|  |  |
| --- | --- |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Radio |
| Religious Preference | Ordinal |
| 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:

{(hhh)(hht)(hth)(thh)(htt)(tht)(tth)(ttt)}

Number with 2 heads: 3

Total number: 8

probability is 3/8 = 0.375 = 37.5%

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3.

Ans:

Total Possible outcome = 62  = 36

A)0/36 = 0

B)Number of favorable outcomes = 6/36 = 1/6 =0.166

C)Number of favorable outcomes = 20/36 = 5/9 = 0.555

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:

There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7.

This leaves 6 balls with 2 blue.

The probability of the second ball not being blue assuming that the first wasn’t is 4/6.

The probability that neither ball drawn was blue is

(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

Ans:  
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

= 3.09

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 also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

**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  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

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



**Ans:**



Histogram:-

Chick weight data is right skewed or positively skewed.---- Yes

More than 50% Chick Weight is between 50 to 150. ---- Yes

Most of the chick weight is between 50 to 100. --- Yes

Box plot:-

-The data is right skewed.

- There are outliers at upper side

**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?

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

**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?

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

Ans:

-it shows the data is normally distributed.

-it is symmetric.

- And has zero skewness.

-mean = median = mode

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

Ans:

-that means it is a positive skewed.

-it is right skewed when the

-and it can be treated with squareroot or mod.

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

Ans:

-negative skewed

-left skewed

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

Ans:

The values are more concentrated towards the right side and the left tail is spread out.

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

Ans:

The data points are more concentrated towards the right-hand side of the distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans:  
a. The data is not actually equally distributed across the plane. There might be outliers influencing the data. Median of the data is 14.7(app x) 25 percent of the data lies between 0-10 50 percent of the data lies between 10-18 25 percent of the data lies after 18-20 a

b. The data will be left skewed since whisker length on the upper quadrant ishigher than the data on the lower quadrant. Median will be greater than the mean since data is left skewed.

c. IQR is the inter quartile range.

IQR is the inter quartile range.

Here Q1 = 10

Q2 = 14.7

Q3 = 18

IQR = Q3 – Q1 = 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:

-Both are equally distributed and symmetrical.

-in camparison of boxplot 1 and boxplot 2, data is widely spread in boxplot 1.

-right skewed

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)

c. P (20<MPG<50)

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

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

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

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

Ans:

Z-SCORE 90% confidence interval is **1.6448536269514722**

Z-SCORE 94% confidence interval is **1.8807936081512509**

Z-SCORE 60% confidence interval is **0.8416212335729143**

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

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

Ans:

T-SCORE 95% confidence interval is **2.059538552753294**

T-SCORE 96% confidence interval is **2.1665866344527562**

T-SCORE 99% confidence interval is **2.787435813675851**

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs

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

**ANSWER:**

Solved in notebook (BASIC STAT LEVEL 1) attached with this docs