**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 | 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 | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Ans:**

Event: Three coins are tossed.

The total number of combinations are 23= 8.

Sample space= {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

Total number of combinations for two heads and one tail

= { HHT, HTH, THH}=3

Therefore the [Probability](https://www.cuemath.com/data/probability/) of getting two heads and one tails is given by

P(Two heads and One Tail) = Number of desired outcomes / Total number of

combinations

= 3 /8 = 0.375

The Probability that two heads and one tail are obtained is **0.375**.

**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:**

Event: Two dice are rolled

Total number of combinations are 62=36

Sample space={(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)}

**Event a)** the sum is equal to 1

favorable outcome (sum equal to 1) = 0

therefore probability=0/36 =**0**

**Event b)** the sum is less than or equal to 4

favorable outcomes (sum is less than or equal to 4) ={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6

therefore probability=6/36= **1/6**

**Event c)** the sum is divisible by 2 and 3

favorable outcomes {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}=6

therefore 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 balls = (2 red+3 green+ 2 blue balls)= 7 balls

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

∴n(E)= Number of ways of drawing 2 balls out of (2 red + 3 green) balls.  
 =5C2  
 =(5×4)/(2×1) =10  
Favorable outcomes= Number of ways of drawing 2 balls out of 7  
 = 7C2  
 = (7×6)/(2×1) = 21

Therefore, P(E)=n(E)/n(S)=10/21= **0.4762**

**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:** Let x= Candies count

P= Probability

Expected number of candies for a randomly selected child

E(X) =

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

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

= 3.090

Therefore, expected number of candies for a randomly selected child is **3.090**.

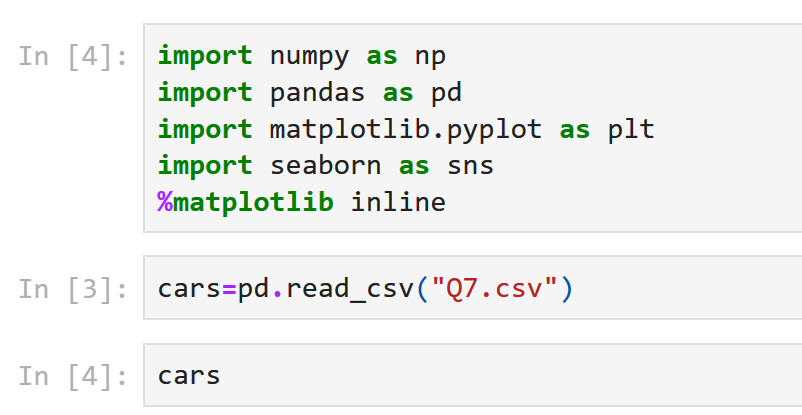
**Q7)** Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

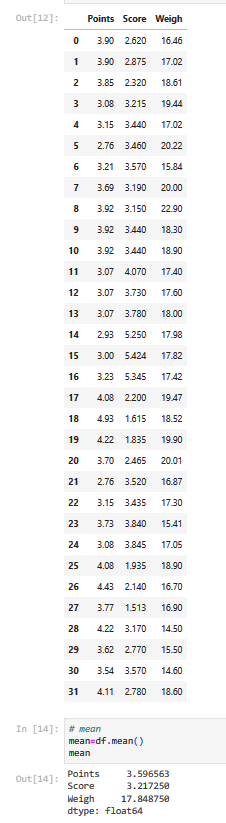
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

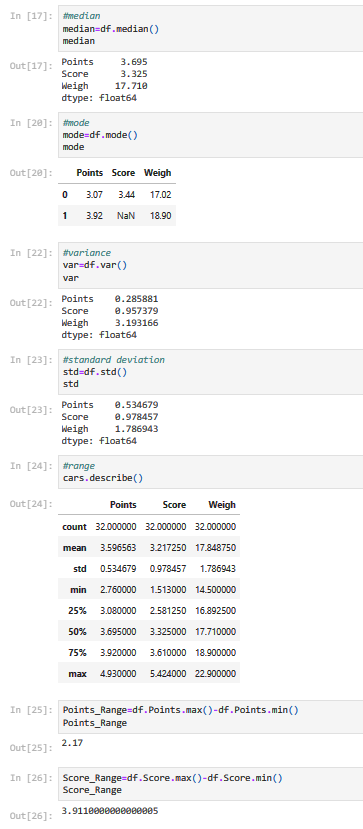
**Use Q7.csv file**

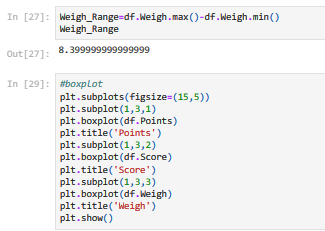
**Ans:**

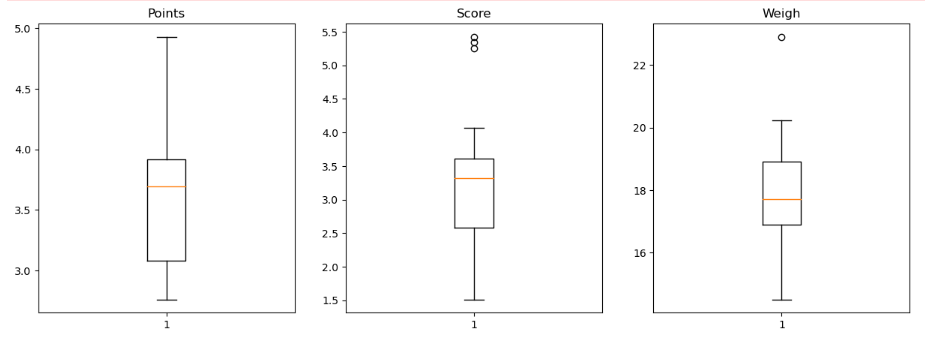
****

****

****

****

****

****

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard Deviation** | **Range** |
| **Points** | 3.596563 | 3.695 | 3.07, 3.92 | 0.285881 | 0.534679 | 2.17 |
| **Score** | 3.217250 | 3.325 | 3.44 | 0.957379 | 0.978457 | 3.9110 |
| **Weigh** | 17.848750 | 17.710 | 17.02, 18.90 | 3.193166 | 1.786943 | 8.3999 |

**Comments:**

Points and Weigh have two modes.

Values of mean, median, mode of variable weigh are greater than the values for variables points and score.

From boxplot it is seen that outliers are present for score and 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:** Expected Value  =  ∑ (random variable \*probability)

E(X) =  ∑x P(x)

1. Given: The weights (X) of 9 patients (in pounds),

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

Probability of selecting each patient = 1/9

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

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

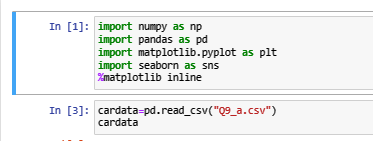
= 145.33

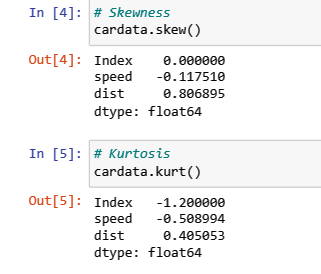
E(X) = Expected Value of the Weight of the randomly chosen patient = 145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

****

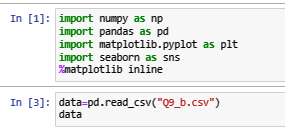
|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.117510 | -0.508994 |
| **Distance** | 0.806895 | 0.405053 |

Speed data is Skewed left since skewness is negative and the distribution is platykurtic.

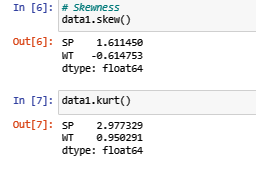
Distance is skewed right since it is positive and the distribution is leptokurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

****

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.611450 | 2.977329 |
| **Weight** | -0.614753 | 0.950291 |

SP data is Skewed right since skewness is positive and the distribution is leptokurtic.

Weight is skewed left and the distribution is leptokurtic.

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



**Histogram:**

Chick weight data is right skewed or positively skewed.

More than 50% weith is between 50 to 150.

Most of the chick weight is between 50 to 100.

Mean > Median>Mode



The data is right skewed.

There are 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?

**Ans:**

We are given the standard deviation for the sample. therefore the **t-distribution** is used here.

**Given**:

* Sample **mean**\overline{x} = 200.
* Sample **standard deviation**  s = 30.
* Sample **size** n = 2000.

The **interval** is:

In which **t** is the critical value for the two-tailed confidence interval.

Considering a **94%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 1.8916**, hence:

The **94%** confidence interval is **(198.73, 201.27).**

Considering a **96%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.0673**, hence:

The **96%** confidence interval is **(198.61, 201.39).**

Considering a **98%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.3452**, hence:

The **98%** confidence interval is **(198.43, 201.57).**

**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:**

**Mean=** 41

**Median=**40.5

**Variance=** 25.5294

**Standard deviation=** 5.0526

Average marks of the students are 41 with variance equal to 25.5294.

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

**Ans:** when mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.

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

**Ans:** The nature of skewness when mean > median right skewed i.e. positively skewed.

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

**Ans:** The nature of skewness when mean > median left skewed i.e. negatively skewed.

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

**Ans:** positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

**Ans:** Negative Kurtosis means the curve will be flatter and broader.

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



**Q)** What can we say about the distribution of the data?

**Ans:**  The image represents data that is a not a [normal distribution](https://www.simplypsychology.org/normal-distribution.html), i.e. data is not symmetrically distributed.

**Q)** What is nature of skewness of the data?

**Ans:** The median is closer to the top of the box, and the whisker is shorter on the upper end of the box, therefore the distribution is negatively skewed (skewed left).

**Q)** What will be the IQR of the data (approximately)?   
**Ans:** IQR is the inter quartile range.

 The interquartile range (IQR) is the box plot showing the middle 50% of scores and can be calculated by subtracting the lower quartile from the upper quartile.

Here Q1 = 10

Q2 = 14.7

Q3 = 18

 IQR = Q3 – Q1 = 8(approx)

**Q19)** Comment on the below Boxplot visualizations?



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

**Ans:** Here there is a representation of 2 box plots.

Medians of the two data are equal. i.e. there is not likely difference between two groups.

Interquartile range for boxplot 2 is greater than for boxplot 1. i.e. the box for group 2 is longer than for group 1. This shows that data 2 is more dispersed than data 1.

The overall spread i.e the extreme values at the end of two whiskers shows the range of scores. Data 2 has larger range than data1. This indicate data 2 has wider distribution, that is, data 2 is more scattered data.

Also both data are symmetric in nature.

**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)

c. P (20<MPG<50)

**Ans:**

Given: mean(MPG)=34.42208,

sd(MPG)=9.131445

1. To find P(MPG>38)

P(MPG>38)= 1 – pnorm(38,mean(MPG),sd(MPG))

= 0.330

= 33%

1. To find P(MPG<40)

P(MPG<40)

=pnorm(40,mean(MPG),sd(MPG))

=0.7293499

=72.3%

1. To find P (20<MPG<50)

 P (20<MPG<50)= pnorm(50,mean(MPG),sd(MPG)) –  pnorm(20,mean(MPG),sd(MPG))

=0.955 -0.057

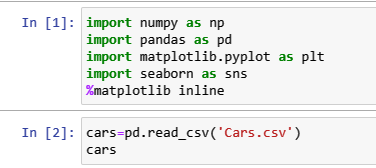
=0.8988689

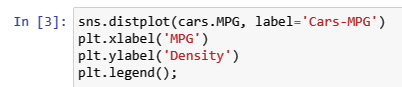
**Q 21)** Check whether the data follows normal distribution

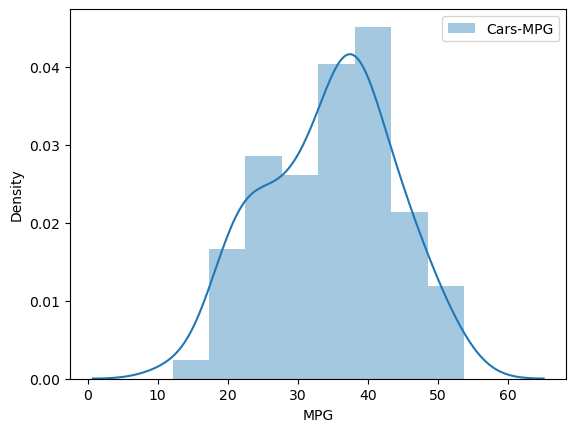
1. Check whether the MPG of Cars follows Normal Distribution

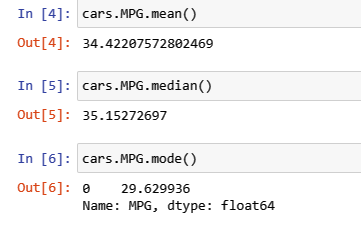
Dataset: Cars.csv

**Ans:**

****

****

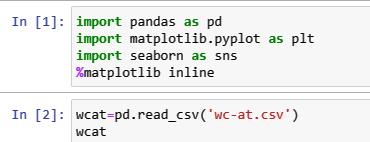
****

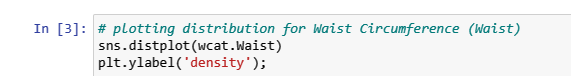


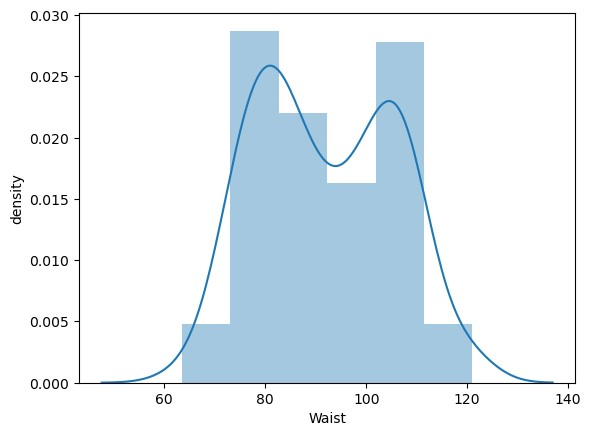
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

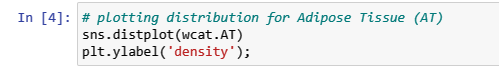
Dataset: wc-at.csv

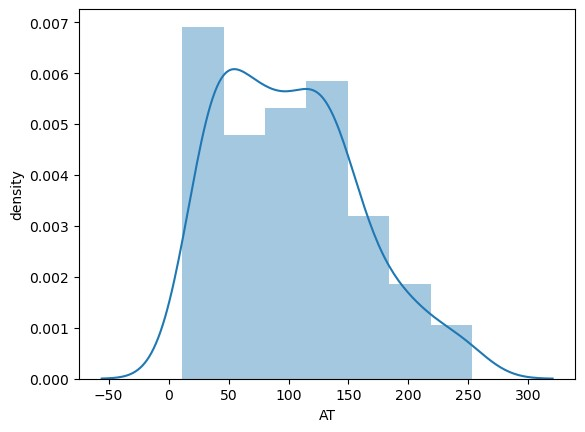
**Ans:**

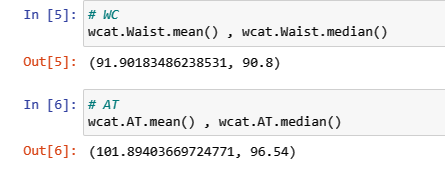
****

****



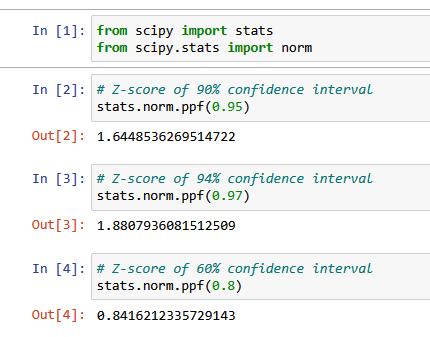
****



****

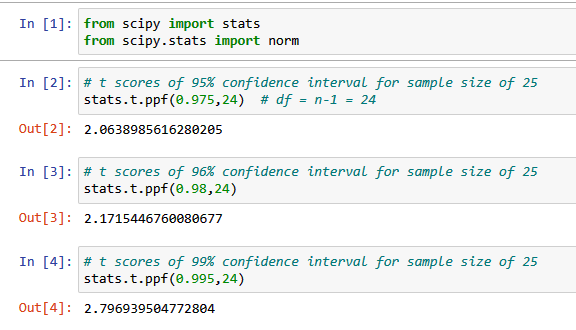
**Q 22)** Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.

**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.

**Ans:**



The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3216.