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

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

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

3/8

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

1. Equal to 1 > 0
2. Less than or equal to 4 >1/6
3. Sum is divisible by 2 and 3> 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?

10/21

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 (expected number of candies-3.09) for Child A as well as Child b are =0.815

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.

Answer No1:-

Points(mean-3.5965,median-3.695,mode- 3.07,3.92,var=0.28588,std=0.53467,range=2.17(min=2.76,max=4.93)

Score(mean=3.21725,median=3.325,mode=3.44,var=0.9573,std=0.9784,range=3.911(min=1.513,max=5.424)

Weigh(mean=17.84875,median=17.71,mode=17.02 ,18.90,var=3.193166,std=1.786943,range=8.4(min=14.5 ,22.9)

Answer No2:-

Comments on mean median and mode:

Mean, Median and Mode are the measurements of central tendency of our dataset.

Comments on variance, Standard deviation, Range:-

Variance, Standard deviation and Range are defined as the measures of dispersion of our datasets or our data.

**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?

Expected value or mean: 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness- -0.1139548(cars speed) and Skewness-** **0.7824835(distance)**

**Cars speed data is negatively skewed and distance data is positively skewed which means at positively skewed data mean should be more than median and at negatively skewed data mean should be less than median.**

**Kurtosis-** **2.422853(cars speed) and Kurtosis-** **3.248019(cars distance)**

**Its refer that cars speed has lesser flatter peak than cars distances. Here the peak is flatness in both cases.**

**Here cars distance peak has greater divergence than three .Here cars speed peak has lesser than divergence of three .**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness-** **1.581454(SP) Skewness-** **-0.6033099(Weight)**

**Weight data is negatively skewed data. SP data is positively skewed data. Positively skewed data mean will be more than median and negatively skewed data mean is less than median here.**

**Kurtosis:- 5.723521(SP) Kurtosis:- 3.819466(Weight)**

**Here in both the cases the peak has been more than the divergence peak value which is three. Both values are more than three here**

**Here the SP column peak has more flat peak as compare to the weight flatter peak here.**

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

Positively skewed data from Positive skewness of the data (Right Skewed) which is denoted as the mean of this histogram will be more than the median value.



1-Outliers



2-Maximum(Q3+1.5\*IQR)

3-Upper quartile Range(Q3)

4-Median(Q2)

5-Lower quartile Range(Q1) 6-Minimum(Q1-1.5\*IQR) 7-IQR=Q3-Q1 7-Whiskers.

**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%(205.76,194.24),98%(207.11,192.89),96%(206.3,193.7)

**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:41,40.5,25.52941,5.05264
2. What can we say about the student marks?

At students marks we can clearly visualize that mean of the data is greater than median of the data so it’s a right skewed distribution and skewness will become positive.

Q13) What is the nature of skewness when mean, median of data are equal? The data will be perfectly skewed data.

Q14) What is the nature of skewness when mean > median ? The nature of the skewness is Right Skewed or Positively Skewed Data.

Q15) What is the nature of skewness when median > mean?-> The nature of the skewness is left skewed data or negatively skewed data.

Q16) What does positive kurtosis value indicates for a data ? Its indicates sharpness of the peak value

Q17) What does negative kurtosis value indicates for a data? Its indicates flatness of the peak or curve.

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



What can we say about the distribution of the data? -> Here the whisker is shorter at the end and distribution of the data’s is left skewed distribution.

What is nature of skewness of the data? -> Left Skewed Distribution.

What will be the IQR of the data (approximately)?   
  
  
IQR=Q3-Q1=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.

The Box plot 1 designed with range = 3, Second one range is = 1.5

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)
  2. P(MPG<40)

c. P (20<MPG<50)

Answer-> a. P(0.3475), b. P(0.7293) c. P(0.8049)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No it will not follows the normal distribution because skewness is negative it’s a left skewed dataset. IN case of normal distribution it must be perfectly symmetrical about its mean.

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

Dataset: wc-at.csv

No it will not follows the normal distribution because skewness is positive

It’s a right skewed dataset. IN case of normal distribution it must be perfectly symmetrical about its mean

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval ->90%(1.645),94%(1.88),60%(0.8416)

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25->95%(2.063),96%(2.171),99%(2.796)

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 will be 0.6778 or 67.78%.