|  |  |
| --- | --- |
| 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 | Qualitative |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Qualitative |

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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

🡺Coin has two outcomes {H , T}

No of outcomes=2\*2\*2=8

Two heads and one tail outcomes = {HHT,HTH,THH}

Probability=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

🡺 a) 0

b) 1/6

c) 5/36

🡺a) Total outcomes=6\*6=36

Sum of two dice can never be a 1, so probability is 0

b) outcomes ={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6 outcomes

Probability = 6/36=1/6

c) Total number of outcomes= 36

Outcomes=sum is divisible by (2,3)

Sum should be divisible by (2,3)

Outcomes={(1,5),(3,3),(4,2),(5,1),(6,6)}

Outcomes=5

Probability=5/36

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?

🡺sum of numbers of balls=2+3+2=7

No of ways in which 2 balls can be drawn out 7=7C2=21

=7\*6/2\*1=21

No of red and green balls =5

No of ways in which 2 balls can be drawn from green and red balls=5=5C2=10

=5\*4/2\*1=10

So probability of not drawing a blue balls =probability of red/green bal

=10/21

=0.48

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

🡺 The expected number of candies=sum of (each candy\*probability)

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

=  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**

🡺mean:-

Points:- 3.596563

Score:- 3.217250

Weight:- 17.848750

🡺MEDIAN:-

POINTS:-3.695

SCORE:-3.325

WEIGHT:-17.710

🡺STANDARD DEVIATION:-

POINTS:- 0.534679

Score:- 0.978457

Weight:- 1.78694

RANGE🡺

POINTS:-[2.76,4.93]

SCORE:-[1.513,5.424]

WEIGHT:-[14.5,22.9]

Variance🡺

POINTS:-0.285881

SCORE:-0.957379

WEIGHT:-3.193166

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?

🡺Expected random value of weight is:-

145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

🡺SKEW OF SPEED :- ,-0.117510

SKEW OF DIST:- ,0.806895

KURTOSIS OF SPEED:-,-0.508994

KURTOSIS OF DIST:-,0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

🡺SKEW OF SP:-1.61140

SKEW OF WT:- ,-0.614753

KURTOSIS OF SP:-2.977329

KURTOSIS OF WT:-0.950291

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



🡺HISTOGRAM:

* RIGHT SKEWED DATA
* MEDIAN=200
* MODE=50-100



🡺BOXPLOT

* THERE ARE APPROX 7 OUTLIERS
* LEFT SKEWED

**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?

🡺94%=(198.73798,201.26201)

98%=(198.43902,201.56097)

96%=(198.62193,201.378065)

**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

Variance:-25.529411

Std:-5.0526638285864

1. What can we say about the student marks?

* Average marks is 41 & median is 40.5 students scored less 40.5 & some scored above 40.5

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

🡺No skewness

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

🡺Right skewness

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

🡺Left skewness

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

🡺distrubtion is peaked and possesses thick tails

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

🡺distrubtion has lighter tails than the normal distrubtion

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



What can we say about the distribution of the data?

🡺data range 1 to 19 & there is no outliers in the data

What is nature of skewness of the data?

🡺left skewed

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

🡺Q1=10

Q3=18

IQR=Q3-Q1

=18-10

=8

Approximately:-IQR=8

Q19) Comment on the below Boxplot visualizations?



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

🡺 \* both boxplot has a same median value

\* one boxplot has small range

\* one boxplot has wide range

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)

🡺 a.)0.3475939251582705

b.) 0.729349872151616

c.) 0.8988689169682046

.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

🡺 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

🡺 does not follow normal distribution

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

🡺 z-score of 90%=1.6448536269514722

z-score of 94%=1.8807936081512509

z-score of 60%=0.8416212335729143

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

🡺 t-score of 95%=2.0638985616280205

t-score of 96%= 2.796939504772804

t-score of 99%= 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

* Probability of 18 randomly selected bulbs with average life <260 is

0.6783