Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal (Discrete) |
| Weight of a person | Ratio (continuous) |
| Weight of Gold | Ratio (continuous) |
| Distance between two places | Ratio (continuous) |
| Length of a leaf | Ratio (continuous) |
| Dog's weight | Ratio (continuous) |
| Blue Colour | Nominal (Discrete) |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Ordinal (Discrete) |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal (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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

=> Possible events with equal probability is 8.

HHH,HHT,HTT,THH,THT,TTH,TTT therefore Number with 2 heads is : 3

The number you are looking for is 3/8 = 0.375 = 37.5%

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

=> Thus, two dice are thrown here is n(s)=36

i) the probability of getting the sum of two dice as 1 is zero.

ii) the sum is equal to 4 the possible outcomes are (1,3),(2,2),(3,1) i.e answer is = (3/36)=(1/12).

iii)sum of numbers divisible by 2 and 3 =29 soo i.e.=29/36

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?

=>Soln:

Combination of 7C2 = 7\*6/2\*1= n(a) = 21

Combination of (2+3) is 5C2 = 5 \*4/2\*1= n(b)=10

probability p(c) = 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

=> soln:

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.

**=>** Points Score Weigh

mean 3.596563 3.217250 17.848750

std 0.534679 0.978457 1.786943

median 3.695 3.325 17.710

variance 0.731217 0.9891 1.336

mode N/A N/A N/A

For points =>Mean=median=mode

This shows the points data is normally distributed.

For score=>Mean=median=mode

This shows the score data is normally distributed.

For Weigh =>Mean=median=mode

This shows the weigh data is normally distributed.

Overall, the given data set has normally distributed data points.

Q8) Calculate Expected Value for the problem below

* The weights (X) of patients at a clinic (in pounds), are

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

Q8) Calculate Expected Value for the problem below

* 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?

=> soln:

Expected Value = ∑ ( probability \* Value ) = ∑ P(x).E(x)

there are 9 patients

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

Expected Value of the Weight of that patient = 145.33

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

Cars speed and distance

Use Q9\_a.csv

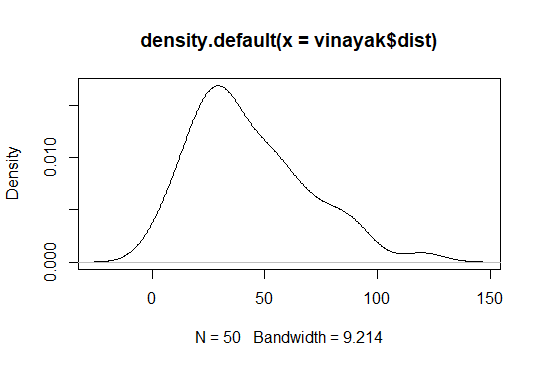
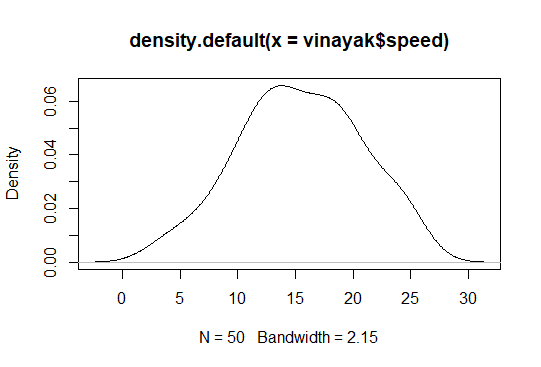
=> speed distance

skewness : -0.1139548 0.7824835

kurtosis : 2.422853 3.248019

For Speed (-0.11395)-Left Skewed

For Distance (0.7824)-Right Skewed



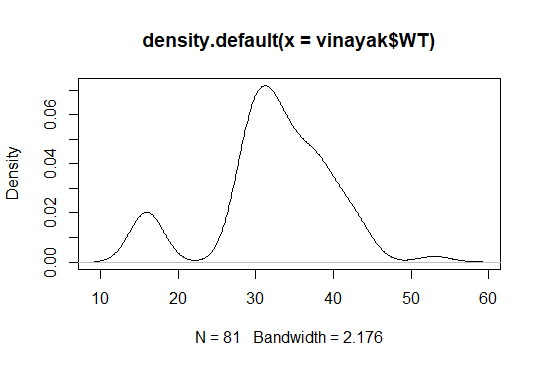
SP and Weight(WT)

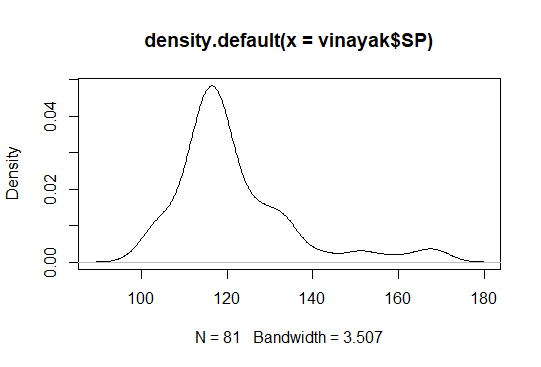
Use Q9\_b.csv

=> SP WT

skewness: 1.5814537 -0.6033099

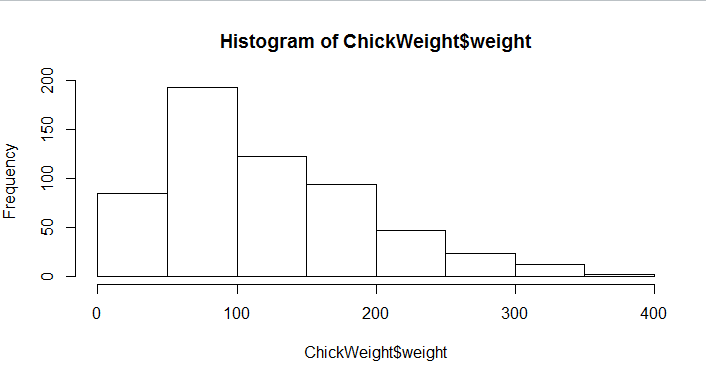
kurtosis : 5.723521 3.819466



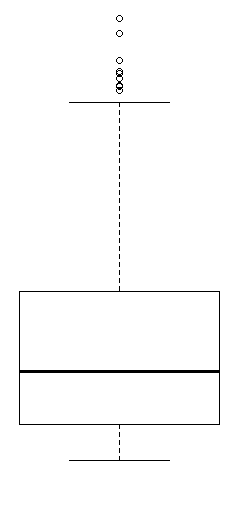


Q10) Draw inferences about the following boxplot & histogram

=>Majority of Chick Weights lies between the weight 50 to 100.

Only a smaller number of chick weights lies between weight 350 to 400.￼

=>We can find the outliers in the data sets which affect the mean and median.



**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?

=>Sample deviation is given then we should do t-distribution

i) Find standard error. The standard error (SE) of the mean is

SE= s/sqrt(n) = 30/sqrt(2000) = 30/44.721 = 0.67

Compute alpha (α) = 0.05

Critical probability (p\*)=0.97

i) 94% =0.94 and Critical probability (p\*)=0.97

t(0.97,1999)= 1.89

margin of error(ME) : ME= criticle value \* standard error

= 1.89\* 0.67 = 1.266

we can 94% confident that the population mean falls within the interval [197.536 202.463]

ii) 96% =0.96 and Critical probability (p\*)=0.98

t(0.98,199)= 2.067

margin of error(ME):ME = criticle value \* standard error

= 2.067 \* 0.67 = 1.3848

we can 96% confident that the population mean falls within the interval [196.7785 203.2215]

iii) 98% =0.99 and Critical probability (p\*)=0.99

t(0.99,199) = 2.345 and

margin of error(ME):ME = criticle value \* standard error

= 2.345\* 0.67 = 1.571

we can 98% confident that the population mean falls within the interval [195.9534 204.0466].

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

=> Soln:

1. Find mean, median, variance, standard deviation.

=>Mean = 41, Median = 40.5, Mode= 41, Variance= 25.52 and

Standard Dev = 5.05

1. What can we say about the student marks?

=>Since the mean=mode=median we can conclude that data is normally Distributed.

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

=> Skewness can be positive, negative or zero, If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness.

If mean = median = mode, the skewness is zero. We can find kurtosis here

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

=> If the mean is greater than the median, the distribution is positively skewed.

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

=> If the mean is less than the median, the distribution is negatively skewed.

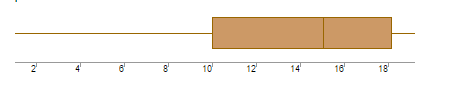
Q16) What does positive kurtosis value indicates for a data ?

=>Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. An extreme positive kurtosis indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean.

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

=>Negative values of kurtosis indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.

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



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

=>A type of distribution in which more values are concentrated on the right side (tail) of the distribution graph while the left tail of the distribution graph is longer. Not a Normal Distribution

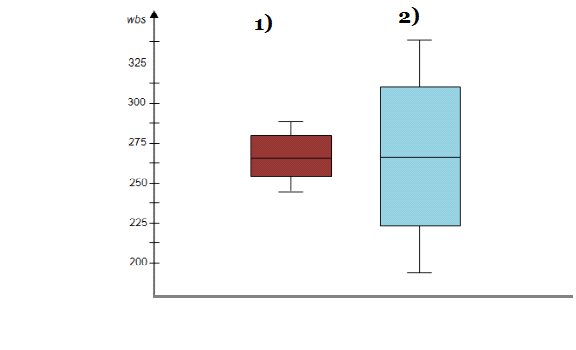
* What is nature of skewness of the data?

=> Nature of skewness of the data is negatively skewed.

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

=>Distribution is negatively skewed, the box plot will show the median closer to the upper or top quartile, says mean<median<mode. 18-8=8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot2.

* Both are Normally Distributed

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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

=> S**oln:**

(a) there are 33 observations in MPG which is greater than 38

(b)there are 61 observations MPG which is greater than 40

(c)there are 69 observation MPG which is 20<MPG<50.

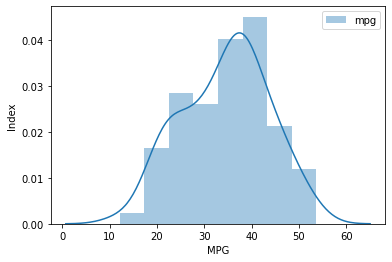
Q 21) Check whether the data follows normal distribution

I) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

=>Ans: Mean=34.42, Median=35.2 and Mode=36.8

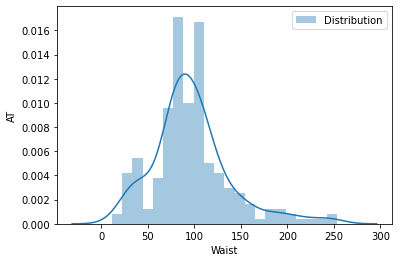
Since mean, median and mode are almost near too equal, we can conclude that MPG of cars follows normal distribution.



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

Dataset: wc-at.csv

=>the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set not follows Normal Distribution i.e. Mean=102 Median=97 and Mode=113 Since mean, median and mode are not to equal. We can conclude that Adipose Tissue does not follow normal distribution.



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

=> Considering the cars.csv and MGP as dataset

I) 90% confidence interval

=>means z= 0.90 and confidence level =0.95

Confidence interval 90% of MGP is (16.523, 52.318), q norm (0.95) = 1.64

II)94% confidence interval

=> means z= 0.94 and confidence level =0.97

Confidence interval 94% of MGP is (14.606, 54.237), qnorm (0.97) = 1.88

III)60% confidence interval

=> means z= 0.60 and confidence level =0.80

Confidence interval 60% of MGP is (22.720, 46.123). qnorm (0.80) = 0.84

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

=>I) 95% confidence interval :

Means t=0.95 and confidence intervel = 0.975

Confidence interval 95% of MGP for sample 25 is [34.0993 54.0567]

II) 96% confidence interval :

Means t=0.96 and confidence intervel = 0.98

Confidence interval 98% of MGP for sample 25 is [33.7211 54.4349]

III) 99% confidence interval :

Means t=0.99 and confidence intervel = 0.995

Confidence interval 99% of MGP for sample 25 is [32.6104 55.5456]

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

=> soln:

t - statistics for the data is given as follows: t= (x-mean)/(s/sqrt(n))

x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

i.e t = - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.