**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: Biswajeet Padhi**

**Batch Id: 280921**

**Topic: Basic Statistics**

**Problem Statements:**

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

Ans) Total number of possible combinations are 23= 8.

Combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

Number of desired outcomes(Two heads and One Tail) =3

Probability (Two heads and One Tail)= 3/8

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

1. Equal to 1

Ans) Possible outcomes of rolling two dice = 62 =36

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (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) |

Number of desired outcomes= 0

Probability of sum is 1 = 0

1. Less than or equal to 4

Ans) Number of desired outcomes = 6

|  |  |  |
| --- | --- | --- |
| (1, 1) | (1, 2) | (1, 3) |
| (2, 1) | (2, 2) |  |
| (3, 1) |  |  |

Probability of sum is Less than or equal to 4 = 6/36

1. Sum is divisible by 2 and 3

Ans) Number of desired outcomes = 6

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

1. Probability of Sum is divisible by 2 and 3= 6/36

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) Total number of balls = (2+3+2)

Number of ways of drawing 2 balls out of 7= = 7!/(2!\*5!)=21

Number of ways of drawing 2 balls out of (2 + 3) balls = = 10

Probability of drawing 2 balls out of which none of the balls drawn is blue= 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 |

**A picture containing shape, arrow

Description automatically generated**

Ans)

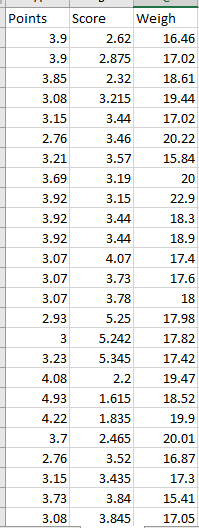
Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12  
 = 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24  
 = 3.090

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.

Ans) In python code all answers

Q6) Calculate Expected Value for the problem below

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

**A picture containing shape, arrow

Description automatically generated**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)

Total number of patient = 9

Probability of selecting each patient = 1/9

Weights (X) = 108, 110, 123, 134, 135, 145, 167, 187, 199

Expected Value = ∑ ( Probability\* Value )

= ∑ P(x)\*E(x)

=(1/9)\*(108+110+123+134+135+145+167+187+199)

=(1/9)\*(1308)

=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]

|  |  |
| --- | --- |
| **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)

Probability of call misdirecting P = 1/200

Probability of call not Misdirecting = 1 – (1/200) = 199/200 =q

Number of ways of getting zero calls misdirected out of 5 calls =

By using the Binomial distribution

P(X = x) = P(x) = ⁿCₓpˣqⁿ⁻ˣ, where q = 1 - p

p can be considered as the probability of a success, and q the probability of a failure.

Probability at least one in five calls reaches wrong number=

1. Probability none of the call reaches the wrong number)

=1 - P(0)

=1-( \*(1/200)0 \* (199/200)5-0

=1-(1\*1\*0.975248753121875)

=0.024751246878125

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 = 2000 , as it has the highest probability of 0.3

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

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

Ans) Yes venture likely to be successful = 60%

Probability of venture being a successful one

= P(1000)+P(2000)+P(3000)

=0.2+0.3+0.1

=0.6

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 earning of business ventures of this kind is the expected returns = 800

Expected value = ∑E(X)P(X)

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

=-200+(-100)+0+200+600+300

=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]

**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

**A picture containing shape, arrow

Description automatically generatedGrading Guidelines:**

**Note: 1. An Assignment submission is considered complete only when successful executable code(s), and documentation explaining the applied solution and results are provided. Failing to submit either of them will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline date will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline date
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of questions asked in assignments are completed. (or)
  + All assignments were submitted, however, after the given deadline
* **A picture containing shape, arrow

  Description automatically generatedGrade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of questions asked in assignments are completed. (or)
  + Less than 80% of questions asked in assignments are submitted after the deadline
* **Grade D: (>= 60 and < 70):** Assignments submitted after the Deadline and with 50% or less of questions
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of questions asked in the assignments are submitted after the deadline (OR)
  + Less than 30% of questions asked in the assignments are submitted before deadline

**Grade F: (< 50):** Copied submission or No submission