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

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

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

1/2

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

1. Equal to 1🡺0
2. Less than or equal to 4🡪1/6
3. Sum is divisible by 2 and 3🡪11/36

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?

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

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*001+2\*0.120

0.015+0.8+1.95+0.025+0.06+0.24

3.09

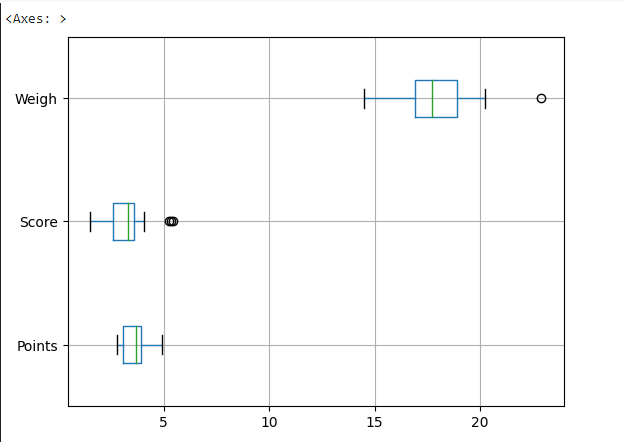
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Points | Score | Weight |  |
| 1 | 3.59 | 3.21 | 17.84 | Mean |
| 2 | 3.69 | 3.325 | 17.71 | Median |
| 3 | 3.92 | 3.44 | 18.9 | Mode |
| 4 | 0.285 | 0.95 | 3.19 | Variance |
| 5 | 0.53 | 0.97 | 1.78 | Standarad Diviation |
| 6 | 2.17 | 3.91 | 8.39 | Range |



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?

108+110+123+134+135+145+167+187+199/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

A number in a row

Description automatically generated with medium confidence

A close-up of a computer code

Description automatically generated

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sol:**

Skewness for speed= -0.1139548, skewness value is negative, so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

And for distance= 0.7824835, right skewed (Positive) slight magnitude to right**.**

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



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.



Medican is less than mean right skewed, and we have outlier on the upper side of box plot and there is less data points.

**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?

Sol: X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval 94% (1-σ/2) = 1-0.03) =0.97

for confidence interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

**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?

Mean= 41, Median= 40.5, variance= 25.52, Standard deviation= 5.05

A graph with green squares

Description automatically generated

A grid with lines and a square

Description automatically generated

From above plot we can say that mean of marks of student is 41 which is slightly greater than median.

#Most of the students got marks in between 41-42, there are two outlier 49,56

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

Normal Distribution

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

Negative Skewed Data

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

Positive Skewed Data

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

K=0

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

K=-1

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



What can we say about the distribution of data?

The data is not actually equally distributed across the plane. There might be outliers influencing the data . Median of the data is 14.7(app x)

25 percent of the data lies between 0-10

50 percent of the data lies between 10-18

25 percent of the data lies after 18-20 appx

What is the nature of skewness of the data?

The data will be left skewed since whisker length on the upper quadrant is higher than the data on the lower quadrant. Median will be greater than the mean since data is left skewed.

Left skewed, median is greater than mean.

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

Q1 = 10

Q2 = 14.7

Q3 = 18

IQR = Q3 –Q1 = 8(approx.)

Q19) Comment on the below Boxplot visualizations?

Here there is a representation of 2 box plots in which box plot 2) is highly distributed across the plane and 1) is slightly less distributed.(variance)

Whiskers in these diagrams also show this.100% of the data is spread across values from 350 in 2 whereas its spread in range 250-290 app x in 1)



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

Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots is equal. And the data spread on both 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)
  2. P(MPG<40)

c. P (20<MPG<50)

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

Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varieshigh in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots are equal. And the data spread in both of them are symmetrical

Q 20) Calculate probability from the given data set for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

a. P(MPG>38)

b. P(MPG<40)

c. P (20<MPG<50)

Solution:

P(MPG>38) = mean (MPG)=34.42208= sd (MPG)=9.131445= 1–

pnorm (38, mean (MPG), sd (MPG)) =0.330=33%

P(MPG<40) =pnorm (40, mean (MPG), sd (MPG)) =0.7293499=

72.3%

P (20<MPG<50) =pnorm (50, mean (MPG), sd (MPG))–pnorm(20,mean(MPG),sd(MPG))=0.955 -0.057=0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

A graph with blue bars

Description automatically generated

A graph with a line

Description automatically generated

Skew: -0.177

Kurtosis: -0.61

From above plot and values, we can say that data is symmetrical, i.e. fairly normally distributed.

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

Dataset: wc-at.csv

A graph of blue bars

Description automatically generated

A graph with a line

Description automatically generated

Skew:

0.1340560824786468

Kurtosis:

-1.1026666011768886

As per above plot and values Waist does not fall in Normal Distribution.

A graph of a bar graph

Description automatically generated with medium confidence

A graph with a line

Description automatically generated

Skew:

0.58

Kurt:

-0.28

As per above plot and values of AT does not fall in Normal Distribution.

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

A screenshot of a computer

Description automatically generated

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

T ((1, alpha),(n-1))

Here n = 25

n-1 = 24

Hence t score values will be:

95%= qt (0.975,24) = 2.063899

96% = qt (0.98,24) =2.171545

99%= qt (0.995,24)= 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

Solution:

Sample size = 18 = n

Sample mean = 260 days = x

Sample standard deviation = s = 90days

=260–270/90/SQRT (18)

= -10/9.487

=-1.054