|  |  |
| --- | --- |
| 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 | Categorical/nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical /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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| 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 | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Nominal |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Sol: total no of combinations are possible=2\*2\*2=8

p(H H T)+P(H T H)+P(T H H)

=1/8+1/8+1/8

=3/8

Binomial distribution

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

Sol:

a)There is no outcomes which corresponds sum is equal to one.

i.e.0/36.probability is 0.

b)6/36=1/6 = 0.167

c) (1,3) (2,2) (3,1)= 3outcomes,5/36 =0.139

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?

SOL: = p(2R,3G,2B)

Total number of balls=2+3+2=7

Then , n(s)= no of ways of drawing 2balls out of 7

7c2=7\*6/2\*1=42/2=21

E=Event of drawing 2balls,none of which is blue

n(e)=no of ways of drawing 2 balls out of (2+3) balls

5c2=5\*4/2\*1=20/2=10

Probability p( E)=n(e)/n(s)=10/21

Ans: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

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

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

=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**

Sol: **points**

Mean=F(X)/N=3.5965625

Median =N+1/2=3.695

Mode=Highest repeated value=3.92,3.07(by model)

Variance=0.2858814

Standard Deviation=0.5346787

Range=Max value-Min value=4.93-2.76=2.17

**Score**

Mean=3.21725

Median=3.325

Mode=3.44

Variance=0.957379

Standard Deviation=0.9784574

Range=3.91

**Weight**

Mean=17.84875

Median=17.71

Mode=17.02

Variance=3.193166

Standard Deviation=1.786943

Range=8.40

**Note:** No case has the variable Mean = Median = Mode

Thus as seen in the graph dataset “score” and “Weigh” has outliers

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?

Sol: mean=1308

Expected value=p(x)/x

=1308/9

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | -0.11 | 2.42 |
| Dist | 0.76 | 3.25 |

* “dist” is positively skewed where as “speed” is negatively skewed

Thus dist has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the graph

Both dist and speed has positive Kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.55 | 5.72 |
| WT | -0.59 | 3.82 |

* “SP” is positively skewed where as “WT” is negatively skewed
* Thus SP has distribution of data concentrated on the left whereas WT has distribution on the right. As seen in the graph
* Both WT and SP has positive Kurtosis

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



ANS :

1. Majority of the Chicks has weight in range 50 – 100, followed by 100 -150 and 150 – 200
2. The data is positively Skewed
3. 3 Categories can be define (Under weight >50; Avg weight 51 – 150; Over weight <150)



Sol: Median is less than mean right skewed and we have outlier on the upper side of the boxplot and there is less data between

Q1 and bottom point. data is positively 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?

**Sol:** X+/-(Z1- α. σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

|  |  |  |  |
| --- | --- | --- | --- |
| Confidence interval | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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?

Sol: Mean= 41, Median= 40.50, variance= 25.53, Standard deviation= 5.05

2.all most students marks between 34-50

Ans:

1. Not normally distributed
2. Data has outlier
3. Majority of the students scored between 35 – 45 Marks

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

Sol: Symmetrical

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

Sol: left skewed: positively

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

Sol: right skewed : negatively

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

Sol: High and narrow peak on central part of the data

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

Sol: wider peak on central part of the data

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



What can we say about the distribution of the data?

Sol: The data is not symmetric. Data is more concentrated towards right side

What is nature of skewness of the data?

Sol: Negatively skewed

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

Sol: IQR of the data18 – 10 =8  
  
Q19) Comment on the below Boxplot visualizations?



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

Sol:

1. Data is Normally Distributed. No Outliers. Center around 262.5. Comparatively, first graph has less range
2. Data is Normally Distributed. No Outliers. Center around 262.5

Comparatively, second graph has more 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)
  3. P (20<MPG<50)

By using command dplyr in ‘R’

1. a=subset P(MPG>38)

show(a)

There are 33 observations in MPG which are greater than 81

= 0.4074074

b) b= subset P(MPG<40)

show(b)

67 observations in MPG which are lesser than 81

= 0.7530864

c) c=subset P (20<MPG<50)

show(c)

69 observations in mpg which are lesser than 81

= 0.8518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol: MPG is 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

SOL: Both AT and Waist doesn’t follow Normal Distribution

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

Sol: Z scores of 90% confidence interval is +1.711,-1.711.

Z score of 94% confidence interval is +1.828, -1.828.

Z scores of 60% confidence interval is+2.492, -2.492.

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

Sol: T Score of 95% +2.060, -2.060.

T score of 96% +2.167, -2.167.

T score of 99% +2.787, -2.787.

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: t-score = -0.4714,  
Degree of freedom = 17   
P(T) = 0.3216725