|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Count data type |
| Results of rolling a dice | Count 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 | Count data type |
| Number of kids | Count data type |
| Number of tickets in Indian railways | Count data type |
| Number of times married | Count data type |
| Gender (Male or Female) | Count data type |

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

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

Ans: Probability =3/8 or 0.375 or 37.5%

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) Probability=0

(b) Probability=1/6

(c) Probability=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: Total number of events= ===21

Interested events===10

Probability that none of the balls 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

Ans: Expected number = E(x)

E(x) =1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | POINTS | SCORE | WEIGHT |
| mean | 3.596563 | 3.21725 | 17.84875 |
| median | 3.695 | 3.325 | 17.71 |
| mode | 3.92 | 3.44 | 17.02 |
| std.dev | 0.534679 | 0.978457 | 1.786943 |
| varience | 0.285881 | 0.957379 | 3.193166 |
| range | 2.17 | 3.911 | 8.4 |

**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?

Ans: EV=Σx/n =145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

**Q 9(a)**

|  |  |  |
| --- | --- | --- |
|  | SP | DIST |
| SKEWNESS | -0.11751 | 0.806895 |
| KURTOSIS | -0.50899 | 0.405053 |

**Q 9(b)**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| SKEWNESS | 1.61145 | -0.61475 |
| KURTOSIS | 2.977329 | 0.950291 |
|  |  |  |

Inferences: As you can see from the above data, there is a huge difference in the kurtosis values when compared with each other. This is due to different equations used by the packages to find kurtosis.

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



Ans: The most of the data points are concentrated 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.



Ans: 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 Estimate= Z

94% confidence =198.74 - 201.26

98% confidence = 198.44 - 201.56

96% confidence = 198.62 - 201.38

**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 |
| MEDIAN | 40.5 |
| VARIANCE | 25.52941 |
| STD DIV | 5.052664 |

1. What can we say about the student marks?

Ans: Students marks between 38-42. Skewness is (1.52) is positive because marks is in left side of plot.

Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.

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

Ans: There is no skewness and data is normalized.

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

Ans: Positive skewness and Right skewed.

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

Ans: Negative skewness and Left Skewed.

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

Ans: It indicates sharper peak and wider tails.

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

Ans: It indicates wider peak and thinner tails.

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



What can we say about the distribution of the data?

Ans: It is not a Normal Distribution

What is nature of skewness of the data?

Ans: It is left skewed and Negative skewness

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

Ans: Inter Quartile Range =Upper Quartile- Lower Quartile

IQR=(Q3-Q1)=18-10=10  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans: 1) The median of the two boxplots are same approximately 260.

2) The boxplots are not skewed in +ve or –ve direction.

3) Outliers doesn’t exist in both of the boxplots.

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)

Ans: a. mean=34.42208

P(MPG>38):

Sd=9.131445

Pnorm=0.652513

=1-P(MPG<38)

=0.35

b.P(MPG<40):

= 0.7294571

c. P (20<MPG<50):

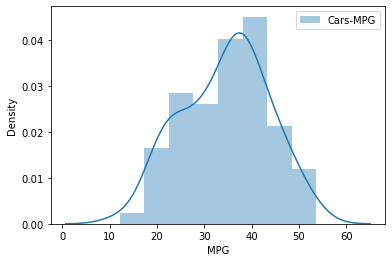
=0.8989178

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Inference: MPG of Cars does follow normal distribution approximately (as mean and median are approx. same)

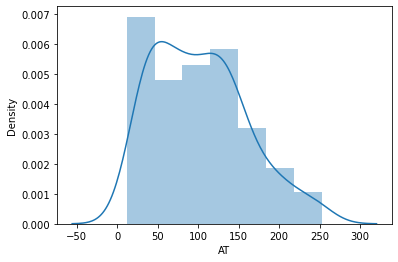
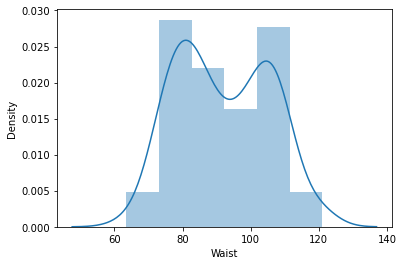


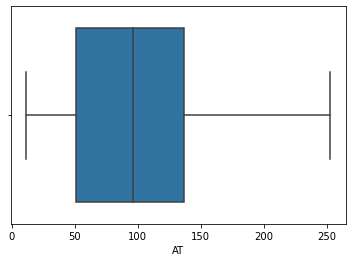
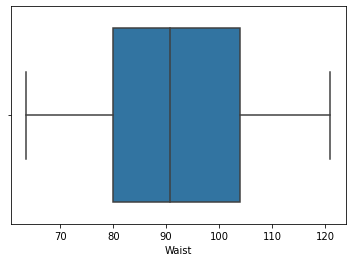
Mean= 34.422075728024666

Median= 35.15272697

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

Dataset: wc-at.csv





Mean>median, both the whisker are of same length, median is slightly shifted towards left. Data is fairly symetrically distributed.

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

Ans: 90% confidence = 1.644854

94% confidence =1.880794

60% confidence =0.8416212

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

Ans: 95% confidence = 2.063899

96% confidence = 2.171545

99% confidence = 2.79694

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: µ=270, =260, SD=90, n=18, df=n-1=18-1= 17

tscore = -0.47

pt(-0.47,17)

Requried probability = 0.32 or 32%