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

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

**Ans** : HHT + HTH + THH

=1/8+1/8+1/8

=3/8

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

1. Equal to 1

**Ans** : No Probability

b)Less than or equal to 4

**Ans** : 1/6

1. Sum is divisible by 2 and 3

**Ans :** 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** : p(2Red,3Green,2Blue)

Red(5/7,) & Green(4/6)

=5/7 ,4/6

= 20/42

= 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

**Ans**: 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.

**Use Q7.csv file**

**Ans**: Using Python

Points:

Mean = 3.596563

Mode = 3.92

Median = 3.695

Range = 2.764 4.93

Variance = 0.28588

Standard Deviation = 0.5346

Score:

Mean = 3.21725

Mode = 3.44

Median = 3.325

Range = 1.513 5.424

Variance= 0.957379

Standard Deviation= 0.9784574

Weigh:

Mean = 17.84875

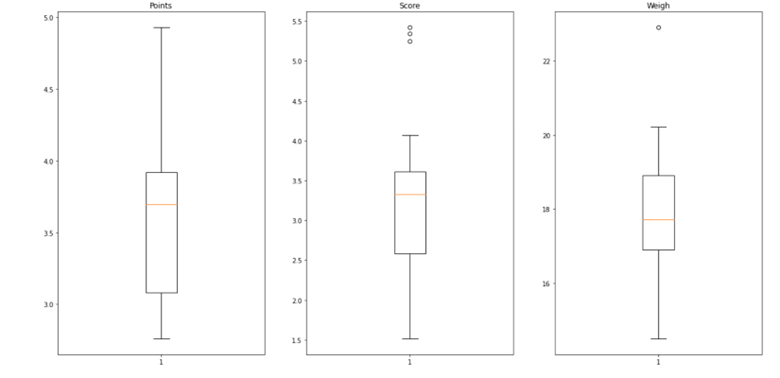
Mode = 17.02

Median = 17.710

Range = 14.5 22.9

Variance=3.193166

Standard Deviation= 1.786943



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

= 1308/9

= 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :** Skewness (speed) = -0.1139548 negative skewness means left skew i.e. data distributed on right side.

Kurtosis (speed) = 2.422853 positive kurtosis # data distribution is wide not peak.

Skewness (distance) = 0.7824835 positive skewness means right skew i.e. data distributed on left.

Kurtosis ( distance) = 3.248019 positive kurtosis # data distribution is wide not peak.

**SP and Weight(WT)**

**Use Q9\_b.csv**

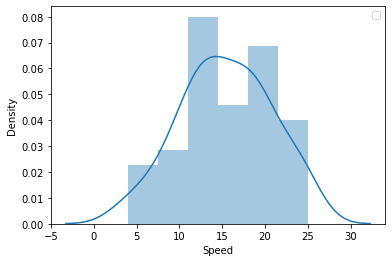
Skewness ( SP) =1.581454 positive so SP is right skewness.

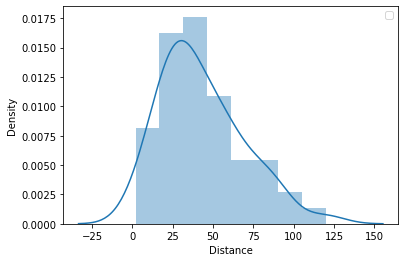
Kurtosis (SP) = 5.723521 positive # data is high peak.

Skewness( Weight) = -0.6033099 negative so WT is left skewness.

Kurtosis(Weight) = 3.819466 positive data is high peak.

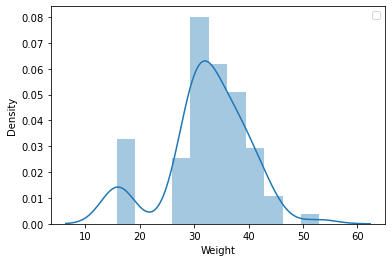
**Inferences:**

****

****

**Chart, histogram

Description automatically generated**

****

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

**Ans** :



50-100 weight having more frequency 180.

350-400 weight having very less frequency 5.

Positive skewness .

Data is right skewed.

Data is not a normal distribution.

0-50 weight having 80 frequency.

100-150 weight having 120 frequency.



**Ans** :

1. 7 Outliers are present in above box plot
2. Positive skewness .i.e. data is right skewed
3. Data is not normally distributed
4. Q1 is smaller than the Q3

**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 don’t have the standard deviation for population. So we have to use the T-distribution to determine the Cl of the given data

`X = 200 pounds, S = 30 pounds, n =2000

Confidence interval for 94% using R getting the

R code : qt(0.97,1999) = 1.88

Substituting values in the equation

200=+-1.88 30/^2000

Hence the confidence interval for 94% is [198,201].

Confidence interval for 96% using R getting the

R code :qt(0.98,1999) = 2.05

Substituting values in the equation

200=+-2.05 30^2000

Hence the confidence interval for 96% is [198.6,201.3].

Confidence interval for 98% using R getting the

R code :qt(0.99,1999) = 2.328

Substituting values in the equation

200=+-2.328 30^2000

Hence the confidence interval for 98% is [198.4,201.4].

**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.

**Ans** : Mean = 41, Median = 40.5, Variance = 25.52941, S.D = 5.052664

1. What can we say about the student marks?

**Ans** : Avg of students marks 41

Students marks range from 34 to 56

Mode is 41

Most of students score is between 35 to 42

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

**Ans** : When the values of mean, median and mode are equal, there is no skewness also you can say the data is in normal distribution.

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

**Ans** : If the mean is greater than the median, the distribution is positively skewed.

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

**Ans** : If the mean is less than the median, the distribution is negatively skewed.

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

**Ans**: A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

**Ans** : A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

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



What can we say about the distribution of the data?

**Ans** : - No outliers

Q1 greater than Q3

Median between 15

Most of data present in range of 10 to 18

Not following normal distribution

Left skewness of data

What is nature of skewness of the data?

**Ans**: Left Skewness

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

**Ans** : IQR=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. Both the plots infer that their data is normally distributed.
2. We can say that box plot 1 is for sample distribution and box plot 2 is for population or sample with larger size.
3. No outliers
4. Q1 is 25%, Q3 = 75%.IQR is 50% for both the box plots. So we can say both the distributions follow normal distribution i.e Mean =median=mode.

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)

**Ans:** There are 33 observations we got that satisfying given condition out of 81 observations therefore, P(MPG>38) = 33/81

* 1. P(MPG<40)

**Ans:** There are 61 observations we got that satisfying given condition out of 81 observations therefore, P(MPG<40)

= 61/81

* 1. P (20<MPG<50)

**Ans :** There are 69 observations we got that satisfying given condition out of 81 observations therefore, P (20<MPG<50)

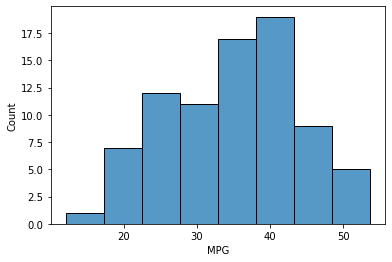
= 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

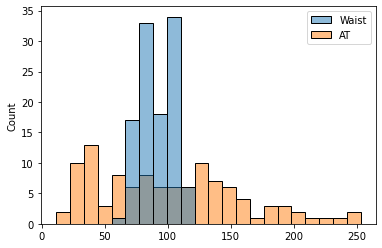
**Ans:**



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** :

qnorm(0.95) #Z score for 90% confidence interval is 1.64485

qnorm(0.97) #Z score for 94% confidence interval is 1.8807

qnorm(0.80) #Z score for 60% confidence interval is 0.841

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

**Ans** :

t score for 95% confidence interval is 2.0638

t score for 96% confidence interval is 2.171

t score for 99% confidence interval is 2.8

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

**Ans :**

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

t = -0.4714045207910317

p value = 0.32167411684460556