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

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

Ans: No of total possible out comes=8

N.o of faviourble out comes=3

P(2H and 1T)=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:Total n.o of possible out comes=36

1. N.o of ways we get sum of 1 is (1,1)

N.o of faviourable out come=0

P(Sum 0f 1)=0\36

B)n.o of ways we get less than or equal to 4 is (1,1)(1,2)(1,3)(3,1)(2,1)(2,2)

n.o of faviourable out comes=6

P(less than or equale to 4)=6\36

=1\6

c)n.o of ways we get sum is divisible by 2 nd 3 = 6

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

P(sum is divisible by 2 nd 3)=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:total n.o of balls=7

n.o of blue balls=2

N.o of non blue ball

The probability of drawig a non blue ball at 1st time=5/7

The probability of drawig a non blue ball at 2nd time=4/6

P(none of the ball blue ball)=(non blue ball at 1st time)\*(non blue ball at 2nd time)

=5/7\*4/6

=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 = 0.02.

Ans: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

=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**

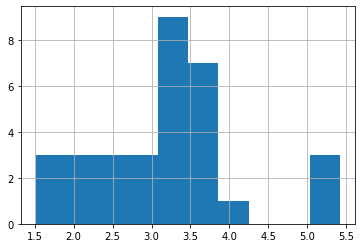
**Ans:Points*:***Mean=3.596 ,Median= 3.695 ,Mode=3.07 ,Standard deviation=0.534,Variance=0.285

**Score:** Mean= 3.217 ,Median= 3.325 ,Mode=3.44 ,Standard deviation=0.978,Variance=0.957

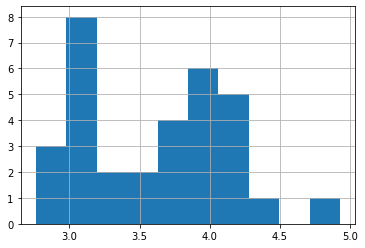
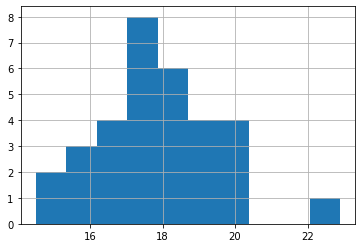
**Weigh:**Mean= 17.84 ,Median=17.71 ,Mode=17.02 ,Standard deviation=1.786,Variance=3.193

Range of Points: 2.17

The data does not follows normal distribution



Score



weight points

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=108+110+123+134+135+145+167+187+199\9

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

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





Ans:Here we can say that the major chick weights falls in the catogory of 50-100gm in X-axis as the max which is 200.

The plot is Right sqewed which show that tere is lesser concentration of chik weights in the 300-400gm

Median < mean it is rightly skewed and we have outlayers in box plot and there is less data points between Q1 and bottom point

**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.

Ans:

Mean: 41.0

Median: 40.5

Variance: 24.111

Standard deviation: 4.910306620885412

1. What can we say about the student marks?

Ans:The student marks are above average with mean=41

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

Ans: Symetrical

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

Ans:Right Skewness

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

Ans:Left Skewness

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

Ans: most of the data will be on Right side

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

Ans; most of the dat will be on left side

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



What can we say about the distribution of the data?

Ans:It is negatively skewed data.it lies towards to the left side

What is nature of skewness of the data?

Ans:negatively Skewed

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

Ans:IQR =Q3-Q1

=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:distribution of data for Boxplot1 with respect to Boxplot2

Mean=median=mode=0

Since the visualizations of both Boxplots mean,median and mode points are center

So the nature of skewness of the data is in the form of symetry

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)

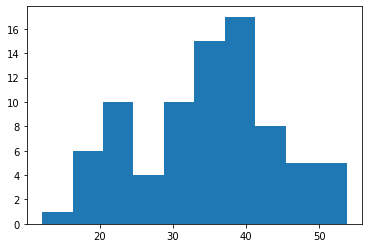
c. P (20<MPG<50)

Ans: MPG, Length: 81, dtype: float64

P(MPG>38): 0.4074074074074074

P(MPG<40): 0.7530864197530864

P(20<MPG<50): 0.8518518518



Q 21) Check whether the data follows normal distribution

a)Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

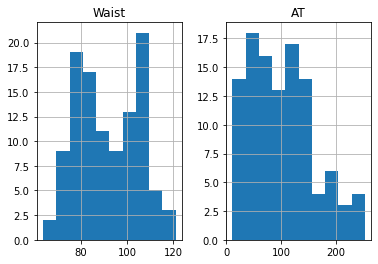
b)Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: Shapiro-Wilk test results:

Statistic: 0.9558579921722412

p-value: 0.0011704873759299517



The data does not appear to be normally distributed.

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

Ans:

Z score for 90% confidence interval: 1.6448536269514722

Z score for 94% confidence interval: 1.8807936081512509

Z score for 60% confidence interval: 0.8416212335729143

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

Ans:

Total score for 95% confidence interval: 2.0638985616280205

total score for 96% confidence interval: 2.1715446760080677

total score for 99% confidence interval: 2.796939504772804

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:

Sample=18=n

Sample mean=260days=x

Sample standard deviation(sd)=x

=260-270/90/sqr(18)

=-10/9.487

=-1.05407