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

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

When 3 coins are tossed, the sample space was 🡪 S = { HHH, TTT, HTT, THT, TTH, THH, HTH, HHT}

Number of elements in the obtained sample space 🡪 n(S) = 8

The probable outcomes of getting two heads and one tail 🡪 {HHT,HTH,THH}

So the Probability of getting two head and one tail 🡪 **3/8**

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

a) Equal to 1

b) Less than or equal to 4

c) Sum is divisible by 2 and 3

1. Probability that the sum is equal to 1 🡪 **Zero** (as the minimum value obtained by rolling two dice is 2)
2. Less than or equal to 4 🡪 outcomes such that sum is less than or equal to 4 🡪 6 outcomes ({(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} )

Probability that sum is less than or equal to 4 🡪 **6/36**

1. Sum divisible by 2 and 3

🡪 Favorable outcomes are 🡪(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)

The probability of getting the sum which is divisible by 2 & 3 is 6/36.

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?

Total number of balls in the bag 🡪 (2 + 3 + 2) 🡪 7  
The number of ways of drawing 2 balls out of 7 🡪 n(s) 🡪 7C2   
🡪 (7 x 6) / (2x 1) = 21

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

The number of ways of drawing 2 balls out of (2 red + 3green) 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

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**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | **Points** | **Score** | **Weigh** |
| Mean | 3.597 | 3.217 | 17.85 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 18.90 |
| Variance | 0.276 | 0.297 | 3.09 |
| Standard Deviation | 0.53 | 0.97 | 1.78 |
| Range | [2.76, 4.93] | [1.513, 5.424] | [14.5, 22.9] |

**Inferences:**

* “Points” & “Score” are negatively skewed and “Weigh” is positively skewed.
* “Points” has low Variance, Standard Deviation, and Range.

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

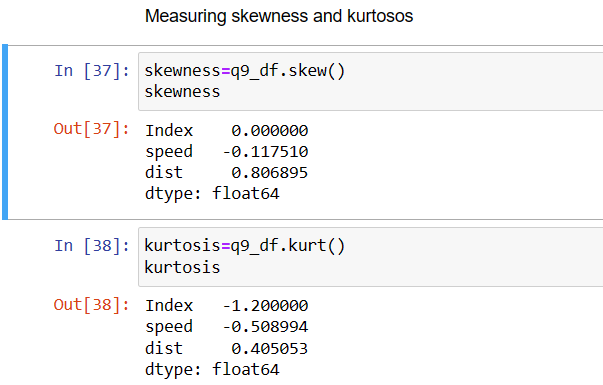
108, 110, 123, 134, 135, 145, 167, 187, 199

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

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

Cars speed and distance

Use Q9\_a.csv



Inferences:

For speed and distance:

Skewness:

speed 🡪 0.117510 🡪 Fairly Symmetrical

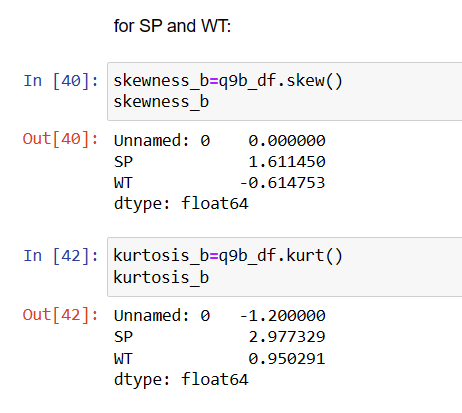
dist 🡪 0.806895 🡪 Moderately Skewed

Kurtosis :

speed -0.508994 🡪 Platykurtic distribution

dist 0.405053 🡪 leptokurtic distribution

Use Q9\_b.csv



Skewness:

SP 1.611450 🡪 Highly Skewed

WT -0.614753 🡪 Moderately Skewed

Kurtosis :

SP 2.977329 🡪 leptokurtic distribution

WT 0.950291 🡪 leptokurtic distribution

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



* Data is Positively Skewed as the tail is to the right.
* The large part of data is concentrated on the left and has a long tail to the right
* Mode of the data is 100 as it has the peak
* It’s unimodal data (has a single mode) and is approximately normal



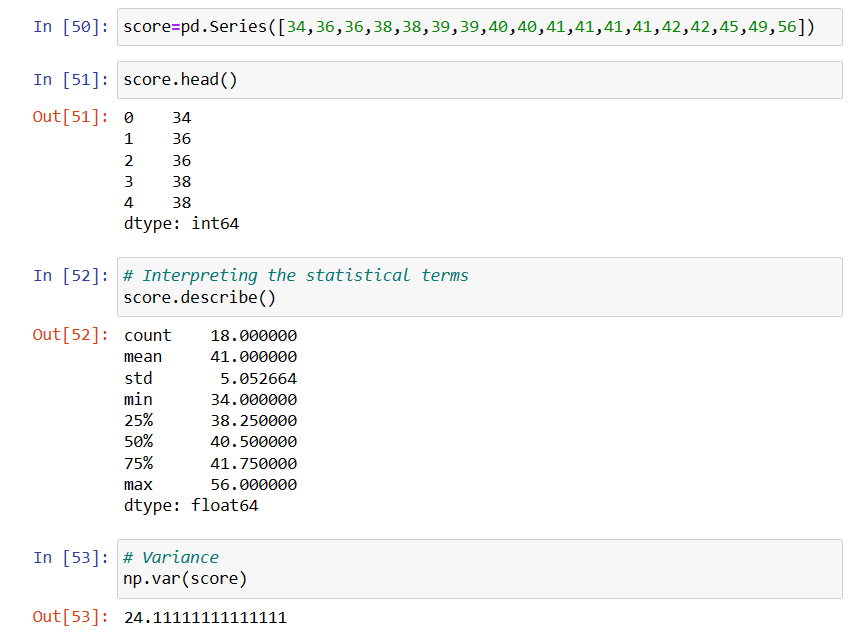
* It has a long tail at the end of the upper quartile.
* Median of the data is close to lower quartile
* Most of the data is concentrated at the first quartile
* The dataset has 7 outliers

**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?

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

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

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



Mean 🡪 41.0

Median 🡪 40.5

Variance 🡪 24.111

Std Deviation 🡪 5.05

1. What can we say about the student marks?

* The average score obtained by the students is 41. The probability of students scoring above 41 is about 57%

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

* Skewness is a measure of asymmetry and it may be zero, positive or negative. If the mean, median and mode of data are equal then there is no skewness.

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

* If the mean > median than the distribution is positively skewed.

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

* If the median > mean than the distribution is negatively skewed.

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

* Positive values of kurtosis indicate that a distribution is peaked and contains thick tails.

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

* A negative kurtosis value in a distribution 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 box plot shown above has no outliers
* Most part of the data is concentrated to the right

What is nature of skewness of the data?

* The skewness is negative as it has long tail towards the left and most part of the data is concentrated to the right.

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

* Inter Quartile Range (IQR) = 8 (approx)

Q19) Comment on the below Boxplot visualization



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

* From the figure, the median of both the boxplots is the same 🡪 (262.5)
* Both Plots shown above has equal distribution of data above and below the median and apparently it’s a normal distribution.
* Kurtosis is negative in 1st plot as it has thinner tails and positive in the 2nd plot.

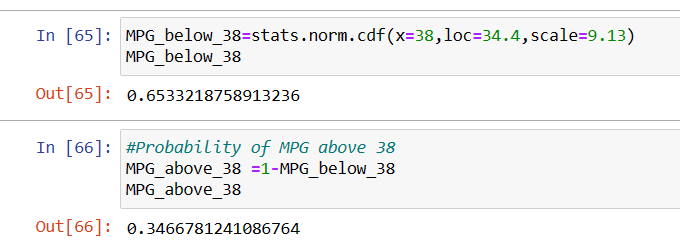
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

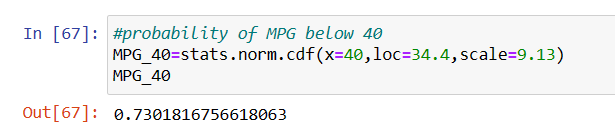
Calculate the probability of MPG of Cars for the below cases.

MPG<- Cars$MPG

1. P(MPG>38)🡪0.346



1. P(MPG<40) 🡪0.730



1. P (20<MPG<50) 🡪0.898

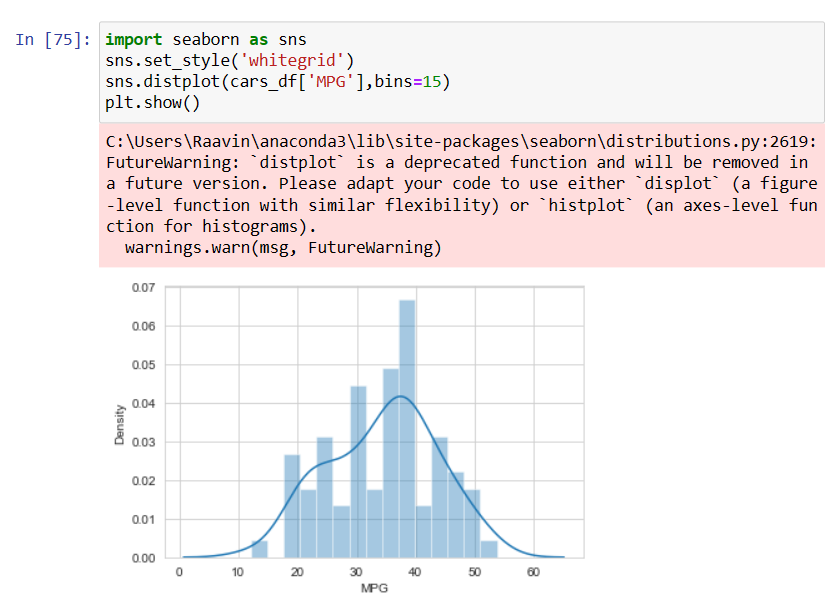


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

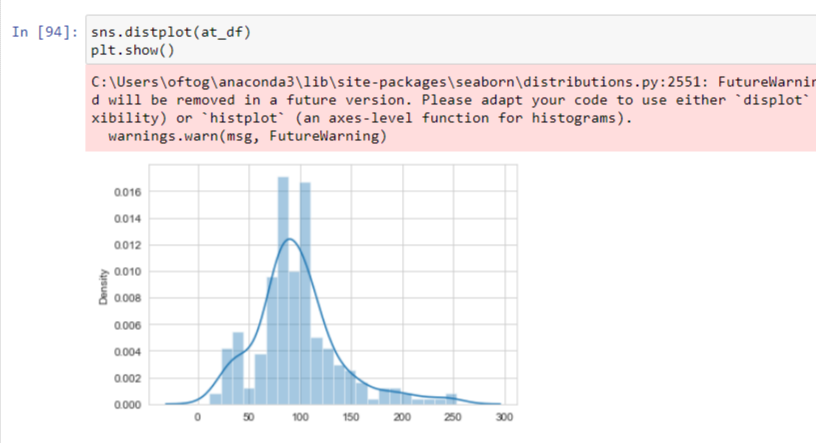
Ans: Yes. The MPG variable of cars\_df follows a Normal Distribution.



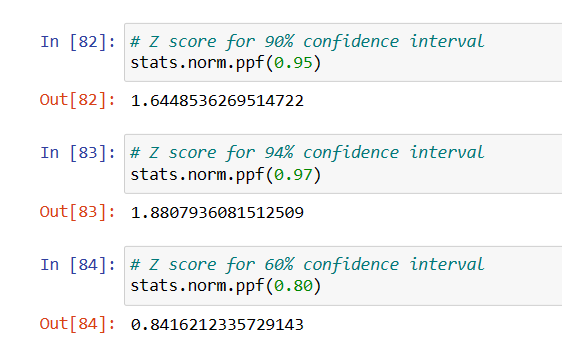
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: By the below graph, it is apparent that it almost follows normal distribution

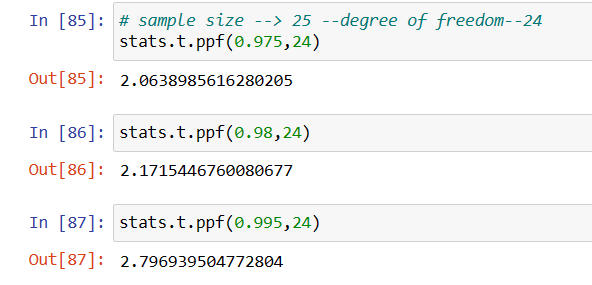


Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

To find t-value for the sample size of 25



Confidence interval for 95% 🡪 2.063

Confidence interval for 96 % 🡪 2.171

Confidence interval for 99% 🡪 2.796

Q 24**)** A Government companyclaims 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 260days

Hint:

R Code 🡪 pt(tscore,df)

df🡪 degrees of freedom

A) The t- statistic value for the above problem is

Population mean 🡪 270

Sample mean🡪 260

Sample std deviation 🡪 90

Sample 🡪 18

**t= (sample mean – population mean)/ (sample std deviation/np.sqrt(sample))**

t value is 🡪 **-0.4714**

substituting this in the code, we get 🡪 stats.t.cdf(- 0.471,df=17)

0.3218140331685075

The probability that an average bulb lasts less than 260 days 🡪 0.3218140331