Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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

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

ANS) all possibilities = 2n = 23 = 8

favorable outcomes = THH, HTH, HHT

favorable outcomes / all possibilities = 3 / 8

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

1. Equal to 1

ANS) Minimum value for each dice is 1. So 1 + 1 = 2 is the minimum sum. Hence, for sum = 1, probability is 0

1. Less than or equal to 4

ANS) all possibilities = 6n = 62 = 36

favorable outcomes = (1,1), (1,2), (2,1), (2,2), (3,1), (1,3)

favorable outcomes / all possibilities = 6 / 36 = 1/6

1. Sum is divisible by 2 and 3

ANS) all possibilities = 6n = 62 = 36

sums divisible by 2 and 3 = 6, 12

outcomes where sum is 6 = (1, 5), (5, 1), (2, 4), (4, 2), (3, 3)

outcomes where sum is 12 = (6, 6)

favorable outcomes / all possibilities = 6 / 36 = 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?

ANS) All possibilities are where we get any combination of 2 balls from 7 balls. Not drawing blue balls means getting any combination of 2 balls from the remaining 5 balls.

all possibilities = 7 choose 2 = 7 \* 6 / 2 = 21

favorable outcomes = 5 choose 2 = 5 \* 4 / 2 = 10

favorable outcomes / all possibilities = 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) 0.015 \* 1 + 0.2 \* 4 + 0.65 \* 3 + 0.005 \* 5 + 0.01 \* 6 + 0.12 \* 2

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

= 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)

For Points:

- Mean : 3.597

- Median : 3.695

- Mode : 3.07, 3.92

- Variance : 0.2858814

- Standard Deviation : 0.5346787

- Range : 2.17

Most values are between 3 and 4. No outliers

For Score:

- Mean : 3.217

- Median : 3.325

- Mode : 3.44

- Variance : 0.957379

- Standard Deviation : 0.9784574

- Range : 3.911

3 outliers above 4.5

For Weigh:

- Mean : 17.85

- Median : 17.71

- Mode : 17.02, 18.9

- Variance : 3.193166

- Standard Deviation : 1.786943

- Range : 8.4

One outlier above 22.

Points and Score have a distinct negative correlation. PFA notebook

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 is the sum of product of value and its probability. Here each value has the same probability. So we only need to take the mean.

(108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199) / 9 = 145.34

PFA notebook

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

Cars speed and distance

Use Q9\_a.csv

ANS)

For Speed :

- skewness : -0.11751

- Kurtosis : 2.422853

For Distance :

- skewness : -0.508994

- Kurtosis : 0.405053

Distance has an outlier above 100

There is a strong correlation of 0.806895 between distance and speed.

SP and Weight(WT)

Use Q9\_b.csv

ANS)

For SP :

- skewness : 1.61145

- Kurtosis : 2.977329

Many outliers at the higher end.

For WT :

- skewness : -0.614753

- Kurtosis : 0.950291

6 outliers above 150 for SP. 1 outlier above and 50 and 9 below 20 for WT.

There is no correlation between SP and WT.

PFA notebook

Q10) Draw inferences about the following boxplot & histogram



ANS) The distribution is right-skewed i.e. most chicks weigh less.



ANS) There are many outliers at the higher end. The distribution is right-skewed i.e. most data points have a small 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) 94% Confidence Interval: 198.738 – 201.262

98% Confidence Interval: 198.438 – 201.562

96% Confidence Interval: 198.621 – 201.379

pfa notebook

**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)

mean : 41

median : 40.5

variance : 25.52941

standard deviation : 5.052664

Except for two outlier high scores, the student consistently scores approximately between 35 to 45.

pfa notebook.

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

Ans) If mean and median are the same, the distribution is normal and has no skewness i.e. skewness = 0

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

Ans) If the mean is greater than the median, the distribution is left-skewed i.e. it has a negative skew.

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

Ans) If the mean is greater than the median, the distribution is right-skewed i.e. it has a positive skew.

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

Ans) A positive kurtosis indicate that the frequency of data is higher around the mean and drops sharply away from it. It has a higher peak and less dispersion around it. The higher the kurtosis, the smaller the dispersion.

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

Ans) A negative kurtosis indicate that the frequency of data is not concentrated around the mean. The distribution has a flattened peak and data is concentrated somewhat equally across the distribution.

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



What can we say about the distribution of the data?

Ans) It is an Exponential distribution

What is nature of skewness of the data?

Ans) It is negatively skewed i.e. left-skewed.

What will be the IQR of the data (approximately)?   
Ans) IQR = 8 approx.

Q19) Comment on the below Boxplot visualizations?



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

Ans) Data from Boxplot 1 has higher kurtosis, smaller range, smaller variance, but is a normal distribution like Boxplot 2 with coinciding medians.

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) 11/27

* 1. P(MPG<40)

Ans) 61/81

c. P (20<MPG<50)

Ans) 23/27

PFA notebook

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans) Yes. PFA notebook

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. PFA notebook

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

Ans) 90% : 1.645

94% : 1.8808

60% : 0.842

pfa notebook

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

Ans) 95% : 2.06389

96% : 2.1715

99% : 2.7969

pfa notebook

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) 32.167%

pfa notebook