|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

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 |  | ratio |
| Blood Group |  | Nominal |
| Time Of Day |  | Interval |
| Time on a Clock with Hands |  | Interval |
| Number of Children |  | Nominal |
| 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?

{HHH, HHT,HTH,THH,THT,TTH,HTT,TTT}

=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

Total possible out comes=36

1. 0
2. {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

P(x<=4) =6/36=1/6

1. {(1,5),(3,3),(2,4),(4,2),(5,1),(6,6)}

=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?

Total balls=7,

=5c2/7c2

=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

Expected number of candies for randomly selected chaild

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=3.090

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.



Mean values***:*** *Points=3.596563, score=3.21725, weight= 17.84875*

Median values: *points= 3.695, score=3.325, weight= 17.71*

*mode values: points=3.92, score=3.44 , weight=17.02*

*standard deviation values*: *points=0.5346787, score=0.9784574, weight= 1.786943*

*variance values*: *points= 0.2858814, score= 0.957379, weight=3.193166*

*range values*: *points=2.76 4.93, score=1.513 5.424, weight= 14.5 22.9*

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 value=(108+110+134+135+145+167+187+199)/9=145.33

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

**Cars speed and distance**



**SP and Weight(WT)**



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



* Most of the data points are fall in the rage 50-100 with frequency 200.
* Long tail is towards right so it is right skewed or positive skewed



* Median is less than mean
* Outliers on the upper side of boxplot and there is less data points between Q1 and lower extreme

**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 ?

Sol: x±z1-α σ/sqrt(n)

At CI= 94%, Zcrit =1.881, [198.62, 201.26]

CI=96%, Zcrit =2.05, [198.62, 201.37]

CI=98%, Zcrit =2.33, [198.43, 201.56]

**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?

Mean=41, median=41, variance=24.111, sd=4.91

If mean=median=mode, so data is said to be symmetrical

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

Ans: symmetric

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: Data is normally distributed

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

Ans: Data has lighter tails and a flatter peak

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



What can we say about the distribution of the data?

nonsymmetric

What is nature of skewness of the data?

Negatively skewed

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

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: -comparing boxplot1 and boxplot2 have the same median value.

* IQR represent the length of the box. If we compare the IQR of two is greater then IQR of one

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)=33/81
  2. P(MPG<40)=61/81

c. P (20<MPG<50)=69/81

R code: MPG=sample(MPG)

A=subset(MPG, MPG>38)

B=subset(MPG, MPG<40)

C=subset(MPG, MPG>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

ANS: conducting shapiro test w=0.97797, p value=0.1764

Kurtosis value=-0.70546404

So, MPG data 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

Waist: shapiro test w= 0.95586, p value= 0.00117

Kurtosis value= -1.141846

AT: shapiro test w= 0.95234, p value= 0.000654

Kurtosis= -0.3760059

If p values <0.05, so wc-ac data not follow normal distribution

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

Z score of 90% confidence interval is 1.65

94% confidence interval is 1.88

60% confidence interval is 0.85

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% confidence interval is 1.96

96% confidence interval is 2.5

99% confidence interval is 2.47

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

sol: population mean=270

sample mean=260

sample size=18

sd=90

t score=-0.47, pt value is 0.3221639 i.e. 32%