|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Ordinal |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| 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: 3 times

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

1. Equal to 1

Ans: Zero

1. Less than or equal to 4

Ans: 16.66% or 0.1666

1. Sum is divisible by 2 and 3

Ans: 16.66% or 0.1666

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: 47.61% or 0.4761

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

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:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| No. of sample | 32 | 32 | 32 |
| Mean | 3.6 | 3.2 | 17.85 |
| Median | 3.7 | 3.3 | 17.71 |
| Variance | 0.29 | 0.96 | 3.19 |
| Std | 0.53 | 0.96 | 1.76 |
| Range | 2.17 | 3.91 | 8.4 |
| N-2deviation | 2.54 | 1.3 | 14.33 |
| N+deviation | 4.66 | 5.14 | 21.37 |

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:

108+ 110 + 123+ 134 + 145 + 167 + 187 +199 / 9

= 1308/9

= 145.333

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

Code :

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

# Skewness

data.skew()

Ans =>

speed = -0.117510

distance = 0.806895

Inference: Speed is left skewed distribution and Distance is right skewed distribution

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

Code:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

data=pd.read\_csv('Database/Q9\_b.csv')

data2=data.iloc[:,1:]

data2

Ans =>

data2.skew()

SP = 1.611450

WT = -0.614753

# Kurtosis

data2.kurt()

SP = 2.977329

WT = 0.950291

Inference: SP is right skewed distribution and WT is left skewed distribution

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





Ans: From the given graph we can infer that the chick weight is positively skewed. Therefore the mean is less than the median, hence the median is less than the mode.

**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:

The confidence intervals are constructed as follows;

(i) The 94% confidence interval;

C.I=(200±1.26)

Upper confidence limit=200+1.26=201.26

Lower confidence limit=200-1.26=198.74

The 94% confidence interval is (198.74, 201.26)

(ii) The 98% confidence interval;

C.I=(200±1.56)

Upper confidence limit=200+1.56=201.56

Lower confidence limit=200-1.56=198.44

The 98% confidence interval is (198.44, 201.56)

(iii)The 96% confidence interval;

C.I=(200±1.37)

Upper confidence limit=200+1.37=201.37

Lower confidence limit=200-1.37=198.63

The 96% confidence interval is (198.63, 201.37)

**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 = 24.11

Standard Deviation = 4.91

2)

There are two outliers in the given data

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

Ans: When the mean and median of data are equal, the distribution is symmetric and hence the distribution has zero skewness.

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

Ans: When mean > median, we get Right Skewed Distribution (Positively Skewed Distribution)

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

Ans: When median > mean, we get Left Skewed Distribution (Negatively Skewed Distribution)

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

Ans: A Positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution

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

Ans: A Negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

Ans: The data has Asymmetrical Distribution.

What is nature of skewness of the data?

Ans: This data has negative skewness

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

Ans:

IQR = Q3 - Q1

IQR = 18 - 10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Ans: Boxplot 1 has negative kurtosis while the Box plot 2 has positive kurtosis.

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) = 33/81
  2. P(MPG<40) = 61/81

c. P (20<MPG<50) = 69/81

Code:

MPG ,-c (Cars$MPG)

MPG

Sample(MPG)

a = subset(MPG,MPG>38)

b = subset(MPG,MPG<40)

c = subset(MPG,MPG>20 & MPG <50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

mean

=> 34.422075728024666

median

=> 35.15272697

Mode

=> 29.629936

Skewness = -0.17794674747025727

Kurtosis = -0.6116786559430913

The given values indicate that the data is fairly symmetrical hence the MPG of cars follows 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

Ans

a.

mean = 91.902

median= 90.8

Kurtosis = -0.3760059

Weight of Waist Circumference follows Normal Distribution

b.

mean = 91.902

median = 90.8

kurtosis = -1.141846

Weight of Adipose Tissue does not follow Normal Distribution

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

Ans:

If the corresponding shape follows normal distribution; 90% CI covers 90% normal curve. Hence, probability of value outside said area is less than 0.117510

Since the normal curve is symmetric, half of the area in the left tail while the other half is in the right tail.

Therefore, Area of each tail of the curve = (1 – CI) / 2 = 0.05

There the Z scores are:

(CI 90%) = 1.645

(CI 94%) = 1.56

(CI 60%) = 0.26

Therefore, we get that the z score at 90 % confidence interval is 1.645, at 94 % confidence interval is 1.555 and at 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:

t scores = ( sample mean - population mean ) / ( std deviation - ( sample size - 1 )

t scores CI95 % for n = 25

= 2.063

t scores CI96 % for n = 25

=1.828

t scores CI99 % for n = 25

=2.797

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:

Mean = 270 days

sample size = 18

sample mean = 260

deviation sample = 90 days

Z = (x – s) / s

s = 90/sq.rt(18)

Therefore

Z=(260-270)/21.2

Z=-0.47

Calculating the p-value using p value calculator

p-value=0.3192.

Hence, we conclude that there is a 31.92% chance that a random sample of 18 light bulbs would have an average lifespan of less than 260 days.