Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Nominal data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Nominal data |

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal data |
| High School Class Ranking | Ordinal data |
| Celsius Temperature | Interval Data |
| Weight | Ratio data |
| Hair Color | Nominal data |
| Socioeconomic Status | Ordinal data |
| Fahrenheit Temperature | Interval data |
| Height | Ratio data |
| Type of living accommodation | Nominal data |
| Level of Agreement | Ordinal data |
| IQ (Intelligence Scale) | Ordinal data |
| Sales Figures | Ratio data |
| Blood Group | Nominal data |
| Time Of Day | Interval data |
| Time on a Clock with Hands | Interval data |
| Number of Children | Ratio data |
| Religious Preference | Nominal data |
| Barometer Pressure | Interval data |
| SAT Scores | Interval data |
| Years of Education | Ordinal data |

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

Ans: The probability of getting two heads and one tail is 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: a) the probability of getting sum “equal to 1” is 0.

b) the probability of getting sum “less than or equal to 4” is 6/36 i.e., 1/6.

c) the probability of getting “sum is divisible by 2 and 3” is 1/6.

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 number of balls = (2 + 3 + 2) = 7. Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7 = 7C2​ = 21

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

n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls = 5C2​ = 10

∴P(E)= n(E)​/n(S)=10/21

∴The probability that none of the balls drawn is blue 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: Expected number of candies for a randomly selected child = ∑XiPi

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

= 3.09

Expected number of candies for a randomly selected child is 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**

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Point** | **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 |
| **SD** | 0.534679 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 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: The expected weight of the randomly selected patient is average of given weight. i.e., ∑ (weight of patients)/no of patient.So, the answer is 145.33.

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

**Car’s speed and distance Use Q9\_a.csv**

**SP and Weight (WT) Use Q9\_b.csv**

Ans:

Skewness: -0.1140 (Car’s speed) and 0.7825 (distance)

Kurtosis: -0.5771 (Car’s speed) and 0.2480 (distance)

Skewness: 1.5814 (SP) and -0.6033 (Weight)

Kurtosis: 2.7235 (SP) and 0.8194 (Weight)

Interpretation:

In cars speed and distance data, cars speed is negatively skewed and the value of kurtosis is negative i.e., it is less peaked than normal curve therefore the curve is platykurtic. while distance is positively skewed and it is more peaked than normal curve hence it is leptokurtic curve.

For SP and Weight data, SP is positively skewed and the value of kurtosis is negative i.e., it is less peaked than normal curve therefore the curve is platykurtic. while WT is positively skewed and its kurtosis value is positive therefore it is more peaked than normal curve hence it is leptokurtic curve.

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





Ans: The given histogram has right skewed and tail is on right side. So it means Mean> Median.

The boxplot has outliers on the 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:

Sample size = n = 2000

Population size = N = 3000000

Sample mean = X = 200

Sample SD = 30

Now we calculate CI for 94%,98% and 98%.

For 94% Confidence Interval is (198.72, 201.26)

For 96% Confidence Interval is (198.62, 201.37)

For 98% Confidence Interval is (198.44, 201.56)

**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: 1) Mean = 41

Median = 40.5

Variance = 25.52

Standard deviation = 5.05

2) here we do not have outliers and the data is right skewed i.e., mean is greater than median.

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

Ans: When mean and median of data are equal the no skewness is present.

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

Ans: skewness and tail are towards right.

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

Ans: skewness and tail are towards left.

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

Ans: Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

Ans: Negative Kurtosis means the curve will be flatter and broader.

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



What can we say about the distribution of the data?

Ans: The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

Ans: The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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.

Ans: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker 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)

c. P (20<MPG<50)

Ans:

* 1. P(MPG>38) = 1-stats.norm.cdf (38, loc = data.MPG.mean(), scale= data.MPG.std ()) =0.3475
  2. P(MPG<40) = stats.norm.cdf (38, loc = data.MPG.mean(), scale= data.MPG.std ()) = 0.7293
  3. P (20<MPG<50) = stats.norm.cdf (20, loc = data.MPG.mean(), scale= data.MPG.std ()) - (1-stats.norm.cdf (50, loc data.MPG.mean(), scale= data.MPG.std ())) = 0.0131

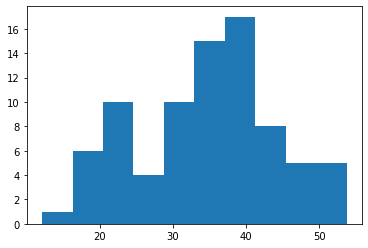
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: The MPG of cars does not follow Normal Distribution it has negatively skewed distribution.

Histogram:



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

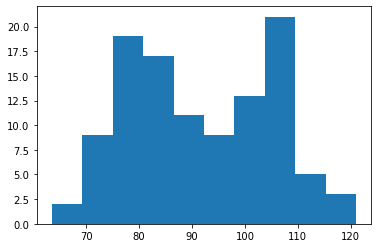
Dataset: wc-at.csv

Ans: The AT and Waist from wc-at data set both have positively skewed distribution.

Histogram of AT:



Histogram of Waist:



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

Ans:

stats.norm.ppf (1-alpha/2)

Z score of 90% is 1.6448

Z score of 94% is 1.8807

Z score of 60% is 0.8416

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

Ans:

stats.t.ppf (P%, df = 24)

t score of 95% is 1.7108

t score of 96% is 1.8280

t score of 99% is 2.4921

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 – statistic is

t =

x = sample mean = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = sample size = 18

t =

t =

t =

t =

t = - 0.471

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

(Using stats.t.cdf (-0.471, 17))

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3218.