|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 | Nominal |
| 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?

Solution:

Total possible outcome - HHH, HHT, HTH, THH, HTT, TTH, TTH, TTT

P(2Head,1Tail)=3/8 =0.375.

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1. No probability like that
2. Less than or equal to 4
3. Sum is divisible by 2 and 3.

Solution:

1. Total Possibilities are =36

P(sum=1) = 0/36=0

1. n(<=4) = (1,1)(1,2)(1,3)(2,1)(2,2),(3,1) = 6

P(<=4) = 6/36 = 1/6 =0.17

1. n(sum is divisible by 2 & 3) = (1,1)(2,4)(3,3)(4,2)(5,1)(6,6)=6

p(Sum is divisible by 2&3) = 6/36 = 1/6 =0.17

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?

Solution:

Total ball is 2 red +3 green +2 blue= 7 ball.  
It can be written as 7C2,  
So None of the ball is blue that is 7- 2 (blue ball) is = 5 ball.

It can be written as 5C2

Probability = 5C2/7C2 = (5\*4/1\*2) / (7\*6/1\*2)

= 20/42

=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

Solution:

Expected No of Candies for randomly selected child

=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.

Solution: 

**Use Q7.csv file**

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?

Solution: 9 patient, 9 observations, randomly chosen one patient out of 9= 1/9

Expected value is =1/9\*[108+110+123+134+135+145+167+187+199]

=1308/9 = 145.33

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

**Cars speed and distance**

**Solution:**

Skewness for speed = -0.1105533, value in negative so it is left skewed.

Skewness for distance = 0.7591268, value in positive so it is right skewed.

Kurtosis for speed = 2.422853, Negative kurtosis because value <3.

Kurtosis for Distance = 3.248019, Positive kurtosis because value >3 also more data in tails than peak

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Solution:**

Skewness for SP = 1.552258, value in positive so it is right skewed.

Skewness for WT = -0.5921721, value in negative so it is left skewed.

Kurtosis for SP = 5.723521,Positive kurtosis because value >3.

Kurtosis for WT = 3.819466, Positive kurtosis because value >3.

**Use Q9\_b.csv**

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



Solution: The highest data point in the range of 50-100 with frequency 200

Lowest data point lies between 350-400 with frequenct range of 0-10

As noticed long tail is distributed in right side so it is right side skewed, so it is positive skewness.



Solution: Here we found the presence of more outliers after upper extreme point.

Median value is less than mean value since right skewed, also there is less data points in lower side than upper side

**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?

Solution: n=2000, degrees of freedom = 2000-1=1999

Mean X=200, Standard deviation of Sample =30

Confidence interval 94%=1.88, Interval Estimate = (134.85, 265.1492)

Confidence interval 96%=2.055, Interval Estimate = (130.153, 269.8464)

Confidence interval 98%=2.32, Interval Estimate = (122.651, 277.3487)

**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.

Solution : Mean = 41, Median=40.5 , Variance = 25.52941, Standard Deviation = 5.052664.

1. What can we say about the student marks?

Solution:

When try to plot boxplot for this found two outliers 49 and 56, there is no huge variance in his marks maintaining his marks average of 41.

Mean>median so skewness is right skewed.

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

Perfectly symmetrical, skewness is 0

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

Right skewed data,

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

Left skewed data

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

Sharper peak, flatter tail- kurtosis value >=3

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

Flatter peak thinner tail – kurtosis value <3

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



What can we say about the distribution of the data?

Distribution is left skewed, median>mean, 50% of the people are above 10 and no outliers here.

What is nature of skewness of the data?

Left skewed, median>mean, skewness is -ve (<0)

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

IQR=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.

Solution: By observing both, distribution of Boxplot 2 is symmetrical means mean and median are equal. When compare to boxplot 1 with 2, boxplot 2 has higher whisker size also boxplot2 follows normal distribution as distributed symmetrically.

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)

1-pnorm(38,mean(Cars$MPG),sd(Cars$MPG))

=0.3475939

* 1. P(MPG<40)

pnorm(40,mean(Cars$MPG),sd(Cars$MPG)) = 0.7293499

* 1. P (20<MPG<50) = pnorm(50,mean(Cars$MPG),sd(Cars$MPG))-pnorm(20,mean(Cars$MPG),sd(Cars$MPG))

= 0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution. No its not following Normal distribution Mean (34.4) < median(35.15), also skewness is -0.17 negative so left skewed not distributed symmetrical. Also plotted boxplot not distributed equally.

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

Dataset: wc-at.csv

Solution: Data: AT

No, not following normal Distribution. Mean = 101.894, median= 96.54, mean>median right skewed, skewness = 0.57, positive right skewed.

Data: Waist, Mean = 91.90183, median= 90.8, Skewness = 0.132, mean>median, skewness also positive right skewed not following normal distribution.

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

Solution: Z score for 90% = 1.65, 94% = 1.88, 60% = 0.842

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

Solution: T score 95% = 2.06, 96% = 2.17, 99% = 2.7969

stats.t.ppf(0.975,24)

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

Solution: n=18 , mean x=260, Standard Deviation s=90, Population mean= 270, p(x<=260)=?

t-score = (260-270)/(90/sqr.root(18))

=-10/21.21

=-0.471

Pt(-0.471,18) = 0.321