|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data type |
| Results of rolling a dice | Discrete data type |
| Weight of a person | Continuous data type |
| Weight of Gold | Continuous data type |
| Distance between two places | Continuous data type |
| Length of a leaf | Continuous data type |
| Dog's weight | Continuous data type |
| Blue Color | Discrete data type |
| Number of kids | Discrete data type |
| Number of tickets in Indian railways | Discrete data type |
| Number of times married | Discrete data type |
| Gender (Male or Female) | Discrete data type |

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

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

Answer 3. The Probability of getting two heads and one tails in the toss of three coins simultaneously is 3/8 or 0.375.

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

1. Equal to 1

Answer a) Probability is 0

1. Less than or equal to 4

Answer b) Probability is 1/6

1. Sum is divisible by 2and 3

Answer c) 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?

Answer 5) Probability 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 6) 3.09

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

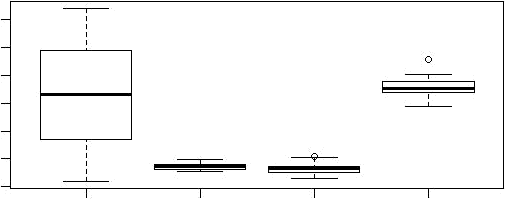
* For Points,Score,Weigh>

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

**Use Q7.csv file**

Answer 7)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.891 | 3.54 | 17.43 |
| Variance | 0.285 | 0.957 | 3.19 |
| Standard  Deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.76,4.93 | 1.513,5.424 | 14.5,22.9 |

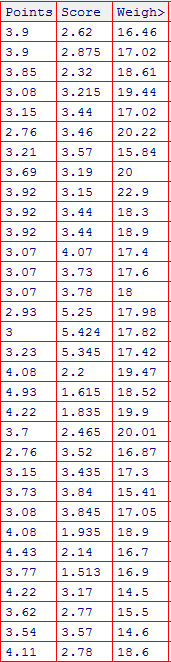
Points Score Weigh

0

5

10

20



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 8) Q8 <- read.csv("D://data sets//csv//Q8.csv")

> mean(Q8$x) [1] 145.3333

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

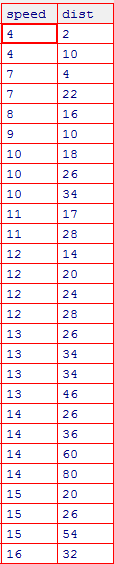
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

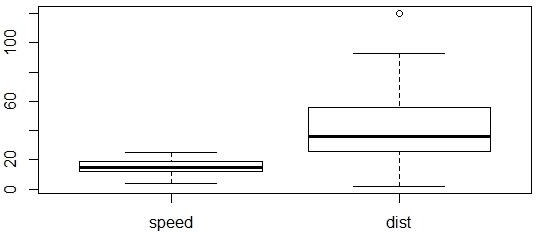
**Use Q9\_b.csv**

Answer 9)

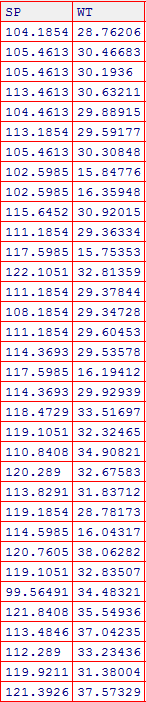


Solution:

|  |  |  |
| --- | --- | --- |
|  | Car speed | Distance |
| **Skewness** | -0.1139548 | 0.7824835 |
| **Kurtosis** | 2.422853 | 3.248019 |



SP and Weight(WT)



Solution:

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.581454 | -0.6033099 |
| Kurtosis | 5.723521 | 3.819466 |

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



Answer 10) The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

The boxplot has outliers on the maximum side.

**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 11)

|  |  |  |
| --- | --- | --- |
| **confidence interval** | **Z value** | **Range** |
| confidence interval94% | 1.880794 | 198.74,201.26 |
| confidence interval96% | 2.053749 | 198.62,201.38 |
| confidence interval98% | 2.326348 | 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?

Answer 12)

1)

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| Standard deviation | 5.05664 |

2) Mass of students marks between 38-42.

Skewness (1.52) is positive because mass of marks in left side of plot.

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

Answer 13) Data is normalized and there is no skewness.

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

Answer 14) Negative Skewness implies mass of the Distribution concentrated on right side.

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

Answer 15) Positive Skewness implies mass of the Distribution concentrated on left side.

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

Answer 16) Positive kurtosis value indicates that thinner peak and wider tails.

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

Answer 17) 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?

Answer: Not normally distributed.

What is nature of skewness of the data?

Answer: Negative skewness

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

Answer: 10-18

Q19) Comment on the below Boxplot visualizations?



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

Answer 21) First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)

Answer a) 1-pnorm(38,34.422,9.13144)= 0.3475908

* 1. P(MPG<40)

Answer b) pnorm(40,34.422,9.13144)= 0.7293527

* 1. P (20<MPG<50)

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

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) MPG of cars follows 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 b) Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

Answer 22)

|  |  |
| --- | --- |
| Confidence interval | Z scores |
| 60% | 0.8416212 |
| 90% | 1.644854 |
| 94% | 1.880794 |

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

Answer 23)

|  |  |
| --- | --- |
| Confidence interval | T scores |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 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 24) 0.32