|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

Total Outcome - HHH – TTT – HTT – THH – HHT – TTH – HTH – THT

2H,1T Probability – HHT,THH,HTH

So probabilities is 3/8

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

1. Equal to 1 - 0
2. Less than or equal to 4 - 1/6
3. Sum is divisible by 2 and 3 - 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?

7c2 = 7\*6/1\*2 = 21

5c2 = 5\*4/1\*2 = 10

Probability = 5c2/7c2

= **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

Answer:

Expected number of candies for a randomly selected child

= 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

= **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.

**Use Q7.csv file**

Answer:

**Mean**

Points 3.596563

Score 3.217250

Weigh 17.848750

**Median**

Points 3.695

Score 3.325

Weigh 17.710

**Standard Deviation**

Points 0.534679

Score 0.978457

Weigh 1.786943

**Variance**

Points 0.285881

Score 0.957379

Weigh 3.193166

**Mode**

Points 3.07,3.92

Score 3.44

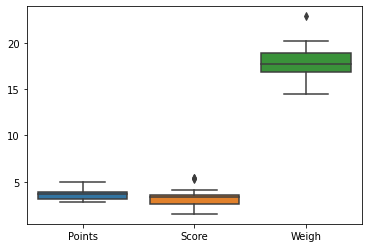
Weigh 17.02,18.90

**Range**

Points 2.76,4.93

Score 1.513,5.424

Weigh 14.5,22.9



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?

Answer:

Probability of selecting each patient = 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)

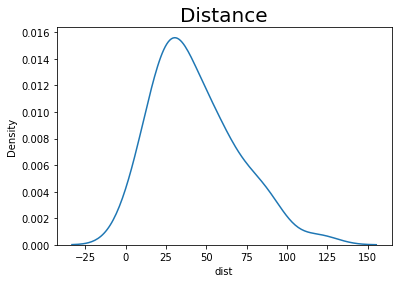
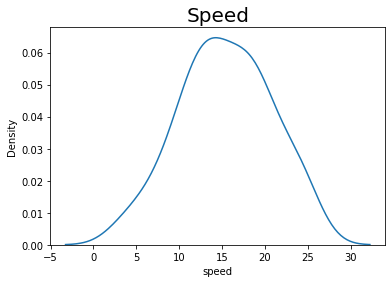
= (1/9) (1308)

= **145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

**Skew**

speed - -0.117510

distance - 0.806895

**Kurtosis**

speed -0.508994

dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skew**

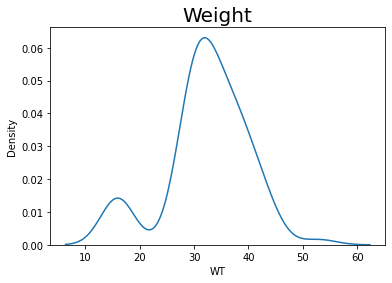
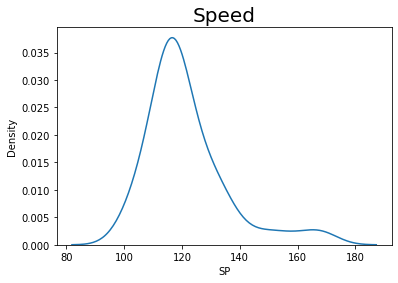
SP 1.611450

WT -0.614753

**Kurtosis**

SP 2.977329

WT 0.950291

****

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





**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?

confidence interval 94% = **(198.738325292158,** **201.261674707842)**

confidence interval 98% = **(198.43943840429978,** **201.56056159570022)**

confidence interval 96% = **(198.62230334813333,** **201.37769665186667)**

**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.

Mean = 41.0

Median = 40.5

Variance = 24.11111

Standard deviation = 4.91031

1. What can we say about the student marks?

From this data the average mark is 40.5 and one of the

student mark is 56 which means variance is 24.11.

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

Answer – no skweness

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

Answer – right side skew

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

Answer – left side skew

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

Answer - Positive kurtosis is a thinner peak and wider tails.

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

Answer - Negative kurtosis is a wider peak and thinner tails.

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



What can we say about the distribution of the data?

Answer – not normal distributed data

What is nature of skewness of the data?

Answer – Left skew

What will be the IQR of the data (approximately)?   
  
Answer - 10-18

Q19) Comment on the below Boxplot visualizations?



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

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.

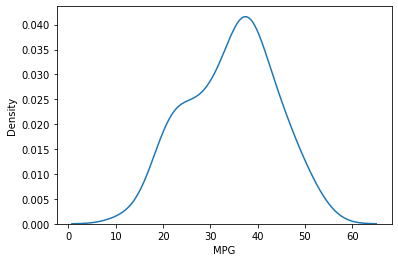
MPG <- Cars$MPG

* 1. P(MPG>38) – **0.34759**
  2. P(MPG<40) – **0.72935**

c. P (20<MPG<50) – **0.01312**

Q 21) Check whether the data follows normal distribution

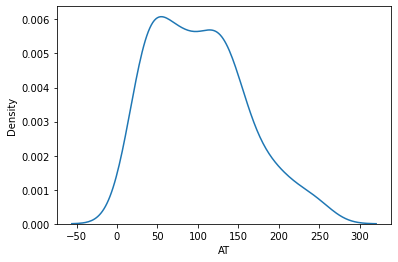
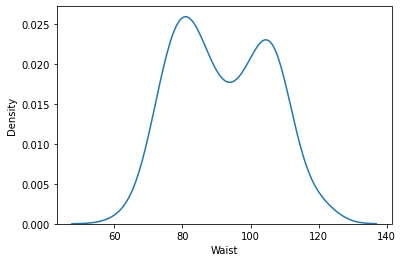
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

This figure represents little left skew

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

Dataset: wc-at.csv



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

**Confidence interval**  **Z scores**

60% - 0.8416212

90% - 1.644854

94% - 1.880794

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

**Confidence interval**  **T scores**

95% 2.063899

96% 2.171545

99% 2.79694

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

**T score**

t=(260-270)/(90/18\*\*0.5) = -0.4714045207910317

## **Find P(X>=260)**

p\_value =round(1-stats.t.cdf(abs(-0.4714),df=17),5) = **0.32167**