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

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

N(S) when the coin tossed thrice = (Possible outcome with one trail)No. of Trial = 23 =8

I.e S={(HHH),(HHT),(HTH),(HTT),(TTT),(TTH),(THT),(THH)}

A={(HHT),(HTH),(THH)}

P(2 heads and 1 Tail)=n(A)/n(S)=3/8

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

1. **Equal to 1**

**Solution:** N(s)=(Possible outcome with one trail)No. of Trial =62=36

Since no two number add to make the sum =0, n(A)=0

P(Sum Equal to 1)= n(A)/n(s)=0/36=0

1. **Less than or equal to 4**=

**Solution:** n(s)=(Possible outcome with one trail)No. of Trial =62=36

(Sum<=4)={(1,3),(3,1),(2,2),(1,1),(2,1),(1,2)}

P(Sum<=4)= n(B)/n(s)=6/36=1/6

1. **Sum is divisible by 2 and 3=**

**Solution:** n(s)=(Possible outcome with one trail)No. of Trial =62=36

(Sum Divisible by 2 and 3)= {(1,5), (5,1),(2,4), (4,2), (3,3), (6,6)}

P(Sum Divisible by 2 and 3)= n(C)/n(s)=6/36=**1/6**

**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**=n(s)=Total possible outcomes when 2 random balls drawn from a bag with 7 total balls=7C2 =(7!/((7-2)!(2!)=21

N(a)=Possible outcome when No Blue ball to be appeared from a bag containing 2 red, 3 green and 2 blue balls =5C2 =(5!/((5-2)!(2!)=10

P(A) =n(a)/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

**Solution:** Expected value of a random variable x =E(x) = ∑x P(x)=

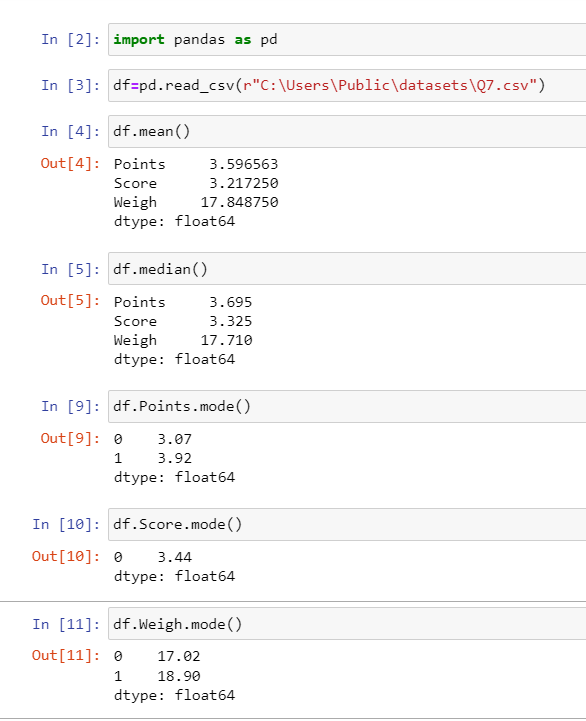
(1x0.015) + (4x0.2) + (3x0.65) + (5x0.005) + (6x0.01) + (2x0.12)=**3.09**

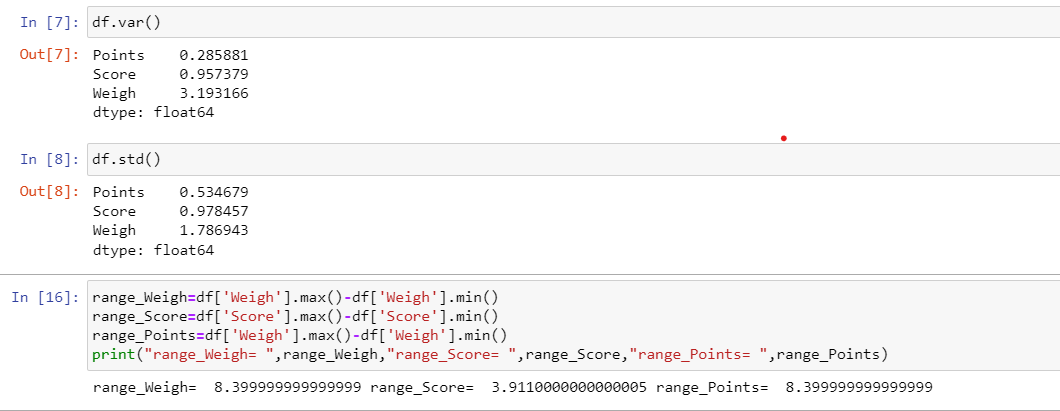
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

**Solution:**



****

**Use Q7.csv file**

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= P(x).E(x)=(1/9)\*( 108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199)=145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution:**

**Inference: Since skewness of “speed” is negative, the data is said to be slightly left skewed where as “Distance” is right skewed.**

**Kurtosis**

speed -0.508994

dist 0.405053

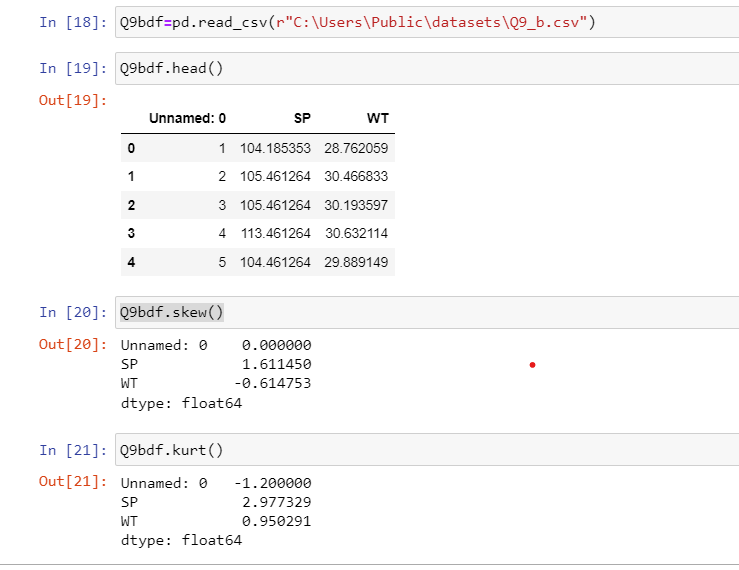
**Inference: Since both the columns have kurtosis<3, data is said to be Platykurti**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution:**

**Code:**

****

Inference: Since skewness of “SP” is positive, the data is inferred as right skewed, skewness of “Weight” is negative, the data is inferred as slightly right skewed.

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



**Solution**: By going through the graph, we can infer the data is right skewed and hence mean<median.

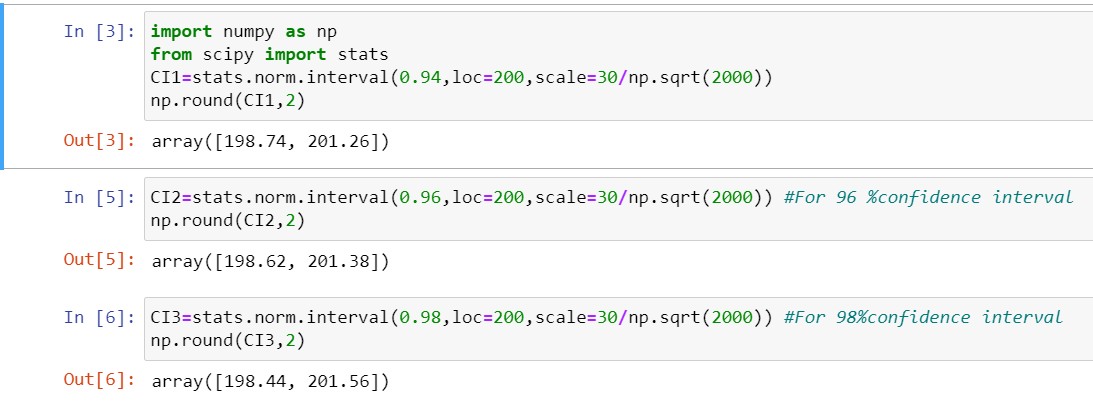


**Solution:** Data contains outliers in the upper end. i.e above 3rd quartile. Data distribution is right skewed.

**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:**

**Code:**

****

**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**

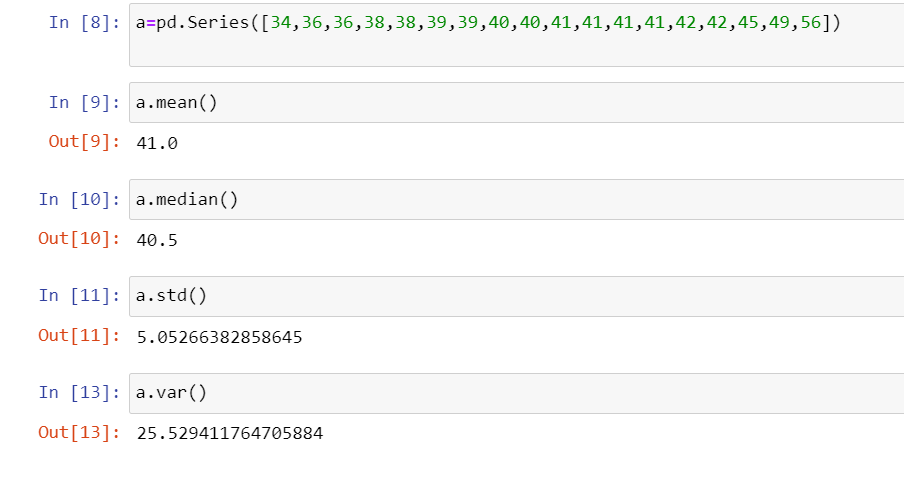
1. Find mean, median, variance, standard deviation.

**Solution:** Mean: 41.0

Median: 40.5

Standard Deviation: 5.05

Variance: 25.529



1. What can we say about the student marks?

**Solution: In student marks data, mean is greater than median. Hence we can infer that the data is positively skewed**

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

**Solution:** When mean, median are equal, data is said to be normally distributed. Hence no skewness

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

**Solution:** Data is said to be positive skewness.

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

**Solution:** Data is said to be in negative skewness.

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

**Solution:** When a peak of the distribution curve is relatively high, then the curve is said to be positive kurtosis (Leptokurtic). It indicates that the data distribution has higher degree of peakness

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

**Solution:** When a peak of the distribution curve is relatively low (flat), then the curve is said to have negative kurtosis (Mesiokurtic). It indicates that the data distribution has higher degree of peakness

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



What can we say about the distribution of the data?

**Solution:** In the boxplot, we can infer that median of data is approximately at 15.

What is nature of skewness of the data?

**Solution:** Distribution is left skewed

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

**Solution:** IQR=(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.

**Solution:**  Median of the Boxplot is almost equal to median of Boxplot i.e 262.

But data distribution range in Boxplot is small in comparison with boxplot 2.

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)

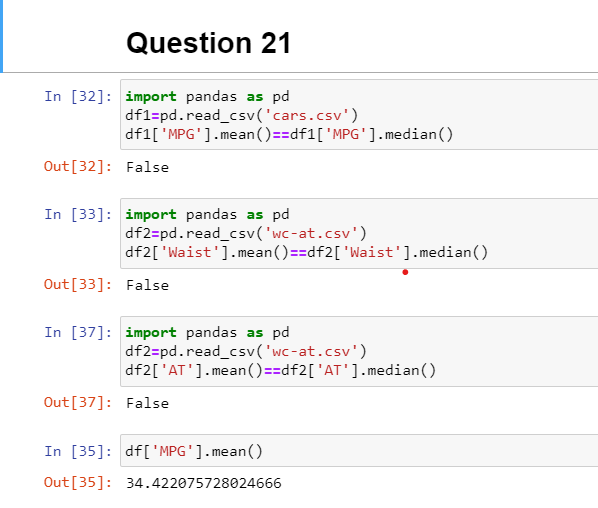
**Solution:**



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution:** 

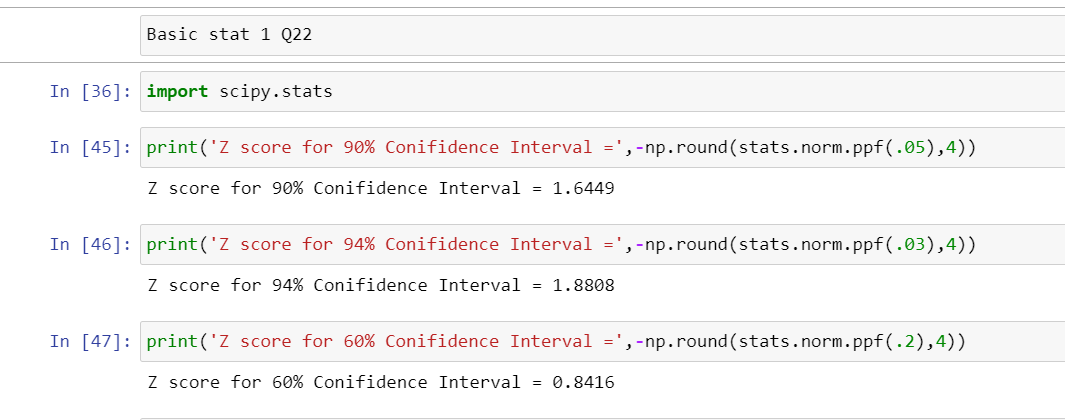
Conclusion: MPG of Cars does not follow Normal Distribution

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

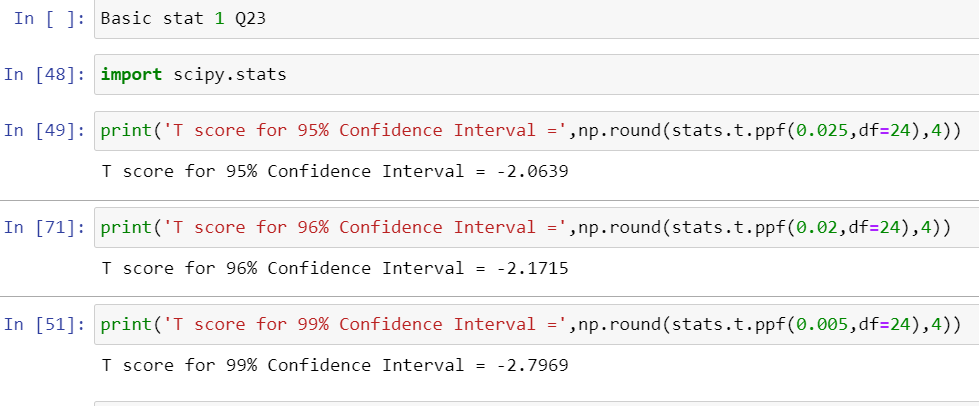
Dataset: wc-at.csv

Conclusion: Both Adipose Tissue (AT) and Waist Circumference(Waist) of wc-at data does not follow Normal Distribution

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



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

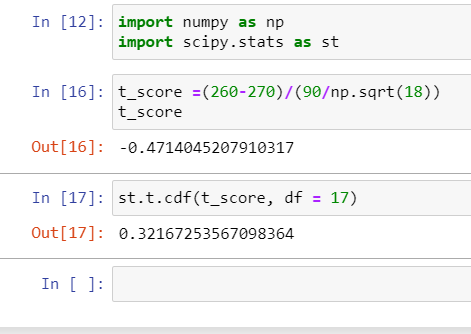


Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom



Ans: 32%