|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | discrete |
| Results of rolling a dice | discrete |
| Weight of a person | continous |
| Weight of Gold | continous |
| Distance between two places | continous |
| Length of a leaf | continous |
| Dog's weight | continous |
| 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 | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nomainal |
| 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 | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

When three coins are tossed, there are 2\*\*3 =8 possible outcomes

(HHH,HHT,HTH,HTT,THH,THT,TTH,TTT)

Two heads and one tail: 3(HHT,HTH,THH)

Probability that two heads and one tail are obtained = 3/8

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
4. Possibility of getting sum equal to 1=0(minimum sum will be two(1,1))

Probabilityof getting sum equal to 1=0/36=0

1. Possibility of getting sum equal to 4=6((1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1))

Probabilityof getting sum equal to 4=6/36=1/6

1. Possibility of getting Sum divisible by 2 and 3=10 ( (1, 5), (2, 4), (3, 3), (4, 2), (5,1), (2, 6), (4, 4), (6, 2), (3, 6), (6, 3))

Probabilityof getting sum equal to 10=10/36

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?

The total number of ways to draw 2 balls out of 7 = 7C2 = 21

Number of ways to draw 2 balls without getting any blue balls= 5C2=10

The probability of drawing 2 balls without getting any blue balls = 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

Expected number of candies for a randomly selected child, E[X]=(0.015\*1)+(0.20\*4)+(0.65\*3)+(0.005\*5)+(0.01\*6)+(0.120\*2)

=0.015+0.8+1.95+0.025+0.06+0.24

=3.14

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**

Answers :

Points

Mean=3.5965625, Median=3.695, Mode=3.92, Variance=0.285881351, Std dev=0.534678736, Range= 2.17

Score

Mean=3.21725, Median=3.325, Mode=3.44, Variance=0.957378968, Std dev=0.978457443, Range= 3.911

Weigh

Mean=17.84875, Median=17.71, Mode=17.02, Variance=3.193166129, Std dev=1.786943236, Range= 8.4

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

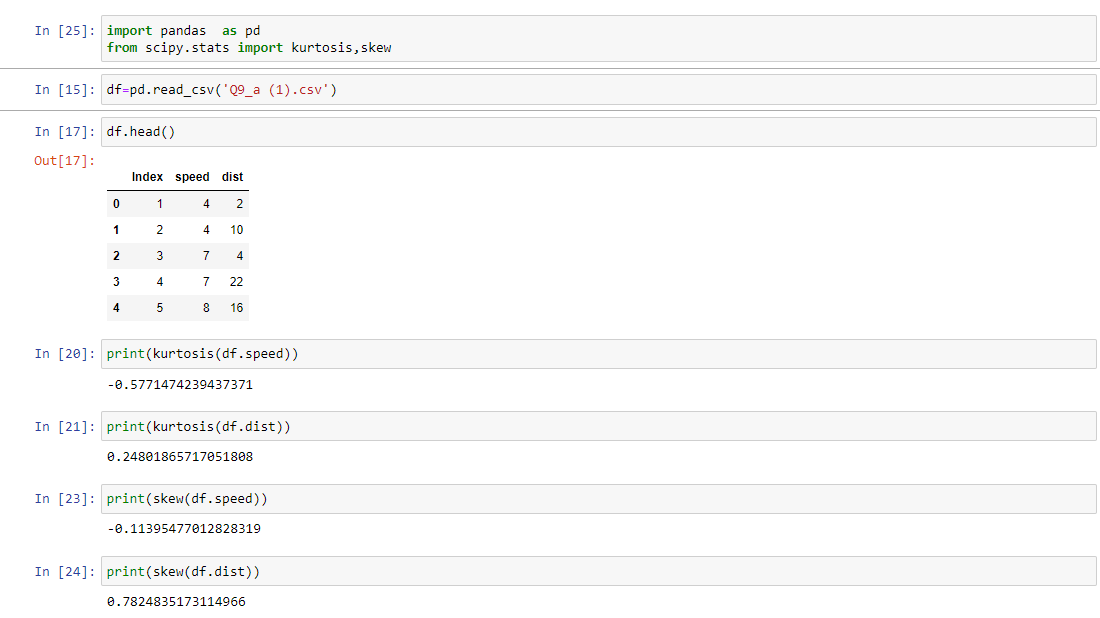
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Expected value=145.33333333333334

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

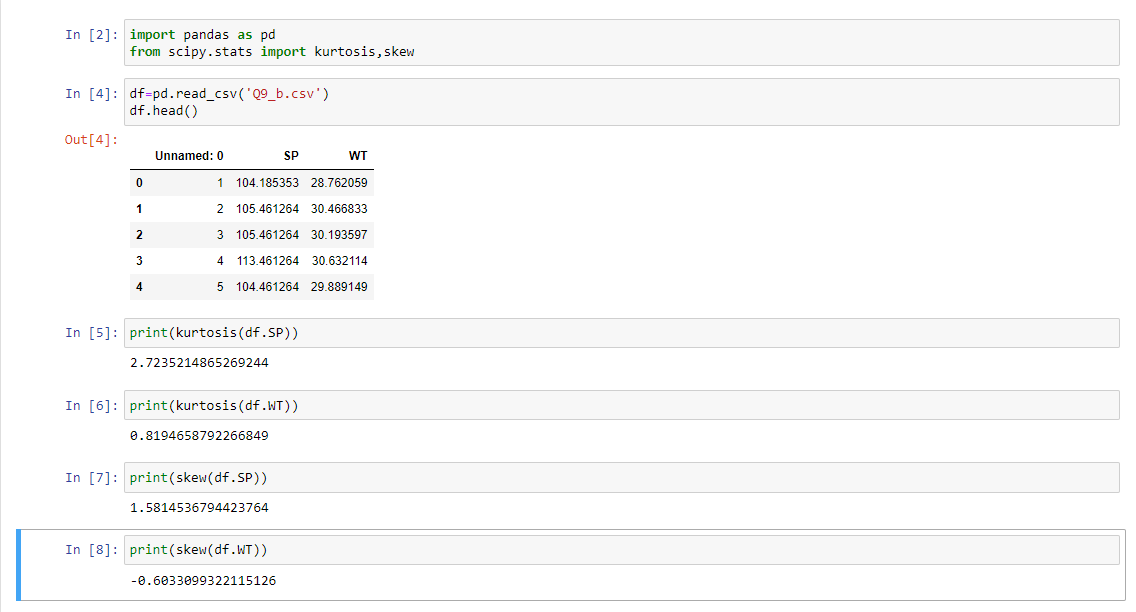
**Cars speed and distance**

**Use Q9\_a.csv**



**SP and Weight(WT)**

**Use Q9\_b.csv**



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

Histogram:

Here the following histogram follows right skewed distribution. Here mean>median. Here 350- 400 contains less frequency and 50-100 contains more frequency.

BoxPlot:

Here the following boxplot follows right skewed distribution. Here mean>median.The upper fence is greater than the lower fence



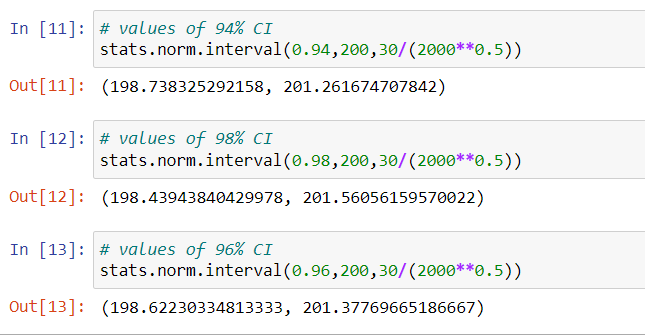
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Answer: given,

Sample size of men,n=2000

Average weight of sample,x\_bar=200

Sample standard devaition,s=30



**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, Variance= 24.11111111111111, Standard Deviation=4.910306620885412

1. What can we say about the student marks?

* Most of the student got mark as 41
* Less student mark is 34 and highest mark is 56
* Average of the student marks is 41
* 4 student got 41 student mark

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

It will be a Normal Distribution,Skewness will be zero,If we separate curve by median,both side will be symmetrical

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

It will be positively skewed or right-skewed distribution has a long tail. Tail will be on the right side of the distribution.If we separate curve by median,both side will be asymmetrical

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

It will be negatively skewed or left-skewed distribution has a long tail.Tail will be on the left side of the distribution.If we separate curve by median,both side will be asymmetrical

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

Positive value of kurtosis indicate that distribution is peaked and possesses thick tails. Extremely positive kurtosis indicates a distribution where more numbers are located in the tails of the distribution instead of around the mean.

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

A distribution with a 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 given box plot is a type left skewed(negetively skewed ) distribution
* 25 th percentile of the data having value 10,75 th percentile of the data having value 18
* Here the lower fence very high and upper fence is low

What is nature of skewness of the data?

* Left skewed or negetively skewed

What will be the IQR of the data (approximately)?   
 IQR= Q3- Q1

=18-10

=8

Approximate value for IQR will be 8

Q19) Comment on the below Boxplot visualizations?



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

* Both Boxplots are normal distributed. Mean=median. Upper fence and lower fence are symmetrical in both boxplot

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)

Count of MPG of Cars, N = 81

Count of MPG > 38 = 33

Probability = Number of observations with MPG > 38/N = 33/81

* 1. P(MPG<40)

Count of MPG of Cars, N = 81

Count of MPG < 40 = 61

Probability=Number of observations with MPG <40/N

=61/81

* 1. P (20<MPG<50)

Count of MPG of Cars, N = 81

Count with 20< MPG < 50 = 69

Probability=Number of observations with20< MPG < 50/N

=69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

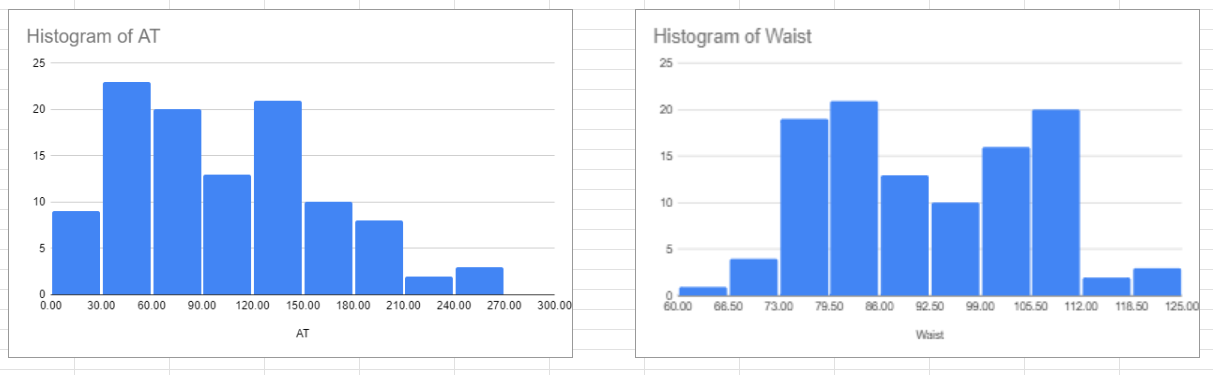
MPG of cars follows left skewed distribution.It is having long tail in left side.

Here mean is less than median.

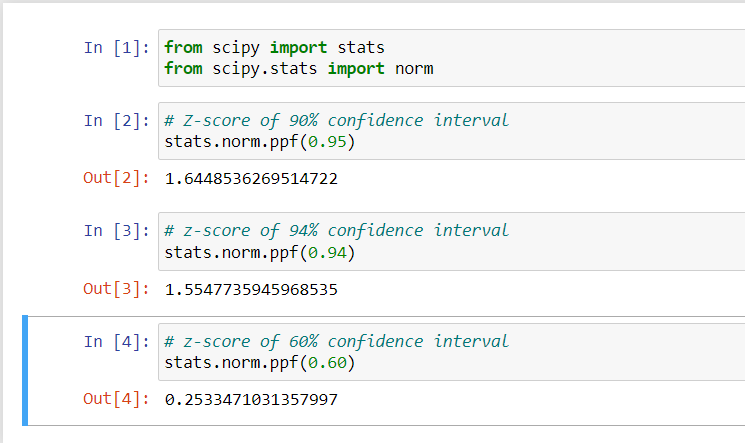
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Both are following normal distribution

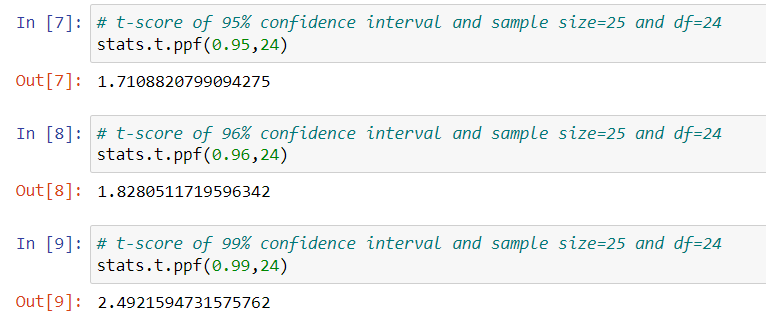


Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

Given sample size=25, degree of freedom=n-1=25-1=24



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

Given x\_bar=sample mean=260

μ =population mean=270(claimed)

n= sample size=18

S=sample standard deviation=90

And

df=degree of freedom=n-1=18-1=17

Assume Null Hypothesis is: Ho = Avg life of Bulb >= 260 days

Alternate Hypothesis is: Ha = Avg life of Bulb < 260 days

finding t-scores at x=260; t=(x\_bar-μ)/(s/sqrt(n))

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

=-0.4714045207910317

Finding P(X>=260) for null hypothesis

P=p(abs(tscore,df))

=0.32167411684460556

So the probability will be 0.32167411684460556.