|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

1. 3/8

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

1. Equal to 1
2. 0
3. Less than or equal to 4
4. 1/6
5. Sum is divisible by 2 and 3
6. 5/36

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?

1. 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

1. 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

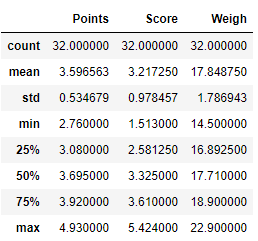
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Answer:**

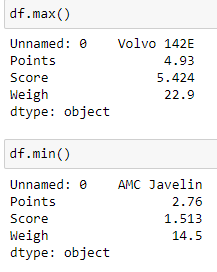
Mean, standard deviation is present in the below diagram extracted from the python.



Variance: (Points: 0.285, Score: 0.957378, Weigh: 3.193165)

Mode: (Points: 3.92, Score: 3.44, Weigh: 17.02)

Median: (Points: 3.695, Score: 3.325, Weigh: 17.710



Range= Highest – Lowest

Points Range: 2.17, Score Range: 3.911, Weigh Range: 8.4

Comments & Inferences:

* This data shows the Performance & attributes of different vehicles. Different Vehicle based on their performances are granted points and then including their weight, scores are given.
* From the data it is clear that Lincoln Continental Car has scored highest while Lotus Europa scored the lowest.
* Honda Civic has scored more number of Points and Valiant has scored the lowest.
* Merc 230 is the highest weighing Car and Ford Pantera L is the lowest Weighing Car.

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?

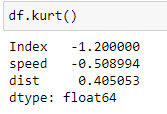
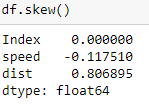
1. (1/9)\*(1308) = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**A)**



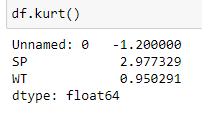
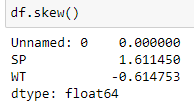
Speed has –ve skewness which says that more number of data points are on the higher range. It also has –ve kurtosis which says that the data point values are more evenly distributed

Distance has +ve skewness which says that more data points are present in the lower range. It also has +ve kurtosis which says that Data values for distance are not evenly distributed and has heavy tails i.e., outliers.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**A)**



SP has +ve skewness which tell us that more data points are present in the lower range. It has +ve kurtosis which tells that it has many outliers and the data is not evenly distributed.

WT has –ve skewness which says that more number of data points are present towards the higher range. It has +ve kurtosis which tell us that it has many outliers or heavy tails but not as much as SP.

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



1. The above box plot has positive skewness. More number of data is present on the lower range. Mean is less than median here.



1. Outliers are present. It is having positive skewness. More number of data values are present in the lower range.

**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?

1. n=2000, Mean = 200, S=30

This is solved using the python code stats.norm.interval

94% interval= CI= (198.738325292158, 201.261674707842)

98% interval= CI= (198.43943840429978, 201.56056159570022)

96% interval= CI= (198.62230334813333, 201.37769665186667)

**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.

**A)** Mean = 41

Variance = 25.5

Std. Deviation = 5.05

1. What can we say about the student marks?
2. Not much deviation of the marks of student from each other based on Standard deviation.

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

1. Normal Distribution and skewness = 0

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

1. Negative skewness

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

1. Positive Skewness

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

1. It says outliers are present and are more influential on the data.

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

1. Not much outliers or no outliers present and no affect on the data dispersion knowing.

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



What can we say about the distribution of the data?

**A)** It has –ve skewness. Most of the data points are present on the higher range. Median is present in between 14 and 16.

What is nature of skewness of the data?

**A)** The nature of the skewness is Negative

What will be the IQR of the data (approximately)?   
  
A) IQR=8  
  
Q19) Comment on the below Boxplot visualizations?



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

**A)** Number 2 data set is widely distributed than Number 1. Number 2 it seems like Mean and Median almost overlap whereas in Number 1 median is on lower side which say that more number of data set is on higher side whereas in Number 2 it looks like the data is evenly distributed.

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)

**A)** a. 33/81

b. 61/81

c. 70/81

I have done this by using excel and not the python

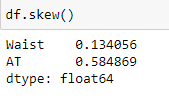
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. The skewness of the MPG data is -0.17. So it is almost zero. Ideally if the skewness is completely zero we can say it is normally distributed. But here the value is almost zero the data set is almost normally distributed.
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

1. Adipose Tissue(AT) is almost having zero skewness and can say it is almost normally distributed.

Waist is having some skewness which is away from zero. So it is not normally distributed.

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

A) As per the left Z table the below values are given

90% CI = 1.65

94% CI = 1.88

60% CI = 0.85

For Right Z table the values may differ.

From the z distribution table and graph from some particular website

For information on Z table : ( <https://www.statisticshowto.com/tables/z-table/> )

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

A) As sample size is less than 30, we should use the T scores to calculate the confidence interval

<https://www.statisticshowto.com/probability-and-statistics/confidence-interval/>

95% CI = 1.711

96% CI = 1.828

99% CI = 2.492

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

1. Mu=270, x bar = sample mean = 260, sample deviation=s= 90

n=18

t= -0.4714

Reference website: <https://www.statisticshowto.com/probability-and-statistics/t-distribution/t-score-formula/>

Degrees of freedom = n-1=18-1=17

The p-value is .479567 this value is from an online calculator.

So 1-0.47957=0.52043

So 52.043% Confidence that the CEO claims are true

And 47.957% that the 18 random selected bulbs will have life less than 260 days.