|  |  |
| --- | --- |
| 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 | discrete |
| 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) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans:- total number of event=2^3=8

Event 2 head and 1 tail=3

Probability=3/8=0.375=37.5%

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

a)Ans:-total number of event =6\*6=36

event with sum 1=0

probability=0/36=0

b)Ans:-event with sum equal to 4 and less=6

probability=6/36=1/6=0.1666=16.66%

c)Ans:-Event with sum divisible to both 2 and 3=6

probability=6/36=1/6=0.1666=16.66%

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 ball=7

Blue ball=2

Probability of 1st ball not blue ball=5/7

Probability of 2nd ball not blue ball=4/6

Total probability of both ball not being blue=(5/7)\*(4/6)=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 0.015\*1+0.4\*2+0.65\*3+0.005\*5+0.01\*6+0.12\*2 = 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**

**Ans:- mean** Points 3.596563

Score 3.217250

Weigh 17.848750

**Median**

Points 3.596563

Score 3.217250

Weigh 17.848750

**Mode** Points:

3.07

3.92

Score:

3.44

Weigh:

17.02

18.90

**Variance :-**

Points 0.285881

Score 0.957379

Weigh 3.193166

**Standard deviation:-**

Points 0.534679

Score 0.978457

Weigh 1.786943

**Range:-**

Points 2.17

Score 3.9110

Weigh 8.3999

**Comment:- dataset is balanced. there are some outlier is score**

**and weigh on higher side, points and score are left**

**skewed and weigh is slightly right skewed.**

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 weight of random patient will be mean of all the data=145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Ans:- skewness**

Speed:- -0.117510 (skewness is approximately symmetric more weight on right tail)

Dist:- 0.806895 (skewness is moderate and right skewed more weight on left tail)

**Kurtosis**

Speed:- -0.508994(more data on tail than normal distribution distribution is more flattered)

Dist:- 0.405053(less data on tail comparing to normal distribution)

**Use Q9\_b.csv**

**Ans:- skewness**

SP:- 1.611450(skewness is high and right skewed more weight on left tail)

WT:- -0.614753(skewness is moderate and left skewed more weight on right tail)

**Kurtosis**

SP:- 2.977329(very less data on tail comparing normal distribution)

WT:- 0.950291(less data on tail)

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



Ans:- data is right skewed, mean>mode,there are some outlier on higher side on weight..

**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%= (198.738325292158, 201.261674707842)

For 98%=(198.43943840429978, 201.56056159570022)

FOR 96%=(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=41.0, median=40.5, variance=24.111, std=4.91

1. What can we say about the student marks?

Ans:-average marks obtained in class is 41 and maximum marks is 56 and minimum is 34. most frequent marks obtained that mode is 41.

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

Ans:-distribution of data is symmetric..

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

Ans:-right skewed,more weight on left tail.

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

Ans:-left skewed

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

Ans:- less data on tail, less variance, more data near to mean. data has more peak than normal distribution

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

Ans:- more data on tail, more variance, less data near to mean. data has less peak than normal distribution

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



What can we say about the distribution of the data?

Ans:-distribution of data is not normal distribution. more data weight is on right tail.

What is nature of skewness of the data?

Ans:- data is left skewed.

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

Ans:-The Inter Quantile Range =Upper quartile – Lower Quartile = 18 –10=8

Q19) Comment on the below Boxplot visualizations?



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

Ans:- median of both boxplot is same.

There are no outlier in both box plot.

Range of blue box plot data is bigger

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:- 0.3475939251582705

* 1. P(MPG<40)

Ans:- 0.27065012378483844

* 1. P (20<MPG<50)

Ans:- **calculate this we will calculate below 50 probability then subtract below 20 probability from it. We will get = 1.2430968797327613e-05**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:- after looking at histogram of MPG we can say data is 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

Ans:- data AT and Waist both are not normally distributed

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval  
Ans:- For 90% z score = z\_90=np.round(stats.norm.ppf((1+0.90)/2),3)=1.645

For 94% z score = z\_94=np.round(stats.norm.ppf((1+0.94)/2),3)=1.96

For 60% Z score= z\_60=np.round(stats.norm.ppf((1+0.60)/2),3)=0.842

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

Ans:-T score for 95% Confidence Interval =np.round(stats.t.ppf((0.975,df=24),4)= 2.063

T score for 96% Confidence Inteval

= np.round(stats.t.ppf(1.96/2,df=24)=2.71

T score for 99% Confidence Interval

=np.round(stats.t.ppf(1.99/2,df=24),4))T score for 95% =2.796

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:-t score=(Sample mean-mean)/(standard deviation/square rootof sample size))

T score= **-0.471**

Probability=stats.t.cdf(tscore,df)= **0.3216**