|  |  |
| --- | --- |
| 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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Interval |
| 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?

Soln. S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let event, E= probability that two heads and one tail are obtained

P (E) = 2/8 = 0.25.

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

1. Equal to 1

P (E) = 0

1. Less than or equal to 4

P (E) = 6/36 = 0.167

1. Sum is divisible by 2 and 3

P (E) = 5/36 = 0.138

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?

Soln. - P (E) = (5/7)\*(4/6) = 20/42 = 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

Soln. – Expected value = (1\*0.015) + (4\*0.20)……. (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.

Soln. -

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **score** | **Weigh** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.92 | 3.44 | 18.9 |
| **Variance** | 0.2858814 | 0.957379 | 3.193166 |
| **Standard deviation** | 0.5346787 | 0.9784574 | 1.786943 |
| **range** | [2.76 , 4.93] | [1.513 , 5.424] | [14.5 , 22.9] |

**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?

Soln. – Here mean is the expected value. So, mean is: 145.34.

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Dist** |
| **Skewness** | -0.1139548 | 0.7824835 |
| **kurtosis** | 2.422853 | 3.248019 |

**Inferences** – Since skewness of speed is negative so we can say that mass of the distribution is concentrated on right side and positive kurtosis implies that mass of the distribution is available in center.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | 1.581454 | -0.6033099 |
| **Kurtosis** | 5.723521 | 3.819466 |

**Inferences** – Here mass of the distribution of WT is concentrated on right hand side and of SP on left hand side, and for kurtosis mass is concentrated in center for both the distribution.

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



**Inferences** – It is positively skewed since long tail continues in right side and most values are concentrated on left side.



**Inferences** – It is skewed to the right. The top whisker is slightly longer then the bottom whisker and center line is towards the bottom whisker.

**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?

Soln. – C.I. of 94% is [201.26, 198.73]

C.I. of 96% is [201.37, 198.62]

C.I. of 98% is [201.55, 198.44]

**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.

Mean = 40.88889 Variance = 27.39869

Median = 40.5 Standard deviation = 5.234376

1. What can we say about the student marks?

* We get high variance because in some tests score is more then average.

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

Ans:- when mean and median are equal there is no skewness as skewness is the measure of asymmetry of the distribution.

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

Ans:- positive in nature.

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

Ans:- negative in nature.

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

Ans:- Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

Ans:- Negative kurtosis value indicates that the distribution has 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?

* Mass of the distribution is concentrated on right.

What is nature of skewness of the data?

* Negative skewness

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

* IQR = I3 – I2 = 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: distribution of bloxplot2 have high variance and high IQR as compared to boxlot1 but both have same median.

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) = 0.63307
  2. P(MPG<40) = 0.5266673

c. P (20<MPG<50) = 0.1427594

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:- Yes it 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

Ans:- No it doesn’t follow normal distribution.

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

Ans:- Z90% = 1.644 Z94% = 1.88 Z60% = 0.841

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

Ans:- t95% = 1.98 t96% = 2.17 t99% = 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

Soln:- tscore = (260-290)/[90/sqrt(18)] = -0.4714

Using r-code – pt(-0.4714,14)

Probability is 0.322

Histogram and Boxplot for sepal length of iris data



