|  |  |
| --- | --- |
| 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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ordinal |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans: Probabilities : HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

Now in the case we are getting 3 outcomes that meet our condition, HHT, HTH,THH.

So the probability is 3/8 that is 0.375

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: When two dice are rolled, 36 possible outcomes we will get.

1. The probability of getting sum equal to 1 is zero.
2. The possible combinations whose sum are less than or equal to 4 is

(1,1) → 2 , (1,2) →3, (1,3) →4,(2,1) →3,(2,2) →4,(3,1) →4. So the probability is 6/36 that is 1/6

1. We need to find all the combinations that result in a sum of 2, 4, 6, 8, 10, or 12.

Possible combinations and their sums:

(1, 1) → 2, (1, 2) → 3, (1, 3) → 4, (1, 4) → 5, (1, 5) → 6, (1, 6) → 7,

(2, 1) → 3, (2, 2) → 4, (2, 3) → 5, (2, 4) → 6, (2, 5) → 7, (2, 6) → 8,

(3, 1) → 4, (3, 2) → 5, (3, 3) → 6, (3, 4) → 7, (3, 5) → 8, (3, 6) → 9,

(4, 1) → 5, (4, 2) → 6, (4, 3) → 7, (4, 4) → 8, (4, 5) → 9, (4, 6) → 10,

(5, 1) → 6, (5, 2) → 7, (5, 3) → 8, (5, 4) → 9, (5, 5) → 10, (5, 6) → 11,

(6, 1) → 7, (6, 2) → 8, (6, 3) → 9, (6, 4) → 10, (6, 5) → 11, (6, 6) → 12.

So the probability is 18/16 that is 1/2

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 balls = 2 red + 3 green + 2 blue = 7 balls

It can be written as 7C2, so the none of the ball is blue that is 7-2 =5 balls or 5C2

So 5C2/7C2 = (5!/3!\*2!) **// (7!/5!\*2!) = 10/21**

So the 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 | 2to | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Ans:**  Expected number of candies for a randomly selected child  
= 1\*0.05 + 4\*0.20 + 3\*0.65 +5\*0.005 + 6\*0.01 + 2\*0.120

=3.09

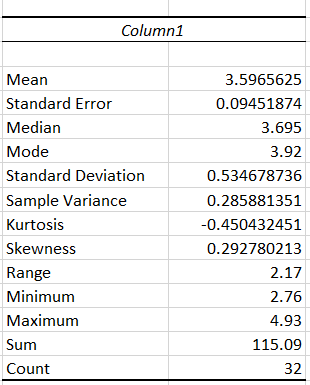
3.09 is the expected number of candies for a randomly selected child.

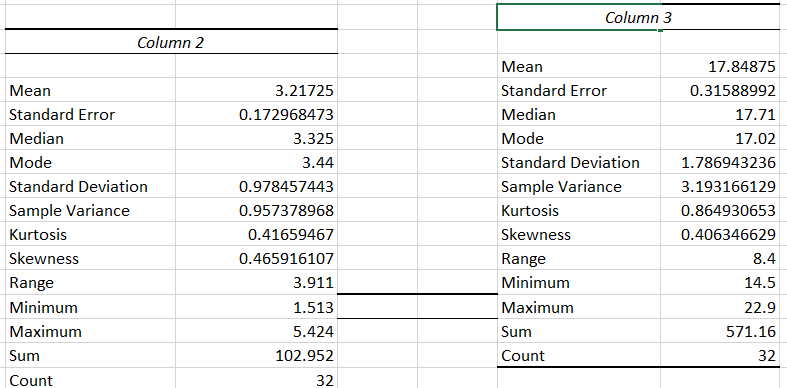
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

****



Colum 1 has data for points, column 2 has data for score and column 3 has data for weigh.

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 : Expected Value (mean) = sum of all the weight/ no of patients

= 1098/9

= 122

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:** Speed:

* Skewness: -0.11751 indicating a mild left-skewed distribution.
* Kurtosis: -0.508994 which is less than 3, indicating a platykurtic (light-tailed) distribution.

These values suggest that the Speed has a distribution that is slightly skewed to the left and has lighter tails compared to a normal distribution.

Distance:

* Skewness: 0.806895 indicating a right-skewed distribution.
* Kurtosis: 0.405053 which is less than 3, indicating a platykurtic (light-tailed) distribution.

These values indicate that the Distance has a distribution that is slightly skewed to the right and has lighter tails compared to a normal distribution.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sp**

* Skewness: 1.61145 indicating a significantly right-skewed distribution.
* Kurtosis: 2.977329 which is almost equal to 3, indicating a Mesokurtic distribution.
* **Weight:**
* Skewness: -0.614753 indicating a mild left-skewed distribution.
* Kurtosis: 0.950291 which is less than 3 but close to 3, suggesting a distribution that is closer to a normal distribution in terms of kurtosis but still slightly platykurtic (light-tailed).

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



Ans: This histogram is Right-Skewed or positively skewed, we can say that there are some outliers present in the data in right hand side.



By looking at the box plot we can say there are outliers present in 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?

Ans: here we have sample mean =200 pounds

Sample size = 2000, population is 3,000,000, sd = 30

**Confidence interval for 94%:**

x̄ ± zα/2(σ/√n)

CI =(198.6851, 201.314)

**Confidence interval for 98%:**

CI =(198.436, 201.563)

**Confidence interval for 96%:**

CI = (198.624, 201.375)

**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?

Ans:

* The mean score is 41.
* The median score is 41, which is very close to the mean, indicating that the data is relatively symmetrically distributed.
* The variance is **25.529412**, indicating a moderate amount of variation in the scores.
* The standard deviation is approximately 5.0526638, which quantifies the amount of variation in the scores.

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

Ans: If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness.

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

Ans: If the mean is greater than the median, the distribution is positively skewed.

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

Ans:If the mean is less than the median, the distribution is negatively skewed.

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

Ans: A positive kurtosis value indicates that a dataset has heavier tails and a more peaked central peak compared to a normal distribution.

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

And: Negative kurtosis indicates that a dataset has lighter tails and is less peaked than a normal distribution

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

Ans: Data is left skewness or negative skewness

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

Ans: IQR = Q3 – Q1 = 18 – 10 = 8

Q19) Comment on the below Boxplot visualizations?



Ams: 1] Here boxplot is symmetric, as median is exact at the equidistance from the Q3 and Q1. There are no outliers are present in 1st boxplot. 50% of the data is ranging from [250 – 275].

2] Here boxplot is symmetric, as median is exact at the equidistance from the Q3 and Q1. There are no outliers are present in 2nd boxplot. 50% of the data is ranging from [225 – 300].

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

Ans) Median is same for both of the boxplot. Boxplot 1 is symmetric to boxplot 2.

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)

c. P (20<MPG<50)

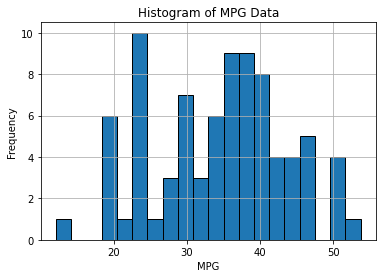
Ans:

1. To calculate the probability that the MPG is greater than 38, we have to count the number of cars with MPG greater than 38 and divide it by the total number of cars in the dataset. Which gives 0.3950(39.50%)
2. To calculate the probability that the MPG is less than 40, we have to count the number of cars with MPG less than 40 and divide it by the total number of cars in the dataset. Which gives 0.7530(75.30%)
3. o calculate the probability that the MPG falls between 20 and 50, we have to count the number of cars with MPG in that range and divide it by the total number of cars in the dataset. Which gives 0.8518(85.18%)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Ans: . It appears to be somewhat bell-shaped, but it's not a perfect bell curve. Therefore MPG does not follow 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

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

Ans:  For 90% confidence interval:

We have the significance level at 5 % ( if it is a two tailed test)

that is:

α = 5 % = 0.05

z at α = 0.05 from the z table will be:

z = 1.645.

For 94 % confidence interval, we get:

We have the significance level at 3 % ( if it is a two tailed test)

that is:

α = 3 % = 0.03

z at α = 0.03 from the z table will be:

z = 1.555.

For 60 % confidence interval, we get:

We have the significance level at 20 % ( if it is a two tailed test)

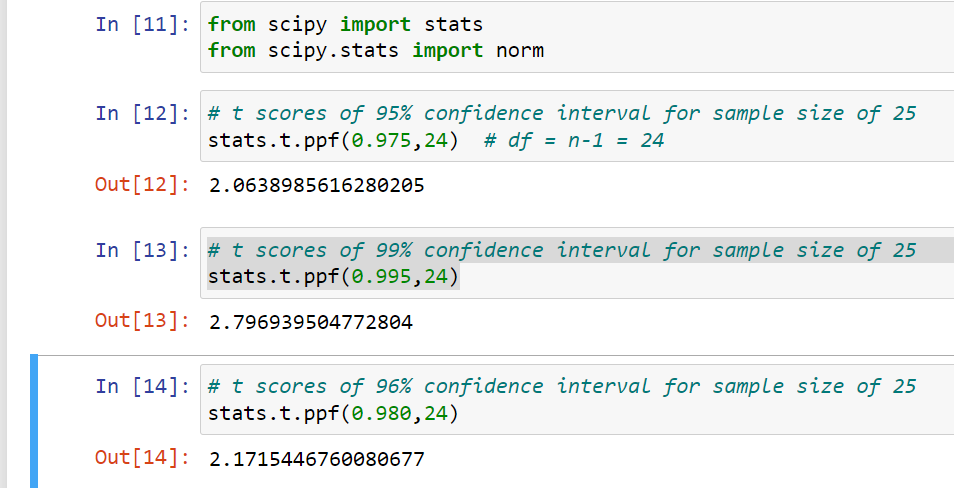
that is:

α =20 % = 0.2

z at α = 0.2 from the z table will be:

z = 0.253

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



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:

* Null Hypothesis (H0): The CEO's claim is true, and the average bulb life is 270 days.
* Alternative Hypothesis (H1): The CEO's claim is not true, and the average bulb life is not 270 days.

H0: μ = 270 H1: μ ≠ 270 (two-tailed test)

* σ (standard deviation of the population) = 90 days
* n (sample size) = 18 bulbs

now to find z value:

z = (X - μ) /σ / sqrt(n)

= (260-270)sqrt(18)/90

= Z-Score: -0.47140

Let choose significance level is 0.05,

If p < alpha(SI) = reject the null hypothesis

If p> alpha(SI) = fail to reject the null hypothesis.

P-Value: 0.63735

Here p > alpha(SI)

* So we fail to reject the null hypothesis. The CEO's claim is true, and the average bulb life is 270 days.