|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete(count) |
| Results of rolling a dice | Discrete(count) |
| 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(categorical) |
| Number of kids | Discrete(count) |
| Number of tickets in Indian railways | Discrete(count) |
| Number of times married | Discrete(count) |
| Gender (Male or Female) | Discrete(binary) |

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 | Nominal |
| Celsius Temperature | Nominal |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

If 3 coins are tossed then te no. ofpossible outcomes

are:- HHH,HHT,HTH,THH,THT,TTH,HTT,TTT

i.e.-8

Now the expected outcomes are:- HHT,HTH,THH

i.e.-3

We know,

Probability= No. of expected outcomes/Total no. of possible outcomes

=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

Ans.- Two dices are rolled.

The no. of possible outcomes are:-[11,12,13,14,15,16,21,22,23,24,25,26,31,32,33,34,35,36,41,42,43,44,45,46,51,52,53,54,55,56,61,62,63,64,65,66]

Total no. of outcomes=36

1. Probability of the outcome of the event where the sum of no. is 1 is ‘zero’. This is an impossible event i.e. the sum of 2 no.in the events will never be 1.
2. Sum is less than or equal to 4

i.e. Interested events are=[11,12,13,21,22,31]

Probability=6/36

Probability=0.16666

1. Probability of the events whose no. of sum is divisible by 2 and 3

Interested events=[11,12,13,15,21,22,24,26,31,33,35,36,42,44,45,46,51,53,54,55,62,63,64,66]

Probability=24/36

Probability=0.666666

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 balls in the bags are 7

Probability for none of the drawn balls will blue is

Probability=5/7

Probability=0.71428

Q6) Calculate the Expected number of candies for a randomly selected child

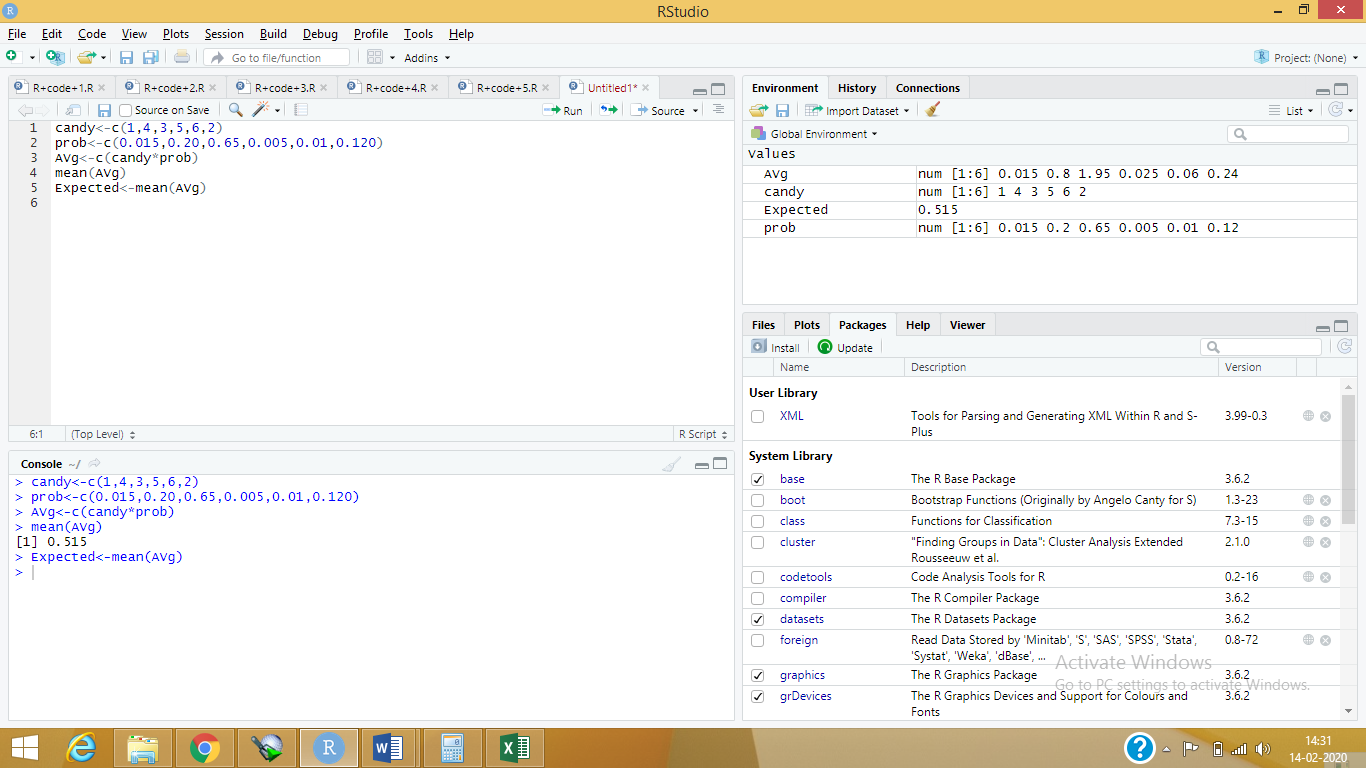
Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| 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:-



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.



Answer:-

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Variance | 0.2858814 | 0.957379 | 3.193166 |
| Standard Deviation | 0.5346787 | 0.9784574 | 1.786943 |
| Range | 2.76 - 4.93 | 1.513 - 5.424 | 14.5 - 22.9 |

Mode of the given data cannot be calculated because the given data is continuous in nature and to calculate the mode the data needs to be discrete.

Means and medians of the given data are almost same so we can say that the data is symmetrical in nature.

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:-

There are 9 patients in total.

So the probability for each patient being chosen at random is 1/9,

i.e.0.111111111111

Expected value = Sum of((1/9)\*Weight of each patient)

=16.14815

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

**Cars speed and distance**



Answer:-

Skewness:-

Speed = -0.1139548

Distance = 0.7824835

From the values of the skewness found by the analysis we can say that the speed has the negative skewness and distance has positive skewness with the values as shown above.

Kurtosis:-

Speed =2.422853

Distance = 3.248019

From above numerical figures we can say that the kurtosis of speed and distance has positive values as shown above.

We have also inferred this information from histogram plotted in the R-studio.

1. **SP and Weight(WT)**



**Answer:-**

Skewness:-

SP = 1.581454

WT = -0.6033099

From the values of the skewness found by the analysis we can say that the SP has the positive skewness and WT has negative skewness with the values as shown above.

Kurtosis:-

SP =5.723521

WT = 3.819466

From above numerical figures we can say that the kurtosis of speed and distance has positive values as shown above.

We have also inferred this information from histogram plotted in the R-studio.

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



In this fig the graph has wider peaks and thinner tails extending towards the right side hence

From this fig we can say that it has positive skewness and kurtosis. Also most of the chicks have their weights concentrated between 0 to 200



By looking at the figure we can say that there are lots of outliers are present in the data towards the upper extreme. Also the IQR of the data is close to the lower extreme.

Also we can say that the given data might have the positive Kurtosis.

**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:-

94% :[201.2623910556929, 198.7376089443071]

96% :[201.3785962570268, 198.6214037429732]

98% :[201.5618139516784, 198.4381860483216]

**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.

Mean:- 41

Median:- 40.5

Variance:- 25.52941

Standard Deviation:- 5.052664

1. What can we say about the student marks?

From the above values we can conclude that the distribution of marks in between students is mostly symmetric also the data of the given records is standard normalized data with few outliers including. It’s varies mostly in the range of 35 to 45.

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

Answer:- Data is perfectly symmetrically distributed.

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

Answer: - When mean>median we can say that data is positively skewed.

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

Answer: - When median>mean, we can say that data is negatively skewed

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

Answer: - It indicates that data has sharp peaks. If we talk about frequency , then a particular element’s frequency is quite higher than frequency of other element.

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

Answer: - It indicates that data has wider peaks and thinner tails.

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



What can we say about the distribution of the data?

First 25% of the data is sparsely distributed in the range of 1 to 10, the next 50% of the data is moderately distributed in the range of 10 to 18 and the last 25% of the data it heavily concentrated in between 18 to 19.

What is nature of skewness of the data?

The data is negatively skewed.

What will be the IQR of the data (approximately)?   
The inter quartile range is the range in the graph from the point Q1 to Q3

i.e.in above fig it is from 10 to 18.

Q19) Comment on the below Boxplot visualizations?



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

Answer: - Data range of boxplot1 is from 240-280, while the data of boxplot2 is from 200-350.

Both the box plot has same median and symmetrically distributed data.

Also the skewness of the given data of the boxplots is almost zero.

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)

Answer:-

1. P(MPG>38) = 0.4074
2. P(MPG<40) = 0.753

c. P (20<MPG<50) = 0.852

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:- From the data the mean and median of MPG are not equal.

Also the QQ-plot, histogram of the MPG does not follows the normal distribution of the data.

Hence we can say that the data of the MPG is not 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

Answer:- From the data the mean and median of WC and AT are not equal.

Also the QQ-plot, histogram of the WC and AT does not follows the normal distribution of the data.

Hence we can say that the data of the WC and AT is not normally distributed.

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

Answer:-Z-scores

90% Confidence level = 1.644854

94% Confidence level = 1.880794

60% Confidence level = 0.8416212

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

Answer:-

t-scores:

95% confidence level = 2.059539

96% confidence level = 2.166587

99% confidence level = 2.787436

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:-

0.4557641 % is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.