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

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

The probability of getting two head and one tail is 0.375

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

1. Equal to 1

As there is no combination who sum is equal to1.

The probability is 0.

1. Less than or equal to 4

The probability is 1/6

1. Sum is divisible by 2 and 3

The probability is 7/36

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?

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 |

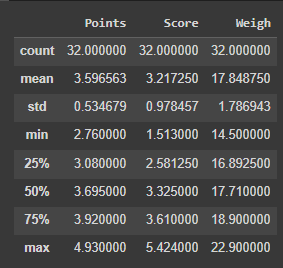
Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

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

* For Points,Score,Weight

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



From the above we can see that all columns have 32 entries with no null value.

Point Std is 0.53 showing there is less variability compared to other columns

Mode:

Point: has two modes 3.07 and 3.92 from which 3.07 is the most occurring

Score : has one mode that is 3.44

Weigh: has two mode 17.02 and 18.90 s. 17.02 has most occurence

**Use Q7.csv file**

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?

For finding the expected value we will take the average of all weights

That is mean =108+110+123+134+135+145+167+187+199/9

Mean=145

There the Expected value of the weight chosen at random is 145

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness:

Speed = -0.117510

Since speed has negative value it is left/negative skewed

Distance =0.806895

Distance has positive value so its right/positive skewed

Kurtosis:

Speed=-0.508994

A negative value kurtosis show us there are less outliers

Dist=0.405053

Since Distance has a positive kurtosis it has more outliers compared to speed

**SP and Weight (WT)**

**Use Q9\_b.csv**

Skewness:

SP = -1.611450

Since SP has positive value, it is right/positive skewed

Weight=0.806895

Weight has negative value so it left/negative skewed

Kurtosis:

SP=2.977329

SP has a high numbers of outliers

WT=0.950291

WT has comparatively less numbers of outliers then Sp

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



It is positively skewed so, it will have high positive outliers



The boxplot shows us that are some outliers and most of the data has higher value then median.

**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?

94% confidence interval= 143.47-256.42

96% confidence interval= 138.38-261.61

98% confidence interval=130.20-269-79

**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.

Mean=41

Median =40.4

Variance=25.52

Standard Deviation=5.05

1. What can we say about the student marks?

Since mean and median is approximately same it appears to show that student marks have symmetrical distribution.

Standard deviation tells us that there is variability in data

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

It tells us that data distribution is symmetric

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

It is positive skewed

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

It is negative skewed

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

It has more extreme values(outliers)

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

There are few extreme values(outliers)

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



What can we say about the distribution of the data?

The distribution is assymetric

What is nature of skewness of the data?

IT is negative skewed

What will be the IQR of the data (approximately)?   
IQR will be approximately 8

Q19) Comment on the below Boxplot visualizations?



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

Boxplot1 and Bloxplot2 seems to have approximately same value for median.

And they both follow normal distribution

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= 34.7%

* 1. P(MPG<40)

Ans=72.9%

* 1. P (20<MPG<50)

Ans=89.9%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes it follows normal distribution

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

Dataset: wc-at.csv

Ans: No they do not follow normal distribution

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

Ans:

Z-score for 90 = 1.64

Z-score for 94=1.88

Z-score for 60=0.84

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

Ans:

Z-score for 95% = 2.06

Z-score for 96% = 2.17

Z-score for 99% = 2.79

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

Probability =32.2%