Sol:** Expected value = (-2000\*0.1) + (-1000\*0.1) + (1000\*0.2) + (2000\*0.3) + (

3000\*0.1) = 800

As the expected value is positive, venture is likely to be successful.

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]

**Sol:** Long-term average earning of business ventures = $ 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]

**Sol:** Risk measure for this venture is Probability where X is negative = **0.2**

**20% risk.**

**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