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

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 | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | ratio |
| Years of Education | Ordinal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?P(x)=3/8+3/8 =6/8 = 3/4

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

* Equal to 1 =0
* Less than or equal to 4 = 6/36 = 1/6
* Sum is divisible by 2 and 3 =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?

= 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

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

= 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. points mean=3.6



data.mean()

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

data.median(

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

data.mode()

data.var()

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

data.std()

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

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?

108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 = 1308

1308/9 =145.333

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

**Cars speed and distance**



**SP and Weight(WT)**



# Importing necessary libraries

import pandas as pd

import numpy as np

# importing data set using pandas

data=pd.read\_csv("C:/Users/elcot/Downloads/Q9\_a.csv")

# Finding skewness

data.skew()

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

# Finding kurtosis

data.kurt()

Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

# visualization purposes

import matplotlib.pylab as plt

# Histogram

plt.hist(data['speed'])

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



# Histogram

plt.hist(data['WT'])

#Boxplot

plt.boxplot(data['WT'])

**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?
* mean =738/18
* =41
* median= 40,41
* variance= (40-34)/18 =6/18 =1/3
* standard deviation= 0.33

Q2 . Ans =average

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

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

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

Q16) What does positive kurtosis value indicates for a data ?heavier tails and a sharper peak than the normal distribution

Q17) What does negative kurtosis value indicates for a data? lighter tails and a flatter peak than the normal distribution

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



What can we say about the distribution of the data? left

What is nature of skewness of the data? left skewness

What will be the IQR of the data (approximately)? 10

Q19) Comment on the below Boxplot visualizations?



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

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) =30/81 =0.37
* P(MPG<40) =59/81 =0.81

c. P (20<MPG<50) =69/81 =0.85

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

plt.hist(data['MPG'])

not normal distribution

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

Dataset: wc-at.csv

not normal distribution

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