|  |  |
| --- | --- |
| 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 | Continuous |
| 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 | 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| 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 – There are 8 Total outcomes of 3 coins tossed from which 3 outcomes are there where the probability is two heads and one tail(HHT,HTH,THH),So the probability is 3/8.

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

1. Equal to 1

Ans - Probability is 0 because minimum sum value after 2 dice rolled is 2(1+1)

1. Less than or equal to 4

Ans – Probability is 6/36 = 1/6

1. Sum is divisible by 2 and 3

Ans - Probability is 6/36 = 1/6 (3,3),(1,5),(5,1),(4,2),(2,4),(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?

Ans – There are 21(7c2) total outcomes of two ball drawn from total 7 balls, so the probability of drawing no blue ball is 10(5c2), therefore 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

Ans – Expected number of candies for a randomly selected child is 3.09(1\*.015+4\*.20+3\*.65+5\*.005+6\*.01+2\*.12)

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 - Points – mean=3.59,median=3.6,mode=3.92,range=2.76-4.93

Score – mean=3.21,median=3.325,mode=3.44,range=1.513-5.424

Weigh – mean=17.24,median=17.71,mode=17.02,range=14.4-22.9

Mean and median is almost same so that means there are no outliers in the Data.

**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?

Ans – expected value = probability\*value. The probability of drawing 1 patient is 1/9 because there are total 9 patients

(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 = **145.3**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans –** Skewness = -0.11 # it means the data is left skewed and longer tail on left side of the distribution.

Kurtosis = -0.50 # it means that the distribution is flat and has thin tails. So there are less chance of outliers.

Skewness = 0.80 #it means the data is right skewed and longer tail on right side of the distribution.

Kurtosis = 0.40 # it means that the distribution is peaked and possess thick tails. So there are more chances of outliers.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans – Skewness = 1.611

Kurtosis = 2.97

Skewness = -0.61

Kurtosis = 0.95

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



Ans – By the help of histogram graph we can conclude that the data is right skewed and the tail on the right side is more. We can also conclude that there may be outliers present.

By the help of boxplot graph we can conclude that there are outliers present on the higher side of the mean value.

**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 – N =3,000,00 , n = 2000 , x =200 , s = 30

Here standard deviation of population is not given so that’s why we have to use t-test.

CI = x(+-)\*t\*(s/n1/2)

Using python we found the value of t-test, stats.t.ppf(0.97,1999) = 1.88

Therefore, CI = (198.738325292158, 201.261674707842)

For CI = 98% , CI = (198.43943840429978, 201.56056159570022)

For CI = 96%, CI = (198.62230334813333, 201.37769665186667)

**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.

Ans – Mean = 41, Median = 40.5, Variance = 25.52, Standard Deviation = 5.05

1. What can we say about the student marks?

Ans- Most of the students scored around 41 marks

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

Ans- There will be no skewness present as mean is equal to median.

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

Ans – When mean is greater than median then the nature of skweness is right skew.

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

Ans- When median is greater than mean the nature of skewness is left skew.

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

Ans – It means the distribution is peaked and there is a chance that there are values at the tail of the distribution.

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

Ans – It means that the distribution is less peaked and most of the data is around mean.

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



What can we say about the distribution of the data?

Ans – We can found out that the distribution is not normally distributed and there is skewness present. Most of the data values is between 10 and 15.

What is nature of skewness of the data?

Ans – It is a right skewed data.

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

Ans – The IQR of the data is around 15.

Q19) Comment on the below Boxplot visualizations?



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

Ans – By viewing those two boxplot we can state that in 1st boxplot the values are ranged in between 250 and 275, but in the case of boxplot 2 the values range between 300 and 225. The median of both those data is about 267.

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)

Ans - p = 0.3474

* 1. P(MPG<40)

Ans – p = 0.7294

* 1. P (20<MPG<50)

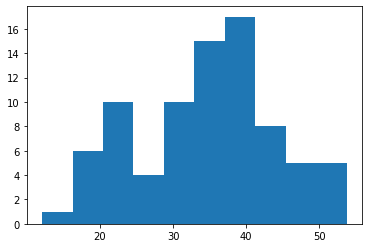
Ans – p = 0.8989

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 does not follow 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

Ans – Both AT and Waist does not follow normal distribution.





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

Ans – 90% = 1.64 stats.norm.ppf(0.95)

94% = 1.88 stats.norm.ppf(0.97)

60% = 0.84 stats.norm.ppf(0.8)

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

Ans – 95% = 2.06 stats.t.ppf(0.975,24)

96% = 2.17 stats.t.ppf(0.98,24)

99% = 2.79 stats.t.ppf(0.995,24)

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 – x = 260, n = 18, s =90

tscore = 260-270/(90/180.5) = -0.47

Stats.t.cdf(-0.47,17) = 0.3221

So there is 32.21% probability that the 18 randomly selected bulb have an average life of less than 260 days.