|  |  |
| --- | --- |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| 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 | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans3: Outcomes= {HHH, HHT, HTT, HTH, THT, THH, TTH, TTT}

Total outcomes=8

2H & 1T outcomes=3

Probability [2H & 1T] =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

Ans4:Outcomes= {(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)}

1. There are 36 outcomes and no outcome would be equal to 1
2. Sum of two dice with sum less than or equal to 4 are 6

Probability=6/36=1/6=0.167

1. Sum is divisible by 2 and 3 are 24

Probability= 24/36 =0.66

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?

Ans5: No. of outcomes = 7c2=7! / [2! \*(7-2)!] =21

Required outcomes = 5c2= 5! / [2! \*(5-2)!] =10

Probability= Required outcomes/No. of outcomes=10/21=0.476

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

Ans6: Expected number of candies for randomly selected child=

1\*0.015+4\*0.2+3\*0.65+5\*0.005+6\*0.01+2\*0.12=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**

**Ans7: For Points**: Mean= 3.596563; Median= 3.69; Variance= 0.285881;

Standard Deviation= 0.534679; Range= 2.17; Mode= 3.07;

**For Score**: Mean= 3.21725; Median= 3.215; Variance= 0.957379;

Standard Deviation= 0.978457; Range= 3.911; Mode= 3.44;

**For weigh**: Mean= 17.84875; Median= 17.6; Variance= 3.193166;

Standard Deviation= 1.786943; Range= 8.4; Mode= 17.02;

**Comment**: Mean, Median and Mode are measure of central tendency in data. Variance, Standard Deviation and Range are measure of variability in data with the help of both, calculated business decisions are taken. For example, from above values, we can conclude an average point that a car is getting is 3.596563, average score that a car is having is 3.21725 with all these values car dealer would decide which car to buy for his dealership to improve his sale.

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?

Ans8: Total weight=108+110+123+134+135+145+167+187+199=1308

1308/9=145.33

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**Ans(9\_a): Speed {Skewness:** [**-0.11750986144663393] it is negatively skewed**

**Kurtosis: [**-**0.5089944204057617] kurtosis is near to zero so it is a mesokurtic (normal) distribution curve}**

**Distance {Skewness:** [**0.8068949601674215] it is positively skewed**

**Kurtosis:** [**0.4050525816795765] kurtosis is near to the zero so it is a mesokurtic (normal) distribution curve}**

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans(9\_b): SP: {Skewness:** [1.6114501961773586] **it is positively skewed**

**Kurtosis:** [**2.9773289437871835] kurtosis is more than 2 so the curve is leptokurtic normal distribution curve}**

**Weight (WT): {Skewness:** [-0.6147533255357768] **it is negatively skewed**

**Kurtosis:** [**0.9502914910300326] kurtosis is near to the zero so it is a mesokurtic (normal) distribution curve}**

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



Ans10: Above histogram is positive skewed



Ans10: This boxplot has 7 outliers

**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?

Ans11:

#confidence interval-->94%= [198.74 201.26]

#confidence interval-->98%= [198.44 201.56]

#confidence interval-->96%= [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.
2. What can we say about the student marks?

Ans12: mean: 41, median: 40.5, variance: 24.11111111111111,

Standard deviation: 4.910306620885412

Average marks obtained by student is 41, with their median point being 40.5, with a deviation of 4.91 from the center and variance being squared for the deviation from the center.

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

Ans13: Mean=Median that means zero skewness

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

Ans14: Mean>Median is Positive Skewness

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

Ans15: Median>Mean is Negative Skewness

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

Ans16: Positive kurtosis value indicates peaked distribution of histogram curve

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

Ans17: Negative kurtosis suggests that flatter distribution of histogram curve

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



a.What can we say about the distribution of the data?

Ans18a: Data seems to distributed between 10 to 18, median being approx. 15.5

b.What is nature of skewness of the data?

Ans18b: Data seems to be Negative Skewed

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

Ans18c: IQR=8  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans19: 1st box plot data clustered, in 2nd box plot data spread thin compare to 1st boxplot. Both have same median values. Whisker length of 2nd boxplot is more compare to 1st boxplot.

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 20a: 34.7% of chance that mpg would be more than 38

* 1. P(MPG<40)

Ans 20b: 72.9% chance that mpg is less than 40

* 1. P (20<MPG<50)

Ans 20c:89.8% chance that mpg is between 20 and 50

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans21a: from the histogram graph we can see that data is normal distribution because it has is fairly symmetric bell curve.

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

Dataset: wc-at.csv

Ans21b: (Waist) seemed to be normal distribution; however, (AT) is positively skewed from the histogram plot.

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

Ans22:

#Z scores of 90% confidence interval= 1.28

#Z scores of 94% confidence interval=1.55

#Z scores of 60% confidence interval=0.25

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

Ans23: degrees of freedom=n-1=25-1=24

# t scores of 95% confidence interval=1.71

# t scores of 96% confidence interval=1.83

# t scores of 99% confidence interval=0.26

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

Ans24: mean=270days, sample=18=n, xbar=260days, sd=90days

#h0=mean>260

#h1=mean<=260

#mu=270days, sample=18=n,df=n-1=17, xbar=260days, sd=90days

# Ho = Avg life of Bulb >= 260 days

# H1 = Avg life of Bulb < 260 days

# Find t-scores at xbar=260; t=(s\_mean-P\_mean)/(s\_SD/sqrt(n))

t= (260-270)/ (90/18\*\*0.5)

t=-0.4714045207910317

pvalue= 0.321 since pvalue>0.05#ho is accepted and h1 is rejected