|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete (countable) |
| Results of rolling a dice | Discrete (countable) |
| 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 (nominal) |
| Number of kids | Discrete (countable) |
| Number of tickets in Indian railways | Discrete (countable) |
| Number of times married | Discrete (countable) |
| Gender (Male or Female) | Discrete (nominal) |

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 | Discrete (nominal) |
| High School Class Ranking | Discrete (ordinal) |
| Celsius Temperature | Continuous |
| Weight | Continuous |
| Hair Color | Discrete (nominal) |
| Socioeconomic Status | Discrete (nominal) |
| Fahrenheit Temperature | Continuous |
| Height | Continuous |
| Type of living accommodation | Discrete (ordinal) |
| Level of Agreement | Discrete (Ordinal) |
| IQ (Intelligence Scale) | Discrete (ordinal) |
| Sales Figures | Continuous |
| Blood Group | Discrete Nominal |
| Time Of Day | continuous |
| Time on a Clock with Hands | continuous |
| Number of Children | Discrete (countable) |
| Religious Preference | Discrete (nominal) |
| Barometer Pressure | Continuous |
| SAT Scores | Discrete (countable) |
| Years of Education | Discrete (countable) |

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

obtained?

1. Sample = {HHH, HHT, HTH, HTT, TTT, TTH, THT, THH}

= 3/8

= **0.375**

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

1. Equal to 1

A: - **zero**

1. Less than or equal to 4
2. = {(1,1), (2,1), (3,1), (1,2), (2,2), (1,3)}

= 6/36

= **0.167**

1. Sum is divisible by 2 and 3

A) Favorable outcomes = (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Therefore,

Number of favorable outcomes = 5

Total favorable outcomes = 36

Probability = Number of favorable outcomes / Total favorable outcomes

**Probability = 5/36**

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?

Red = 2, Green = 3, Blue = 2

Total number of balls 🡺 2+3+2 = 7

N(s) = 2 balls are drawn random out of 7

ncr = n!/(n-r)!\*r!

= 7!/(7-2)!\*2!

= 21

Lets us consider N(e) = Events of drawing 2 balls, none of the balls drawn is blue

N(e) = number of ways of drawing 2 balls is out of (2 + 3) = 5

ncr = n!/(n-r)!\*r!

= 5!/(5-2)!\*2!

= 10

Probability that none of the balls drawn is blue) = N(e) / N(s)

= 10 / 21

= **0**.**4762**

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 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

**Expected number of candies for a randomly selected child = 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

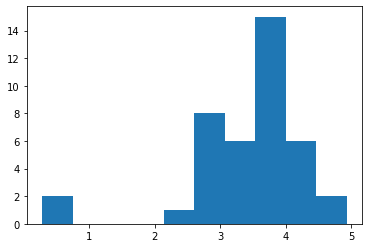
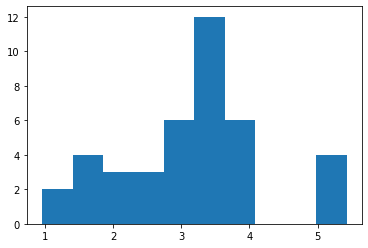
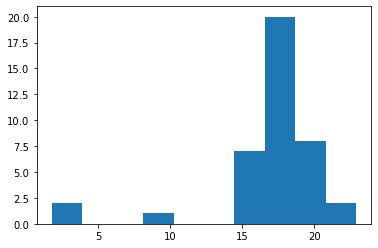
also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**A)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | Points | Score | Weigh | | Mean | **3**.**58** | **3.231355** | **17.82452** | | Median | **3.69** | **3.325** | **17.71** | | Mode | **3.92** | **3.44** | **17.02** | | Variance | **0.285881** | **0.957379** | **3.193166** | | Standard Deviation | **0.534679** | **0.978457** | **1.786943** | | Range | **2.17** | **3.911** | **8.4** |  |  | | --- | |  | |

HISTOGRAM FOR THE GIVEN DATA SET

**1. POINTS 2. SCORE 3. WEIGHT**

**MEAN < MEDIAN < MODE in** all the figures

i.e. Fig 1, fig 2 & fig 3 all the histogram are negatively skewed

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?

A) Expected Value = ∑ (probability \* Value)

∑ P(x). E(x)

Total patients = 9

Probability of selecting each patient = 1/9

Expected Value = ∑ (each person probability) \*(each person’s weight)

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)

**Expected Value of the Weight of that patient = 145.33**

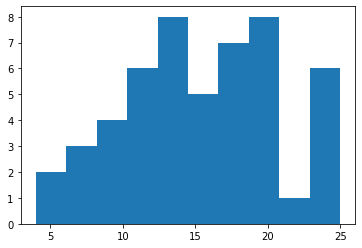
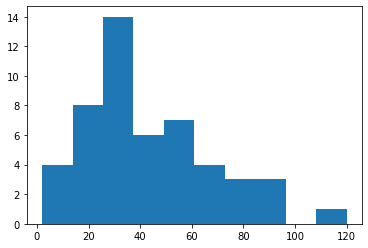
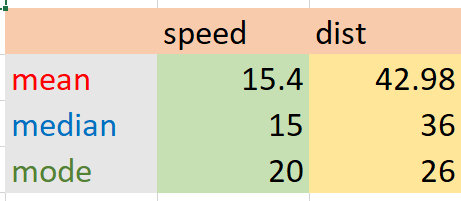
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Car’s speed and distance**

**Use Q9\_a.csv**

**A)** kurtosis**: [2.42285258 3.24801866]**

skewness**: [-0.11395477 0.78248352]**

**  **

**SPEED DISTANCE**

From figure speed

* MEAN < MEDIAN < MODE
* Negatively skewed

From figure distance

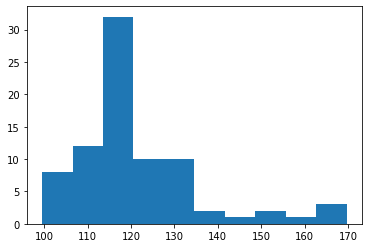
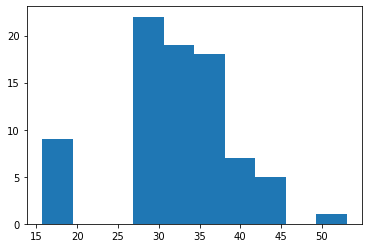
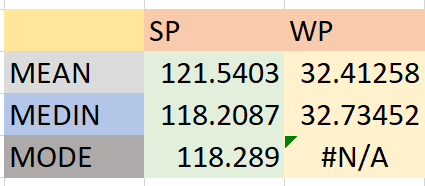
* MEAN > MEDIAN > MODE
* Positively skewed

**SP and Weight (WT)**

**Use Q9\_b.csv**

**A)** kurtosis**: [5.72352149 3.81946588]**

skewness**: [ 1.58145368 -0.60330993]**

**  **

**SP WT**

From figure speed

* MEAN > MEDIAN > MODE
* Positively skewed

From figure distance

* MEAN > MEDIAN > MODE
* Positively skewed

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



**Histogram: -**

* Chick Weight is taken on X - axis and frequency is taken on Y – axis
* Chick Weight is right skewed and positively skewed
* 50 % of the chick weighs lies between 50 to 150
* Chick Weight at 100 is having more frequency (i.e., 200)
* A histogram in which most of the data is falling on to the right of the graphs peak is known as **right -skewed** histogram. It is also known as a **positively skewed** histogram
* The relationship between the mean, median and mode will be **mean > media > mode**



**Boxplot: -**

* By using boxplot, we will identify the out liars
* Out liars mean **extrema** **values** which are very far from the remaining data points
* The longer part of the box is on the right of the median, so the data is **right** **skewed** in the above figure
* Seven out liars are present at the upper whisker

**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)** **confidence interval for 94% = 143.5762 256.4238**

**confidence interval for 98% = 130.2096 269.7904**

**confidence interval for 96% = 138.3875 261.6125**

**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 | **41.11111** |
| Median | **40.5** |
| Variance | **24.10458** |
| Standard Devotion | **4.909641** |

1. What can we say about the student marks?
2. **positively** **skewed**

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

A) **Symmetrical** **distribution**

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

A) **positively** **skewed**

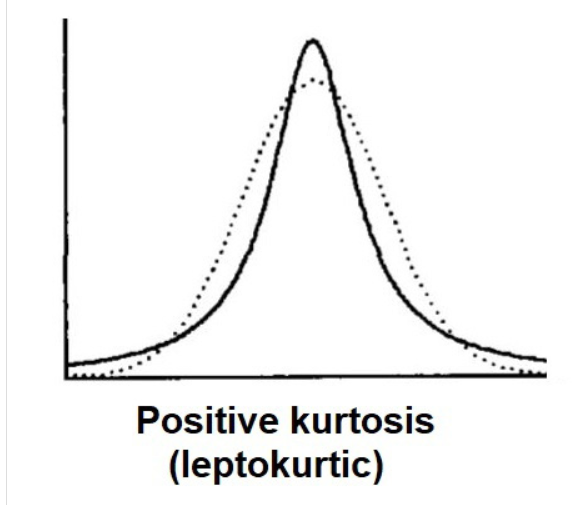
Q15) What is the nature of skewness when median > mean?

1. **positively** **skewed**

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

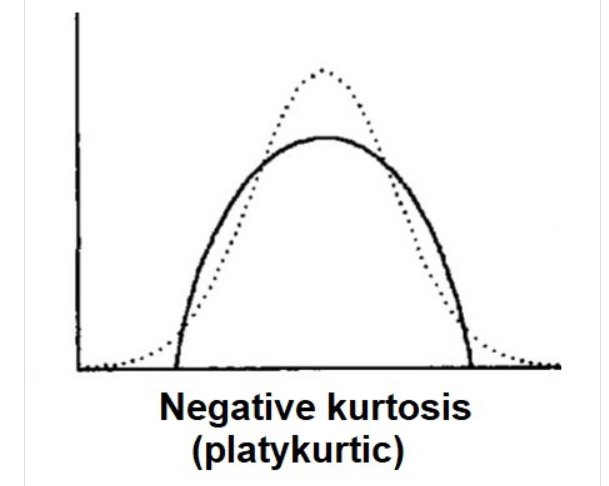
A) **Positive kurtosis value indicates that the distribution is packed and**

**possesses the thick tail**



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

1. **Negative kurtosis value indicates that the distribution is flat and has thin tales**



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



What can we say about the distribution of the data?

1. **The data is not normally distributed**

What is nature of skewness of the data?

A) **Positively skewed**

**Mean > median**

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

**= 18 – 10**

**IQR = 8**  
  
Q19) Comment on the below Boxplot visualizations?



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

1. **The box plot 1 is designed with range of 3**

**The box plot 2 is designed with range of 1.5**

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

Dataset: Cars.csv

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

MPG <- Cars $MPG

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

C. P(20<MPG<50)

**A) The probability of P (MPG<38) is 0.3475907861423393**

**The probability of P(MPG<40) is 0.691935760963533**

**The probability of P (20<MPG<50) is 0.8988689146142506**

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. **Mean = 34.4**

**Median = 35.1**

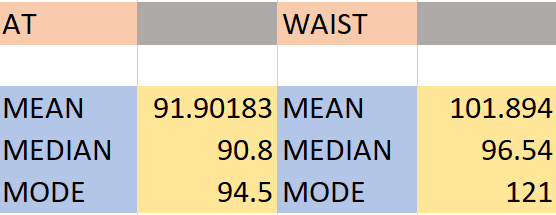
**Mode = 29.6**

* It is not following normal distribution because the **median is greater than mean**

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

Dataset: wc-at.csv

**A)**

****

* Both **AT** & **WAIST** are not following normal distribution because the **mean greater than median greater than mode**

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

60% confidence interval

1. Z score of 90% confidence interval is **1.645**

Z score of 94% confidence interval is **1.8808**

Z score of 60% confidence interval is **0.8416**

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval,

99% confidence interval for sample size of 25

1. T score of 95% confidence interval is **1.711**

T score of 96% confidence interval is **2.064**

T score of 99% confidence interval is **3.467**

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

1. t - statistics for the data is given as follows:

t = (x – μ)/ (s/ √ 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

FORMULA:

**t = (x – μ)/ (s/ √ n)**

t = (260 – 270)/ (90/ √ 18)

t = -10 / (90/ 3√ 2)

t = -10 / (30/ √ 2)

t = -1 \* √ 2 / 3

t = - 0.471



W.K.T degrees of freedom = n - 1,

= 18 -1

= 17

So, t – distribution with 17 degrees of freedom

**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