|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Discrete |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Continuous |
| 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 | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ordinal |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans:-

{HHH,THH,TTH,TTT,HTT,HHT,THT,HTH}

The probability of getting Two heads and One tail is 3/8 = 0.375

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:-

*(1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)*

*(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)*

*(3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)*

*(4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)*

*(5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)*

*(6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)*

a) Equal to1:-They don’t have any probability

b) Less than or equal to 4:-(1, 1) (1, 2) (1, 3) (2, 1) (2, 2) (3, 1)

7/36=0.194

c) Sum is divisible by 2 and 3:- (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

5/36=0.138

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:-

=(5/7)\*(4/6)

=20/42

=10/21

=0.47

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:- The 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\*0.120

=0.015+0.80+1.95+0.025+0.06+0.24

=3.09

The 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**

Ans:-

Mean Points 3.596563 Mode Points 3.07

Mean Score 3.217250 Mode Score 3.44

Mean Weigh 17.848750 Mode Weigh 17.02

Median Points 3.695

Median Score 3.325

Median Weigh 17.710

Variance Points 0.285881

Variance Score 0.957379

Variance Weigh 3.193166

Standard Deviation Points 0.534679

Standard Deviation Score 0.978457

Standard Deviation Weigh 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

=145.3

The Weight of that patient is 145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans:-Skewness

Index 0.000000

speed -0.117510

dist 0.806895

Kurtosis

Index -1.200000

speed -0.508994

dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans**:-** Skewness

SP 1.611450

WT -0.614753

Kutosis

SP 2.977329

WT 0.950291

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



Ans:- The skewness of the data is positively skewed

The most chick weight between 50 to 100



Ans:- lower whisker is smaller than the upper whisker

Size of IQR is thin

**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:-

The 94% confidence interval is(198.73, 201.27).

The 96% confidence interval is (198.61, 201.39).

The 98% confidence interval is (198.43, 201.57).

**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)Find mean, median, variance, standard deviation.

Mean=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=738/18

=41

Median=40+41/2

= 81/2

=40.5

Variance=434/18-1

=434/17

=25.529

Standard deviation=sqrt of variance

=5.05

2)What can we say about the student marks?

Ans:- Average score of student is 41

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

Ans:- skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.

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

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

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

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

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

Ans:- Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Ans;- Negative kurtosis indicate that a distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

-The min value of data is 2

-Lower limit is -2 (lower limit = Q1 – 1.5(IQR)

-Upper limit is 30(Upper Limit = Q3 + 1.5(IQR)

-50% of data lies between 10 to 18.3

-Most of the data is present between 10 to 15.2(Which is from first quartile to second quartile )

What is nature of skewness of the data?

-Negatively skewed

What will be the IQR of the data (approximately)?   
 =8

Q19) Comment on the below Boxplot visualizations?



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

Data 1 has lower IQR then data 2

Mean is common in both data which is 262.5

There are no outliers in both data

Both the data is normally distributed

2nd data is big compare to 1st (which means the upper limit and lower limit is bigger in 2nd data)

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 = 0.347598

* 1. P(MPG<40)

Ans : 0.729353

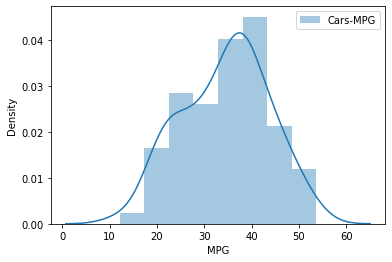
* 1. P (20<MPG<50)

Ans : 0.013119

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

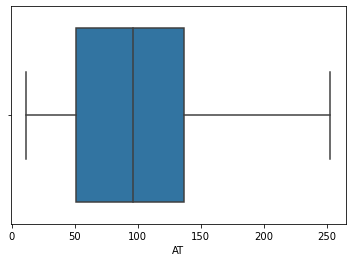
Dataset: Cars.csv



-From above we can say the data is fairly distribution

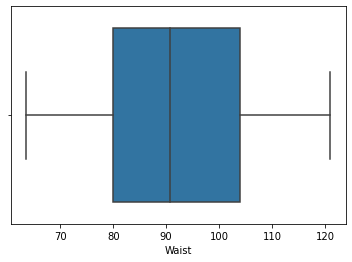
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



-AT is positively skewed

(ii)



**-Waist fairly normal distributed**

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

Ans:-

90% confident interval= 1.645

94%confident interval= 1.880

60%confident interval= 0.253

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

Ans:-

95% confidence interval=1.710

96% confidence interval=1.828

99% confidence interval=2.492

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:-

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

degree of freedom=18-1=17

t=0.471