|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numeric Discrete Data XD |
| Results of rolling a dice | Numeric Discrete Data |
| Weight of a person | Numeric Continuous Data |
| Weight of Gold | Numeric Continuous Data |
| Distance between two places | Numeric Continuous Data |
| Length of a leaf | Numeric Continuous Data |
| Dog's weight | Numeric Continuous Data |
| Blue Color | Categorical Data |
| Number of kids | Numeric Discrete Data |
| Number of tickets in Indian railways | Numeric Discrete Data |
| Number of times married | Numeric Discrete Data |
| Gender (Male or Female) | Categorical Data |

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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

**ANS:**- Sample set {HHH,THH,TTH,TTT,HTT,HHT,THT,HTH}

Favorable outcome:- {THH, HHT,HTH}

Probability is 3/8 or 37%

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:-** Sample Set:-

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

1. **None**
2. Favorable outcome={1,1;1,2;1,3;2,1;2;2;3,1}6/36 = 1/6= **16%**
3. Favorable Outcome = {1,5;2,4;3,3;4,2;5,1;6,6}=6/36=**16%**

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 balls= 2+3+2=7 balls

n(S)=Two balls are drawn at random= 7C2= 21

n(A)Number of drawing two balls that are not blue= 2+3=5=5C2=10

Probability (P)=n(A)/n(S)= 10/21= 0.47=**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 |

**Ans:-**

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20 = 4\*0.20= 0.80

Child C- probability of having 3 candies = 3\*0.65= 1.95

Child D- Probability of having 5 candies= 5\*0.005 = 0.025

Child E- probability of having 6 candies= 6\*0.01=0.06

Child F- probability of having 2 candies=2\*0.120=0.240

0.015+0.80+1.95+0.025+0.06+0.240=**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.

Ans:-

|  |  |  |
| --- | --- | --- |
| **Mean** | **Median** | **Mode** |
| Text  Description automatically generated with medium confidence |  | Points 3.92  Score 3.44  Weigh 17.02 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Variance**  **SD**  **Range** | 0.28  0.53  2.17 | 0.95  0.97  3.911 | 3.193  1.78  0.399 |

**Inference**:- Points and score have low variance that means majority of data is near the mean where as weigh has high variance so its data is away from the mean.

**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:-** Expected value (E) = summision of Probability(P)\*value(V)

P of each patients= 1/9 as there are 9 patients

E=P\*V=(1/9)( 108+110+ 123+ 134+ 135, 145+ 167+ 187+ 199)

E=145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:-** Speed negative skew, Distance positive skew

Speed low Peakedness , distance High peakedness

**Inference**:- Speed and distance are inversely proportional to each other

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans** SP Positive skew, Weight Negative skew

Sp and Weight both have High peakedness

**Inference:-** SP and WT both have outliers and both have majority of data in the middle thus giving high peakedness

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



**Inference for HISTOGRAM**

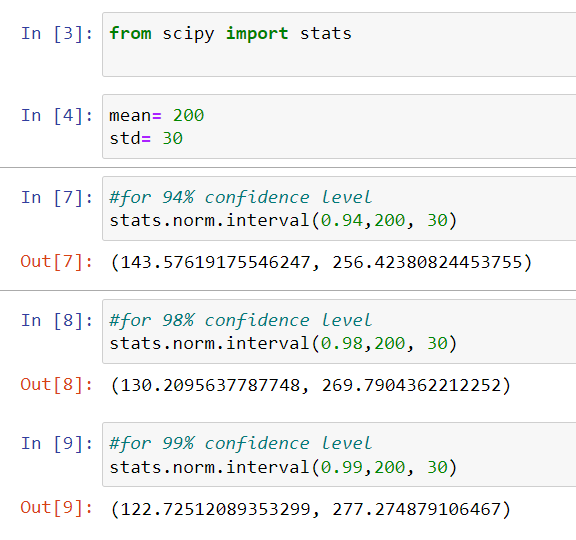
1. Histogram is Right skewed.
2. Maximum data lies between 50 to 100
3. Outliers exists with major variance(400)



1. Outliers exists very far away from the mean.
2. Median is near Upper Quartile and away from Lower Quartile.

**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:-**

****

**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?

Ans:- 1)mean:- 41 , Median:- 40.5, Variance:- 25.25, SD:- 5.05

2 The average marks of the students is 41. There is very less difference between Mean and median so the data has

No outliers.

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

**ANS:-** Zero Skew

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

**ANS:-** Positive Skew

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

**ANS:-** Negative skew

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

**ANS:-** Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. A distribution where more of the values are located in the tails of the distribution rather than around the mean.

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

**ANS:-** A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

**ANS:-** Box plot is not normally distributed. Majority of the data is between 10 to 18. When the median is closer to the top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is negatively skewed. There are no outliers.

What is nature of skewness of the data?

**ANS:-** Left Skew

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

**ANS:-** IQR 8

Q19) Comment on the below Boxplot visualizations?



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

**ANS:-** 1. Median is same for both boxplots

2.Whiskers are equal for each boxplot

3. standard deviation is also same for each box plot

4. They are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

5. There are no outliers .

Q 20) Calculate probability from the given dataset for the below cases

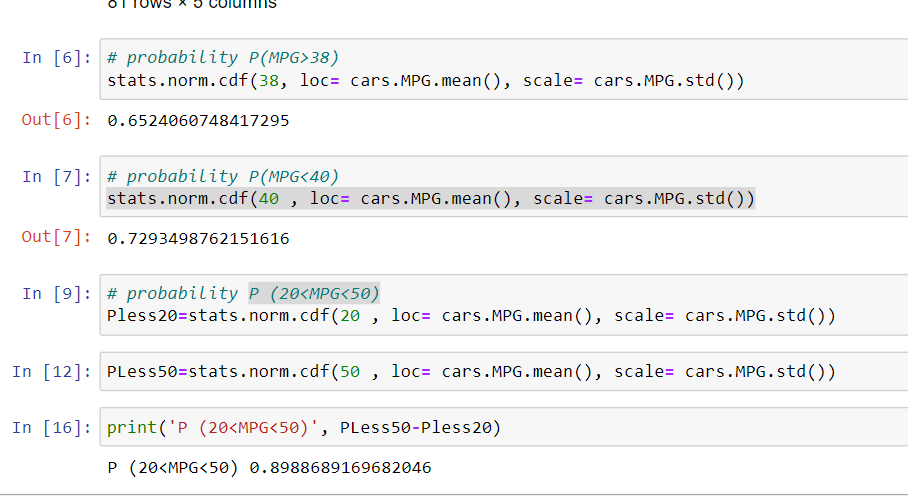
Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**ANS:-**



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:- Yes**

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

Dataset: wc-at.csv

**Ans:- Waist -No , AT- No**

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

**Ans:-**90% confidence interval

1+0.90

1.90/2=0.95

**1.65**

94% confidence interval

1+0.94

1.94/2=0.97

**1.89**

60% confidence interval

1+0.60

1.60/2=0.8

**0.85**

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

**ANS:-** 95% confidence interval

α=1-0.95= 0.05

df=n-1= 25-1= 24

α/2=0.05/2=0.025

t value= **2.064**

96% confidence interval

α=1-0.96= 0.04

df=n-1= 25-1= 24

α/2=0.04/2=0.02

t value= **2.492**

99% confidence interval

α=1-0.99= 0.01

df=n-1= 25-1= 24

α/2=0.01/2=0.005

t value= **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:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

sd=90

n=18

mu=260

x=270

tscore=-0.47

df=17

probability= pt(tscore,df)

Probability=**0.32**