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

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

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

Ans: 3/8

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: a):0

b):1/6

c):1/3

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: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

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.

**Ans: Mean:**

Points 3.596563

Score 3.217250

Weigh 17.848750

**Median:**

Points 3.695

Score 3.325

Weigh 17.710

**Standard Deviation:**

Points 0.534679

Score 0.978457

Weigh 1.786943

**Range:**

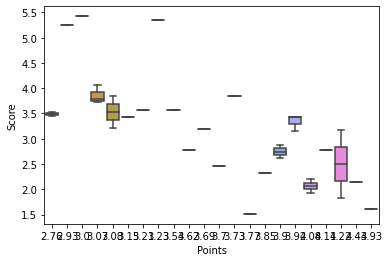
Points (2.760000,4.930000,)

Score (1.513000, 5.424000)

Weigh (14.500000, 22.900000)

**Comments:**

* While plotting the graph between Points and Score I see that the points are inversely proportional to each other
* While seeing the boxplot between Score and Points can see that the density of data points between Score 2.0 to 4.0 is maximum.



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?

Ans: 0

**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**

**Ans: Skewness(Use Q9\_a.csv):**

Index 0.000000

speed -0.117510

dist 0.806895

**Kurtosis:**

Index -1.200000

speed -0.508994

dist 0.405053

**Skewness(Use Q9\_b.csv):**

SP 1.611450

WT -0.614753

**Kurtosis:**

SP 2.977329

WT 0.950291

From the above observation I concluded that in 1st dataset(Q9\_a.csv ) the mean of speed is greater than its median and in 2nd dataset (Q9\_b.csv ) the mean of Weight(WT) is greater than its median.

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



Ans:In the given histogram and the boxplot the inferences what we get that the most of the observaation lie around(50%) in the range(50,200) and the outliers lie in (0,50)and from (200,400).

**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?

Ans:

For 94% confidence interval is (1996.520 ,2003.479)

For 96% confidence interval is (1995.651 , 2004.349)

For 98% confidence interval is ( 1995.066, 2004.934)

**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?

Ans:

1. Mean: 41.00, median :40.5, std:5.0526
2. The range of students marks lie between(34,56) whereas 25% observation lie between(34,38.25) ,50% observation lie between(34,40.5) and 75% observation lie between (34,41.75) so we can say that most students marks lie between(34,41.75).

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

Ans: skewness =0

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

Ans: skewness<0

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

Ans: skewness>0

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

Ans: the +ve kurtosis indicates that it the distribution has higher tales than the normal distribution.

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

Ans: the -ve kurtosis indicates that it the distribution has lighter tales than the normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans:

-> There is negative skew because median is shorter on the lower end of the box . so it’s a –ve skewness.

-> the inter quartile range is (10,18.2).

Q19) Comment on the below Boxplot visualizations?



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

Ans: Boxplot1:

The minimum and maximum value is(237,287)

IQR range is (255,280)

Skew=0

Boxplot2:

The minimum and maximum value is (150,350)

IQR range is (225,310)

Skew =0(mean = median)

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)

Ans:

a)0.69204

b)0.652

c)0.89

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: No it doesn’t follow Normal Distribution because (mean, median!=1).

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

Dataset: wc-at.csv

Ans:No ,the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set doesn’t follows Normal Distribution.

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

Ans: zscore for (90% confidence interval )=1.645

zscore for (94% confidence interval )=1.88

zscore for (60% confidence interval )=0.84

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

Ans t scores for(95% confidence interval)=26.96

t scores for(96% confidence interval)=27.05

t scores for(99% confidence interval)=27.57

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: qt(0.95,17)=1.73

Problality using t-distribution =1.73\*90/sqrt(18)

= 36.69