|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | ordinal |
| 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 | nominal |
| Number of kids | discrete |
| Number of tickets in Indian railways | discrete |
| 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 | interval |
| Hair Color | nominal |
| Socioeconomic Status | ordinal |
| Fahrenheit Temperature | interval |
| Height | interval |
| Type of living accommodation | ordinal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | ratio |
| Sales Figures | ratio |
| Blood Group | nominal |
| Time Of Day | nominal |
| Time on a Clock with Hands | interval |
| Number of Children | Ordinal |
| Religious Preference | nominal |
| Barometer Pressure | interval |
| SAT Scores | interval |
| Years of Education | ordinal |

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

-total number of possible outcomes is 23 =8  
event=(HHT,THH.HTH)  
probability=>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

* Total number of possible outcomes= 62 =36

1. p(sum=1)=0
2. p(sum<=4)  
   event((1,1)(1,2)(1,3)(2,2) (2,1)( 3,1))

p(sum<=4)=6/36=0.166  
c)p(sumis divisible by 2 and 3)=  
event((1,5)(2,4)(3,3)(4,2)(5,1)(6,6))

p(sumis divisible by 2 and 3)=6/36=0.166

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?

Total number of ball is =7  
total number of event=21

No blue ball=10

Probability(no blue)=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

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected values |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.20 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.120 | 0.24 |
| Total expected candies |  |  | 3.085 |

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.5365 | 3.2172 | 17.8487 |
| Medain | 3.695 | 3.325 | 17.710 |
| variance | 0.285881 | 0.957379 | 3.1 |
| Standard Deviation | 0.53467 | 0.9784 | 1.7869 |
| Range | 2.170 | 3.911 | 8.400 |

**For points and score, weigh mean and median is nearly same so that data is distributed normally. weigh has high range means there is huge difference between max and min point in respective of the points and weigh. In points std is less so that the data is cluster near to mean in respective of others they have high std.**

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

1. The weights (X) of patients at a clinic (in pounds), are

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

Random variables {108, 110, 123, 134, 135, 145, 167, 187, 199

}

Sum =108+110+134+123+135+145+167+187+199=1308

Expected values=1308/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness speed =-0.11395477**

**distance=0.78248352**

**Kurtosis speed =-0.57714742**

**distance=0.24801866**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness sp=1.58145**

**wt=-0.6033**

**Kurtosis sp=2.7235**

**wt=0.819465**

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





On the basis of boxplot and histogram we concluded that Its right skewed data set so that most number of data points on lower side and median will be on the left of the mean of the data set and the histogram has log tail so that there some of the data point in higher side has outliers and this outliers we can see in the box plot as well.

**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 of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

94% Confidence Interval: (198.7376089443071, 201.2623910556929)

98% Confidence Interval: (198.4381860483216, 201.5618139516784)

96% Confidence Interval: (198.6214037429732, 201.3785962570268)

**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

Median=40.5

Var=24.111

Std=4.91

1. What can we say about student in the exam is the student marks?

1] there is some deviation in the marks so that sum of the students scores high marks and some of the student score low marks in respective of the other student

2] mean and median approx. similar so that the data is equally distributed both side of the mean points so that scores of the students in the exam equally distributed almost % student got low marks and 50% students got high marks.

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

No skewness (zero skewness)

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

Right skewness (positive skew)

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

Left skewness (negative skew)

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

It means that it has shaper tip and less tail means that deviation in data less rather than mean so that all data points is near to mean only

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

It means that it has large spread and less mean it means that it has serval data point has little bit high distance from the mean

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



What can we say about the distribution of the data?

It has high left skewness means it has high value data point in the higher side of the data and data has more number of data point whose value little bit small than mean more on the lower side of the data rather than the higher side of the data

What is nature of skewness of the data?

Left skewness

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

The boxplot 1 has low variance and box plot 2 has high vaiance both of them has similar mean

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)=99.99%(0.99)
  2. P(MPG<40)=0.010%(0.0010)

c. P (20<MPG<50)=0.0129(0.00012)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Mean of mpg 34.422

Median of mpg 35.1527

As per the Q\_Q plot and mean and median nearly same we conclude that MPG of Cars 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

As per the Q-Q plot we conclude that adipose tissue and waist deviates from normal distributions circumference

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

Z score(90%) 1.645

Z score(94%) 1.880

Z score(60%) 0.253

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

t score(95%) 2.064

t score(96%) 2.171

t score (99%) 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

probability we got for 18 randomly selected bulbs would have an average life of no more than 260 days is 32.167