|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

**Q1)** Identify the Data type for the Following:

**Q2)** Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

**Q3)** Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Ans:** Three coins are tossed, and there are a total of 23 potential combinations, which equals eight.

HHH, HHT, HTH, THH, TTH, THT, HTT, and TTT are examples of these combinations.

There are three different combinations that each have two heads and one tail: HHT, HTH, and TTH= 3.

As a result, the probability of simultaneously tossing three coins and getting two heads and one tails can be given by:

Probability of two heads and one tail is = 3/8

**Q4)** Two Dice are rolled, find the probability that sum remains

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

**Ans:** Total number of outcomes when two dice are rolled=6\*6=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)} = 36

1. Equal to 1: 0% Probability
2. Less than or equal to 4: 6/36 = 1/6
3. Sum is divisible by 2 and 3: 6/36 = 1/6

**Q5)** 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 events= 21

Interested events= 10

Probability that none of the balls is blue = 10/21= 0.47

**Q6)** 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)

|  |  |  |
| --- | --- | --- |
| 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 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Ans:** Expected number = E(x) =μ\_x=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= 3.09

**Q7)** Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

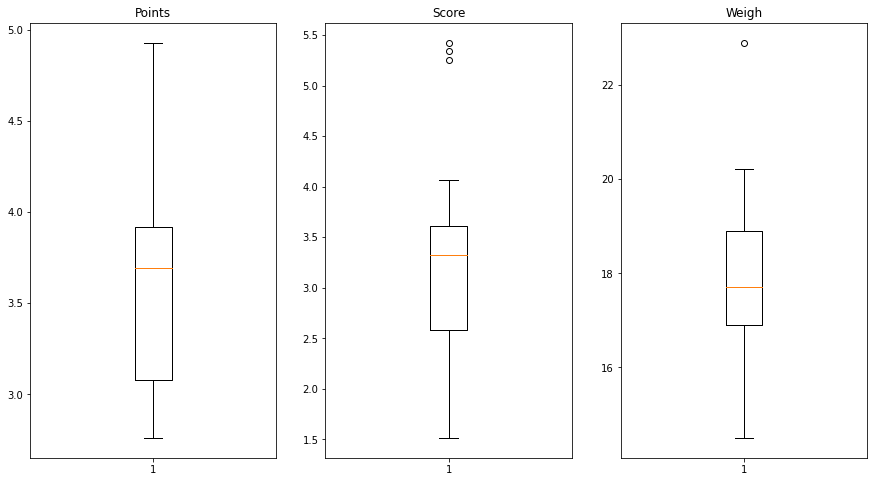
* For Points,Score,Weigh>

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

**Ans:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Std. Dev** | **Range** |
| **Points** | 3.596563 | 3.695 | 3.92 | 0.285881 | 0.534679 | 2.17 |
| **Score** | 3.217250 | 3.325 | 3.44 | 0.957379 | 0.978457 | 3.91 |
| **Weight** | 17.848750 | 17.710 | 17.02 | 3.193166 | 1.786943 | 8.4 |

* The distribution is symmetric and is having zero skewness, because mean is almost equal to the median.
* The score and Weigh column contains some outliers.



**Q8)** 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:** Ans: EV=Σx/n =(108+110+123+134+135+145+167+187+199)/9=145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**SP and Weight(WT)**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.117510 | -0.508994 |
| **Distance** | 0.806895 | 0.405053 |
| **SP** | 1.611450 | 2.977329 |
| **Weight** | -0.614753 | 0.950291 |

**Q10) Draw inferences about the following boxplot & histogram**



**Ans:**

1. Most of the chick weight lies in between 50 - 100
2. The data is positively Skewed



**Ans:** The above boxplot suggests that the distribution has lots of outliers towards upper extreme.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Ans:**

n=2000

x̅=200

s=30

SE=

CI = SE \* tvalue

|  |  |  |  |
| --- | --- | --- | --- |
|  | **94%** | **98%** | **96%** |
| **Lower** | 201.261674707842 | 201.56056159570022 | 201.3776966518666 |
| **Upper** | 198.738325292158 | 198.43943840429978 | 198.62230334813333 |

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

**Ans:**

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52941 |
| Deviation | 5.052664 |

1. What can we say about the student marks?

**Ans:** 1. Average of students get 41 marks.

2. There is outlier.

**Q13)** What is the nature of skewness when mean, median of data are equal?

**Ans:** Skewness is totally 0 and perfectly symmetric bell-shaped curve.

**Q14)** What is the nature of skewness when mean > median ?

**Ans:** Right Skewed (tail on the right side).

**Q15)** What is the nature of skewness when median > mean?

**Ans:** left Skewed (tail on the left side).

**Q16)** What does positive kurtosis value indicates for a data ?

**Ans:** High peak at center of data (sharp peak) and less variation.

**Q17)** What does negative kurtosis value indicates for a data?

**Ans:** Wide peak at center of data (Broad peak) and more variation.

**Q18)** Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**Ans:** It is not a Normal Distribution

What is nature of skewness of the data?

**Ans:** It is left skewed

What will be the IQR of the data (approximately)?   
  
**Ans:** Inter Quartile Range =Upper Quartile- Lower Quartile => 18-10= 8.

**Q19)** Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Ans:** a) The median of the two boxplots are same approximately 260.

b) Outliers doesn’t exist in both of the boxplots.

**Q 20)** Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

1. P(MPG>38)

**Ans:** 0.347593925

b. P(MPG<40)

**Ans:** 0.72934987

c. P (20<MPG<50)

**Ans:** 1.24309687973e-05

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** MPG have Normal Distribution as visible on the graph(shown in Assignment-1.ipynb, Q 21)

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Ans:** Adipose Tissue (AT) and Waist Circumference (Waist) follow Normal Distribution as visible on the graph(shown in Assignment-1.ipynb, Q 21)

**Q 22)** Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Ans:**

|  |  |
| --- | --- |
| **90%** | 1.644854 |
| **94%** | 1.880794 |
| **60%** | 0.8416212 |

**Q 23)** Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Ans:**

|  |  |
| --- | --- |
| **95%** | 2.063899 |
| **96%** | 2.171545 |
| **99%** | 2.79694 |

**Q 24)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Ans:** µ=270, x ̅=260, SD=90, n=18, df=17

t-score = (x ̅-µ)/(s/√n) = -0.4714

Degree of freedom = 17

P(t) = 0.3216725

Required probability = 0.32=32%