**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| 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 |

**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) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

Ans:- Total Number of possible Combination :- HHH , **HHT** , **HTH** , HTT , **THH** , THT , TTH , TTT => 8 Combination

two heads and one tail :- HHT , HTH , THH

So the Probablility is => **3/8 = 0.375**

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

1. Equal to 1 **= 0**
2. Less than or equal to 4 =

Probability of getting Sum as 4 = (1,1) , (1,2) , (1,3) , (2,1) , (2,2) , (3,1)

Total number of Combination = 36

probability that sum is = 6 / 36 **= 0.1666**

1. Sum is divisible by 2 and 3

Probability of getting Sum is divisible by 2 and 3 = (1,5) , (2,4) , (3,3) , (4,2) , (5,1) , (6,6)

probability that sum is = 6/36 = **0.166**

**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 :- Total Number of Balls = (2 +3+2) = 7

Drawing 2 balls out of 7 = 7C2 = (7x6)/(2x1) = 21

Number of ways drawing two balls red and green = (2+3) = 5

Drawing 2 balls out of 5 = 5C2 = (5x4)/(2x1) = 10

= 10 / 21 = **0.4761**

**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:- Probability = 0.015 + 0.20 + 0.65 +0.005 + 0.01 + 0.120 **= 3.09**

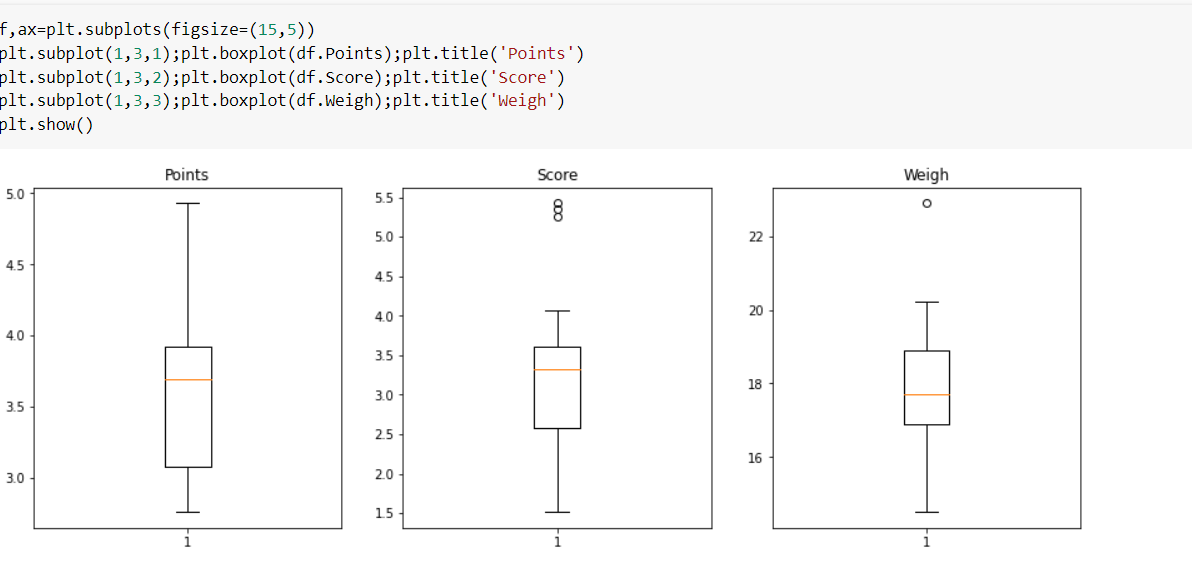
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* **For Points,Score,Weigh>**

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

**Ans :- **

|  |  |  |  |
| --- | --- | --- | --- |
|  | **POINTS** | **SCORE** | **WEIGH** |
|  |  |  |  |
| **MEAN** | 3.596563 | 3.217250 | 17.848750 |
| **MEDIAN** | 3.695 | 3.325 | 17.710 |
| **MODE** | 0 3.07  1 3.92 | 0 3.44 | 0 17.02  1 18.90 |
| **VARIANCE** | 0.285881 | 0.957379 | 3.193166 |
| **STANDARD DEVIATION** | 0.534679 | 0.978457 | 1.786943 |
| **RANGE** | 2.17 | 3.9110000000000005 | 8.399999999999999 |

****

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 patientsProbability of selecting each patient = 1/9Expected 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

**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**

**Ans:- **

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



Ans :- The Histogram peak has right skew and tail is on right . Mean > Median We have outliers on the higher side.



Ans:- The boxplot has outliers on the maximum 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?

Ans:- For 94% confidence interval Range is [ 198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

**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.



|  |  |
| --- | --- |
| **MEAN** | 41.0 |
| **MEDIAN** | 40.5 |
| **MODE** | 41 |
| **VARIANCE** | 25.529412 |
| **STANDARD DEVIATION** | 5.052664 |
| **RANGE** | 22 |

1. What can we say about the student marks? -

**Ans:- From above plot we can say that mean of marks of student is 41 which is slightly greater than median. Most of the students got marks in between 41-42, there are two outlier 49,56.**

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

**Ans :- If the distribution is symmetric , then the mean is equal to median , and the distribution has zero Skewness**

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

**Ans:- If mean is greater than median , the distribution is positively skewed**

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

**Ans :- If mean is less than median , the distribution is negatively skewed**

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

**Ans:- Positive Kurtosis indicates a distribution where more of the number are located in the tails of the distribution instead of around the mean. (thick tails)**

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

**Ans:- Negative Kurtosis indicates a distribution is flat and has thin tails**

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



What can we say about the distribution of the data?

Ans:- The above Boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

Ans:- 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)?   
Ans:- The Inter Quantile Range = Upper Quartile – 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.

Ans:- First there is no Outliers. Second both the box plot shares same median that is approximately in the 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)

Ans:- The Mean of MPG is :- 34.42207572802469 = df.MPG.mean()

Standard Deviation is :- 9.131444731795982 = df.MPG.std()

1. **The probability of MPG of Cars P(MPG>38) :**

1 - stats.norm.cdf(38 , df.MPG.mean(), df.MPG.std()) = 0.34759392515827137

1. **The probability of MPG of Cars P(MPG<40) :**

stats.norm.cdf(40 , df.MPG.mean(), df.MPG.std()) = 0.7293498762151609

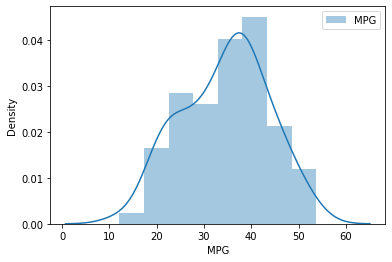
1. **The probability of MPG of Cars P(20<MPG<50) :**

stats.norm.cdf(50 , df.MPG.mean(), df.MPG.std()) - (1 - stats.norm.cdf(20 , df.MPG.mean(), df.MPG.std())) = 0.01311646961052304

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

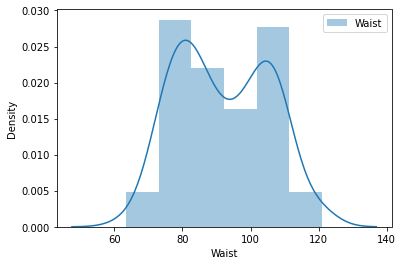
Dataset: Cars.csv



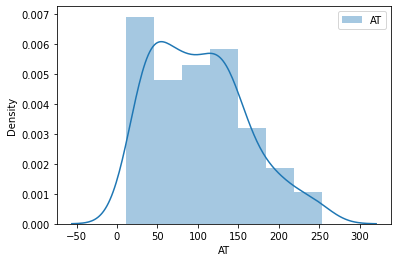
Ans :- The MPG of Cars is distributed Normally

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

Dataset: wc-at.csv



Waist Circumference(Waist) is distributed Normally



Adipose Tissue (AT) is distributed Normally

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

|  |  |
| --- | --- |
| **confidence interval** | **Z score** |
|  |  |
| 90% | 1.645 |
| 94% | 1.88 |
| 60% | 0.841 |

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

|  |  |
| --- | --- |
| **confidence interval** | **t score** |
|  |  |
| 95% | 2.060 |
| 96% | 2.1715 |
| 99% | 2.7874 |

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

t=(260-270)/(90/18\*\*0.5) = -0.4714045207910317

p\_value=1-stats.t.cdf(abs(-0.4714),df=17) = 0.32167411684460556

p\_value=stats.t.sf(abs(-0.4714),df=17) = 0.32167411684460556