|  |  |
| --- | --- |
| 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?

Answer :-

When three coins are tossed the possibility of getting two head and one

tail are as follows :-

(T,H,H) ; (H,T,H) ; (H,H,T) .

= 1/8 + 1/8 + 1/8

n(P) = 3/8 OR 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

Answer :-

a) Equal to 1:- There is no possibility of getting sum equal to one. Therefore P(a) = 0.

b) Less than or equal to 4 :-

b{(1,1) ; (1,2); (1,3) ; (2,1) ; (2,2) ; (3,1) } = n(b) =6.

P(b) = 1/6.

c) Sum is divisible by 2 and 3 :- n(P) =5/36.

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?

Answer :-

Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space

(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)=(7×6)(2×1)

n(S)=21

E = Event of 2 balls, none of which is blue

n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)=(5×4) / (2×1)

n(E)=10

P(E)=10/21.

Therefore the probability of getting no blue ball is 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

Answer :-

Expected number of candies for a randomly selected child

=  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

=  3.09.

Expected number of candies for a randomly selected child is 3.09.

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

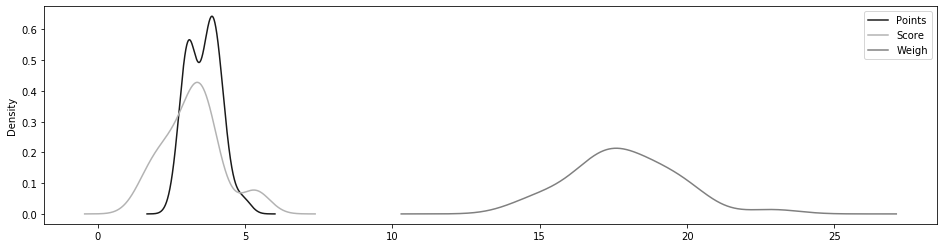
* For Points,Score,Weight>

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

**Use Q7.csv file**

**Answer :-**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weight** |
| **Mean** | **3.5965** | **3.2172** | **17.8487** |
| **Median** | **3.695** | **3.325** | **17.71** |
| **Mode** | **3.92** | **3.44** | **17.02** |
| **Variance** | **0.2858** | **0.9573** | **3.1931** |
| **Standard Deviation** | **0.5346** | **0.9784** | **1.7869** |
| **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?

Answer :-

Expected Value  =  ∑ ( probability  \* Value )

  = ∑ P(x)\*E(x)

there are 9 patients

Probability of selecting each patient = 1/9

E(x):-  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x):-  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)\*108 + (1/9)\*110  + (1/9)\*123 + (1/9)\*134 + ( 1/9)\*135 + (1/9)\*145 + (1/9\*167 + (1/9)\*187 + (1/9)\*199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33 pounds

Expected Value of the Weight of a patient is = 145.33 pounds.

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

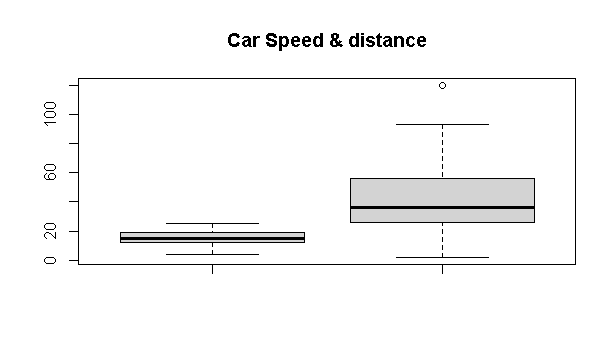
**Cars speed and distance**

**Use Q9\_a.csv**

**Answer :-**

* **Skewness & Kurtosis of given data of car speed and distance =**

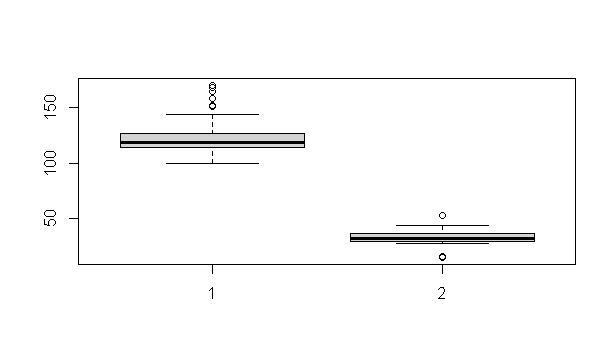
|  |  |  |
| --- | --- | --- |
|  | **speed** | **distance** |
| **skewness** | **-0.1139548** | **0.7824835** |
| **Kurtosis** | **2.422853** | **3.248019** |

****

**SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | **1.5814537** | **-0.6033099** |
| **Kurtosis** | **5.723521** | **3.819466** |

**Use Q9\_b.csv**

****

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



**Answer :-**

Histogram:-

Chick weight data is right skewed or positively skewed,

More than 50% Chick Weight is between 50 to 150.,

Most of the chick weight is between 50 to 100.



Boxplot :-

The data is right skewed,

There are outliers at upper side. 7 outliers are present in above box plot.

**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?

Answer :-

Sample mean = 200

Sample SD = 30

n = 2000

* Average weight of an adult male in Mexico with confidence interval 94% :-

CI range between - 198.738325292158 to 201.261674707842.

* Average weight of an adult male in Mexico with confidence interval 98% :-

CI range between - 198.43943840429978 to 201.56056159570022.

* Average weight of an adult male in Mexico with confidence interval 96% :-

CI range between - 198.62230334813333, 201.37769665186667.

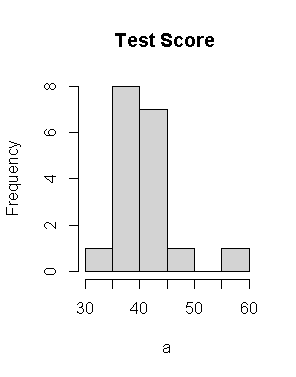
**Q12)** Below are the scores obtained by a student in tests

**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?

Answer:- 1)

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| Standard deviation | 5.0526 |



2)Avg of student marks 41.

The student markes range from 34 to 56.

Mode is 41

Most of students score is bw 35 to 42

2) Mass of students marks between 38-42. Skewness is 1.52 is positive because mass of marks in left side of histogram.

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

Answer :-

If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness. Therefore, if the mean and median are equal there will be No Skewness.

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

Answer :-If the mean is greater than the median, the distribution is positively skewed.

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

Answer :- If the mean is less than the median, the distribution is negatively skewed.

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

Answer :-

Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

Answer :-

Negative values of kurtosis indicate that a distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

Answer:- The data is not in Normally Distributed.

What is nature of skewness of the data?

Answer:- The data is negatively-skewed.

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

Answer :- The IQR of the data 10 to 18 approximately.

Q19) Comment on the below Boxplot visualizations?



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

Answer :-

Both the plots infer that their data is normally distributed.

We can say that box plot 1 is for sample distribution and box plot 2 is for population or a sample with larger size.

No outliers

Q1 is 25%, Q3=75%. IQR is 50% for both the box plots. So we can say both the distribution s follow normal distribution ( mean=median=mode).

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)

Answer:-

P(MPG>38) = 0.3475908

P(MPG<40) = 0.7293527

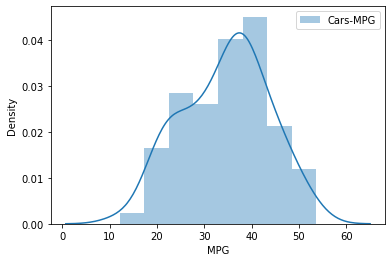
P (20<MPG<50) = 0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

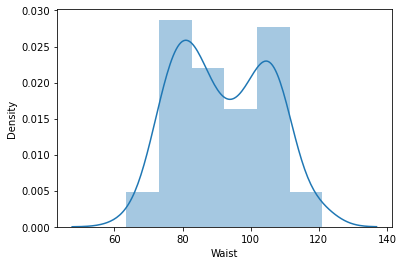
Answer :-a) Answer:- MPG of cars not following normal distribution



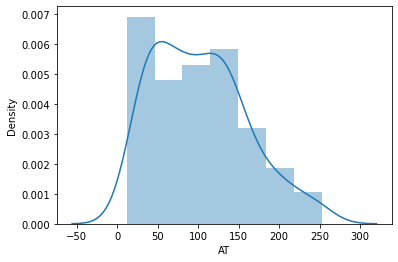
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer :- Waist & AT does not follow normal Distribution



Adipose Tissue (AT) Normal Distribution :-



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

Answer :-

* Z scores of 90% confidence interval :- 0.8416212
* Z scores of 94% confidence interval :- 1.644854
* Z scores of 60% confidence interval :- 1.880794

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

Answer :-

* T scores of 95% confidence interval :- 2.063899
* T scores of 96% confidence interval :- 2.063899
* T scores of 99% confidence interval :- 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

Answer :-

Population mean =270 days

Sample mean = 260 days

Sample SD = 90 days

Sample n = 18 bulbs

df = n-1 = 17

t= {(260-270) / (90/√18)}

t=(-1 \* √2) / 3

t= - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

`