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

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

**ANS:** HHT -1/8

HTH -1/8

THH -1/8

Probability= number of favorable outcomes/Total number of outcomes

=1/8+1/8+1/8=3/8

=3/8

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

1. Equal to 1

There is no probability that sum is equal to one because minimum sum of dices is two.

Probability= number of favorable outcomes/Total number of outcomes

Probability=0/36

1. Less than or equal to 4

Outcomes=(1,3),(2,2),(3,1)

Probability= number of favorable outcomes/Total number of outcomes

=3/36

=1/12

1. Sum is divisible by 2 and 3

Probability= number of favorable outcomes/Total number of outcomes

Outcomes=(1,5),(2,4),(3,3),(6,6),(5,1),(4,2) Probability=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:** Number of balls=2+3+2=7

7C2=7\*6/2\*1=21

No blue balls=5C2=5\*4/2\*1=10

Probability= number of favorable outcomes/Total number of outcomes

probability=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

**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

= 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.

**Use Q7.csv file**

**ANS:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | **POINTS** | **SCORE** | **WEIGH** | | **MEAN** | 3.596 | 3.217 | 17.84 | | **MEDIAN** | 3.695 | 3.325 | 17.71 | | **VARIANCE** | 0.28588 | 0.9573 | 3.193 | | **STANDARD DEVIATION** | 0.534 | 0.978 | 1.786 | | **RANGE** | 2.76,4.93 | 1.51,5.42 | 14.5,22.9 | |

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:** EXPECTED VALUE= (108+110+123+134+145+167+187+199)/9=145.33

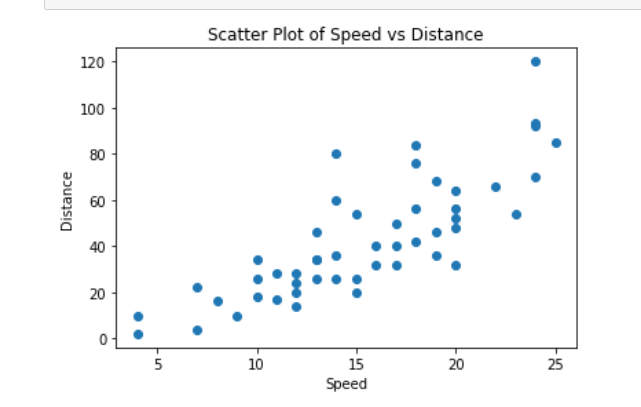
=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **SPEED** | **DIST** |
| **Skewness** | -0.1175 | 0.8068 |
| **Kurtosis** | -0.5089 | 0.4050 |

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.611 | -0.614 |
| **Kurtosis** | 2.977 | 0.950 |

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



**Sol:** The most of the data points are concerated in the range 50-100 with frequency 200. And least range of weight is 400 somewere around 0-10. So the expected value the above distribution is 75. Skewness- we can notice a long tail towards right so it is heavily right skewed.



**Sol:**

Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

**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:**

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Zvalue** | **Range** |
| 94% | 1.880 | 198.74,201.26 |
| 96% | 2.054 | 198.62,201.38 |
| 98% | 2.326 | 198.43,201.56 |

**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. | |

1. What can we say about the student marks?

**ANS:**

|  |  |
| --- | --- |
| **MEAN** | 41 |
| **MEDIAN** | 40 |
| **VARIANCE** | 24.1 |
| **STANDARD DEVIATION** | 4.91 |
|  |  |

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

ANS:Data is normalized and there is no skewness

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

ANS:Negative Skewness

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

ANS:Positive Skewness

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

ANS:positive kurtosis indicates that thinner peak and wider tails

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

ANS:Neagtive Kurtosis indicates that wider peak and wider tails

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



What can we say about the distribution of the data?

Not normally Distrubuted

What is nature of skewness of the data?

Negative skewness

What will be the IQR of the data (approximately)?   
10-18

Q19) Comment on the below Boxplot visualizations?



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

ANS:The above the two box plots are in normal distrubution

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: 1-cars.cdf(38)

0.3475838101730251

* 1. P(MPG<40)

ANS: 1-cars.cdf(38)

0.729362470706113

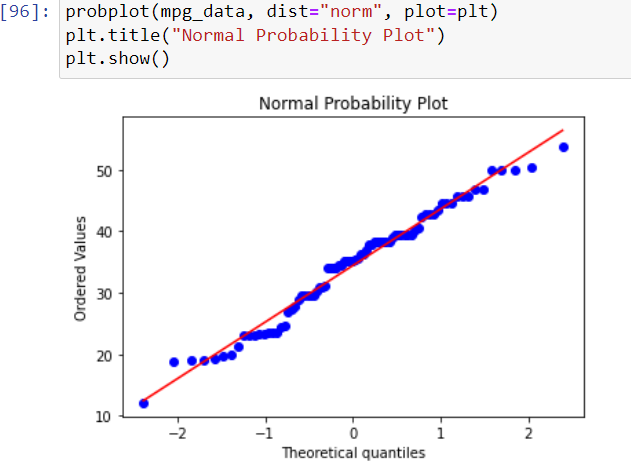
* 1. P (20<MPG<50)

0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

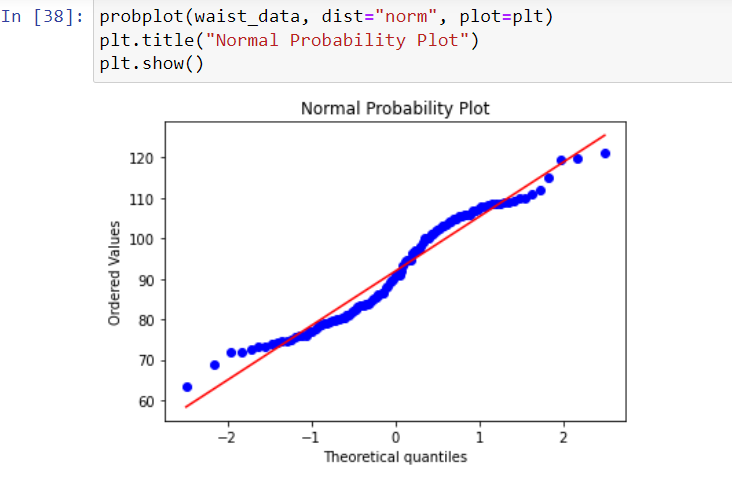
Dataset: Cars.csv

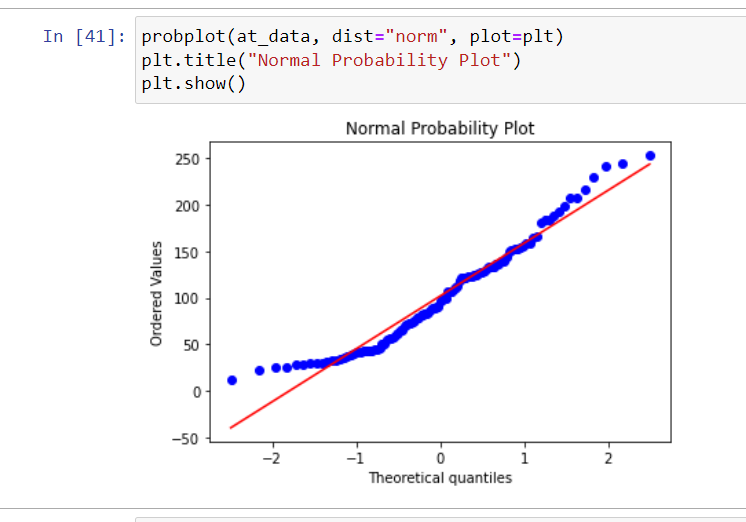


**:** **:** It follows the normal distribution the bar gaph plotted is symmetrical so it is Normal Distribution

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

Dataset: wc-at.csv





**:** Adipose Tissue and waist does not follow normal distribution is has bell curve non symmetric data curve.

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

|  |  |
| --- | --- |
| **CONFIDENCE INTERVAL** | **VALUES** |
| 90% | 1.644 |
| 94% | 1.880 |
| 60% | 0.841 |

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

|  |  |
| --- | --- |
| **CONFIDENCE INTERVAL** | **VALUES** |
| 95% | 2.063 |
| 96% | 2.171 |
| 99% | 2.796 |

T-Score for 95% confidence interval-2.0638985

T-Score for 96% confidence interval-2.171544

T-Score for 95% confidence interval-2.796939

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:** probability using T-Distribution is=(1.73)\*90/sqrt(18)=36.69