|  |  |
| --- | --- |
| 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 |

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 | 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 | Interval |
| Years of Education | Ratio |

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

Ans:-

Total number of events= {hhh, hht, htt, ttt, tth, thh, hth, tht} =8

Interested events=3

Probability=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

Ans:-

Total number of outcomes when two dice are rolled=6\*6=36.

(1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6)

(2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6)

(3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6)

(4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6)

(5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6)

(6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6)

a) Equal to 1 = 0% probability

b) Less than or equal to 4= 6/36 = 1/6

c) sum is divisible by 2 and 3

{ 2 3 4 5 6 7

3 4 5 6 7 8

4 5 6 7 8 9

5 6 7 8 9 10

6 7 8 9 10 11

7 8 9 10 11 12}

Probability=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?

Ans:-

total number of events= ===21

Interested events===10

Probability that none of the balls is blue =10/21=0.47

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:-

Expected number = E(x) ==1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= **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.

Ans:-  #measures of central tendencies

# for points

> mean (ex1$points)

[1] 3.596563

> median (ex1$points)

[1] 3.695

> get mode (ex1$points)

[1] 3.92

> Var (ex1$points)

[1] 0.2858814

> sd(ex1$points)

[1] 0.5346787

> range (ex1$points)

[1] 2.76 4.93

> #For Score

> mean(ex1$score)

[1] 3.21725

> median(ex1$score)

[1] 3.325

> get mode(ex1$score)

[1] 3.44

> var(ex1$score)

[1] 0.957379

> sd(ex1$score)

[1] 0.9784574

> range(ex1$score)

[1] 1.513 5.424

> #For Weight

> mean(ex1$weight)

[1] 17.84875

> median(ex1$weight)

[1] 17.71

> get mode (ex1$weight)

[1] 17.02

> var(ex1$weight)

[1] 3.193166

> sd(ex1$weight)

[1] 1.786943

> range(ex1$weight)

[1] 14.5 22.9

**Inferences:**

The mode usually just gives the class (ex: numeric). To get mode, first we have to write a function to get mode. In the above code, get mode (it can be any name) is the name of the function to find the mode.

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:-

EV= Σx/n ==145.33

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

**(a) Cars speed and distance**

Ans**: ## Skewness For speed and distance**:

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

**Skewness Inference:**

1. Speed distribution is left skewed (negative skewness)

2. Distance distribution is right skewed (positive skewness)

**## Kurtosis For Speed & Distance:**

Index -1.200000

Speed -0.508994

Dist 0.405053

dtype : float64

**Kurtosis Inference:**

1. Speed distribution is platykurtic (negative kurtosis i.e. flatter than normal distribution)

2. Distance distribution is leptokurtic (positive kurtosis i.e. peaked than normal distribution)

**(b) SP and Weight(WT)**

Ans**: ## Skewness for SP & Weight:**

SP 1.611450

WT -0.614753

dtype: float64

**Skewness Inference:**

1. WT distribution is left skewed (negative skewness)

2. SP distribution is right skewed (positive skewness)

**## Kurtosis For SP & Weight:**

SP 2.977329

WT 0.950291

dtype: float64

**Kurtosis Inference:**

SP & WT distribution both are leptokurtic (positive kurtosis i.e. peaked than normal distribution)

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



Ans:  **Histogram:**

* Chick weight data is right skewed or positively Skewed – Yes
* More than 50% chick weight is between 50 to 150 – yes
* Most of the Chick Weight is between 50 to 100 - Yes



Ans: **Boxplot:**

* The data is right Skewed
* There Are outliers at upper 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?

Ans: Graphical user interface, text, application, email

Description automatically generated

**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?

Ans:

Graphical user interface, text, application, email

Description automatically generated

2). Inference:

"1. There are 2 Outliers in Student's marks: 49 and 56"

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

Ans: No skewness, symmetric

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

Ans: [positively skewed distribution](https://www.wallstreetmojo.com/positively-skewed-distribution/)

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

Ans: Negatively skewed distribution

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

Ans: peakness (sharp peak) and less variation.

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

Ans: less peakness (Broad peak) and more variation

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



What can we say about the distribution of the data?

Ans: it is not a Normal Distribution

What is nature of skewness of the data?

Ans: It is left skewed

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

Ans: Inter Quartile Range =Upper Quartile- 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.

Ans: 1) The median of the two boxplots are same approximately 260.

2) The boxplots are not skewed in Positive or Negative direction.

3) Outliers doesn’t exist in both of the boxplots.

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

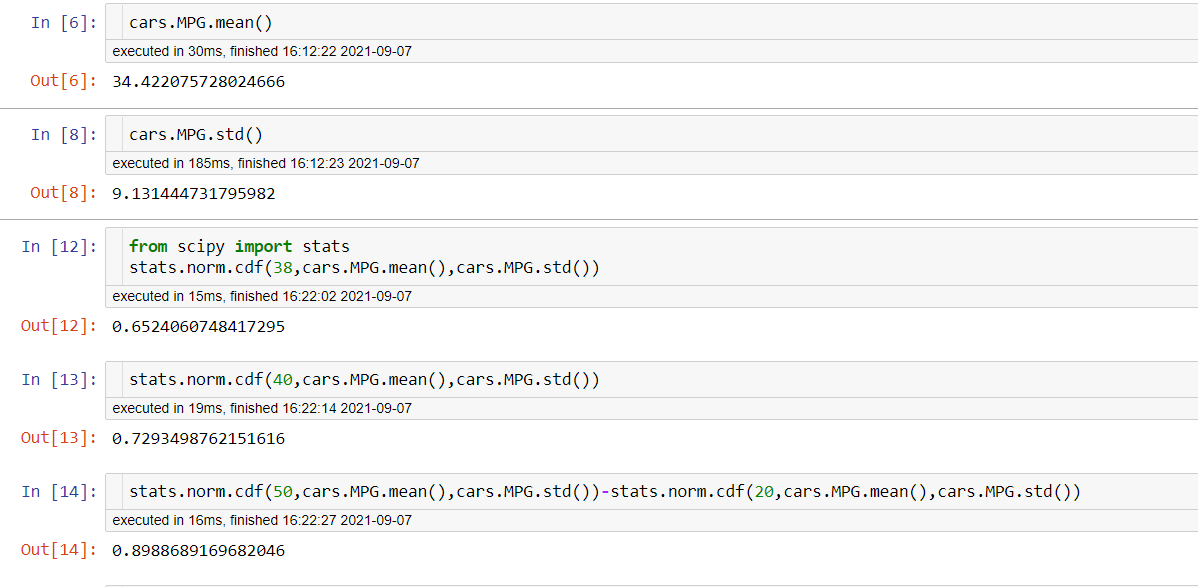
Data \_set: Cars.csv

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

MPG <- Cars$ MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans :

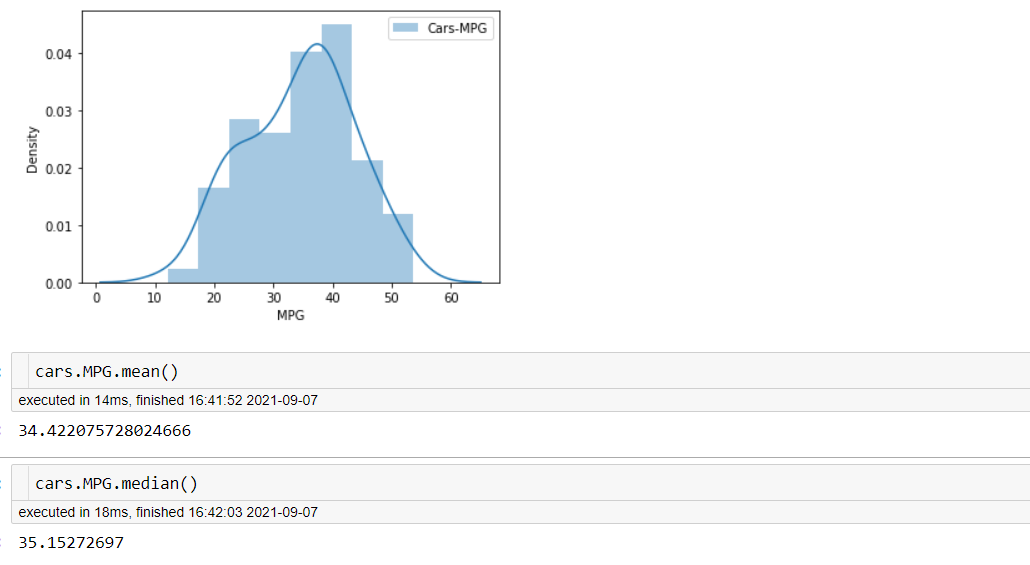


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

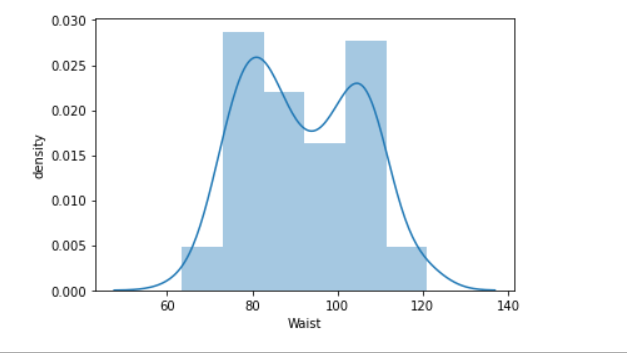


Inference: MPG of Cars does follow normal distribution approximately (as mean and median are approx. same)

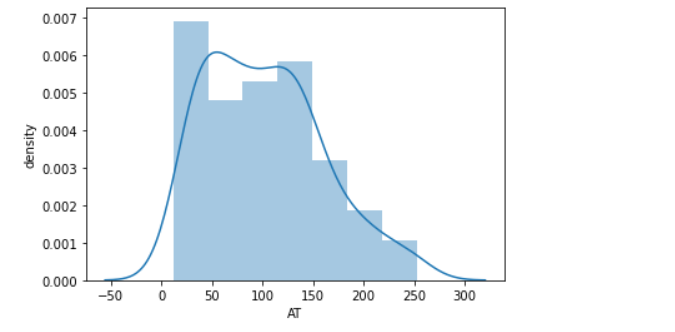
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: **# plotting distribution for Waist Tissue (WT)**



**# plotting distribution for Adipose Tissue (AT)**

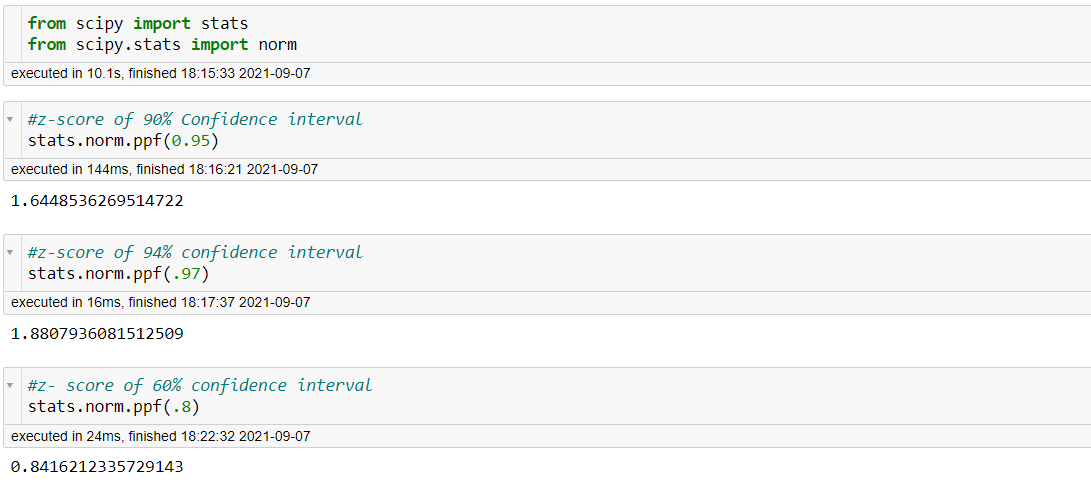


**# Inference:**

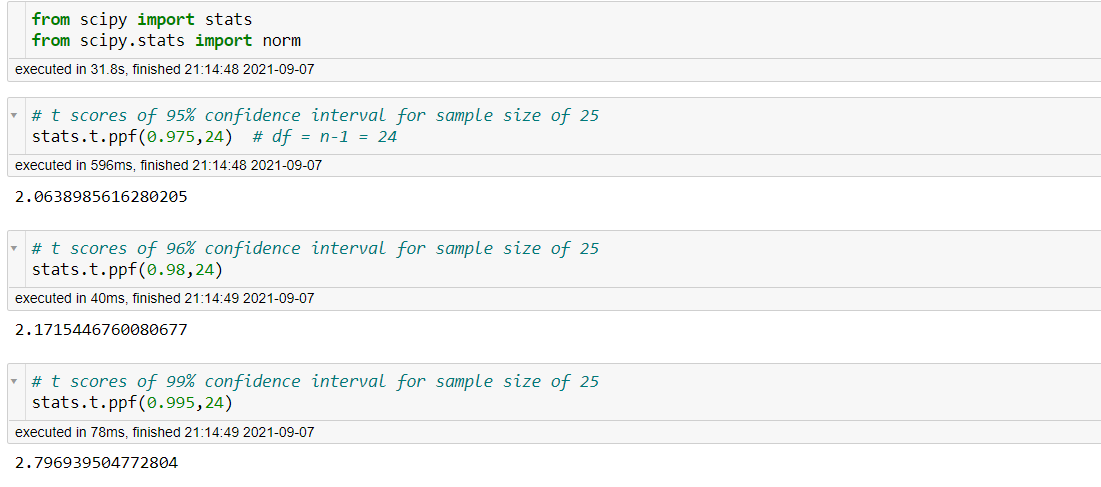
Both the Adipose Tissue (AT) and Waist Circumference (Waist) data set do not follow the normal distribution approximately (as mean and median of both the data are approximately different)

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

Ans:

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

Ans :



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 (t-score, df)

df 🡪 degrees of freedom

Ans: 