|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Nominal |
| Blue Color | Discrete |
| Number of kids | Ordinal |
| Number of tickets in Indian railways | Ordinal |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | 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?

ANS: Possible outcomes= {HHH,HHT,HTT,THH,THT,TTH,HTH,TTT}

Probability = (number of favorable outcome) / (total number of outcome)

Probability of getting two heads and one tail is 3/8.

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

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

ANS: possible outcome = 36

1. The probability of sum is equal to 1 is 0.
2. the sum is equal to 4 the possible outcomes are (1,3),(2,2),(3,1) therefore n( b) = 3/36 = 1/12
3. their sum is less than 13. here the total sample space will come . n(c)= 36/36 = 1.

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: total number of balls=7

Total number of balls  
= (2 + 3 + 2)  
= 7  
Let S be the sample space  
Then, n(S) = Number of ways of drawing 2 balls out of 7

N(S)=7C2n(S)=(7×6)(2×1)n(S) =21

Let E = Event of 2 balls, none of which is blue  
∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2n(E)=(5×4)(2×1)n(E)=10∴P(E)=n(E)n(S)=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

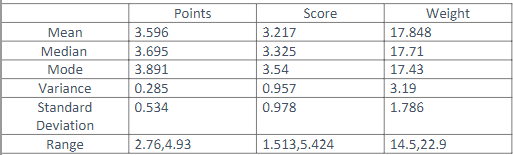
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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:

****

1) since the mean, median and mode for point, score and weight all three variables are approximately equal hence they show symmetrical frequency curve symmetrical distribution of data.

2)There are no outliers in this data.

3)Since the standard deviation value is low hence the spread is small. All the values are near the mean value.

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: Expected Value = summation of all the values / number of values =1308/9 =145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

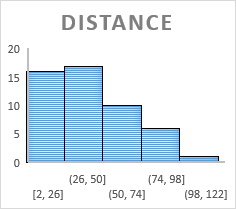
ANS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Dist* |  |  | *speed* |  |
|  |  |  |  |  |
| Mean | 42.98 |  | Mean | 15.4 |
| Standard Error | 3.644340314 |  | Standard Error | 0.747785847 |
| Median | 36 |  | Median | 15 |
| Mode | 26 |  | Mode | 20 |
| Standard Deviation | 25.76937749 |  | Standard Deviation | 5.287644435 |
| Sample Variance | 664.0608163 |  | Sample Variance | 27.95918367 |
| Kurtosis | 0.405052582 |  | Kurtosis | -0.50899442 |
| Skewness | 0.80689496 |  | Skewness | -0.117509861 |

Inference for distance 3rd and 4 th moments ;-

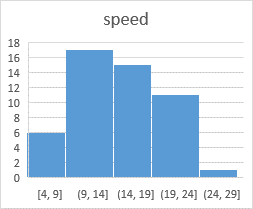
Since the skewness is positive number hence this is right skewed with a value of skewness = 0.80689

Kurtosis is the measure of tailedness it also is a positive value = 0.4050 show the graph will be show right tailedness with a value = 0.4050



Inference for speed 3rd and 4 th moments ;-

Since the skewness and kurtosis values are –ve number hence this is slightly left skewed with a slight left tailedness .but since the values are -0.5 and -0.11 we can consider this to be a normally distributed data following the bell curve.



**Histograms for speed**

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| *SP* | | *WT* | |
|  |  |  |  |
| Kurtosis | 2.977329 | Kurtosis | 0.950291 |
| Skewness | 1.61145 | Skewness | -0.61475 |

|  |  |
| --- | --- |
|  |  |

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



**Histogram**: In this histogram right skewness present also called positive skewness. skewness is a measure of lack of symmetry of the [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution) of a [real](https://en.wikipedia.org/wiki/Real_number)-valued [random variable](https://en.wikipedia.org/wiki/Random_variable) about its mean. This is asymmetrical distribution of data with a positive skewness hence in this case mean >median>mode



**Boxplot**: Inupper extreme outliers are present. The right whisker - line above is larger than the lower whisker line hence the data is having more variance is very high and data shows the spread of data on the right of the Q3 because more data lies in 25% on the right than that in the left 25% **.**

**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?

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

ANS:

1)MEAN=41, Median=40.5, variance=25.52, Standard deviation=5.05664

2) **According to empirical rule**

68 % people fall in the one std deviation which is (41-5.0526 ) = 35.9474 to (41 +5.0526)=46.0526

95% people fall in the second std deviation which is (41-2\*5.0526) =30.8948 to (41+2\*5.0526) = 51.10526

99.5% people fall is the third std deviation which is ( 41-3\*5.0526)= 25.8422 to (41+3\*5.0526)=56.1578

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

ANS : If both mean and median equal then zero skewness present.

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

ANS: if mean>median then right skewness present also called as positively skewed.

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

ANS: if median>mean the left skewness present also called as negatively skewed.

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

ANS: Positive kurtosis value indicates that thinner peak and wider tails

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

ANS: Negative kurtosis value indicates that 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? ANS: Not normally distributed

What is nature of skewness of the data? ANS: Negative skewness

What will be the IQR of the data (approximately)? ANS : Approximately 10-18

Q19) Comment on the below Boxplot visualizations?



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

ANS:

Both the box plot have the same median value is around approximately 250

Boxplot 1) The IQR for this boxplot is between 253 to 275(approx.),

Q1=253 ,Q4 =275 (approx.)

Boxplot 2)The IQR for the boxplot is between 225 to 320(approx.) Q1=225,Q4=320 (approx.)

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)
  2. P(MPG<40)

c. P (20<MPG<50)

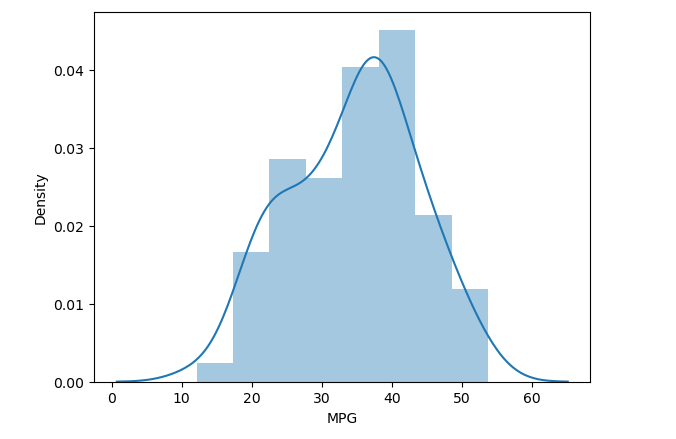
Ans: 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

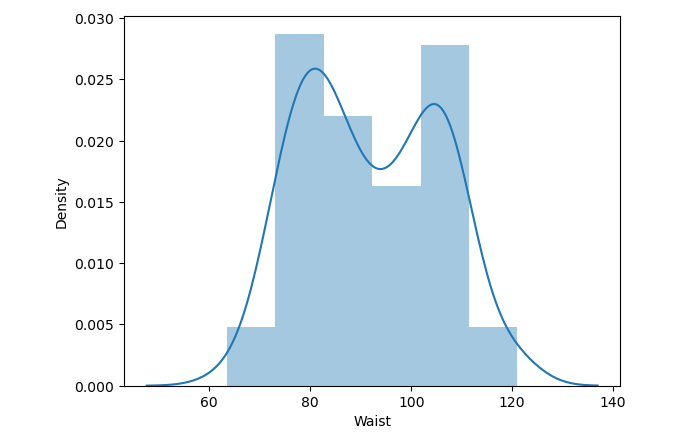
ANS: Follows normal distribution

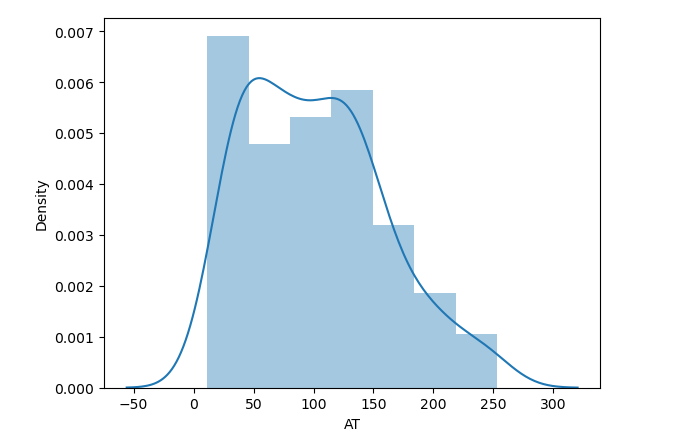


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

ANS: Neither of them follows normal distribution.





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

ANS:

|  |  |
| --- | --- |
| Confidence Interval | Z-Score |
| 60% | 0.253 |
| 90% | 1.645 |
| 94% | 1.555 |

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

ANS: 

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:

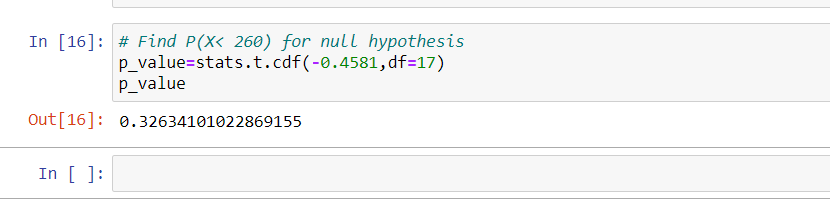
population mean =270, sample size n=18 ,sample average = 260 ,SD=90.

Since the sample size is less than 30 hence we will use the t test.

Manual calculation(t statistics ) t= (VAL-EV)/Sd

260-270/21.828 = -0.4581

t .cdf(-0.4581, 17) = 0.6783



Hence the p-value = 32.63%

Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.63%