|  |  |
| --- | --- |
| 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 | Continuous |
| Number of times married | Discrete |
| Gender (Male or Female) | Continuous |

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 :-** If Three Coins are tossed , There are 2\*2\*2 =8

i.e P={HHH,HHT,HTH,HTT,TTT,TTH,THT,THH}

Therefore, {HHT,HTH,THH} The probability of getting two head 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 :-** Two Dice are rolled Outcome are **S= 6\*6=36**

i.e

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

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

**S =** (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)

1. Sum is Equal to 1 = 0

**P(a)** = n(a)/S = 0/36 = **0**

1. Less than or equal to 4

n(b) ={ (1, 1) ,(1, 2) ,(1, 3), (2, 1) , (2, 2), (3, 1)} = **6**

**P(b)** = n(b) /S = 6/36 =**1/6**

1. Sum is divisible by 2 and 3

**n(c)** ={(1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (6, 6)}= **6**

**P(c)** = n(c) /S = 6/36 =**1/6**

**(note:- P = Probability)**

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?3.9

**Ans :-** A bag contains 2 red, 3 green and 2 blue balls,two balls are drawn at random i.e

Total NO. Of balls = 2+3+2 =**7**

* Number of ways of drawing 2 balls out of 7 (X)= (7\*6)/(2\*1) = 42/2 = **21**
* Number of balls other than blue = 5
* Number of ways of drawing 2 balls out of 5 (Y)= (5\*4)/(2\*1) = 20/2 = **10**

**Probability(P)** = n(Y)/n(X) = **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

**Ans :- Given Data,**

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Therefore, 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\*0.10 = **3.09**

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

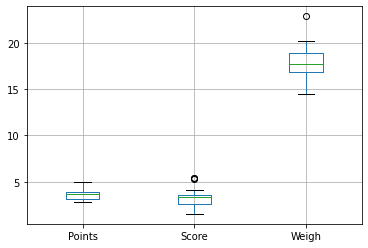
* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans :- Using .csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.6950 | 3.325 | 17.71 |
| Mode | 3.891 | 3.54 | 17.43 |
| Variance | 0.285 | 0.957 | 3.19 |
| Std. Deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.71 | 3.911 | 8.34 |



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 :-** Given Data : The weights(X) of patients at a clinic (in pounds), are 108, 110, 123, 134, 135, 145, 167, 187, 199.

One of the patients is chosen at random.

Expected Value = ∑ (probability \* value)

∑ P(x).E(x)

There are 9 patients

Probability of selecting each patient = 1/9

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

P(x) 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 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)

= 145.3333

**Mean =145.334**

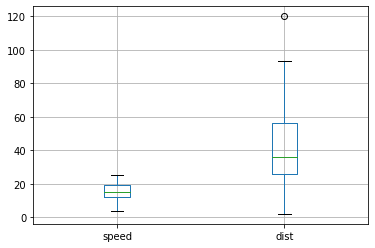
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :-**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **distance** |
| **Skewness** | -0.117510 | 0.806895 |
| **Kurtosis** | -0.508994 | 0.405053 |

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **distance** |
| **Skewness** | -1.611450 | -0.614753 |
| **Kurtosis** | -2.977329 | 0.950291 |

**Therefore, =Skewness for speed =** - 0.117510 , skewness value is negative so it is left skewed .since magnitude is slightly greater than 0 it is slightly left skewed

And for distance **=** 0.806895 , right skewed (positive) slight magnitude to right.



**Ans :- Given Data**  **,**

* The most of data points are concerated in the range 50-100 with frequency 200.
* Least range of weight is 400 some were around 0-10.
* So the expected value the above distribution is 75.
* Skewness is can notice a long tail towards right so it is heavily right skewed.



Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points

**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 :- here,**

|  |  |  |
| --- | --- | --- |
| Confidence interval | Z Value | Range |
| Confidence interval 94% | 1.880794 | 198.74 , 201.26 |
| Confidence interval 98% | 2.326348 | 198.43 , 201.56 |
| Confidence interval 96% | 2.053749 | 198.62 , 201.38 |

**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=24.11**

**Std. Deviation =4.910**

1. What can we say about the student marks?

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 skewed.

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

**Ans :-** Left skewed.

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

**Ans :-** The data is normally distributed And Kurtosis value is 0.

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

**Ans :-** The distribution of the data has lighter tails And a flatter peaks than the normal distribution.

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



What can we say about the distribution of the data?

**Ans :-** Let’s assume box plot is about ages of the students in a school. 50% of the people are above 10 yrs old and remaining are less . And student whos age above 15 are approx. 40%.

What is nature of skewness of the data?

**Ans :-**Left skewed ,median is greater than mean.

What will be the IQR of the data (approximately)?   
**Ans :-** Approximately = -8

Q19) Comment on the below Boxplot visualizations?



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

**Ans :- Above** plot by observing both the plots whiskers level is high in boxplot 2 ,mean and median are equal hence distribution is symmetrical.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

**Ans :- given data,**

* confidence interval(CI) =90% =0.90

Significant S = 1+CI/2 = 1+0.90/2 =1.90/2 =0.95

Using Z tables for z value

**Z value** =**1.645**

* CI =94% =0.94

S =1+CI/2=1+0.94/2=0.97

Using Z tables for z value

**Z value** =**1.882**

* CI =60% =0.60

S =1+CI/2=1+0.60/2=0.8

Using Z tables for z value

**Z value** =**0.842**

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

**Ans :- n** = 25

**Degree of freedom (df)** = 25-1 = **24**

* Confidence interval 95% , Using T table t value

**T Value =2.064**

* Confidence interval 96% , Using T table t value

**T Value =2.172**

* Confidence interval 99% , Using T table t value

**T Value =2.797**

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 :- Given data,**

Sample mean = 260 day, Sample std.deviation = 90 , n=18

df=n-1=18-1=**17**