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

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

**Ans :** P(two heads and one tail)=3/8=0.375

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

Possible Outcomes

N= {(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. Equal to 1

**Ans :** P =0

1. Less than or equal to 4

the possible outcomes are (1,1),(1,2),(1,3) (2,1),(2,2), (3,1)

**Ans :** P = 6/36 = 0.167

1. Sum is divisible by 2and 3

possible outcomes are

{(1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6), (3,1),(3,3),(3,5),(3,6),(4,2),(4,4),(4,5),(4,6),(5,1),(5,3),(5,4),(5,5), (6,2),(6,3),(6,4),(6,6)}

**Ans :** P = 2/3 = 0.66

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 no of balls=7 balls

n(S)= 2 balls drawn out of 7

7C2=21

n(S1)= 2 balls drawn non of is blue

5C2=10

P(S1)=7C2/5C2 = 21/10

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 :**

**E=Expected number of candies for a randomly selected child**

P(E)= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

P(E)=3.09

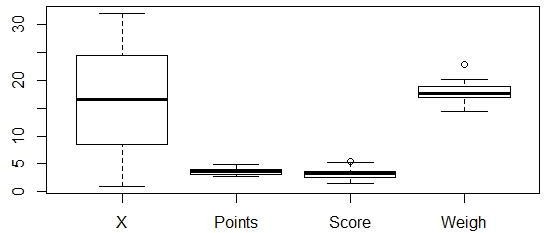
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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:

|  |  |  |  |
| --- | --- | --- | --- |
|  | POINT | SCORE | WEIGHT |
| Mean | 3.597 | 3.217 | 17.849 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.920 | 3.440 | 17.020 |
| Var | 0.286 | 0.957 | 3.193 |
| Std | 0.535 | 0.978 | 1.787 |
| Range | (2.76,4.93) | (1.1513,5.424) | (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?

=Mean(weight) = 145.33

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

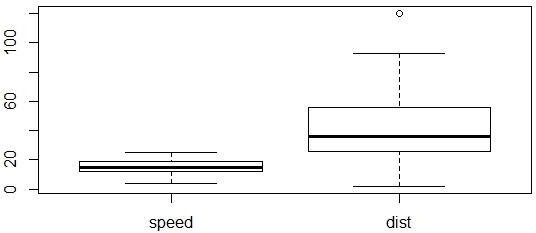
**Cars speed and distance**

**(Use Q9\_a.csv)**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | speed | dist |
| Skewness | -0.114 | 0.782 |
| Kurtosis | -0.577 | 0.248 |

**Inferences for data :**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans :**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.581 | -0.603 |
| Kurtosis | 2.723 | 0.819 |

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



**Ans :**

1.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.

4.From the boxplot it is clear that it has several outliers

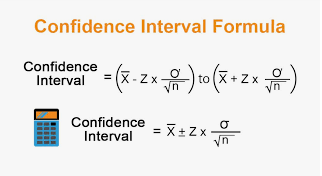
**Q11)**Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans :

n=2000

X = 200

y = 30



1)Confidence Interval = 94%

**Ans:**

Z Value : 1. 880794

Range : (1.9874, 201.26)

2) Confidence Interval = 96%

**Ans:**

Z Value : 2.053749

Range : (198.62,201.38)

3) Confidence Interval = 98%

Ans:

Z Value : 2.326348

Range : (198.43,201.56)

**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?

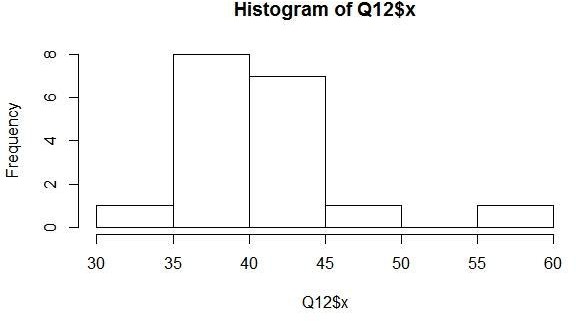
**Ans :**

Mean : 41

Median : 40..5

Variance : 25.53

Standard Deviation : 5.05



The mean and median are almost same and The changes in each point from mean is 25.53 so the students of this class are not good enough in studies or their teacher is not teaching them properly

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

**Ans**: Skewness can be positive, negative or zero when the value of mean, median are equal there is no skewmess

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

**Ans** : The distribution is Negative skewed.

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

**Ans** : The distribution is Positive skewed.

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

**Ans** : Positive kurtosis value indicates that thinner peak and wider.

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

**Ans** : Negative kurtosis value indicates that wider peak and thinner tails.

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



What can we say about the distribution of the data?

**Ans** : Not normally distributed

What is nature of skewness of the data?

**Ans** : Negative skewness

What will be the IQR of the data (approximately)?   
**Ans** : IQR between 10-18

Q19) Comment on the below Boxplot visualizations?



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

**Ans :** The mean of the both box plot is same but the IQR range for both is not same and 1st dataset have less data as compare to 2nd dataet

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)

**Ans :** a.

1-pnorm(38,34.422,9.13144)=0.34759

probability of MPG of cars more than 38 is 0.34759(34%)

**Ans :** b.

pnorm(40,34.422,9.13144)=0. 7293

probability of MPG of cars less than 40 is 0.7293(73%)

**Ans :** c.

pnorm(50,34.422,9.13144)-(1-pnorm(20,34.422,9.13144))=

0.01311

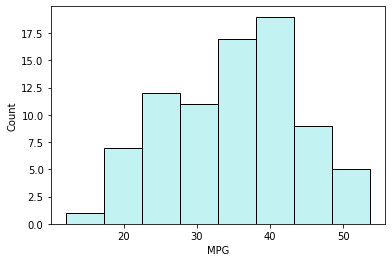
probability of MPG of cars between 20 and 50 is 0.01311(1.3%)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

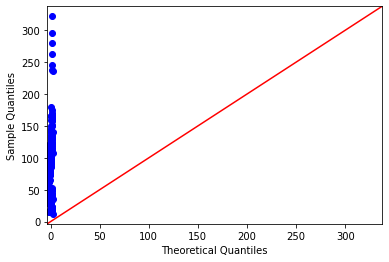
Ans.



From above histogram we can say that data is normally distributed

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

Ans.



Data is not Normally Distributed

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

Ans.

1. stats.norm.ppf(0.95)

At 90% C.I, Z Score is 1.644

1. stats.norm.ppf(0.97)

At 94% C.I, Z Score is 1.880

1. stats.norm.ppf(0.8)

At 60% C.I, Z Score is 0.841

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

Ans.

1. stats.t.ppf(0.975,24) # df = n-1 = 24

At 95% C.I, T Score is 2.063

1. stats.t.ppf(0.98,24)

At 96% C.I, T Score is 2.171

1. stats.t.ppf(0.995,24)

At 99% C.I, T Score is 2.796

Q 24**)**A Government companyclaims 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 :**

t = x- μ/ s/sqrt.n

x=260

μ =270

s=90

n=18

t=(260-270/90)/squrt.18

t=(-10\*squrt.18)/90

t=(-1\*squrt.2)/3

t=-0.471

from scipy import stats

stats.t.cdf(- 0.471,17)

Probability of bulb lasting less than 260 days is 0.3218 or 32.1%

Hence, if a bulb lasts 270 days then there is 32% chance if a bulb picked up from 18 randomly selected bulb would be less then or equal to 260 days.