Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| **Activity** | **Data Type** |
| Number of beatings from Wife | Discrete data type |
| Results of rolling a dice | Discrete data type |
| Weight of a person | Continuous data type |
| Weight of Gold | Continuous data type |
| Distance between two places | Continuous data type |
| Length of a leaf | Continuous data type |
| Dog's weight | Continuous 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) | Categorical data type |

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

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

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

**Ans**: If three coins are tossed; there are 2\*2\*2=8 possible outcomes.

Those are: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

So, three outcomes give two heads and a tail (HHT, HTH, THH).

Therefore, the probability of getting two heads and a tail is 3/8.

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

**Ans**: a) the sum is equal to 1 is zero because they start’s with (1,1) .... likewise. other than in the dice we are not having zero.  
b) the sum is equal to 4 the possible outcomes are (1,3), (2,2), (3,1) therefore n(b) = 3/36.   
c) the sum is divisible by 2 and 3 the possible outcomes are (1,5), (3,3), (4,2), (5,1), (6,6) Therefore n(c)= 5/36.

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?

**Ans**: Total number of balls = (2 + 3 + 2) = 7  
Let S be the sample space.  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
=7C2​  
=(2×1)(7×6)​  
=21  
Let E = Event of drawing 2 balls, none of which is blue.  
∴n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
=5C2​  
=(2×1)(5×4)​  
=10  
∴P(E)=n(S)n(E)​=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 Weight**

**Mean** 3.596563 3.21725 17.84875

**Median** 3.695 3.32517.71

**Mode** 3.92 3.44 17.02

**Standard deviation** 0.534679 0.978457 1.786943

**Range**  2.17 3.911 8.4

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 = ∑ (probability \* Value)

 ∑ P (x). E (x)

There are 9 patients

Probability of selecting each patient=1/9.

Expected Value = (1/9)108+ (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9

= (1308)/9

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**Ans: Skewness Kurtosis**

**Speed -**0.11751 -0.50899

**Distance** 0.806895 0.405053

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans: Skewness Kurtosis**

**SP** 1.61145 -0.61475

**Weight (WT)** 2.977329 0.950291

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



**Ans**: Thefrequency 200 andmost of the data points are concerned in the range 50-100 with frequency 200 and least range of weight is 400 somewhere 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

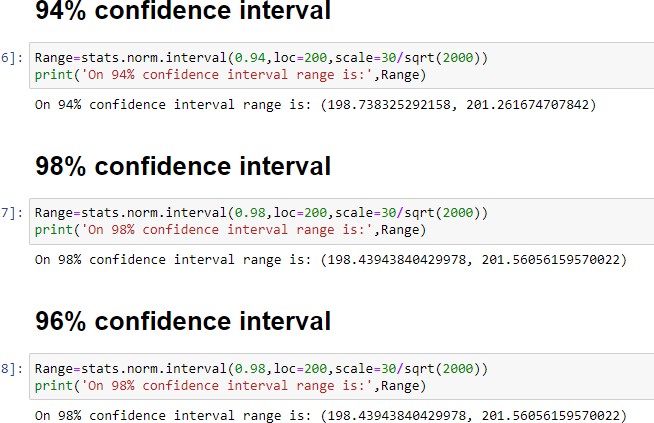


* 1. **Ans**: The box plot shows that greater density data are present towards the left.

Histogram is right or positively skewed and outlier are upper side of boxplot.

**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**:



94% confidence interval:

Range:198.73-201.26

96% confidence interval:

Range:198.43-201.56

98% confidence interval:

Range: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.

**Ans:** Mean: 41.0

Median: 40.5

Standard Deviation: 4.9103

Variance: 24.111

2)What can we say about the student marks?

**Ans**: Most of students marks at the range of between 35-45. It is positive skewness.

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

**Ans:** Data is normalized and there is zero skewness.

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

**Ans:** Positive skewness implies mass of the distribution concentrated on Right side.

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

**Ans**: Negative skewness implies mass of the distribution concentrated on Left side.

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

**Ans**: Positive skewness value indicates that thinner peak.

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

**Ans**: Negative skewness value indicates that thicker peak and wider tails.

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



What can we say about the distribution of the data?

**Ans**: Not normal distribution.

What is nature of skewness of the data?

**Ans:** Negative skewness: For a distribution that is negatively skewed, the box plot will show the median closer to the upper quartile.

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

**Ans**: 10-18=8  
  
Q19) Comment on the below Boxplot visualizations?



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

**Ans**: Box plot1: Q1=250

Q3=275

Q2=265(Boxplot 1 is median)

Box plot2: Q1=225

Q2=300

Q3=265(Boxplot 2 is median)

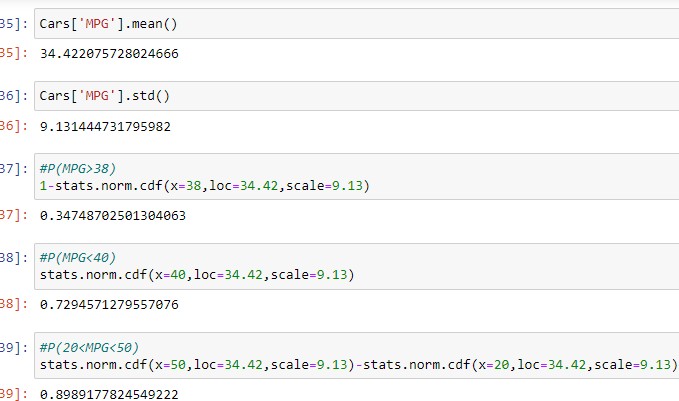
From both boxplots are median.

**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**

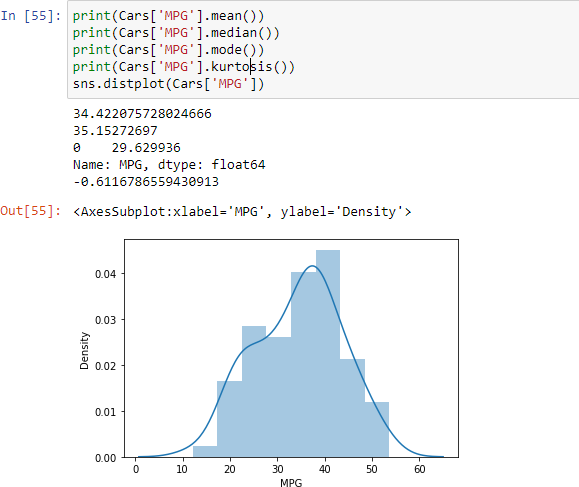


* + P(MPG>38) =0.34
  + P(MPG<40) =0.72
  + P(20<MPG<50) =0.89

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**



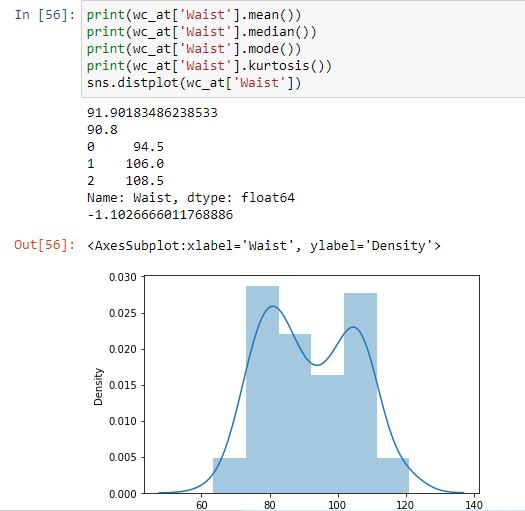
Mean is 34.42, median is 35.15 and mode is 29.62 so this is not normally

distributed.

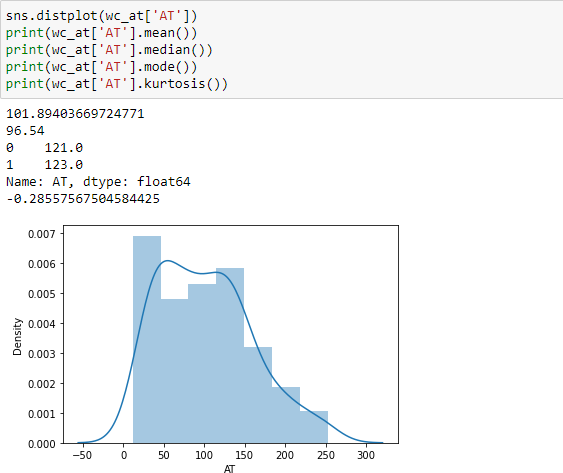
The kurtosis negative and (<3) which means that it is a platykrutic

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

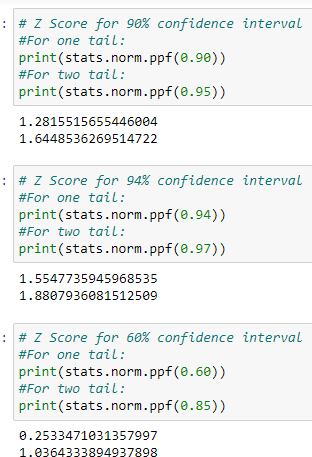
**Dataset: wc-at.csv**



Here the mean is 91.90, median is 90.8 and mode is 108.5 so this is not normally distributed also here the kurtosis is negative and (<3) which signifies that it is a platykrutic

 Here the mean is 101.89, median is 96.54 and mode is 123.0 so this is not normally distributed also here the kurtosis is negative and (<3) which signifies that it is a platykrutic.

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



Z scores of 90% confidence interval: one tail= 1.28,

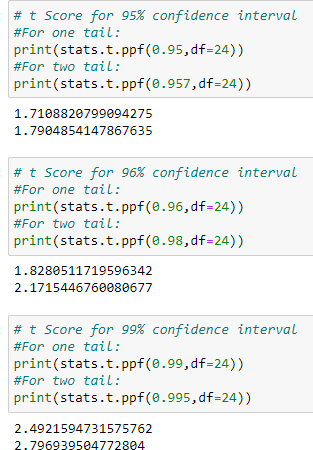
two tail=1.64

Z scores of 94% confidence interval: one tail= 1.55,

two tail =1.88

Z scores of 60% confidence interval: one tail= 0.25, two tail=1.03

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

**Ans**: 

For sample size 25

t scores of 95% confidence interval: one tail= 1.71, two tail=1.79

t scores of 96% confidence interval: one tail= 1.82, two tail=2.17

t scores of 99% confidence interval: one tail= 2.49, two tail=2.79

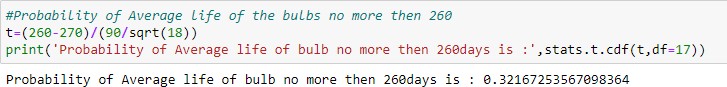
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt (tscore, df)

df 🡪 degrees of freedom

**Ans:**



The probability that 18 randomly selected bulbs would have an avg. life of no more than 260 days is 32.16%.