|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Continuous |
| 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 | Continuous |
| 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 | Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

**Solution:**

* Number of possible combinations are 23 =8

=P(HHT)+P(HTH)+P(THH)

=1/8+1/8+1/8=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**

**Solution:**

* 1) There is no outcomes which corresponding sum is equal to 0

**Which is** **0/36=0.**

2)There are 6 possible outcomes for getting the sum is equal to or less than which are (1,1) (1,2 )(2,1 )(1,3)(2,2)(3,1)=**6/36**

**Which is 1/6**

3) The sum is divisible by 2 and 3 are

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

**WHICH IS 6/36**

**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?**

**Solution:**

* balls =7
* When 2 balls are drawn at random the probability of getting non blue ball is **=10/27**

**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

**Solution:**

Expected number of candies for randomly selected child=**1\*0.15+4\*0.20+3\*0.65+5\*0.005+6\*0.01+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.**

**Use Q7.csv file**

**SOLUTION:**

**For Code Refer “Basic Statistics Level 1.ipynb”**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

**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?**

**Answer:**

Patients Weights(X)= 108, 110, 123, 134, 135, 145, 167, 187, 199

Probability of selecting a patient, p(x)= 1/9.

**Expected weight of that patient = ∑(x.p(x))**

= (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)

= 12+12.22+13.66+14.88+15+16.11+15.55+20.77+22.11

= 145.33

Expected value of the weight of the randomly chosen patient is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution: For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

**b) SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution: For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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



Upper Outlier: There is an outlier on the upper side of the box plot, indicating a data point that is significantly higher than the rest of the data.

Few Data Points Between Q1 and Minimum: There are relatively few data points between the first quartile (Q1) and the lowest data point.

Outliers at the Maximum Side: The outliers are observed at the maximum end of the data, suggesting some unusually high values.

Median < Mean: The median value is lower than the mean value, indicating a right-skewed distribution where the tail of the distribution extends towards higher values.



**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?

**Solution:** **For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

**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?

**Solution:** **For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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

**Answer:** **Symmetrical or Normal skewed.**

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

**Answer:** **Right skewed.**

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

**Answer:** **Left skewed.**

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

**Answer: The Data is normally distributed and kurtosis value is 0.**

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

**Answer: The distribution of the data 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?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
**Answer:**

* 50% of the data is above 10 and remaining is less than 10

In that 50% data which is above 10

60% of the data lies between 10 to 15 and 40% data lies between 15 to 18.

* Left skewed, median is greater than mean.
* Inter Quartile Range = Q3 – Q1 (where Q3 = 10 & Q1 = 18)

= 10-18

= -8.

Q19) Comment on the below Boxplot visualizations?



**Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.**

**Answer:**

* The median of both the boxplot is same i.e., 262.5
* Both the boxplots have equal distribution of data above and below the median. Hence, it’s shows normal distribution.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer : For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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

Dataset: wc-at.csv

**Answer : For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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

**Answer : For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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

**Answer : For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>

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 : For Code Refer “Basic Statistics Level 1.ipynb**

<https://colab.research.google.com/drive/1oP2UC0z7r9xe2bc-z6vkTn_qOqwWD5TZ?usp=sharing>