|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Continuous |
| 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 | Ratio |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Interval\* |
| Level of Agreement | Nominal\* |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal\* |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

ANS. Sample space { HHH , HHT, HTH, THH, TTH, THT, HTT, TTT }

P(2H & 1T) = 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

ANS. Sample space

[(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. Equal to 1 = 0/36
2. Less than or equal to 4 will be

(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)

6/36 = 1/6

1. Sum is divisible by 2 and 3 will be

[ (1,1) (1,2) (1,3) (1,5)

(2,1) (2,2) (2,4) (2,6)

(3,1) (3,3) (3,5) (3,6)

(4,2) (4,4) (4,5) (4,6)

(5,1) (5,3) (5,4) (5,5)

(6,2) (6,3) (6,4) (6,6) ]

24/36 = 2/3

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?

ANS.

Total no of balls 7

two balls drawn 2

so 7C2 is 21

No blue ball should drawn is 2

except blue total no of ball is 5

so 5C2 is 10

Probability is 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

ANS.

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count (X) | Probability P(X) | (X)xP(X) |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.20 | 0.80 |
| 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 |  |  | 3.09 |

EXPECTED VALUE = 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.



ANS

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Meadian | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| STD | 0.534679 | 0.978457 | 1.786943 |
| Max | 4.93 | 5.424 | 22.9 |
| Min | 2.76 | 1.513 | 14.5 |
| Range | 2.17 | 3.729 | 8.4 |

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?

ANS.

we chosen one patient randomly

so prob of each patient chosen randomly = 1/9 = 0.11

|  |  |  |
| --- | --- | --- |
| X | P(X) | XxP(X) |
| 108 | 0.11 | 11.88 |
| 110 | 0.11 | 12.1 |
| 123 | 0.11 | 13.53 |
| 134 | 0.11 | 14.74 |
| 135 | 0.11 | 14.85 |
| 145 | 0.11 | 15.95 |
| 167 | 0.11 | 18.37 |
| 187 | 0.11 | 20.57 |
| 199 | 0.11 | 21.89 |
| TOTAL |  | 143.88 |

**Q9)**

1. **Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

ANS.

SPEED DISTANCE

|  |  |  |
| --- | --- | --- |
| Skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

As we know about skewness higher the magnitude longer will be the tail and +ve or -ve tell position of tail right and left

Here the magnitude of distance data is slightly higher than the speed data hence the tail of distance data also will be slightly longer than the speed data and position of tails will be opposite to each other.

**INFERENCES OF SPEED DATA**

|  |  |
| --- | --- |
| SKEWNESS | KURTOSIS |
| Here the skewness value of speed data is -0.11751. A negative value which indicates negative skewness and left skewed and tail is on the left side or mass of the distribution concentrated on right side.  Also here  Mean<Median<Mode | Here the kurtosis value of speed data is -0.50899. A negative value which indicates negative kurtosis which implies wider peak and thinner tail. |

**INFERENCES OF DISTANCE DATA**

|  |  |
| --- | --- |
| SKEWNESS | KURTOSIS |
| Here the skewness value of distance data is 0.806895. A positive value which indicates positive skewness and right skewed and tail is on the right side or mass of the distribution concentrated on left side.  Also here  Mean>Median>Mode | Here the kurtosis value of speed data is 0.405053 . A positive value which indicates positive kurtosis which implies narrow peak and thicker tail. |

|  |  |
| --- | --- |
|  |  |



1. **SP and Weight(WT)**



ANS.

SP WEIGHT

|  |  |  |
| --- | --- | --- |
| Skewness | 1.61145 | -0.61475 |
| Kurtosis | 2.977329 | 0.950291 |

As we know about skewness higher the magnitude longer will be the tail and +ve or -ve tell position of tail right and left

Here the magnitude of SP data is higher than the WEIGHT data hence the tail of SP data will be longer than the WEIGHT data and position of tails will be opposite to each other.

**INFERENCES OF SP DATA**

|  |  |  |
| --- | --- | --- |
| SKEWNESS | KURTOSIS |  |
| Here the skewness value of distance data is 1.61145 . A positive value which indicates positive skewness and right skewed and tail is on the right side or mass of the distribution concentrated on left side.  Also here  Mean>Median>Mode | Here the kurtosis value of speed data is 2.977329 . A positive value which indicates positive kurtosis which implies narrow peak and thicker tail |  |

**INFERENCES OF WEIGHT DATA**

|  |  |  |
| --- | --- | --- |
| SKEWNESS | KURTOSIS |  |
| Here the skewness value of speed data is -0.61475 . A negative value which indicates negative skewness and left skewed and tail is on the left side or mass of the distribution concentrated on right side.  Also here  Mean<Median<Mode | Here the kurtosis value of speed data is 0.950291 . A positive value which indicates positive kurtosis which implies narrow peak and thicker tail |  |

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



**ANS.**

**Inferences about above histogram**

The above histogram the bin size is 50 where X-axis describe chick weight and Y-axis describe the frequency also the frequency of chick weight is max at 50 to 100. The Data is not Normally Distributed .Apart from this the histogram have tail on right and concentration of mass is on left side which indicated a clearly right skewed histogram and the tail is not much longer hence the magnitude would be not higher value.

**BOX PLOT**



ANS.



Here when we rotate the box plot by 90 degree clockwise we have seen a right skewed graph with a long tail and Multiple outliers. The dispersion of the box plot is not high bcoz the IQR of the box plot is not much stretched and in the box the median is not in the center which implies Q1-Q2 is not equal to Q3-Q2. Data id not Normal.

**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 ?

ANS.

We have

Population = 3,000,000

Sample size n = 2,000

√n =44.72

Sample Mean ͞x = 200

Std dev (σ) = 30

σ/√n=30/44.72 =0.67

**µ = ͞x ± Z(1-α)****σ /√n**

1. Confidence interval for Z(1-α) = 94%

From T table the value of Z94% =1.88

So

= 200 ± 1.88 x 0.67 = [200 ± 1.25]

**The interval will be = [198.5 201.25] for 94%confidence interval**

1. Confidence interval for Z(1-α)  = 96%

From T table the value of Z96% =2.05

So

=200 ± 2.05 x 0.67 = [200 ± 1.37]

**The interval will be = [198.63 201.37] for 96%confidence interval**

1. Confidence interval for Z(1-α)  = 98%

From T table the value of T98%=2.33

So

=200 ± 2.33 x 0.67 = [200 ± 1.56]

**The interval will be = [198.44 201.56] for 98% confidence interval**

**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**

**ANS.**

1. Find mean, median, variance, standard deviation.

ANS.

|  |  |
| --- | --- |
| mean | 41.17647 |
| median | 41 |
| variance | 26.52941 |
| STD | 5.150671 |

1. What can we say about the student marks?

ANS.

There is very less deviation in student mars most of them have average marks or near by average marks.

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

ANS.

The skewness is zero and the distribution of the graph is normal because in normal distribution mean = median = mode.

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

ANS.

The nature of the skewness is “positive” and graph will be right skewed

Coz in right skewed mean > median >mode.

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

ANS.

The nature of the skewness is “negative” and the graph will be left skewed

Coz in left skewed mean < median< mode or median >mean.

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

ANS.

A positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

ANS.

A negative kurtosis value indicates the distribution of the data has lighter tails and flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

* The data of above box plot is decent dispersed or less consistent as the box has a wider width stretch

What is nature of skewness of the data?

* Left skewed as the tail is on left side and the contraction of mass distribution is on right side.

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

* The IQR = Q3  - Q1

And here Q1 =10(app.)  
 and Q3 = 18(app.)

IQR = 18 -10

=8(app.)

Q19) Comment on the below Boxplot visualizations?



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

ANS.



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)

ANS. Here we have mean = 34.42 & SD = 9.13

1. P(MPG>38) =

Z = x-µ/σ = 38-34.42/9.13

=3.58/9.13=0.39

Z score for 0.39 = 0.65

The probability of P(MPG>38) = 1-0.65 =0.35 = 35%

And

pnorm(38,34.42,9.13,F) =0.347 = 34.7%

1. P(MPG<40) =

Z = x-µ/σ =40-34.42/9.13

=5.58/9.13=0.61

Z score for 0.61=0.725

The probability of P(MPG<40)= 72%

And

pnorm(40,34.42,9.13,T)= 0.72= 72%

1. P (20<MPG<50)= P(MPG<50) – P(MPG<20)

P (MPG<20)=

Z = x-µ/σ = 20-34.42/9.13

=-14.42/9.13= -1.57

Z score for -1.57= 0.05821

And

pnorm(20,34.42,9.13,T)=0.057

Lly

P (MPG<50) =

Z = x-µ/σ = 50-34.42/9.13

=15.58/9.13= 1.70

Z score for 1.70= 0.9554

And

pnorm(50,34.42,9.13,T)=0.95

So, P (20<MPG<50) = P(MPG<50) – P(MPG<20)

= 0.9554-0.05821 =0.89 = 89%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS. Yes it is 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

ANS. Data set waist follows almost normal distribution bu not completely

Where as AT a good normal distribution.

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

ANS.

Z scores of 90% confidence interval

qnorm(0.95)=1.64

Z scores of 94% confidence interval

qnorm(0.97)=1.88

Z scores of 60% confidence interval

qnorm(0.80)=0.841

**Sol :2**

**Z Scores**

**90% = 1.28**

**94% = 1.55**

**60% = 0.25**

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

ANS.

Here we have (n-1) =24

t scores of 95% confidence interval

qt(0.975,24)=2.06

t scores of 96% confidence interval

qt(0.98,24) = 2.17

t scores of 99% confidence interval

qt(0.995,24) = 2.79

**Sol :2**

**T Scores**

**95% = 1.71**

**96% = 1.82**

**99% = 2.49**

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

ANS.

We have

Pop mean = 270

Sample size (n) = 18

Sample mean = 26

Sample SD = 90

Tscore = -0.4714

df = 17

pt(-0.4714,17) = 0.3216