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

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

Ans 3/8

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

1. Equal to 1 = Nil
2. Less than or equal to 4 = 1/6
3. Sum is divisible by 2 and 3= 6/36

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 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=3.14

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(points,score,weigh)=3.59,3.21,17.84

Standard Deviation=053,0.97,1.78

Median=3.69,3.32,17.71

Variance=028,0.95,3.19

**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=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=** kurtosis(speed)=-0.577147

Kurtosis(distance)=0.2480186

Skewness(speed,distance)=-0.11395477,0.7824835

Kurtosis(SP,WT)=2.723521,0.819465

Skewness(SP,WT)=1.5814536,-0.603309932

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

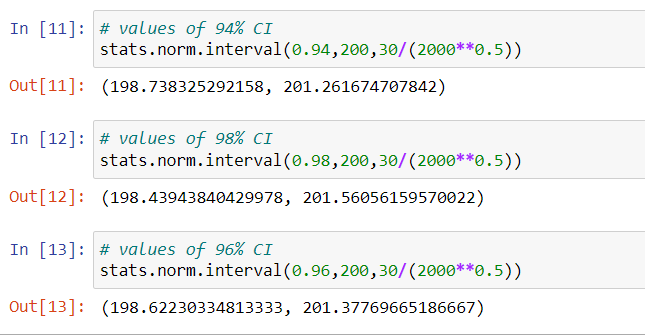


Histogram- The plot is right skewed because mean is greater than median.

Boxplot- The plot is right skewed because mean is greater than the median



**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?



**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=4.10, median=40.5, variance=24.1111,STD=4.9103066

1. What can we say about the student marks?

The least marks scored by a student is 34

The maximum marks scored by a student is 56

Most students scored 41 marks

Average marks is 41

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

It will be a Normal Distribution,Skewness will be zero,If we separate curve by median,both side will be symmetrical

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

When the mean is greater than the median in a distribution, it indicates that the distribution is right-skewed.

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

It will be nnegatively skewed or left-skewed distribution has a long tail.Tail will be on the left side of the distribution. If we separate curve by median, both side will be asymmetrical

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

Positive kurtosis indicates that a probability distribution has heavier tails and a sharper peak than the normal distribution. In statistical terms, kurtosis measures the "tailedness" or the extent to which a distribution deviates from the normal distribution in terms of the distribution of data points.

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

A negative kurtosis value indicates that a probability distribution has lighter tails and a flatter peak than the normal distribution. In statistical terms, kurtosis measures the "tailedness" or the extent to which a distribution deviates from the normal distribution in terms of the distribution of data points.

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



What can we say about the distribution of the data?

The plot is negatively skewed and the lower fence is very high and upper fence is low

What is nature of skewness of the data?

Negatively Skewed

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

IQR=Q3-Q1

=18-10

=8

Q19) Comment on the below Boxplot visualizations?



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

Since mean and median are equal, both the plots are normally distributed.

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)

stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

0.3475939251582705

* 1. P(MPG<40)

stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std()) 0.7293498762151616

c. P (20<MPG<50)

stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())- stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())

1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No, the mpg of cars follows left skewed distribution

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

Dataset: wc-at.csv

Yes, Both are Normally Distributed

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

94%=1.5547

60%=0.2533

90%=1.2815

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

Df=25-1=24

95%=1.71088

99%=2.4921

96%=1.82805

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= xbar=sample mean=260

μ =population mean=270

n= sample size=18

S=sample standard deviation=90

Df=18-1=17

Ho=Avg Bulb life>=260 days

H1=Avg bulb life<260 days

t= (260-270)/ (90/18\*\*0.5)

=-0.4714045207910317

Finding P(X>=260) for null hypothesis

P=p(abs(tscore,df))

=0.32167411684460556

So, the probability will be 0.32167411684460556.