**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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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?**

- Let S be the sample space then

S = {HHH,HHT,HTT,THT,TTH,HTH,THH,TTT}

event E is {HHT,HTH,THH}

Total number of possible combinations = 23 = 8

The probability of two heads and one tail when three coins are tossed simultaneously are

P (Two heads and One tail) = Number of desired outcomes

= 3/8 or 0.375

**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**

* a) If two dices were rolled, then total possible cases =36

Total Favorable outcome (Having sum =1) = 0

The minimum sum is 2 for outcome (1,1).

Hence, probability is 0

* b) The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, 6}

So, when we roll two dice there are 6 × 6 = 36 possibilities.

When we roll two dice, the possibility of getting Sum Less than or equal to 4 is (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1), (1, 1), (1, 2), and (2, 1)

P (Sum less then or equal to 4) = Number of desired outcomes

= 9 / 36 or 0.25

* c) The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, 6}

So, when we roll two dice there are 6 × 6 = 36 possibilities.

When we roll two dice, the possibility of getting Sum is divisible by 2 and 3 is (1 , 5) , (2 , 4) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

P(Sum is divisible by 2 and 3) = Number of desired outcomes

= 6/36 or 0.17

**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?**

* Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

= 7C2

= (7\*2)/(2\*1) = 21

Let E = Event of drawing 2 balls, none of which is blue.

n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.

= 5C2

= (5\*4)/(2\*1) = 10

P(E) = n(E)/n(S) = 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**

* To calculate the expected number of candies for a randomly selected child, we need to multiply each possible count of candies by its corresponding probability and sum up the results. This will give us the probability of number of candies.

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

= 3.09

**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**

* Mean:-

Mean = (Sum of all the values) / (Number of values)

So, the mean for Points, Score, and Weigh are:

Mean of Points = 3.59

Mean of Score = 3.21

Mean of Weigh = 17.84

Therefore, the mean of Points is 3.59, the mean of Score is 3.21, and the mean of Weight is 17.84

* Median:-

The median for Points, Score, and Weigh are:

Median of Points = 3.69

Median of Score = 3.32

Median of Weigh = 17.71

* Mode:-

Mode of Points = 3.07

Mode for Score = 3.44

Mode for Weigh = 17.02

* Variance:-

Variance for Points = 0.28

Variance for Score = 0.95

Variance for Weigh = 3.19

* Standard Deviation:-

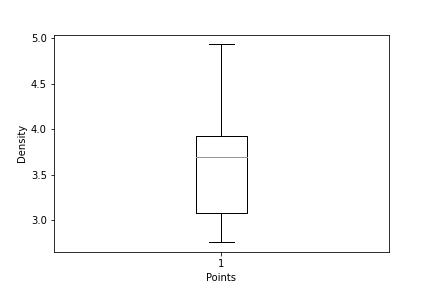
Standard Deviation for Points = 0.53

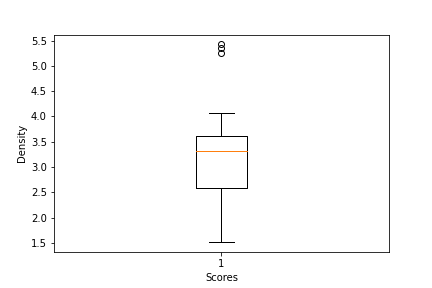
Standard Deviation for Score = 0.97

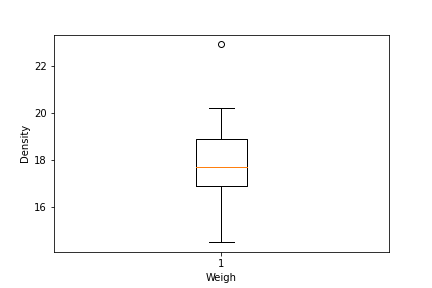
Standard Deviation for Weigh = 1.78

* Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]

Inferences







**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?**

* Expected value = Sum (X \* Probability of X)

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

= 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**

* 9a) Skew

Index 0.000000

speed -0.117510

dist 0.806895

Kurtosis

Index -1.200000

speed -0.508994

dist 0.405053

* 9b) skew

SP 1.611450

WT -0.614753

Kurtosis

SP 2.977329

WT 0.950291

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



* The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



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?**

* 94% = (198.738325292158, 201.261674707842)
* 98% = (198.43943840429978, 201.56056159570022)
* 96%= (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.**
2. **What can we say about the student marks?**

* 1) Mean = 41

Median = 40.5

Variance = 25.529411764705884

Standard Deviance = 5.05266382858645

* 2) We can say that the mean of students' marks is 41 which is slightly greater than the median. Most of the students got marks in between 41-42, there are two outliers 49,56.

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

* Zero

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

* Positive

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

* Negative

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

* In statistics, kurtosis is used to describe the shape of a probability distribution.

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

* In statistics, kurtosis is used to describe the shape of a probability distribution.

Negative kurtosis value indicates that the distribution has lighter tails than the normal distribution

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



**What can we say about the distribution of the data?**

* The above Boxplot is not normally distributed the median is towards the higher value

**What is nature of skewness of the data?**

* 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)?**

* The Inter Quantile Range = Q3 Upper quartile – Q1 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.**

* There are no outliers. Both the box plots share the same median that is approximately in a 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)**

**c. P (20<MPG<50)**

* a) 0.3475939251582705

b) 0.7293498762151616

c) 0.013116469610523374

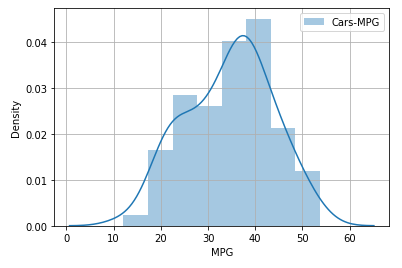
**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

* From the below plot and values we can say that data is fairly symmetrical, i.e fairly normally distributed

MPG of cars follows normal distribution



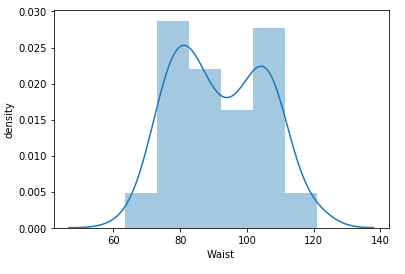
Mean = 34.422075728024666

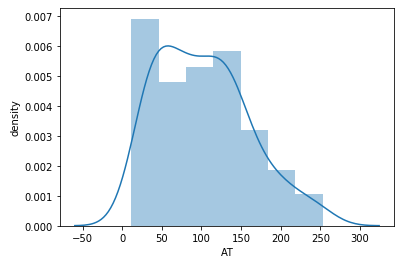
Median = 35.15272697

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

**Dataset: wc-at.csv**







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

* 90% = 1.2815515655446004

94% = 1.5547735945968535

60% = 0.2533471031357997

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

* 95% = 2.0638985616280205

96% = 2.1715446760080677

99% = 2.796939504772804

**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\_score = (x - pop mean) / (sample standard daviation / square root of sample size)

(260-270)/90/np.sqrt(18))

t\_score = -0.471

stats.t.cdf(t\_score, df = 17)

0.32 = 32%