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

**Answer-3/8 or 0.375**

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

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

**Answer-a) 0**

**b)1/6**

**c)2/3**

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?

**Answer-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

**Answer**- **Expected number of candies for a randomly selected child  = 3.09**

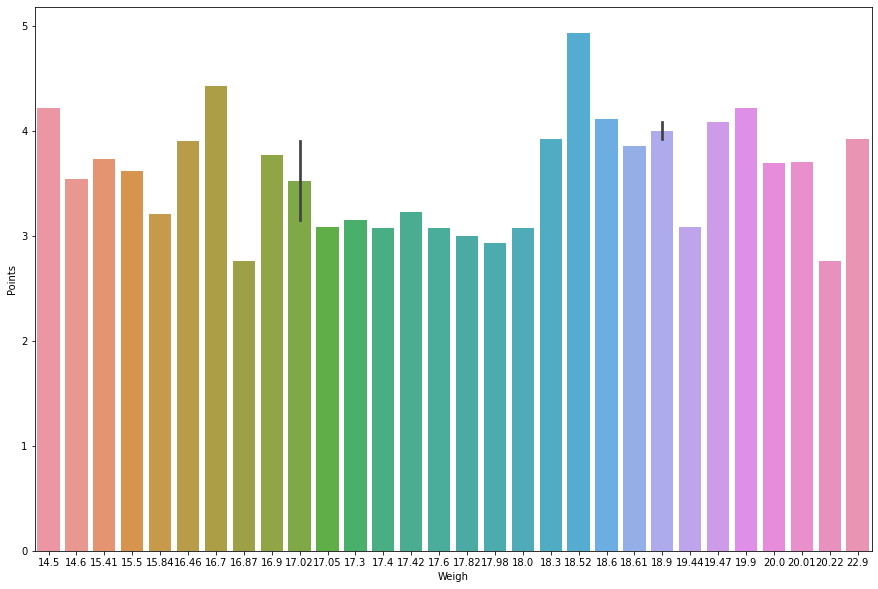
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

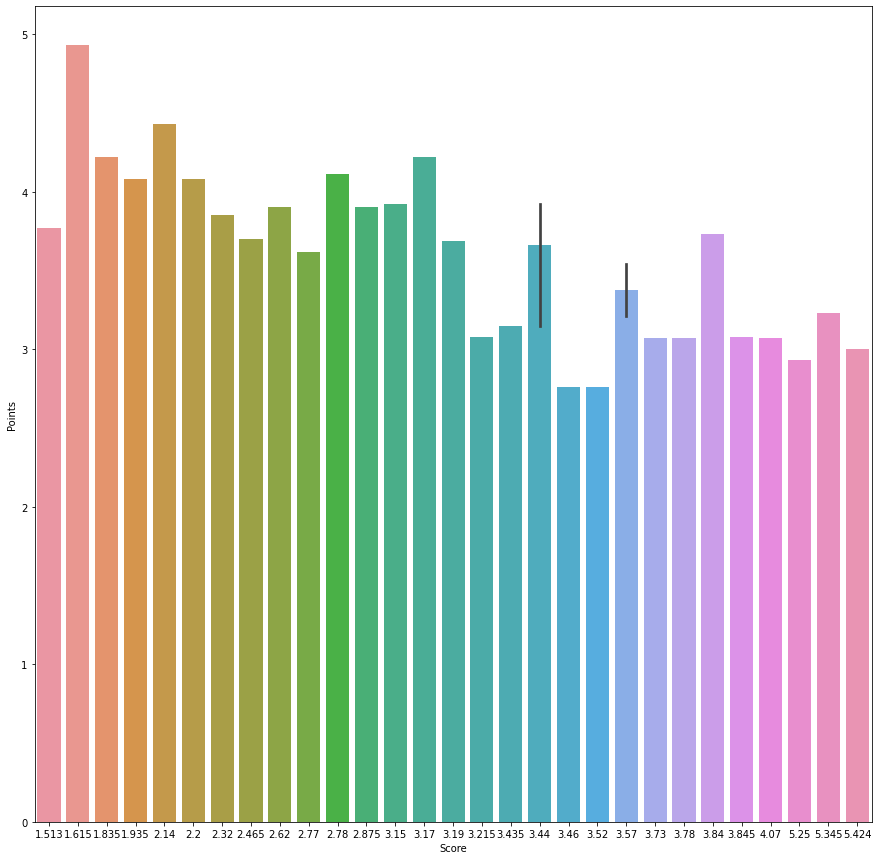
* For Points,Score,Weigh>

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

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
| Mean | 3.59 | 3.21 | 17.84 |
| Median | 3.69 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| varience | 0.276 | 0.927 | 3.09 |
| S.D | 0.52 | 0.963 | 1.75 |
| Range | 2.17 | 3.911 | 8.4 |

**Answer-The graph of weigh and point show that point increases slight as weigh increases and the median of weight has low point .**

****

**Answer-The above graph represent score vs point of cars indicating that there is not big different in point as score increases and it slight descrease while score is higher.**

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?

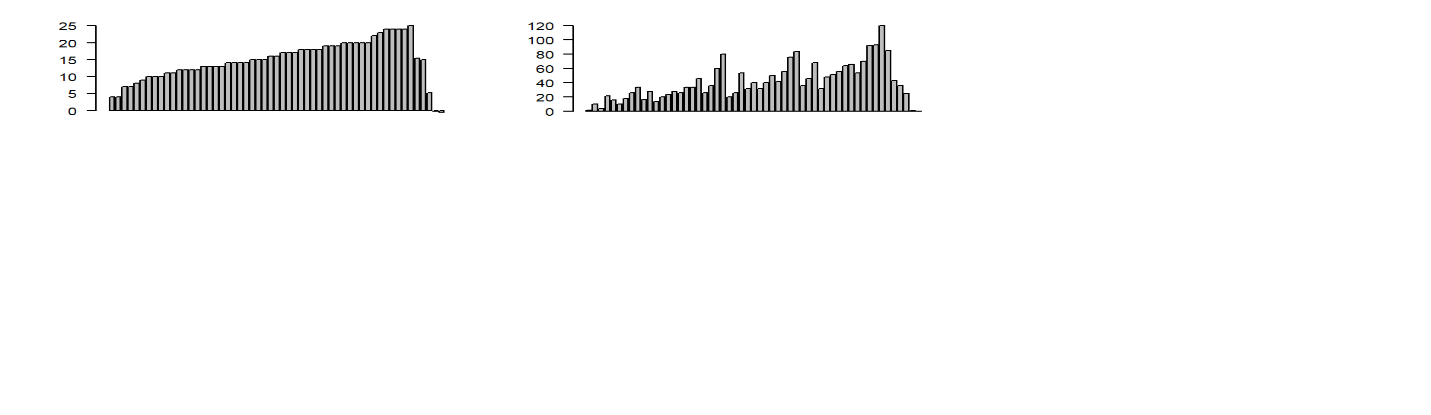
**Answer=Expected Value = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
| skewness | speed-0.11751 | 0.806895 |
| kurtosis | -0.50899 | 0.405053 |

**Inference-Speed-The speed of car in barplot is increase for left to right indicating the present of higher car speed ,going through down the record and the bottom 3 record shown decrease in speed.The mean of speed is slighty towards center with negative skewness**

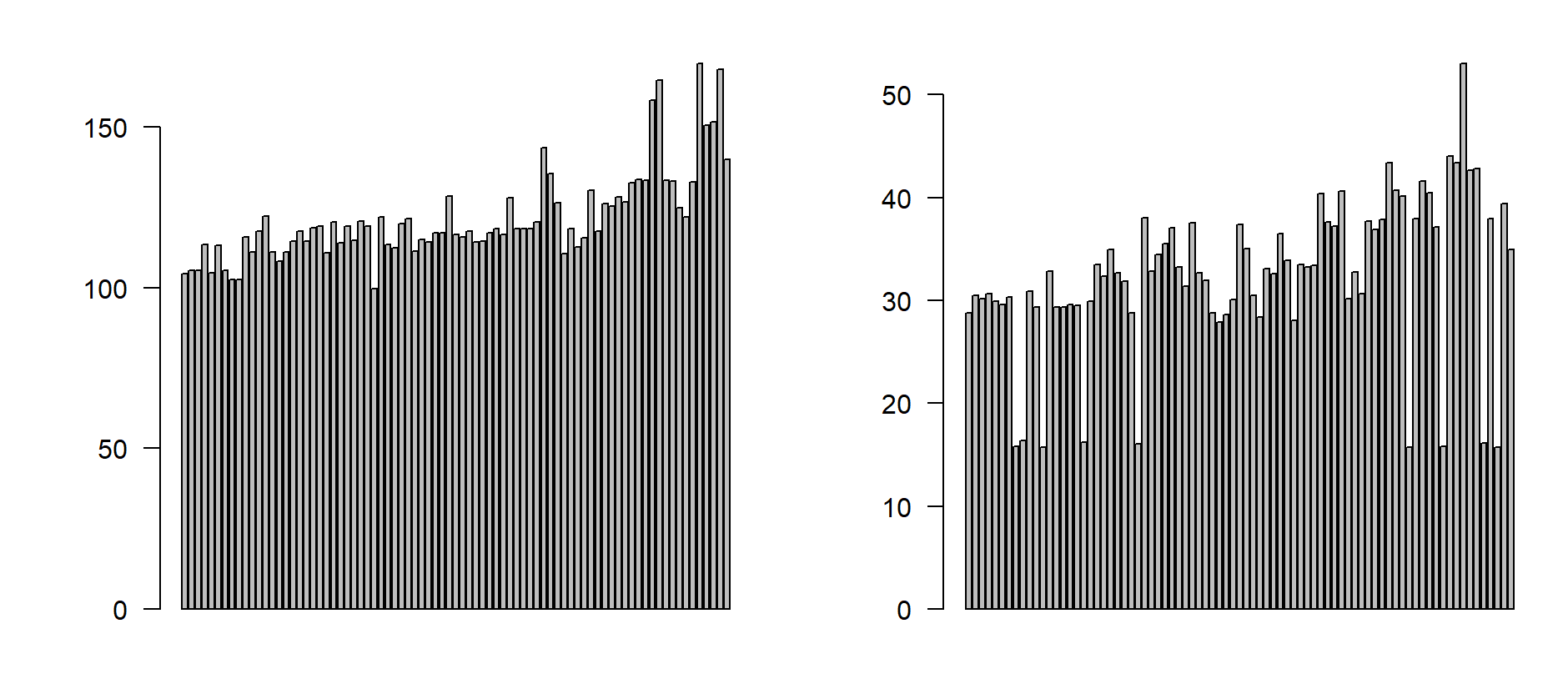
**Distance-The distance plot represent that the distance of car travelled down the record is varying by height of plot going up and down, the record is not in ascending order.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer-**

|  |  |  |
| --- | --- | --- |
| skewness | 1.581454 | -0.60331 |
| kurtosis | 2.977329 | 0.950291 |

s

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



**Answer-The above barchat is having a positive skewness,In the bin interval of 50 to 100 maximum frequency(number are present ,As the weight increases the following frequency decreases.**



**Answer-The above boxplot shown that the median is close to lower quatile ,which represent maximum number present towards lower quatile with positive skewness.The are outlier which are higher than the upper quatile.**

**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?

**Answer- The 94% confidence interval is (198.73, 201.27).**

**The 96% confidence interval is (198.61, 201.39).**

**The 98% confidence interval is (198.43, 201.57).**

**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 |  |  | 40.5 |
| varience |  |  | 24.11 |
| S.D |  |  | 4.91 |

1. What can we say about the student marks?

**Answer-if we consider the bins of interval (38 to 41) maximum student has secured the value.**

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

**Answer-zero.**

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

**Answer-postive.**

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

**Answer-Negative.**

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

**Answer-The maximum value of number given in interval (bins) is maximum compared to other bins.**

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

**Answer-All the values of given data belonging to number of interval (bins) is nearly equal.**

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



What can we say about the distribution of the data?

**Answer-All the given data is concreted towards the right of graph.**

What is nature of skewness of the data?

**Answer-negative**

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

**Answer-8**

Q19) Comment on the below Boxplot visualizations?



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

**Answer-both the box plot median is equal with zero varience and symmetric curve,with 1st box plot all values are close to median and cumulated near median while other graph value are disturbuted far away from median with interval between the value is high.**

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=33/81**

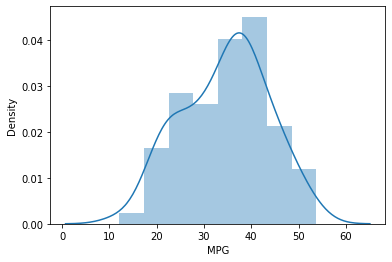
**b=67/81**

**c= 69/81**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

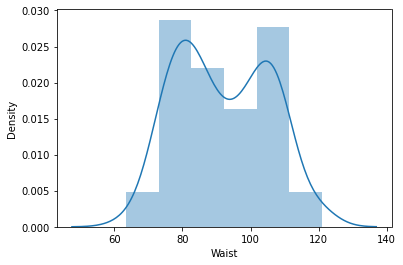
Dataset: Cars.csv



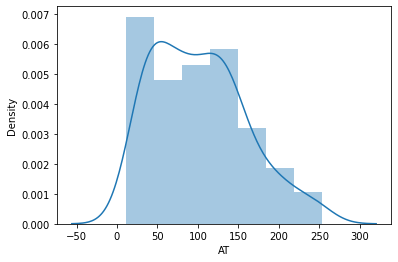
**Answer-The above MPG of cars doesn’t follow the normal disturbution curve because the from peak of curve ,right side =! Left side and mean,mode,median is not equal.**

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

Dataset: wc-at.csv



**Answer-The above Waist of cars doesn’t follow the normal disturbution curve because the from peak of curve ,right side =! Left side and mean,mode,median is not equal.**



**Answer-The above AT of cars doesn’t follow the normal disturbution curve because the from peak of curve ,right side =! Left side and mean,mode,median is not equal.**

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

**Answer- 90%= 1.645**

**94%= 1.880**

**60%= 0.253**

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

**Answer- 95%= 2.064, 96%= 2.172,99%=2.797**

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:

**Answer- t= -0.471, probability for t,(sample of 18 bulbs) is 0.3218**