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

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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ordinal |
| 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?

* **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**

**= 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

**Expected number of candies = 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.

**Mean-**

**Points 3.596563**

**Score 3.217250**

**Weigh 17.848750**

**Median-**

**Points 3.695**

**Score 3.325**

**Weigh 17.710**

**Mode-**

**Points= 3.07 , 3.92 Score=3.44 Weigh=17.02,018.90**

**Standard deviation-**

**Points 0.534679**

**Score 0.978457**

**Weigh 1.786943**

**Varience-**

**Points 0.285881**

**Score 0.957378**

**Weigh 3.193165**

**Range –32**

**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?

**-145.33**

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

**Cars speed and distance**

**Skewness-**

**Speed= -0.11 dist=0.75**

**Kurtosis-**

**Speed=-0.67 dist=0.11**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Skewness-**

**SP=1.55 WT=-0.59**

**Kurtosis-**

**SP= 2..58 WT=0.72**

**Use Q9\_b.csv**

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



**Histrogram shows comparision of chick weight to frequency**

**Histogram is not symmentric.**

**Data is positively skewed.**

**Centre of data for weght group is about 100 and then further values becomes rare.**

**Centre of data for frequency is about100(most values between 0 to 100)**



**Boxplot showing most values are obtained in upper quartile and comparatively less values are obtained in lower quartile with 7 outliers.**

**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 : [134.8973 265.1027]**

**96%=confidence interval : [130.2096 269.7904]**

**98%=confidence interval : [122.7251 277.2749**]

**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.
2. What can we say about the student marks? student having marks with range of 18. maximum marks obtained by a student is 56 and minimum is 34. most marks obtained around 38 to 41.

* **Mean=41**
* **Median=40.5**
* **Variance=25.5**
* **standard deviation=5.05.**

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

* **Zero skewness**

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

* **Positively skewed**

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

* **Negatively skewed**

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

* **positive kurtosis indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean.(sharp peak)**

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

* **Negative values of kurtosis indicate that a distribution is flat and has thin tails. (broad peak)**

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



What can we say about the distribution of the data?

* **This is not normal distribution.**

What is nature of skewness of the data?

* **Negative skewed**

What will be the IQR of the data (approximately)?   
  
**- inter quartile = upper – lower**

**=>18-10=8**

Q19) Comment on the below Boxplot visualizations?



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

* **Medians of both boxplots are same.**
* **There is not much skewness on either side.**
* **There are no outliers.**

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

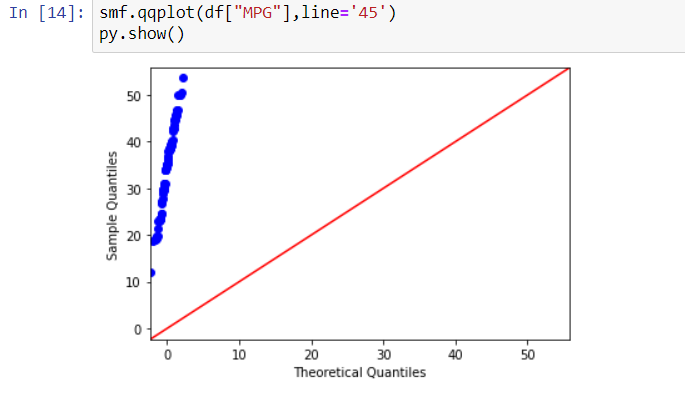
* 1. P(MPG>38)= **0.348**
  2. P(MPG<40)=**0.729**

c. P (20<MPG<50)=**0.214**

Q 21) Check whether the data follows normal distribution

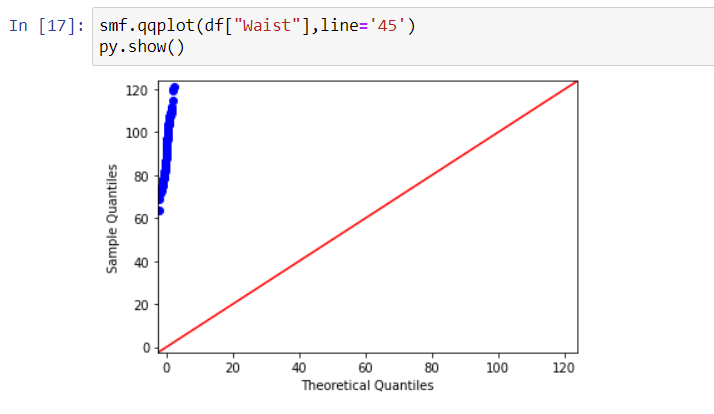
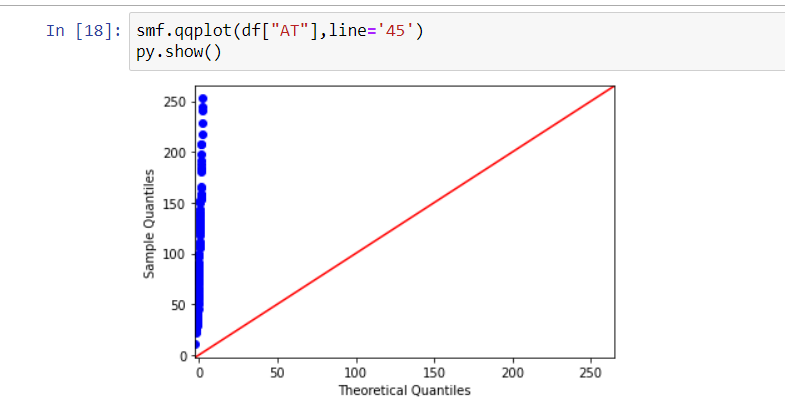
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



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

Dataset: wc-at.csv



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

* **90%= (1.66)**
* **94% = (1.88)**
* **60%= (0.84)**

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

* **95%(2.06)**
* **96% (2.17)**
* **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**) -0.471**

df 🡪 degrees of freedom

probability= **0.32**