Q1) Identify the Data type for the Following:

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

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

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

Ans Total no of chances = 8

no of chances that will be two heads and one tail is {HHT,HTH,THH} i.e is 3

probability that two heads and one tail is 3/8

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

* Equal to 1

**ans** )For this question the probability is zero why? because Dice has 6 faces and has nos from 1 to 6 if sum can never be 1 as min sum of two dice rolled would be 2

* Less than or equal to 4

**ans** ) no of possible outcomes of less than 4 and 4 is {(1,3),(3,1),(2,2),(1,2),(2,1),(1,1)}

total no of outcomes are 36

probability that sum less than or equal to 4 is 6/36 ie 1/6

* Sum is divisible by 2 and 3

**Ans:**

Probability of sum divisible by 2 and 3 is 6/36

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?

**Ans:**

Total no of outcomes are 7c2

Total no of outcomes where they are none of the balls drawn is blue 5c2

Probability = 5c2/7c2

= 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

**Ans:** The expected no of candies for a randomly selected child are

E(x)= 3.125

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

**Ans**

* For Points,Score,Weigh>

1) For points:mean: 3.5965625

median: 3.6950000000000003

mode: 3.07

standard deviation 0.5346787360709716

Range: 2.17

2) For Score: mean: 3.2172500000000004

median: 3.325

mode: 3.44

Name: Score, dtype: float64

standard deviation 0.9784574429896967

Range: 3.9110000000000005

3) For Weigh: mean: 17.848750000000003

median: 17.71

mode: 17.02

standard deviation 1.7869432360968431

range: ​8.39

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

* 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?

Ans:

Expected value of mean=(Sum of all weights)/(Total number of patients)

Total number of patients = 9

Sum of all weights= 108+110+123+134+135+145+167+187+199

= 1308

Expected value of mean= 1308/9

= 145.33

Code:

X=c(108,110,123,134,135,145,167,187,199)

expected\_value=mean(X)

cat("Expected value of weight:", expected\_value)

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

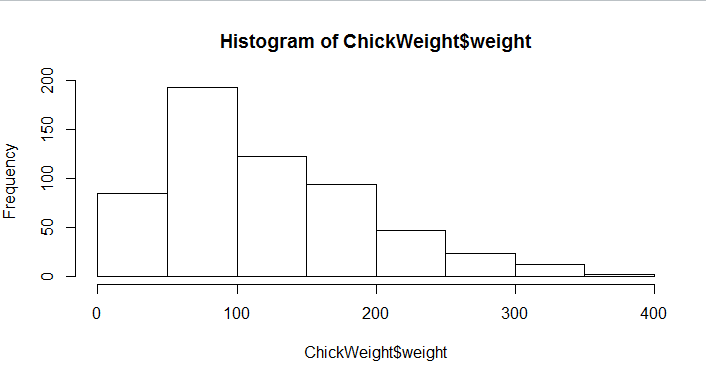
**Cars speed and distance**

**Use Q9\_a.csv**

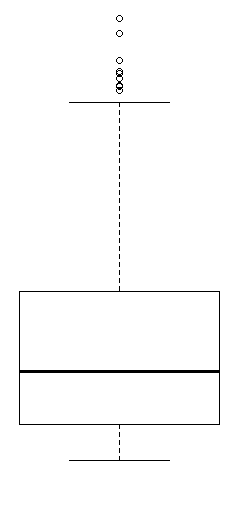
**SP and Weight(WT)**

**Use Q9\_b.csv**

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



The histogram is positively skewed and the graph is about the frequency vs weight. The no of people from 0 to 100 are more than the rest of them 7



Boxplot:

By looking at the boxplot, we can see that there are some outliers present, we also observe that the median is less than mean as there are less data points between Q1 and below points range, hence we can consider it as the right skewed.

**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?

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

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

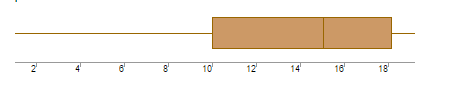
Q14) What is the nature of skewness when mean > median ?

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

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

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

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



I)What can we say about the distribution of the data?

As we can see the box plot visualization we can easily say that the data is distributed between 10 and 18 most the data is right side of box plot graph so it negatively skewed

II)What is nature of skewness of the data?

Ans:

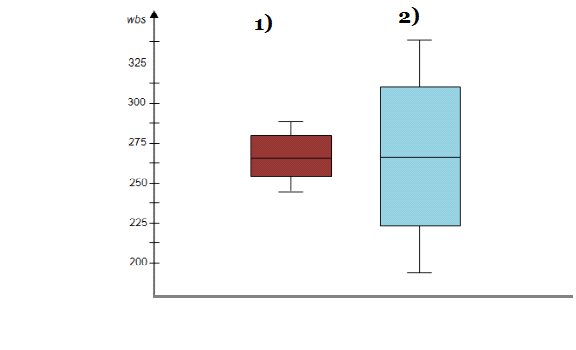
In the boxplot, most of the data was lying in the range of 10-18, hence it will be considered as the left skewed, as the data values in upper quadrant contains more than the low quadrant.

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

IQR(Inter quartile range)=Q3-Q1

= 18.3-10.3  
 = 8 approx

Q19) Comment on the below Boxplot visualizations?



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

Ans:

In the above given 2 boxplots, we can observe that mean and median are equal, even though the whisker level is high in boxplot 2, hence we can consider the distribution is symmetrical.

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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

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

Ans: To solve this problem we use test of hypothesis

H0=The average lifespan of the bulbs is indeed 270 days

H1= The average life span of the bulbs is less than 270 days

Assuming the level of significance( = 0.05

 (mean= 270, std= 90, population mean 260, no of samples=18)

Z= -0.88

Using a standard normal distribution table or a statistical calculator, the critical value at α = 0.05 is approximately -1.645.

If the calculated test statistic falls in the critical region (i.e., less than -1.645), we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

In this case, since the calculated test statistic (-0.8817) does not fall in the critical region (-1.645), we fail to reject the null hypothesis.

Since we failed to reject the null hypothesis, we do not have enough statistical evidence to conclude that the average lifespan of the bulbs is less than 270 days. Therefore, based on the sample data, we cannot support the claim made by the researcher.

Top of Form