|  |  |
| --- | --- |
| 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 | Qualitative |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Qualitative |

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 | Ratio |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ratio |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| 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 | Ratio |

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

Possible combinations = 2^3 = 8

Possible combinations = HHH, HHT, HTT, TTT, THH, TTH, THT, HTH

Combinations have 2 heads and 1 tail = HHT, HTH, THH

X(3)=8 = 3/8

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

1. **Equal to 1:** Probability 0/36 = 0
2. **Less than or equal to 4:** (1,3), (2,2), (3,1) = 3
3. **Sum is divisible by 2 and 3:** 5 rolls produce 6 (1–5), (2–4), (3–3), (4–2), (5–1) = rolls that produce sums that are divisible by both 2 and 3 = 16.67%

Total possible outcomes=6^2=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?

Probability of drawing 2 balls such that none is blue = 10/21

Cal: Total number of balls = 2+3+2= 7

Two balls can be drawn in 7C2 = 21

Number of ways of drawing 2 balls such that none is blue = Number of ways of drawing 2 balls from 2 red and 3 green balls = 5C2 = 10

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

**Calculations:**

Expected number of candies for a randomly selected child

= 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

Expected number of candies for a randomly selected child = 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 Comment about the values/ Draw some inferences.

**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= (108+110+123+134+135+145+167+187+199)/9= 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv :**

Using Moments packages in R , found the skewness and Kurtosis

**> Skewness (speed)**

[1]-0.1139548 negative skewness means left skew i.e.data distributed on right side

**> kurtosis(speed)**

[1]2.422853 positive kurtosis #data distribution is wide not peak

**> skewness(distance)**

[1]0.7824835 postive skewnewss means right skew i.e data distributed on left

**> kurtosis(distance)**

[1]3.248019 positive kurtosis #data distribution is wide not peak

**SP and Weight(WT)**

**Use Q9\_b.csv: s**

**> skewness (SP)**

[1]1.581454 positive so SP is Right skewness.

**> kurtosis(SP)**

[1]5.723521 positive #data is high peak.

**> skewness(WT)**

[1] -0.6033099 negative so WT is left skewness.

**> kurtosis(WT)**

[1] 3.819466 positive data is high peak

Q10) Draw inferences about the following boxplot & histogram.



Ans:

* 50-100 weight having more frequency 180.
* 350-400 weight having very less frequency 5.
* Positive skewness, Data is right skewed.
* Data is not a normal distribution.
* 0-50 weight having 80 frequency.
* 100-150 weight having 120 frequency.

**BOXPLOT**



* 7 Outliers are present in above box plot
* Positive skewness.i.e. data is right skewed
* DATA is not normally distributed
* Q1 is smaller than the Q3

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: Provided:

Standard Deviation = 30

Size = 2000

Mean = 200

Confidence Interval is calculated using the formula:

(Mean+(t\*( Std.Dev /math.sqrt(Sample Size)))): Refer attachement



The **94%** confidence interval is (198.73, 201.27)

The **96%** confidence interval is (198.61, 201.39)

The **98%** confidence interval is (198.43, 201.57)

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.



1. What can we say about the student marks?

By above analysis we can say that average scores of students are 41 greater than median value. We have two outliers i.e 49 and 56.

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

We call this as a Symmetric distribution and had zero skewness

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

The distribution is positively skewed

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

Distribution is right skewed

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

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails

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

* If a distribution has negative kurtosis, it is said to be platykurtic, which means that it has a flatter peak and thinner tails compared to a normal distribution.
* This simply means that more data values are located near the mean and less data values are located on the tails.

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



What can we say about the distribution of the data?

* The above Boxplot is not normally distributed the median is towards the higher value
* 25th percentile value is between 10-15 and 75th percentile value is between 15-18. Upper limit is 18 and the lower limit is 10

What is nature of skewness of the data?

* The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

* The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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.

* First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



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

Dataset: wc-at.csv



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

