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

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

Solution:

The total possible outcomes 2n=23=8

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

No. of favorable outcomes=3

{HHT+HTH+THH}

P(Two head and one tail)=3/8=0.375

Therefore probability of two heads and one tail is 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

Solution:

The total possible outcomes =36

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)}

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?

Total number of balls = (2 + 3 + 2) = 7  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
= 7C2​  
= (2×1)(7×6)​  
= 21  
Let E = Event of drawing 2 balls, none of which is blue.  
∴ n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
=5C2​  
=(2×1)(5×4)​  
=10  
∴P(E)=n(S)n(E)​=2110

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

Solution:

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

=   3.090

=  3.09

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

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?

Solution

Expected Value  =  ∑ ( probability  \* Value )

∑ P(x).E(x)

there are 9 patient

Probability of selecting each patient = 1/9

Ex  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

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**

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



**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?

Here,

Sample mean = 200

Sample standard deviation = 30

Sample size = 2000

The 98% confidance interval is (198.43,201.57).

**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?

Solution:

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

Solution: Symmetric

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

Solution: Positively skewed

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

Solution: Negatively skewed

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

Solution: Positive kurtosis value indicates that distribution is peaked and possesses thick tails

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

Solution: A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution

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



What can we say about the distribution of the data?

Solution: When the median is closer to top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is Negatively Skewed.

What is nature of skewness of the data?

Solution: Negatively Skewed.

What will be the IQR of the data (approximately)?   
Solution: (Q3-Q1) = (18-10)

Q19) Comment on the below Boxplot visualizations?



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

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)

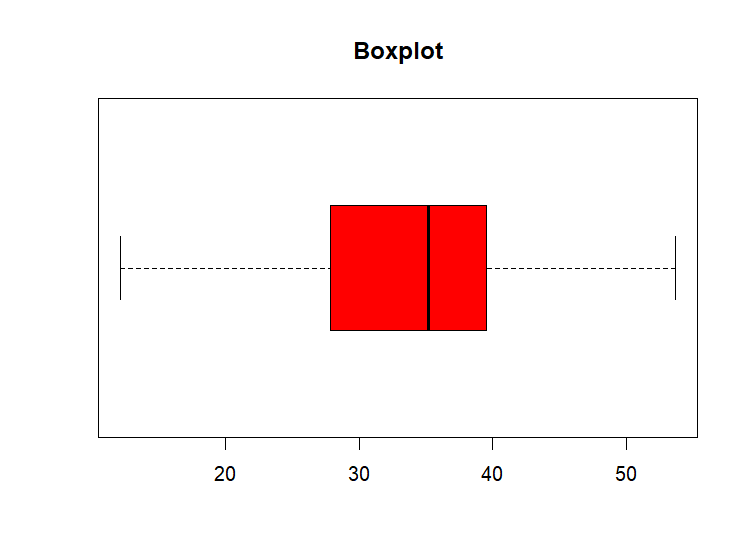
c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution:

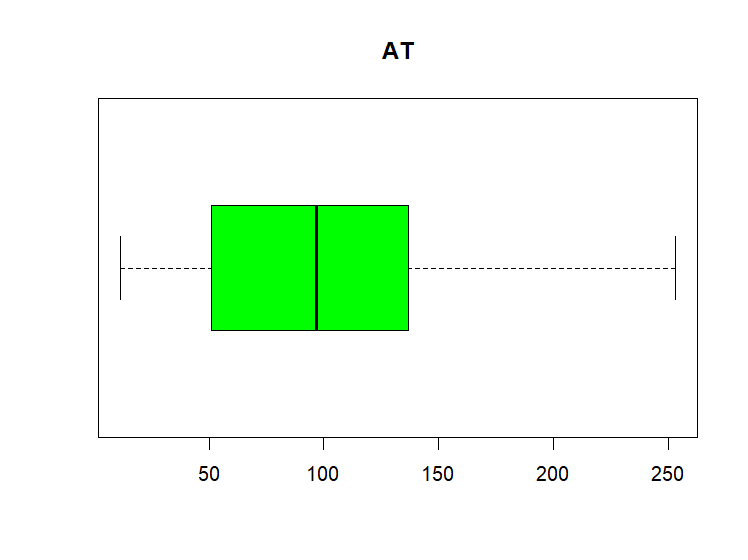


From the above boxplot shows that MPG of cars does not follows Normal distribution. It is Negatively Skewed distribution. Because the median is closer to top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is Negatively Skewed.

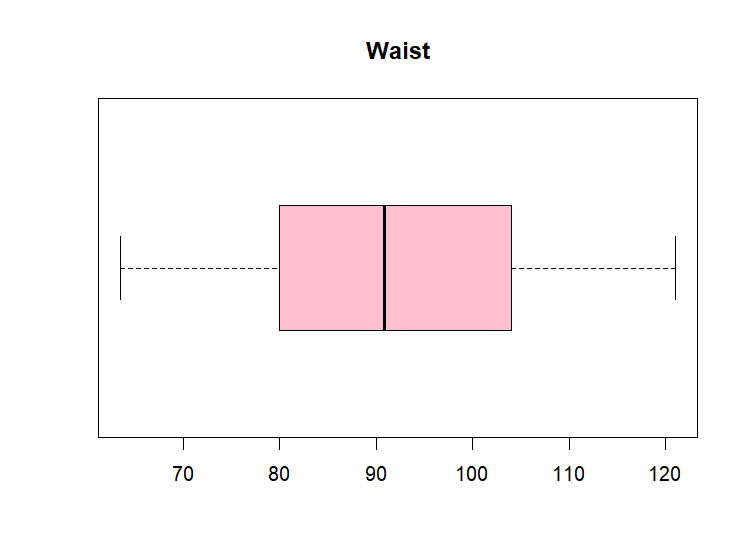
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Solution:



From the above Boxplot shows that AT follows Normal distribution. That is Symmetric distribution because the median is in the middle of the box, and the whisker are about the same of both sides of the box, then the distribution is Symmetric.



From the above boxplot shows that Waist follows Possitively Skewed distribution. It is Posstively Skewed distribution because the median is closer to bottom of the box, and if the whisker is shorter on the lower end of the box, then the distribution is Possitively Skewed.

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

Solution:

qnorm(0.90)

[1] 1.281552

> qnorm(0.94)

[1] 1.554774

> qnorm(0.60)

[1] 0.2533471

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

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:

Here,

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

df=number of freedom=18-1=17

therefore,t=(250-270)/(90/sqrt(18)) = -0.471

i.e t score= -0.417

the probability of the bulbs lasting less than 260 days on average of 0.3218

assuming mean life of the bulbs is 300 days.