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

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

3) Ans: Probability of 2 heads and 1 tail = 3/8

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

4) Ans:

a) P(sum equal to 1) = 0

b) P(sum <= 4) = 6/36 = 1/6

c) P(sum is divisible by 2 and 3) = 6/36 = 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?

5) Ans: P(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

Ans) Expected number of candies = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=0.015+0.8+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.

**Use Q7.csv file**

Ans 7) Mean: Points = 3.596563 ,Score = 3.217250, Weigh = 17.848750

Median: Points =3.695,Score = 3.325,Weigh = 17.710

Mode: points = 3.07 and 3.92 , score = 3.44, weigh = 17.02 and 18.90

Variance: Points = 0.285881, Score = 0.957379, Weigh = 3.19316

Standard Deviation: Points = 0.534679, Score = 0.978457,

Weigh = 1.786943

Range: Points(2.76-4.93), score(1.51-5.424), weigh(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?

Ans 8) mean = 145.33

Expected values = 0.111\*108 + 0.111\*110 + 0.111\*123 + 0.111\*134 + 0.111\*135 + 0.111\*145 + 0.111\*167 + 0.111\*187 + 0.111\*199

= 12 + 12.21 + 13.653 + 14.874 + 14.985 + 16.095 + 18.537 + 20.757 + 22.089

= 145.2

**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 9) a)

SKEWIndex 0.000000

speed -0.117510

dist 0.806895

KURTOSIS : Index -1.200000

speed -0.508994

dist 0.405053

SKEW: Unnamed: 0 0.000000

SP 1.611450

WT -0.614753

KURTOSIS: Unnamed: 0 -1.200000

SP 2.977329

WT 0.950291

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



ANS ) This histogram shows that its right skewed , which means mean>median. The highest freguency of weight is in range 50-100 having 200 frequency.



Ans ) From the above boxplot fig, we can see that there are plenty of outliers which needs to be treated . The data is normally distributed.

**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?

ANS11) 1)200±1.262= 201.262 or 198.738 at 94%

2) 200±1.5618= 201.561 or 198.438 at 98%

3) 200±1.3784= 201.3784 or 198.621 at 96%

**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 12)

1. Mean= 41, median= 40.5 , variance= 24.111, std= 4.91
2. We can say that from 18 students, average marks scored by students is 41, middle marks is 40.5 .

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

Ans 13) When mean, median of data are equal than the nature of skewness would be zero as the data would be symmetric . We called it as perfect symmetric.

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

Ans 14) It will be right skewed (+ve skew) if mean > median

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

Ans 15) It will be left skewed (-ve skew) if median >mean

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

Ans 16) Positive kurtosis indicates that sharp peak for the data, and mostly negative skewed .

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

Ans17) Negative kurtosis indicates that flatten peak for the data , and mostly positive skewed.

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



What can we say about the distribution of the data?

Ans) We can clearly see that the data are not distributed normally which returns skewness factors.

What is nature of skewness of the data?

Ans) From the above fig we can figure it out that its left skewness or say negative skewed.

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

Ans) IQR inter quartile range = 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.

Ans 19) From above boxplot representation we can figure it out that both the boxplot have same median but there inter quartile varies which shows that boxplot 1 has lesser IQR than the boxplot 2 IQR.

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)

Ans: 0.347487

* 1. P(MPG<40)

Ans: 0.7294571

* 1. P (20<MPG<50)

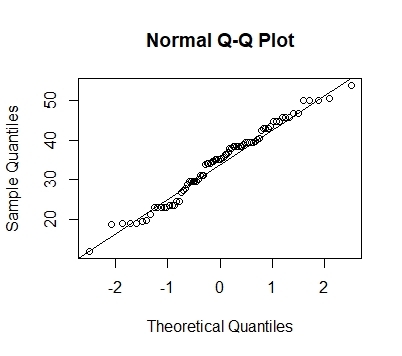
Ans: (0.9428788<MPG<0.956039)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

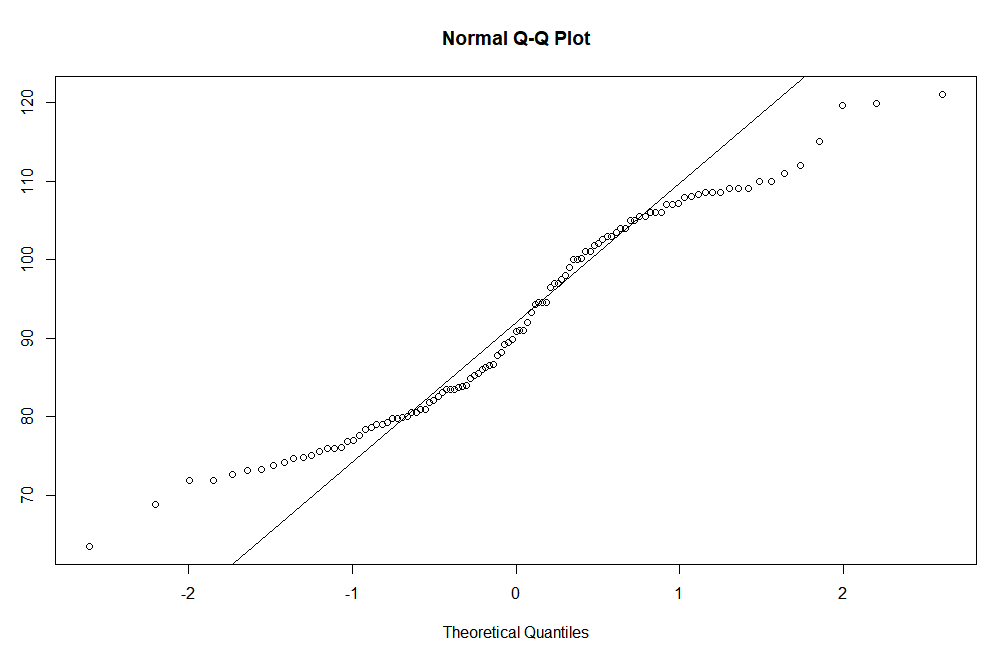
Ans21): a) From the qqplot we can say data is normally distributed



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

Dataset: wc-at.csv

Ans b) It’s normally distributed.



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

Ans 22) z scores at confidence level 90% = 1.6448536269514722

z score at confidence level 95% = 1.8807936081512509

z score at confidence level 60% = 0.8416212335729143

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

Ans 23) t score at confidence level 95% and sample size 25

= 2.0638985616280205

t score at confidence level 96% and sample size is 25

=2.1715446760080677

t score at confidence level 99% and sample size is 25

=2.796939504772804

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 23): The probability that 18 randomly selected bulbs would have an average life of no more than 260 days = 0.05