|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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

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

Ans: Three coins = 2^(3) = 8

(H, H, T), (H, T, H), (T, H, H) = 3

P = 3/8

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

1. Equal to 1

Ans: P(X=1) = 0 / 36 = 0

(Minimum addition starts from (1+1)= 2)

1. Less than or equal to 4

Ans: (X <=4) = (1,1)(1,2)(1,3)(2,1)(2,2)(3,1) = 6

Two Dice = 6^(2) = 36

P = 6/36 = 1/6

1. Sum is divisible by 2 and 3

Ans: (X/2 & X/3) = (1,5)(2,4)(3,3)(4,2)(5,1) = 5

P = 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?

ANS : Total no of balls = 7  
 two balls drawn = 2  
 so 7C2 = 21  
 No blue ball should drawn = 2  
 except blue total no of ball = 5  
 so 5C2=10  
 **=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

ANS: =1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

= 0.015+0.80+1.95+0.025+0.06+0.24

=3.09

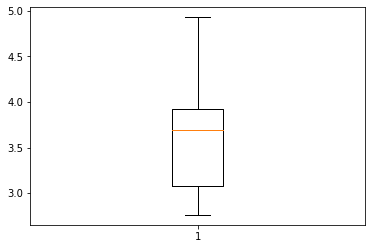
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

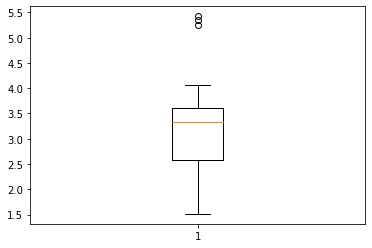
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file **

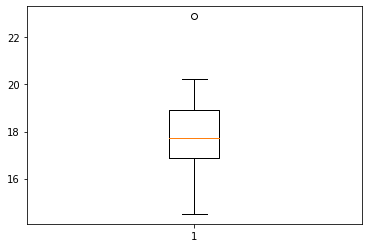
**Points:**

****

**Score::**

****

**Weigh::**

****

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 : mean = (108+110+123+135+145+167+187+199)/9 = 145.33

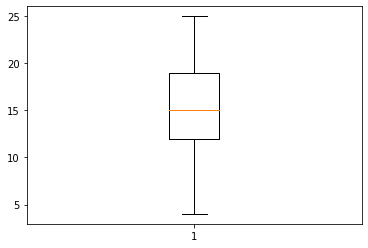
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

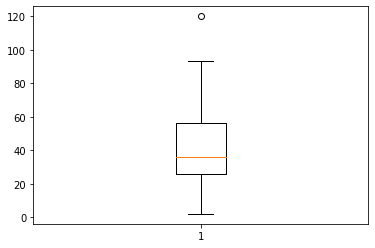
**Use Q9\_a.csv**

****

**Speed::**

****

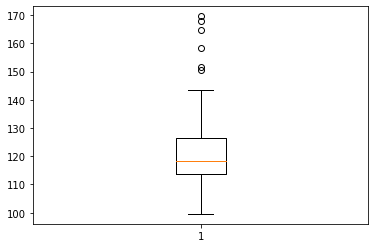
**Dist::**

****

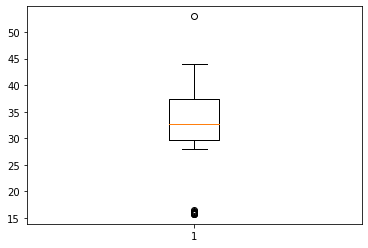
**SP and Weight(WT)**

**Use Q9\_b.csv**

**SP :**

****

**WP:**

****

****

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



1)positive skewness



1. Outliers at Upper side
2. Positive skewness

**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: import numpy as np

import pandas as pd

from scipy import stats

from scipy.stats import norm

import math

#Avg weight of adult mexico with 94% CI

**stats.norm.interval(0.94,200,30/math.sqrt(2000))**

(198.738325292158, 201.261674707842)

#98%CI

**stats.norm.interval(0.98,200,30/math.sqrt(2000))**

(198.43943840429978, 201.56056159570022)

#96%CI

**stats.norm.interval(0.96,200,30/math.sqrt(2000))**

(198.62230334813333, 201.37769665186667)

**stats.norm.ppf(0.94)**

1.5547735945968535

**stats.norm.ppf(0.98)**

2.0537489106318225

**stats.norm.ppf(0.96)**

1.7506860712521692

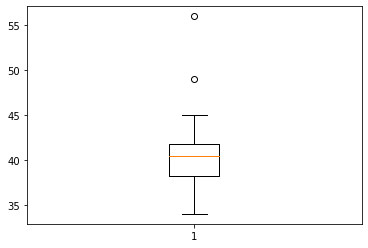
**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

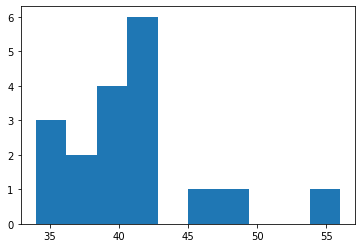
1. Find mean, median, variance, standard deviation.



1. What can we say about the student marks?



Mass of students Marks is between 38-42 &



Positive skewness.

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

ANS:: Data is Normalized and No Skewness.

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

ANS:: Negative Skewness

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

ANS: Positive Skewness

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

ANS: Positive Kurtosis indicates Thinner peak and Wider Tails.

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

ANS: Negative kurtosis indicates Wider peak and Thinner Tails.

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



What can we say about the distribution of the data?

Ans : Negative Skew Distribution (Not normal Distribution)

What is nature of skewness of the data?

Ans: Negative Skewness.

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

Ans : (10 – 18)

Q19) Comment on the below Boxplot visualizations?



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

ANS : The box plot 2 is highly distributed compared to box plot 1. Both the box plots are symmetrical and median is same for both. In box plot 2 the data range is high and it will be difficult to make any predictions.

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)

ANS: P = 33/81

* 1. P(MPG<40)

ANS: P = 61/81

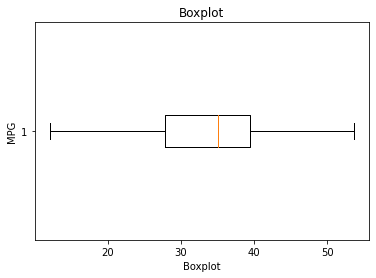
* 1. P (20<MPG<50)

ANS: P = 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



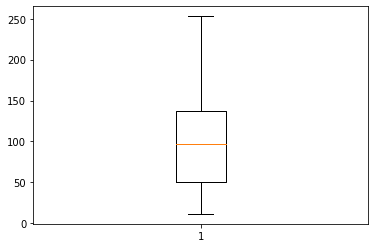
Ans: Negative Distribution( Not Normally Distributed)

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

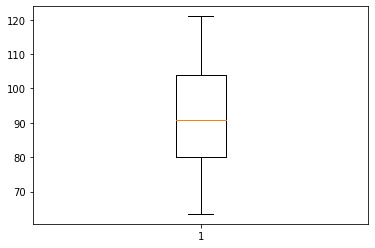
Dataset: wc-at.csv

Ans:

AT: Normally distributed



Waist: Normally distributed



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

Ans: **from scipy import stats**

**from scipy.stats import norm**

**#z-score of 90% confidence interval**

stats.norm.ppf(0.95)

1.6448536269514733

**#z-score of 94% confidence interval**

stats.norm.ppf(0.97)

1.8807936081512509

**#z-score of 60% confidence interval**

stats.norm.ppf(0.80)

0.8416212335729143

1. Z scores of 90% confidence interval = 1.65
2. Z scores of 94% confidence interval = 1.88
3. Z scores of 60% confidence interval = 0.84

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

Ans: from scipy import stats

stats.t.ppf(0.975,df=24)

2.0638985616280205

stats.t.ppf(0.98,df=24)

2.1715446760080677

stats.t.ppf(0.995,df=24)

2.796939504772804

1. t scores of 95% confidence interval = 2.06
2. t scores of 96% confidence interval = 2.17
3. t scores of 99% confidence interval = 2.80

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:

**from scipy import stats**

**from scipy.stats import norm**

**import math**

**#x = 260, mu = 270, n= 18, SD = 90**

**t = (260-270)/(90/(math. sqrt( 18 )))**

**t**

-0.4714045207910317

**#using cdf function**

**p\_value = 1-stats.t.cdf(abs(-0.4714045207910317), df=17)**

**p\_value**

0.32167253567098353

**#or using sf function**

**p\_value = stats.t.sf(abs(-0.4714045207910317), df = 17)**

**p\_value**

0.32167253567098364

**#probability that 18 randomly selected bulbs would have an average life no more than 260 days is 32.17%.**

**#Assuming Significance value alpha(α)=0.05(standard value)**

**#if p\_value < α ----reject Ho(null) and accept Ha(Alternative)**

**#if p\_value > α ----accept Ho and Reject Ho**

**#i,e., therefore: p\_value is > α , then accept Ho, The CEO claims are false and the average life of bulb > 260 days**.