|  |  |
| --- | --- |
| 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 | Interval |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ratio |
| Number of Children | Interval |
| 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?**

=HHH, HHT, HTH, THH, TTT, HTT, THT, HTT

=3/8

Probability=0.375

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

(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)

1. Equal to 1=Probability=0
2. Less than or equal to 4= Probability =6/36=0.166
3. Sum is divisible by 2and 3= Probability =5/36=0.138

**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?**

Sum of number of balls=2+3+2=7

Number of ways to draw 2 Number of balls drawing from 7

n(S)=7C2=(7 X 6)/(2 X 1)=21

n(A)= 5C2=(5 X 4)/(2 X 1)=10

The Probability of Drawing 2 balls, none of which is blue

P(A)= n(A)/ n(S)=21/10=0.476

**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 X 1=0.015

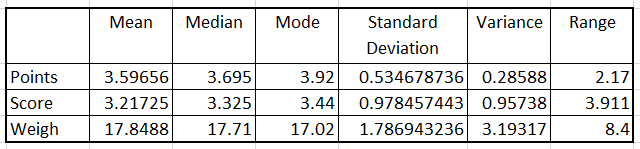
Child B – probability of having 4 candies = 0.20X4=0.8

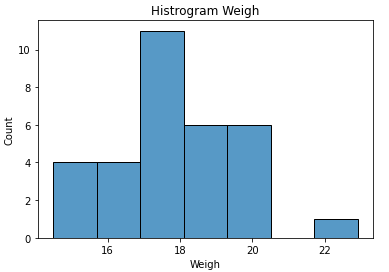
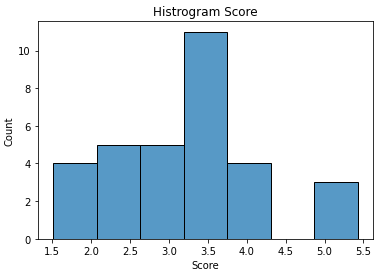
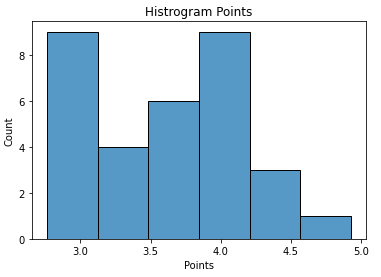
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* For Points,Score,Weigh>

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

Ans:

****

****

1. **Points**: For points is positively skewed, and not having outliers
2. **Score** : For points is negatively skewed, and having outliers
3. **Weigh** : For points is negatively skewed, and having outliers

**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:- Number of Points=9

Probability for 1 patient=1/9

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

=145.33

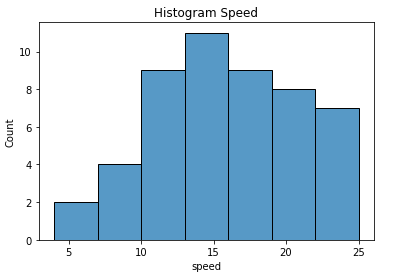
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer:**

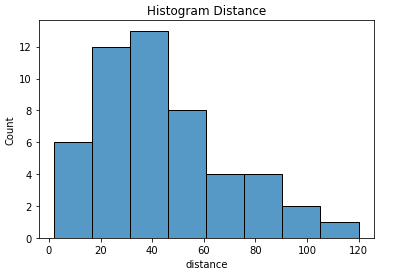
1. **Cars speed:**

****

1. **Skewness:** -0.117510
2. **Kurtosis:** -0.508994

**Inference:** Car Speed data shows that distribution has tail in left side that means distribution is negatively skewed and Kurtosis also shows that distribution is negative kurtosis is means that distribution is flatter

1. **Distance:**



1. **Skewness :** 0.806895
2. **Kurtosis:** 0.405053

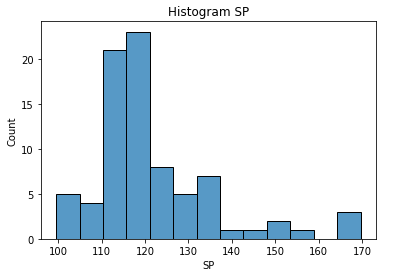
**Inference:** The distance data shows that distribution has tail at right side means distribution is positively Skewed . Kurtosis value is less so normal distribution with heavy tail on right side.

**Que: SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:**

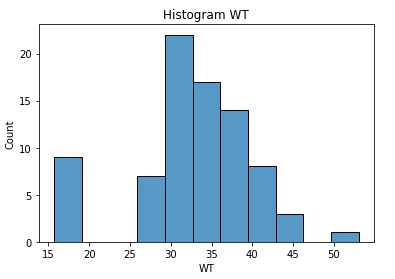
1. **SP:**

****

1. **Skewness :** 1.611450
2. **Kurtosis:** 2.977329

**Inference:** The distribution has tail at right side so distribution is positively skewed and kurtosis value is near to 3 so it is Mesokurtic distribution, Distribution has outliers.

1. **WT:**

****

1. **Skewness :** -0.614753
2. **Kurtosis:** 0.950291

**Inference:** The negative value of skewness shows that distribution is negatively skewed and kurtosis is less than 3 so it is Platykurtic distributions, Distribution has outliers.

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



**Inferences:** Asymmetric Distribution, The tail on right shows that distribution is positively skewed and Histogram also indicate that mean is greater than Median.



**Inferences:** Asymmetric Distribution as higher boundry of boxplot is stretched on right side.boxplot shows outliers on right side so the distribution is positively skewed.

**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:

No. of Samples=2000

Avg. weight =200 pounds

Standard deviation=30

1. For confidence interval = 94%

Z value for 94 % =1.88

Std. error = 30/ =0.670

Margin of error =0.680 \* 1.88 =1.261

So

Confidence interval for 9%=(198.738, 201.26)

1. For confidence interval = 96%

Z value for 96 % =2.05

Std. error = 30/ =0.670

Margin of error =0.680 \* 2.05=1.373

So

Confidence interval for 9%=(198.627, 201.373)

1. For confidence interval = 98%

Z value for 98 % =2.32

Std. error = 30/ =0.670

Margin of error =0.680 \* 2.32=1.561

So

Confidence interval for 9%=(198.439, 201.561)

**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.

Answer:

1. Mean=41
2. Median=40.5
3. Variance=25.52
4. Standard Deviation = 5.05
5. What can we say about the student marks?

Answer: most student score marks are near average value,56 marks is outlier.

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

Answer: mean is equal to median then skewness is zero and distribution is symmetrical.

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

Answer: mean is greater than median then distribution is positively skewed.

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

Answer: mean is greater than median then distribution is negatively skewed.

**Q16) What does positive kurtosis value indicates for adata ?**

Answer: distribution is short tails with values close to mean values

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

Answer: distribution is long tails with values spread out wider to mean values

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



What can we say about the distribution of the data?

Answer: Most of the values lies in between 10 to 18 and having median value 15

What is nature of skewness of the data?

Answer: Distribution is negatively skewed

What will be the IQR of the data (approximately)?   
Answer:IQR od the data =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.

Answer:

* Both 1 and 2 boxplot has Symmetric distribution. boxplot 1 has short tails as campared to boxplot 2
* Range of box plot 1 is less than range of box plot 2.values are very closed in boxplot 1
* IQR of box plot 1 is less than IQR of boxplot 2

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)

Answer:(33/81)\*100=40.74

There are total 33 cars which are mpg is greater than 38 out of 81 cars so probability of mpg > 38 is 40.74%

* 1. P(MPG<40)

Answer:(61/81)\*100=75.30

There are total 61 cars which are mpg is greater than 40 out of 81 cars so probability of mpg > 40 is 75.30%

* 1. P (20<MPG<50)

Answer:(69/81)\*100=85.18

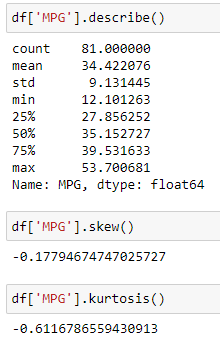
There are total 69 cars which are mpg is greater than 20 to 50 out of 81 cars so probability of mpg > 20 is 85.18%

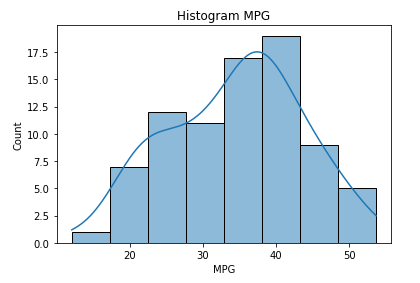
**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

Answer: mean value 34.42 and median value 35.15 is nearly same. from histogram kde distribution is approximately similar, no major skewness in general MPG follows normal distribution





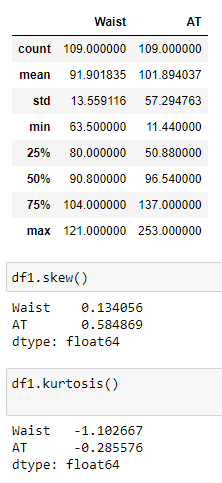
1. **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution**

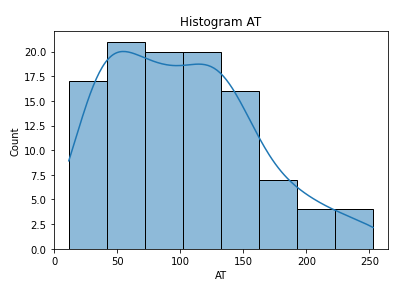
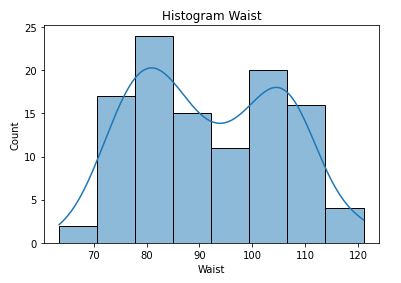
**Dataset: wc-at.csv**

Answer: for waist mean 91.90 is slightly greater as median 90.80 .kurtosis is negative

-1.10 from histogram waist is not normal distribution

For AT mean 101.89 is not same as Median 96.54.and kurtosis is negative is flatter distribution AT is not Normal Distribution





**Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval**

Answer: For 90% =area under curve left =1+0.90/2=0.950

For 94% = area under the curve to left=1+0.94/2=0.970

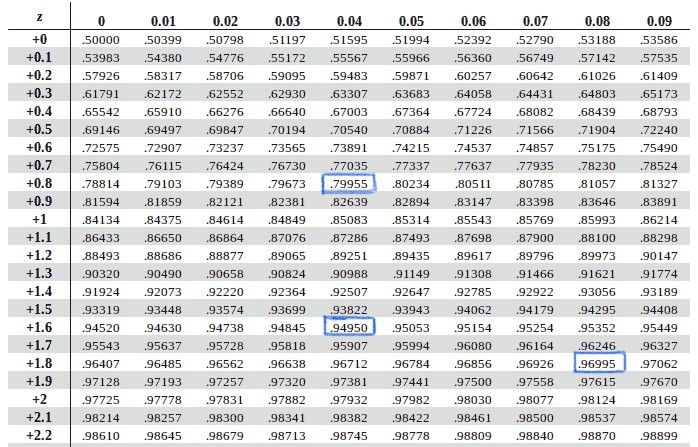
For 60% = area under the curve to left=1+0.60/2=0.8

Z score-

For 90% interval z-score=1.6

For 94 % internal z-score=1.8

For 60 % internal z-score=0.8



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

**Answer: degree of freedom = 25-1 =24**

So t score of 95% confidence interval = right tailed - 1.170

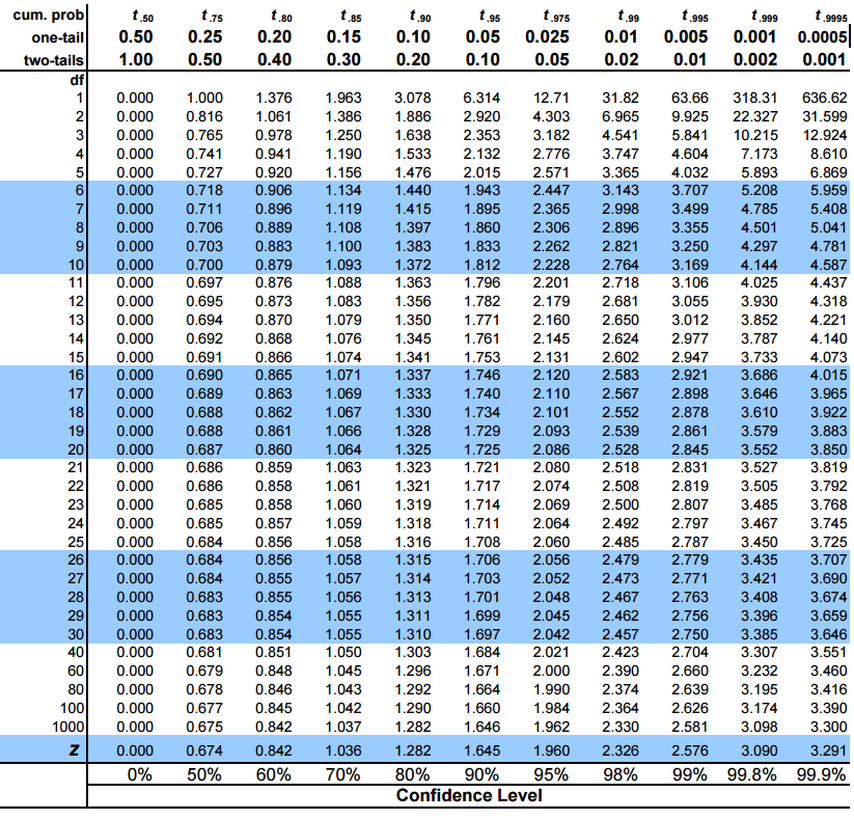
two tailed - +/- 2.063

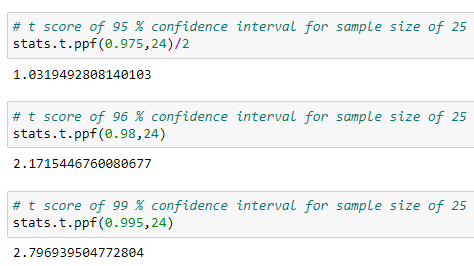
t score of 99 % confidence interval = right tailed – 2.492

two tailed - +/- 2.796

t score of 99 % confidence interval = right tailed – 1.828

two tailed - +/- 2.171

****



**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**

Answer: Population = 270

Sample mean = 260

Standard Deviation = 90

Sample Size = 18

Null Hypothesis = Ho = avg life < 270

Alternative hypothesis = Ha = avg life < 260

So t value = (260-270)/(90/) = -0.471

Degree of freedom = 17

By using t value and df we calculate p value

P value = 0.324

Failed to reject null hypothesis

CEO is true and there is 64% probalility that bulb avg life is not less than 260 days

