|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| 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?

🡺 For above situation sample space will be

S={ HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

And the probability that that two heads and one tail are obtained is 3/8.

**P(x) = 3/8=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

🡺 The sample space will be

S={(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)}

a)Probability = number of favorable outcomes / total no. of outcomes

**P(Sum equal to 1)= 0/36=0**

b)🡺

**P(sum is Less than or equal to 4) = 6/36 = 0.16**

c)🡺

**P(sum is Sum is divisible by 2 and 3)=6/36 = 0.16**

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?

🡺

There are total 7 balls and 2 balls are to be drawn at random therefore the number of elements in in sample space will be

n(S)=7C2 =21.

Now to find the probability that none of the balls drawn is blue:

We will subtract the no. of blue balls from total balls and the sample space for that will be:

n(x)=5C2=10

let E be a event of drawing 2 balls none of them is blue:

**p(E)=10/21=0.476**

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

🡺 Expected number of candies for a randomly selected child is 3.090

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

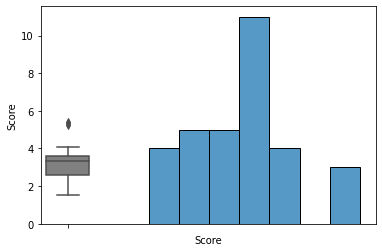
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences

**Use Q7.csv file**

**🡺**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **points** | **scores** | **weigth** |
| **mean** | *3.596563* | *3.217250* | *17.848750* |
| **median** | *3.695* | *3.325* | *17.710* |
| **mode** | *3.891* | *3.54* | *17.43* |
| **variance** | *0.285881* | *0.957379* | *3.193166* |
| **std** | *0.534679* | *0.534679* | *1.786943* |
| **range** | *2.76,4.93* | *1.513,5.424* | *14.5,22.9* |

****

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?

🡺 Expected Value  =  ∑ ( probability  \* Value )

There are 9 patients therefore probability will be 1/9.

Now Expected value = (1/9) (1308)

= 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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

🡺**Skewness for cars speed is -0.1175**

**🡺 Skewness for cars distance is 0.8068**

**🡺Kurtosis for car speed is -0.5089**

**🡺Kurtosis for car distance is 0.4050**

**🡺Skewness for SP is 1.6114**

**🡺Skewness for WT is -0.6147**

**🡺Kurtosis for SP is 2.9773**

**🡺Kurtosis for WT is 0.9502**

**#** *From the value of skewness and kurtosis the data for the car speed is positively skewed which means more datapoints are on left side .it has wider peak and thinner tails.*

#*For car distance the case is opposite the data is negatively skewed and has low peakedness.*

# *For SP data is negatively skewed and has low peakedness.*

# *For WT data is positively skewed and has high peakedness*

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



🡺

1)From the above histogram it is clear that it is positively skewed which means most of the datapoints are on the left side.

2)The column between 50 and 100 has most datapoints.

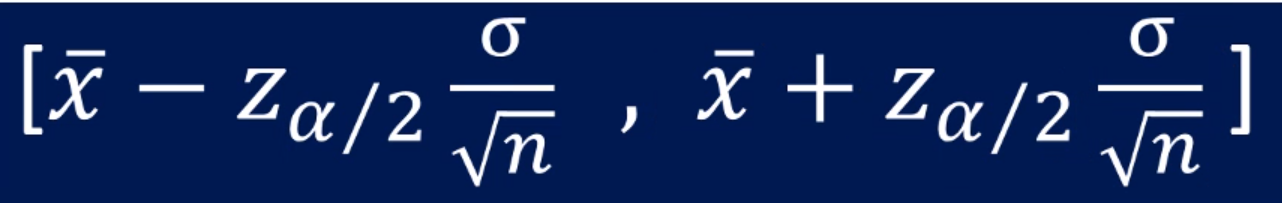
3) Frequency decreases as the weight increases.

🡺

1)From the boxplot it is clear that it has several outliers .

**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?

🡺The value of confidence interval is=



**Where x= sample mean**

**Z=z value**

σ = the population standard deviation

√n = the square root of the population size

Now for 94%, zvalue is(1.89),

CI=[200- (1.89)(0.67082), 200 + (1.89)(0.67082)

**CI=[198,201.26]**

Similarly for 96% , zvalue is (2.05),

**CI = [198.624,201.375]**

Similarly for 98% , Zvalue is (2.362),

**CI=[198.439,201.560]**

**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 is 41

🡺Median is 40.5

🡺Mode is 41

🡺Variance is 25.529

🡺Standard deviation is 5.052

🡺From the student marks we can conclude that he/she is not a bright student ,the mean is 41 which means he/she is hardly just passing his/her subjects .The student must work hard.

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

🡺**Data is normal ,there is no skewness.**

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

🡺**Data is negatively skewed ,most of the data points are on right side.**

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

🡺**Data is positively skewed, most of the data points are on left side.**

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

🡺**It indicates sharper peak and wider tails.**

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

🡺**It indicates flatter peaks and lighter tails.**

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



What can we say about the distribution of the data?

🡺*The data is not normally distributed.*

What is nature of skewness of the data?

🡺*The data is negatively skewed.*

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

🡺*IQR=Q3-Q1*

*=17-12*

*=5*

Q19) Comment on the below Boxplot visualizations?



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

🡺**1)The median of both the datasets is equal to 260.**

**2)The first visualization has very less no. of datapoints.**

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)

🡺***The probability of MPG of cars more than 38 is 0.34759(32%)***

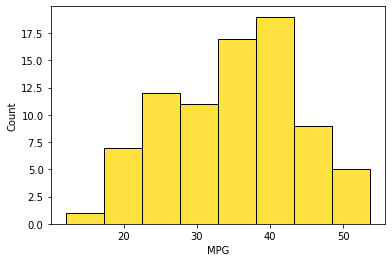
***🡺 The probability of MPG of cars less than 40 is 0.7293(72%)***

***🡺 The probability of MPG of cars between 20 and 50 is 0.01311(1.3%)***

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

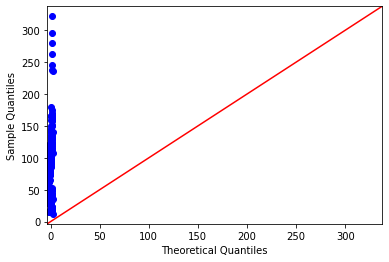
Dataset: Cars.csv

🡺 

From the plotted histogram the data looks 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

🡺 

The data is not normally distributed.

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

🡺

|  |  |
| --- | --- |
| CI | Z score |
| 90% | 1.644854 |
| 94% | 1.880794 |
| 60% | 0.841621 |

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

🡺*The sample size has 25 elements.*

*n = 25, now the degree of freedom is n-1 =24.*

*Using t table we can get the value of tscore.*

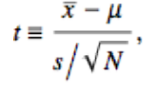
|  |  |
| --- | --- |
| CI | T scores |
| 95% | 2.06390 |
| 96% | 2.171545 |
| 99% | 2.79694 |

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

🡺The formula for t value is 

The t value is -0.47147

*t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218.*

***Hence, if the true bulb life were 270 days, there is a 32% chance that the average bulb life for 18 randomly selected bulbs would be less than or equal to 260 days.***