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

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

HHH,HHT,HTH,THH,TTH,THT,HTT,TTT

P[2-heads & 1-tail] = 3

Total no:of outcomes = 8

3/8=0.375

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

1. Equal to 1=0
2. Less than or equal to 4=3/11=0.2727
3. Sum is divisible by 2 and 3=8/11=0.7272

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?

n(S)=7C2=21

n(E)=5C2=10

P(E)=10/21=0.476

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\*0.120=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 | Score | Weigh |
| Mean | 3.596563 | 3.217250 | 17.848750 |
| Median | 3.695 | 3.325 | 17.710 |
| Mode |  |  |  |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Std Deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.76-4.93 | 1.513-5.424 | 14.5-22.9 |

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

**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?

Mean=145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **skewness** | **kurtosis** |
| **speed** | **-0.117510** | **-0.508994** |
| **distance** | **0.806895** | **0.405053** |

**Chart, histogram

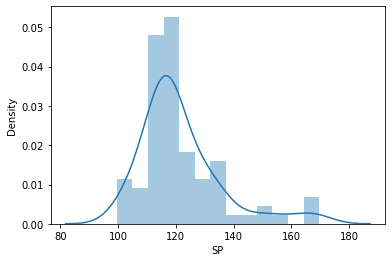
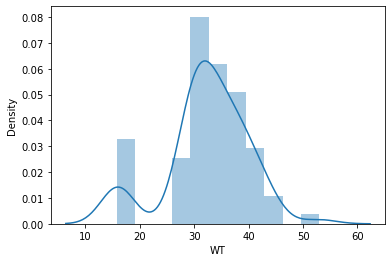
Description automatically generatedChart, histogram

Description automatically generated**

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **skewness** | **kurtosis** |
| **SP** | **1.611450** | **2.977329** |
| **WT** | **-0.614753** | **0.950291** |

** **

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



* Positive skew
* Majority of the data lies between 0-200
* No outliers present in the data



**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?

* For CI=0.94, mean lies between 198.74 and 201.26
* For CI=0.98, mean lies between 198.46 and 201.54
* For CI=0.96, mean lies between 198.63 and 201.37

**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?

* Mean=41
* Median=40.5
* Variance=24.11
* Std deviation=4.91

The marks are very closely spread as all the students have performed well

The mean and median are almost same, therefore it is normally distributed

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

Zero skewness

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

Right skewed

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

Left skewed

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

It indicates that the distribution possess thick tails on both sides

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

It indicates that the data is less in the tail region than the normal distribution

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



What can we say about the distribution of the data?

* Data is normally distributed

What is nature of skewness of the data?

* the distribution is negatively skewed

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

* IQR=Q3-Q1=18-10= 8

Q19) Comment on the below Boxplot visualizations?



* Both have almost same median value
* Boxplot(1) has less variance therefore it is more consistent when compared to Boxplot(2)

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

* 1. P(MPG>38)=0.347
  2. P(MPG<40)=0.729

c. P (20<MPG<50)=0.899

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* Mean=34.42
* Median=35.15
* Mean and median are almost same, therefore MPG follows 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

* the Adipose Tissue (AT) and Waist Circumference(Waist) does not follow normal distribution as they are positively skewed.

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

* for 90% CI, Z = 1.645
* for 94% CI, Z = 1.808
* for 60% CI, Z = 0.84

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

* For 95%CI, t=2.064
* For 96%CI, t=
* For 99%CI, t=2.787

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

* The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.17%