Q1) Identify the Data type for the Following:

Answer:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| 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 | Continuous |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

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

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

Ans: If 3 coins are tossed then, total 8 possibilities are there as follows:

P = {HHH, HHT, HTH, THH, TTT, TTH, THT, HTT}

From, above possibilities 2 heads and 1 tail coming 3 times

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

1. Equal to 1

Ans: sum will not come 1 because in dice min number is 1

1. Less than or equal to 4

Ans: probability is 6/36 = 0.166

P(n) = 36

P = {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. Sum is divisible by 2 and 3

Ans: probability is 24/36 = 0.666

P = {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}

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: probability that none of the ball drawn blue is 4/6 = 0.666

R = 2, G = 3, B = 2

P = {RR, RG, RB, GG, GB, BB}

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: (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 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 : Assignment 1.ipynb file attached for above question**

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 value of the data given in case of list

Mean = (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.3

**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: Q9\_a.csv - For speed Skewness = -0.12 kurtosis = -0.51 - For dist Skewness = 0.81 kurtosis = 0.41

​ Q9\_b.csv - For SP Skewness = 1.61 kurtosis = 2.98 - For WT Skewness = -0.61 kurtosis = 0.95

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

**Ans:** In histogram as the ChickWeight$weight is increasing then frequency also increasing. And it is positively skewed.



In boxplot the points which can observed are whisker is shorter at the bottom end and longer on top, some outliers also present, it is right skewed(or positively skewed)



**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 : Confidence interval = sample mean ± (critical value x standard error)

Standard error = 30 / √2000 = 0.671

For a 94% confidence interval, the critical value is ±1.88.

For a 98% confidence interval, the critical value is ±2.33.

For a 96% confidence interval, the critical value is ±2.05.

94% confidence interval = 200 ± (1.88 x 0.671) = (198.16, 201.84)

98% confidence interval = 200 ± (2.33 x 0.671) = (197.35, 202.65)

96% confidence interval = 200 ± (2.05 x 0.671) = (197.84, 202.16)

**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 =25.52 and Standard Deviation =5.05

1. What can we say about the student marks?

Ans: we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

Q13) What is the nature of skewness when mean, median of data are equal? Ans: Symmetrical Distributiion or Normal Distribution

Q14) What is the nature of skewness when mean >median ? Ans: Positively Skewed

Q15) What is the nature of skewness when median > mean? Ans: Negatively Skewed

Q16) What does positive kurtosis value indicates for a data? Ans: Highly Peaked

Q17) What does negative kurtosis value indicates for a data? Ans: Low Peaked

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



What can we say about the distribution of the data? Ans: Its not normally distributed

What is nature of skewness of the data? Ans: Negatively Skewed.

What will be the IQR of the data (approximately)? Ans: 8 (from 10 to 18)

Q19) Comment on the below Boxplot visualizations?



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

Ans: No outliers present in both plots. Both the box plot shares the same median that is approximately in a range between 260 to 270 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) = 0.348
  2. P(MPG<40) ) = 0.729

c. P(20<MPG<50) = 0.013000000000000012

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Data is normally distributed.

import scipy.stats as stats

# Perform Shapiro-Wilk test on the data

stat, p = stats.shapiro(car\_df['MPG'])

# Interpret the results

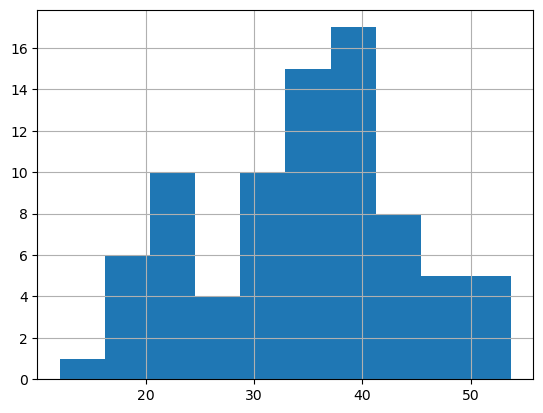
alpha = 0.05

if p > alpha:

print('Data is normally distributed')

else:

print('Data is not normally distributed')



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 is not normally distributed of AT and Waist

import scipy.stats as stats

# Perform Shapiro-Wilk test on the data

stat, p = stats.shapiro(wc\_df['AT'])

# Interpret the results

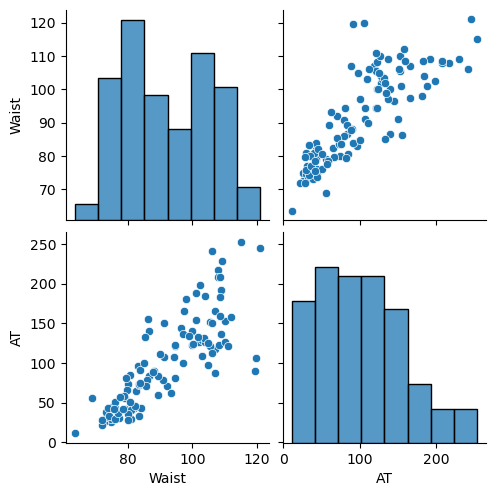
alpha = 0.05

if p > alpha:

print('Data is normally distributed')

else:

print('Data is not normally distributed')



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

Ans : from scipy import stats

print("Z score for a 90% confidence interval:", stats.norm.ppf((1 + 0.9) / 2))

Z score for a 90% confidence interval: 1.6448536269514722

print("Z score for a 94% confidence interval:", stats.norm.ppf((1 **+** 0.94) **/** 2))

Z score for a 94% confidence interval: 1.8807936081512509

print("Z score for a 60% confidence interval:", stats.norm.ppf((1 **+** 0.6) **/** 2))

Z score for a 60% confidence interval: 0.841621233572914

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

Ans: # t score for 95% confidence interval

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))

T score for 95% Confidence Interval = -2.0639

# t value for 94% confidence interval

T score for 94% Confidence Inteval = -1.974

# t value for 99% Confidence Interval

print('T score for 99% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))

T score for 99% Confidence Interval = -2.7969

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 = (sample mean - population mean) / (standard deviation / sqrt(sample size))

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

t = -0.471

To calculate degrees of freedom, df = 18-1 = 17

To calculate cdf used jupyter to execute below code:

from scipy import stats

t\_score = -0.471

df = 17

p\_value = stats.t.cdf(t\_score, df=df)

print(p\_value)

0.321 = 32.1% probability is there if 18 randomly selected bulbs have average life 260 days