|  |  |
| --- | --- |
| 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) | Ordinal |
| 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 | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

**A)** Probability of Two heads and One Tail are=P(2heads/1tail)/P(No of Events)

=> 3/8=0.375=3.75%

**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

**A)** Equal to 1= 0/36=**0**

Less than or Equal to 4= (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)=6/36=**1/6**

Sum of Divisible by 2 and 3=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?

**A)**10/21=5/7=47.6%

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

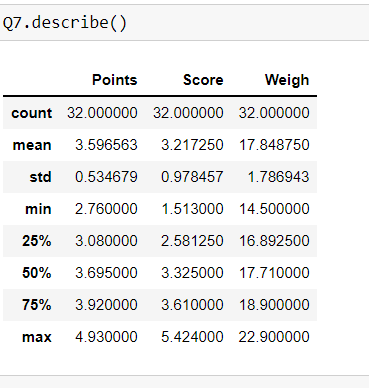
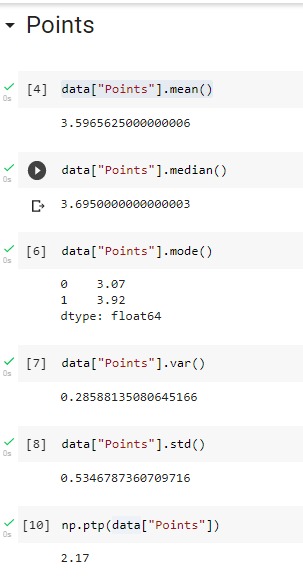
**A)** Expected number of candies = ∑x.p(x) =1×0.015+4×0.20+3×0.65+5×0.005+6×0.01+2×0.120=0.015+0.8+1.95+0.025+0.06+0.24=**3.09**

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

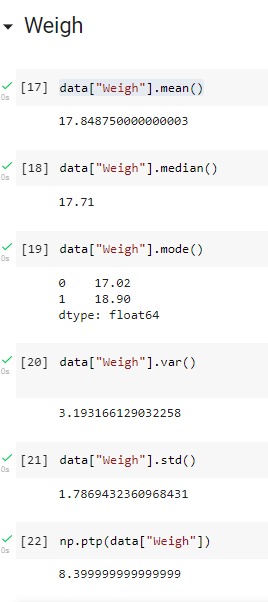
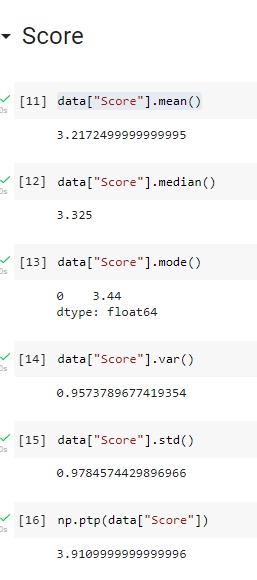
* For Points,Score,Weight

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

**A)**

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Points | Score | Weigh | Inference |
| Mean | 3.59 | 3.21 | 17.84 | More number of observations near mean |
| Median | 3.69 | 3.32 | 17.71 | Median is not influenced by outliers. |
| Mode | 3.07,3.92(bimodal) | 3.44 | 17.02,18.90 | bimodal |
| Range (Max-Min) | 2.17 | 3.91 | 8.4 | It is the spread between highest and lowest values. |
| Variance | 0.28 | 0.94 | 3.16 | Variance of weigh is more, which means the data are far from the mean compared to points and score. |
| Standard deviation | 0.53 | 0.97 | 1.78 | Standard deviation tells how the spread of data around mean. |

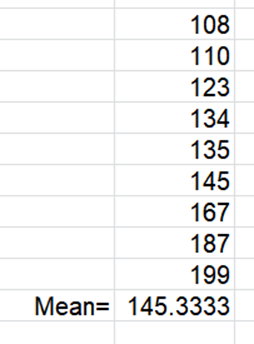
****

**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?

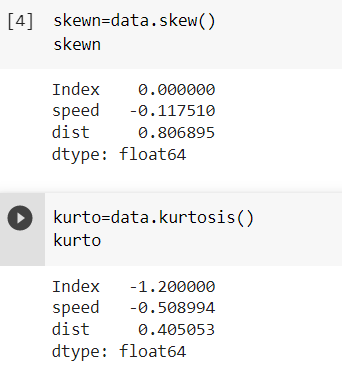


**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

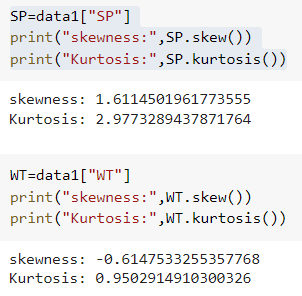
**Use Q9\_a.csv**

**Answer:**

****

**SP and Weight (WT)**

**Use Q9\_b.csv**

****

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



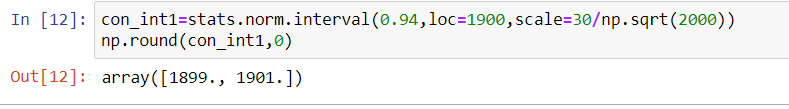


|  |  |
| --- | --- |
| Histogram | Box plot |
| This histogram shows it is right skewed data. | Right skewed data (Whisker is more on the right side of median) |
| Each column represents data within the bin. | Outliers present |

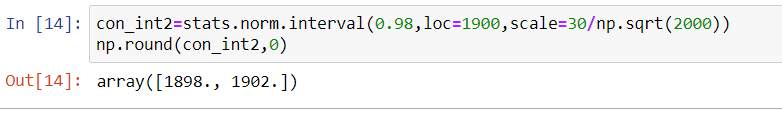
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

**A)**

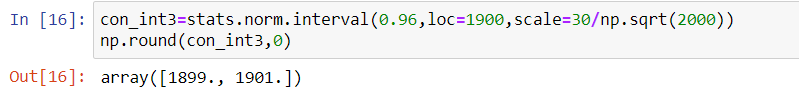
For 94% confidence level:



For 98% confidence level:



For 96% confidence level:

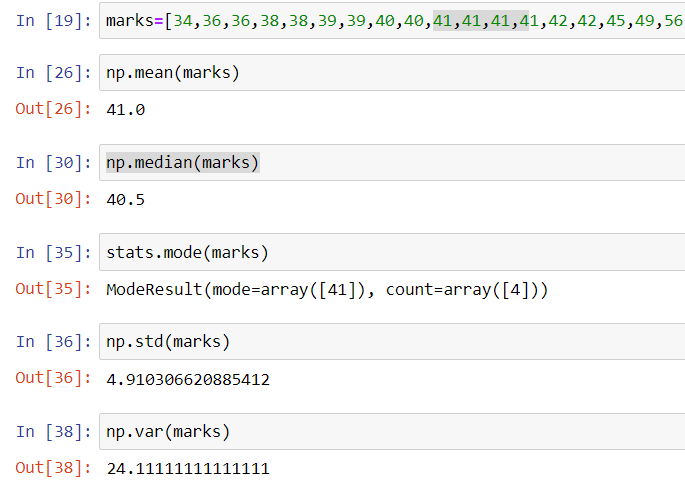


**Q12)** Below are the scores obtained by a student in tests

**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?

A)



2)

* The average student marks are 41
* Median value of student marks is 40.5
* Most of the students got **41** in the test

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

1. When mean= median, the data is normally distributed.

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

**A)  positive/right nonparametric skew**

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

**A)  Negative /left nonparametric skew**

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

**A)**

* Distribution is leptokurtic (taller than normal distribution with same mean)
* Sharp peak
* Flatter tails

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

**A)**

* Distribution is Platykurtic(flatter than normal distribution with same mean)
* Thin nails
* Less peaked

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



What can we say about the distribution of the data?

**A)**

* The data is not normally distributed.
* Median of the data is 15.
* No outliers present

What is nature of skewness of the data?

1. It is left skewed data.

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

**A)**

* Both 1 and 2 are normally distributed
* Median is same for 1 and 2. Median= Q2= 50th percentile= 260
* IQR for boxplot 1 is less compared to boxplot 2, which means the middle 50% of data for boxplot 2 spread more than boxplot 1.

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

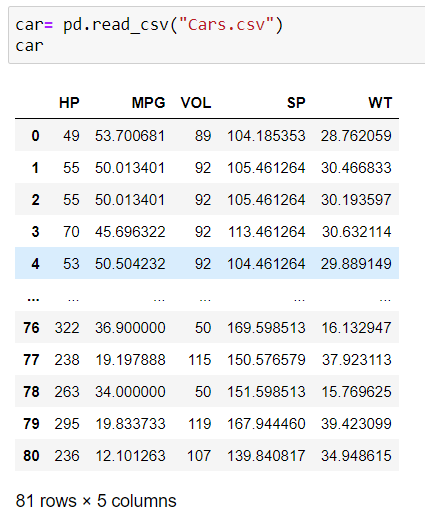
Data \_set: Cars.csv

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

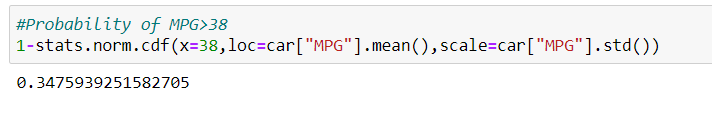
MPG <- Cars$MPG

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

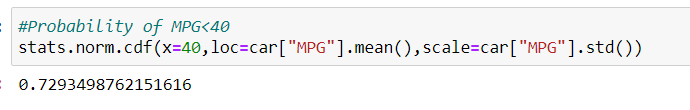
**A)**



a.



b.



c.



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

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**A)**

|  |  |
| --- | --- |
|  | **Mean, median, mode is not same.** |

|  |  |
| --- | --- |
|  | **Median is not on the middle of the box.** |

**“Not normally distributed.”**

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

Dataset: wc-at.csv

**A)**

**AT**

|  |  |
| --- | --- |
|  |  |
|  | * Not normally distributed. * platykurtic * right skewed |

**Waist**

|  |  |
| --- | --- |
|  |  |
|  | * Not normally distributed. * platykurtic * right skewed |

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

|  |
| --- |
| 90% |
| 94% |
| 60% |

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

|  |
| --- |
| 95% |
| 96% |
| 99% |

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

**A)**

Population mean =µ=270

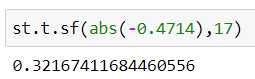
Sample size = n=18

df= n-1=17

Sample mean ==260

Sample standard deviation = s= 90

t score = (260-270)/ (90/)=-0.4714



18 randomly selected bulbs would have an average life of no more than 260 days= 0.3216=32.16%