|  |  |
| --- | --- |
| 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 | Discrete data |
| Distance between two places | Discrete 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Ans: number of possible outcomes = 23

Number of favorable outcomes = 3

Hence probability = 3/8

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

Ans: number of possible outcomes = 36

1. Equal to 1 = zero, because when we roll two dice the minimum outcome is 2
2. Less than or equal to 4 = 6/36 = 1/6, because possible outcomes are (1,1)(1,2)(1,3)(2,1)(2,2)(3,1)
3. Sum is divisible by 2 and 3 = 5/36, because sum is divisible by 2 and 3 i.e. divisible by 6 and possible outcome for this case are (1,5)(2,4)(3,3)(4,2)(5,1)

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: 10/21

Total no of balls = 7

No of balls without containing blue balls = 7-2 = 5

Therefore, Probability for 1st ball is not blue = 5/7

Now, total no of balls remain = 7-1 = 6

No of balls without containing blue balls = 6-2 = 4

Therefore, Probability for 2nd ball is not blue = 4/6

Thus, the probability that none of the balls drawn is blue = (5/7) \* (4/6) = 20/42 = 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: 3.09

Expected number of candies for a randomly selected child

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

=0.015+0.80+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

Ans: in excel:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| mean | 3.596563 | 3.21725 | 17.84875 |
| median | 3.695 | 3.325 | 17.71 |
| mode | 3.92 | 3.44 | 17.02 |
| variance | 0.285881 | 0.957379 | 3.193166 |
| standard deviation | 0.534679 | 0.978457 | 1.786943 |
| max | 4.93 | 5.424 | 22.9 |
| min | 2.76 | 1.513 | 14.5 |
| range | 2.17 | 3.911 | 8.4 |

also Comment about the values/ Draw some inferences.

**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?

Ans: 145.33

Expected value = (probability \* value)

=

There are 9 patients therefore, probability of selecting each patient is 9/81=1/9

Expected value = (1/9)(108)+(1/9)(110)+(1/9)(123)+(1/9)(134)+(1/9)(135)+(1/9)(145)+(1/9)(167)+(1/9)(187)+(1/9)(199)

=(1/9)\*1308 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans: In google colaboratory

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



Ans: The most frequent chick weight is 50-100. Chick weight frequency 0-50 is on left side of the histogram and 100-400 is on the right side. As the most of the data falls to the right of the graph's peak

Outliers

Upper extreme

1.5 IQR

Upper Quartile (75%) Q3

Median (50%) Q2

Lower Quartile (25%) Q1

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?

Ans: Using t-distribution

* 94% confidence interval is (198.73,201.27)
* 96% confidence interval is (198.61,201.39)
* 98% confidence interval is (198.43,201.57)

Given data:

Sample mean of = 200

Sample standard deviation of s = 30

Sample size of n = 2000

The interval is:

In which t is the critical value for the two-tailed confidence interval

* Considering 94% confidence level using a calculator, with 200-1=199 df, the critical value is t = 1.8916, hence:

= 200-1.8916 =198.73

= 200+1.8916 =201.27

The 94% confidence interval is (198.73,201.27)

* Considering 96% confidence level using a calculator, with 200-1=199 df, the critical value is t = 2.0673, hence:

= 200-2.0673 =198.61

= 200+2.0673 =201.39

The 96% confidence interval is (198.61,201.39)

* Considering a 98% confidence level, using a calculator, with 200-1 = 199 df, the critical value is t = 2.3452, hence:

= 200-2.3452 =198.43

= 200+2.3452 =201.57

The 98% confidence interval is (198.43,201.57)

**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.

Ans: Mean = 41

Median = 40.5

Variance = (49+25+25+9+9+4+4+1+1+0+0+0+0+1+1+16+64+225)/(18-1)

= 434/17 = 24.11

Standard deviation = 4.91

1. What can we say about the student marks?

Ans: If we calculate the coefficient of variation (cv), it’s 11.98% i.e. there is less variance in data.

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

Ans: it means the data is equally distributed on left and right side of median, (neither positive nor negative)

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

Ans: It means the majority of data/observations are right to the median i.e. it’s positive

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

Ans: It means the majority of data/observations are left to the median i.e. it’s negative

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

Ans: it indicate narrow peak and narrow tail.

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

Ans: it indicate thick peak and thick tail

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



What can we say about the distribution of the data?

Ans: the data is more distributed to the left side.

What is nature of skewness of the data?

Ans: the data is left skewed.

What will be the IQR of the data (approximately)?   
Ans: whisker = 1.5\*IQR

Therefore,

IQR = whisker/1.5 = 10/1.5 = 6.67

Q19) Comment on the below Boxplot visualizations?



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

Ans:

for boxplot 1: (approx.)

Q2 = mean = 268.75

Q1 = 256.25

Q3 = 281.25

IQR = Q3 – Q1 = 281.25 – 256.25 = 25

Lower extreme = 243.75

Upper extreme = 287.5

for boxplot 2: (approx.)

Q2 = mean = 268.75

Q1 = 225

Q3 = 312

IQR = Q3 – Q1 = 312 – 225 = 87

Lower extreme = 193.75

Upper extreme = 337.5

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: in google colaboratory

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

Ans: in google colaboratory

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

Ans: z= 1.555. therefore, we get that the z score at 90% confidence interval is 1.645, 94% confidence interval is 1.555, 60% confidence interval is 0.253

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

Ans:

M=(2+3+5+6+9)/5 = 5

N -1 = 25-1=24

Confidence level from t table

95% = 1.711

96% = approximately between 1.711 to 2.064

99% = 2.492

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: t-statistics for the data is given as

X = mean of sample bulbs = 260

= population mean = 270

s = standard deviation of the sample = 90

n = number of item in the sample = 18

t = (260-270)/(90/)

t = -0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.