|  |  |
| --- | --- |
| 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 (Ex- low class, middle class, upper class)** |
| Fahrenheit Temperature | **Interval** |
| Height | **Ratio** |
| Type of living accommodation | **Ordinal** |
| Level of Agreement | **Ordinal** |
| IQ(Intelligence Scale) | **Interval** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Ratio** |
| Time on a Clock with Hands | **Ratio** |
| Number of Children | **Ratio** |
| 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?

**Solution –**

**When three coins are tossed the total number of possible combinations are 2^3 = 8.**

**The Probability of getting two heads and one tails in the toss of three coins simultaneously is 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

**Solution –**

1. **0 (Minimum value in each dice is 1. Thus, the sum of both the dice can never be equal to 1)**
2. **1/6**
3. **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? - **10/21**

**Solution -**

**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  
​  
= (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.  
  
= (2×1) (5×4)​  
=10**

**∴P(E)= n(E)/ n(S) ​=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

**Solution:**



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**

**Solution-**



**Below are some inferences, for the given dataset.**

**1) Points - Mean < Median. The value 2.76 from Dodge Challenger is reducing the overall mean value.**

**2) Score - Mean > Median. The value of 5.424 from Lincoln Continental is increasing the overall mean value.**

**3) Weigh - Mean ~ Median.**

**4) "Weigh" has more variability than "points" and "score".**

***Kindly let me know if I missed any point.***

**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?

**Solution –**



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

Car’s speed and distance

Use Q9\_a.csv

**Solution –**



**Speed -> -0.11 skewness implies majority of the data are concentrated on right side of the bell curved. Also -0.50899 implies the data are flat/ spread around the bell curved.**

**Dist-> 0.806 skewness implies majority of the data are concentrated on left hand side of the bell curved. Also 0.405 implies the data have a bit wider peaks and thinner tails compared to speed data.**

SP and Weight (WT)

Use Q9\_b.csv

**Solution –**



**SP - Majority of the data are on the left-hand side, the data has wider peaks and have thinner tails.**

**WT - Relatively more data lies on the right-hand side of the bell curve. The data have close normal peak.**

***Kindly let me know if I missed any point.***

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



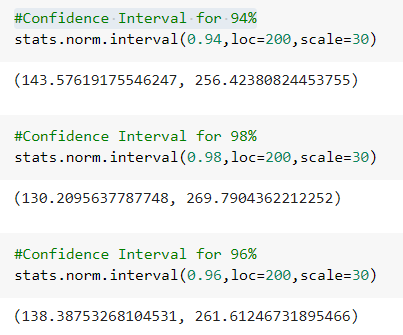
**Answer –**

1. **The above data does not fall under normally distribution.**
2. **The data is positively skewed**
3. **There are more outliers above 3rd quartile.**

***Kindly let me know if I missed any 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?

**Solution –**



**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?

**Solution –**



**The data seems to have 2 outliers – scores 49 and 56.**

**The data seems to be sightly negatively skewed because of 2 outliers. (Mean > median).**

***Kindly let me know if I missed any point.***

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

Answer **- Skewness would be equal to 1. Which is a perfect scenario for a normal distribution.**

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

**Answer- Negatively skewed. Majority of the data or some outlier’s present in the right-side of the bell curve.**

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

**Answer- Positively skewed. Majority of the data or some outlier’s present in the left-side of the bell curve.**

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

**Answer - Positive kurtosis means that it has a sharper peak and heavier tails compared to a normal distribution**

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

**Answer - Negative kurtosis means that it has a flatter peak and thinner tails compared to a normal distribution.**

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



What can we say about the distribution of the data? **Median ~ 15; lower quartile~10; upper quartile~ 18. There are some data points on the left- side which decreases the minimum limit but there is no outliers in the data.**

What is nature of skewness of the data? **Negatively Skewed data**

What will be the IQR of the data (approximately)? - **8**

Q19) Comment on the below Boxplot visualizations?



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

**Answer – The mean of both the data is same. But data 2 have more variability than data 1.**

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.

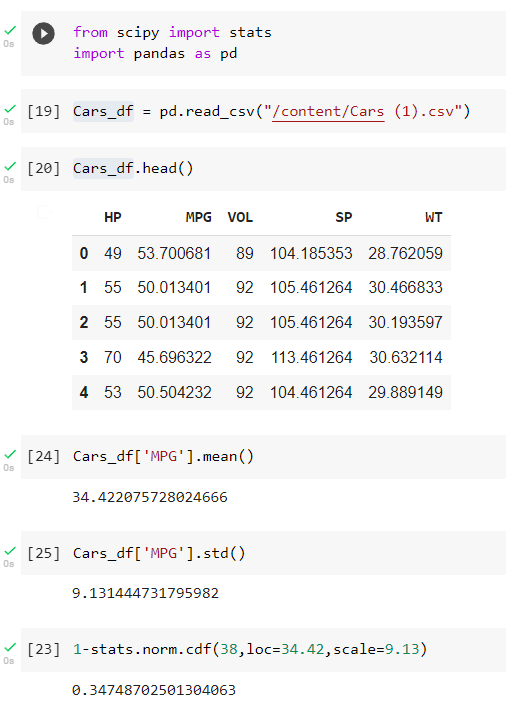
MPG <- Cars$MPG

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

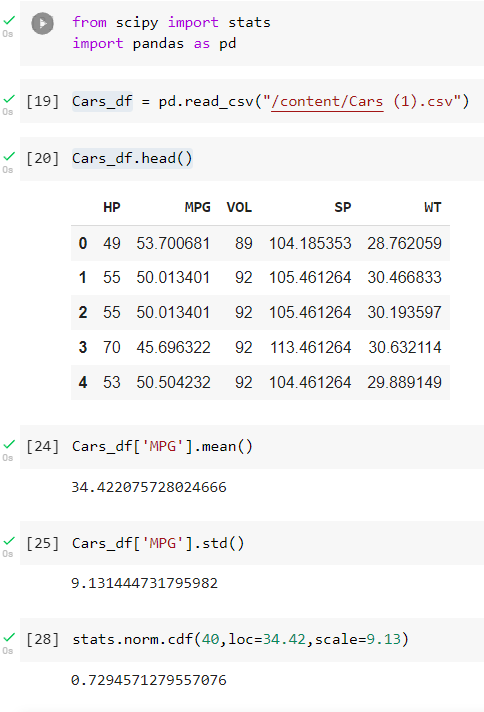
c. P (20<MPG<50)

**Solution-**

1. **P(MPG>38) –> 0.347**



1. **P(MPG<40) -> 0.729**



1. **P (20<MPG<50) -> 0.898**



Q 21) Check whether the data follows normal distribution

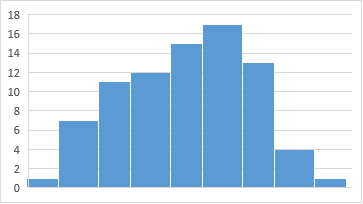
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution -**

**We understand that the data to qualify as Normal Distribution the Mean, Median, Mode should be almost equal and the skewness and kurtosis should be equal to 0.**





**It seems like the data set follows Normal Distribution with little negatively skewness.**

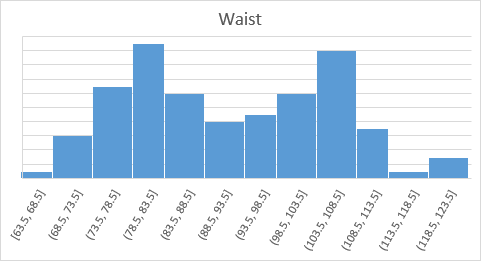
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

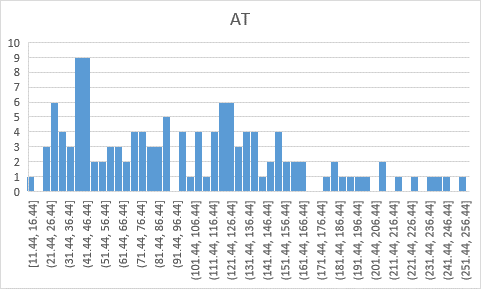
Dataset: wc-at.csv

**Solution-**

**It seems both the data does not follow Normal Distribution. But as per central limit theorem it is understand that the distribution of the sample mean will be normal even if the distribution of data in the population is not normal.**







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

**Answer -**



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

**Answer –**



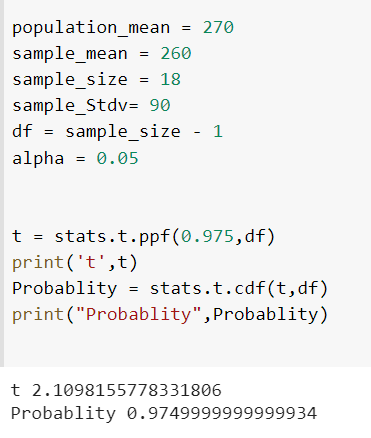
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs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

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Solution –**

****