|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | 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:** if we tossed three coins then,

{ HHT ,

THH,

HTH }.

The probability of two heads and one tail is = 3/8

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

1. Equal to 1

**ANS :** if we rolled two dices, then the probability of getting equal to 1 is = 0

**Probability P(E) = Number of Favourable outcomes n(E) / total number of interested outcomes n(S)**

P(E) = n(E) / n(S) = 0/36= 0

1. Less than or equal to 4

**ANS :** E = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

P(E) = n(E) / n(S) = 6/36 = 1/6

1. Sum is divisible by 2 and 3

**ANS** **:** E = {(1,5),(3,3),(4,5),(5,1),(5,4)}

P(E) = n(E) / n(S) = 5/36

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 balls in a bag = 7

n(S) = 7C2 = 7x6 = 21

2x1

n(E) = 5C2 = 5x4 = 10

2x1

**Probability P(E) = Number of Favourable outcomes n(E) / total number of interested outcomes n(S)**

P(E) = n(E) / n(S) = 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 a candies for a randomly selected child**

**= (1\*0.015) +(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)**

=**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.

**Use Q7.csv file**

**Ans:-**

**Parameters Points Score Weigh**

**Mean** 3.5965633.217250 17.848750

**Median** 3.695000 3.325000 17.710000

**Mode** 3.07 – 3.92 3.325  17.71

**Variance**  0.2858814  0.957379 3.193166

**Std.** 0.534679 0.978457 1.786943

**Range (**2.76 - 4.93) (1.513 - 5.424) ( 14.5 – 22.9)

**Inferences –**

* *“Points” & “Score” are negatively skewed and “Weigh” is positively skewed.*
* *“Points” has low Variance, Standard Deviation, and Range.*

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: =** sum of the weights of patients/Total number of patients

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

**=1308/9**

**= 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans.)**

**Speed Distance**

**Skewness** -0.11750 0.80689

**Kurtosis** -0.508994 0.405053

**Inferences :** In speed there is negative skewness value it shows a longer or fatter tail on the left side of the distribution. Whereas, in Distance there is positive skewness value refers to a longer or a flatter tail on the right.

In speed there is negative kurtosis value that peak is lower. Whereas in distance there is positive kurtosis value, it means peak is higher.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans.)**

**SP Weight**

**Skewness** 1.611450 -0.614753

**Kurtosis** 2.977329 0.950291

**Inferences :** In SP there is positive skewness value refers to a longer or a flatter tail on the right. Whereas, in weight speed there is negative skewness value it shows a longer or fatter tail on the left side of the distribution.

In SP and Weight there is positive kurtosis value, it means peak is higher indicate that a distribution is flat and has thin tails.

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



**Ans.) - inferences about the histogram –** It is right skewed histogram (or) we can say it is positive skewed. And in this histogram we can say that mean is greater than median.



**Ans.)- inferences about the boxplot -** Outliers are too many and far from mean, median and mode.

**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: confidence interval= Z(significance level/2) \* std.population/sample mean**

94%=(198.738325292158, 201.261674707842)

98% **=** (198.43943840429978, 201.56056159570022)

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

**Median =** 40.5

**Variance =** 25.53

**Std. Deviation =**  5

1. What can we say about the student marks?

**ANS:** Mean is greater than median. And it contains no outliers.

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

**ANS:** Symmetric.

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

**ANS:** Right skewed.

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

**ANS:** Left skewed.

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

**ANS:** Implies data has a sharp peak and thick tails.

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

**ANS:** It means a distribution is flat and has a thin tails.

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



What can we say about the distribution of the data?

**ANS:** We can say that there is a big difference between upper quartile and upper extreme compared to the lower quartile and lower extreme.

What is nature of skewness of the data?

**ANS:** Left skewed

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

**ANS: INTER QUARTILE RANGE (**IQR) =Q3-Q1 = IQR=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:** Medium is same for both graph, there is difference between upper limit and lower limit (IQR1-IQR2) and there is no outliers in both the boxplot.

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:** 0.652513

* 1. P(MPG<40)

**ANS:** 0.652513

* 1. P (20<MPG<50)

**ANS:** 0.89891777

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:** From above plot and values we can say that data is fairly symmetrical. so, that is fairly 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:** From above plot and values of (AT) and (Waist) we can say that data is fairly symmetrical. So, that is fairly normally distributed.

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

**ANS: confidence interval= Z(significance level/2) \* std.population/sample mean**

Z scores of:

90% = 1.6448536269514722

94% = 1.8807936081512509

60% = 0.8416212335729143

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

**ANS: confidence interval= t(significance level/2), (n-1) \* std.deviation/sample mean**

T-Score:

95% = 2.0638985616280205

96% = 2.1715446760080677

99% = 2.796939504772804

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:** Probability for 18 Randomly selected bulbs of average life of no more than 260 days is

t = -0.4714045207910317

Probability: 0.32167411684460556