|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Ordinal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

(p)={TTT,THT,TTH,THH,HHH,HTH,HHT,HTT} N(P)=8

Probability of taking two head and one tail.

n(s)={THH,HHT,HTH}=3

N(S)=n(s)\n(p)=3/8

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

P ={(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

p(n)=36

a=sum is equal to one=0

Ans a) Probability of equal to 1 = 0

n(a)=0, p=n(a)/n(p)=0/36

b) sum is less than or equal to 4

n(b)= {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6

p(b)=n(b)/n(p)=6/36=1/6

Ans b)= p(b)=n(b)/n(p)=6/36=1/6

1. sum is divisible by 2 and 3

n(c)= {((1,5),(2,4),(3,3),(4,2), (5,1),(6,6,)}

N(c)=6

Ans c)=P(c)=n(c)/n(p)=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: We are considering that without repetition

P of getting none of the balls drawn is blue

(R\_R)+(R+G)+(G+R)+(G+G)

(2/7\*1/6) + (2/7\*2/6) + (1/6\*3/7) + (3/7\*2/6)

= 0.357143

Probability of getting none of the balls drawn in blue = 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 | XiPi |  |
| A | 1 | 0.015 | 0.015 |  |
| B | 4 | 0.20 | 0.8 |  |
| C | 3 | 0.65 | 1.95 |  |
| D | 5 | 0.005 | 0.025 |  |
| E | 6 | 0.01 | 0.06 |  |
| F | 2 | 0.120 | 0.24 |  |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

EXPECTED no of candies =xipi=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.

Answer=[Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)**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?

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans =** [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

**SP and Weight(WT)**

**Use Q9\_b.csv**

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

**Ans =** [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)



Ans: In the above data, range is highest – Lowest = 400 – 0 = 400

From a given histogram we can infer that ,its is positively skewed, and it is slightly tailed at right, so it is called as right Skew. The mean beings skew to the right of a typical center of the data

This is the assymetrical distribution of the data with a positive skewness hence in this case

Mean>median>mode



Ans: This boxplot shows that there is a presence of outilier in the dataset.

Its show the median is on the left side to the center of the box hence this is positively skewed, it is also showing that wishker of the boxplot if extende at right from that also we can infer that it is positively skewed it means that ,there is presence of outlier in data. From that we can also refer that data is highly spread on the right side of 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:- solution- population count N = 3000000, Sample count =2000, average=200, standard deviation=30.

Explaination : A is 94% confidence interval has two tail

σ/2=6/2=3

so, it can give from 3% to 97% which leaves 94% in the middle,the z-score of 94%

two closest value in the z-table

p(z<Z)=0.97

p(z<1.88)= *0.96995*

*p(Z<1.89)= 0.97062*

[Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

**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 : [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

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

Ans – when mean ,median of data is equal it means it is normally distributed and there is no skewness,

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

Ans- when mean is greater than median it means it is right skewness.

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

Ans- when median is greater than mean it indicate there is presence of left skewness.

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

Ans- positive kurtosis value indicates that there is slight right tailedness.

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

Ans- negative kurtosis value indicate that there is presences of slight left tailedness.

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



What can we say about the distribution of the data?

Ans- data is highly distributed at left hand side.

What is nature of skewness of the data?

Ans- nature of skewness of data is negative because, there is slight left skew and median is greater than mean.

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

Ans – IQR= Q3-Q1

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: Both the boxplot have same median value. Around 256(approx.)

Both the boxplot shows normal distribution ,median is passing through the center of IQR

1st boxplot is between 245 to 280,from that we can say that Q1=245 and Q3=245

2nd boxplot is between 180 to 320 from that we can say that Q1 = 180 and Q3=320

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)

Ans: [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

Q 21)

Check Whether the Adipose Tissue (AT) and Waist Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

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

Ans: [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

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

Ans: [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)

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 : [Statistical Assignment level\_1 - Jupyter Notebook](http://localhost:8888/notebooks/Desktop/Assignment/Statistical%20Assignment%20level_1.ipynb)