|  |  |
| --- | --- |
| 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 | 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| 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 | 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:**

Three coins are tossed then the total number of possible combinations are:

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT = 8

The number of combinations which have two heads, and one tail are:

HHT, HTH, TTH which makes them 3 in number.

Therefore, the Probability of getting two heads and one tails in the toss of three coins simultaneously is defined as:

P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three

coins tossed))

= 3/8

= 0.375

= 37.5%

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

**Ans:**

Number of possible outcomes:

N (Event (Two dice rolled)) = 6^2 = 36

1. Equal to 1

sum is equal to 1 is zero because they start with (1,1)

P (sum is Equal to 1) = ‘0’ zero null.

1. Less than or equal to 4

The sum is less than or equal to 4 the possible outcomes are (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

= 6/36

= 1/6

P (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to

4)) / N (Event (Two dice rolled))

= 6 / 36

= 1/6

= 0.166

= 16.66%

1. Sum is divisible by 2 and 3

The sum is divisible by 2 and 3 the possible outcomes are

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

P (Sum is divisible by 2 and 3) = N (Event (Sum is divisible by 2 and 3)) / N

(Event (Two dice rolled))

= 6 / 36

= 1/6

= 0.16

= 16.66%

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

N (Event (2 balls are drawn randomly from bag) = 7! / 2! \* 5!

= (7\*6\*5\*4\*3\*2\*1) /

(2\*1) \* (5\*4\*3\*2\*1)

= (7\*6)/ (2\*1)

= 21

If none of them drawn 2 balls are blue = 7 – 2 = 5

N (Event (None of the balls drawn is blue) = 5! / 2! \* 3!

= (5\*4) / (2\*1)

= 10

P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) /

N (Event (2 balls are drawn randomly from

bag)

= 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:**

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

=0.015+0.8+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

* For Points,Score,Weight>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans:**

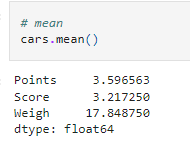
cars=pd.read\_csv("C:\Assignments\Assignment 1 Basic stats level 1/Q7.csv")

**Graphical user interface, text, application, email

Description automatically generated**

Mean for Points = 3.59, Score = 3.21 and Weight = 17.84

cars.mean()



Median for Points = 3.69, Score = 3.32 and Weight = 17.71

cars.median()

Graphical user interface, text, application

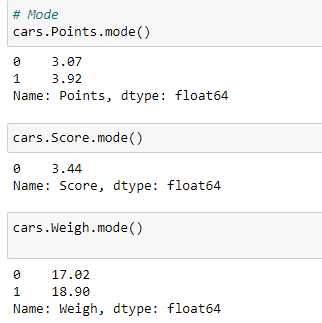
Description automatically generated

Mode for Points = 3.07, Score = 3.44 and Weight = 17.02

cars.Points.mode()

cars.Score.mode()

cars.Weigh.mode()



Variance for Points = 0.28, Score = 0.95, Weight = 3.19

Cars.var()

Graphical user interface, text, application

Description automatically generated

Standard Deviation for Points = 0.53, Score = 0.97, Weight = 1.78

Cars.std()

Graphical user interface, text, application

Description automatically generated

Range [Min-Max] for Points [2.76 – 4.93], Score [1.51 – 5.42] and Weight [14.50 – 22.9]

Points\_Range=cars.Points.max()-cars.Points.min()

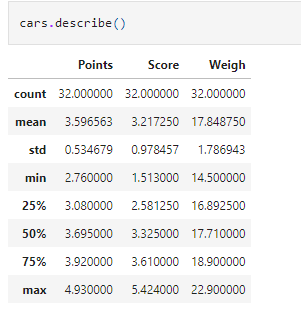
Points\_Range

Score\_Range=cars.Score.max()-cars.Score.min()

Score\_Range

Weigh\_Range=cars.Weigh.max()-cars.Weigh.min()

Weigh\_Range



**Graphical user interface, text, application, email

Description automatically generated**

Draw inferences:

f,ax=plt.subplots(figsize=(15,5))

plt.subplot(1,3,1)

plt.boxplot(cars.Points)

plt.title('Points')

plt.subplot(1,3,2)

plt.boxplot(cars.Score)

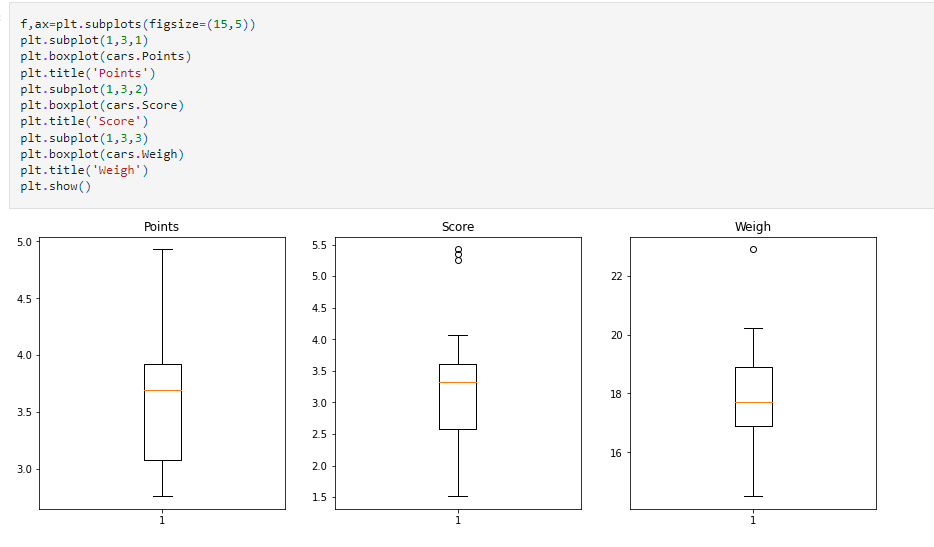
plt.title('Score')

plt.subplot(1,3,3)

plt.boxplot(cars.Weigh)

plt.title('Weigh')

plt.show()



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 = Sum (X \* Probability of X)

= (1/9) (108) + (1/9) (110) + (1/9) (123) + (1/9) (134) + (1/9) (145) +

(1/9) (167) + (1/9) (187) + (1/9) (199)

= 145.33

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

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

**Ans:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

data=pd.read\_csv("C:/Assignments/Assignment 1 Basic stats level 1/Q9\_a.csv")

data.skew()

data.kurt()

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

For Cars Speed Skewness value= -0.12 and Kurtosis value= 0.81

Skewness value = 0.81 and Kurtosis value = 0.41 for Cars Distance

SP and Weight (WT)

Use Q9\_b.csv

**Ans:** data=pd.read\_csv('C:/Assignments/Assignment 1 Basic stats level 1/Q9\_b.csv')**Graphical user interface, text, application, email

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

For SP /WT Skewness = 1.61 Kurtosis = 0.95

Q10) Draw inferences about the following boxplot & histogram



**Ans:** The histogram’s peak has a right skew, and tail is on right.

Mean > Median.

We have outliers on the higher side.



**Ans:** The boxplot has outliers on the maximum 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**: import numpy as np

import pandas as pd

from scipy import stats

from scipy.stats import norm

**Text

Description automatically generated**

For 94% confidence interval Range is [ 198.73 – 201.26]

stats.norm.interval(0.94,200,30/(2000\*\*0.5))

Graphical user interface, text, application

Description automatically generated

For 98% confidence interval range is [198.43 – 201.56]

stats.norm.interval(0.98,200,30/(2000\*\*0.5))

Graphical user interface, text, application

Description automatically generated

For 96% confidence interval range is [198.62 – 201.37]

stats.norm.interval(0.96,200,30/(2000\*\*0.5))

Graphical user interface, text, application

Description automatically generated

**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.

Mean =41

Median =40.5

Variance =25.52

Standard Deviation =5.05

1. What can we say about the student marks?

Ans= Data is slightly skewed towards the right because mean is greater than median.

We don’t have outliers.

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

Ans= No skewness is present we have a perfect symmetrical distribution

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 indicate for a data?

Ans= Positive kurtosis means the curve is more peaked

Q17) What does negative kurtosis value indicate 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 data?

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

What is the nature of skewness of the data?

Ans= The data is 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:** Both the box plot shares the same median that is approximately in a range between 275 to 250.

There are no outliers.

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:** import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats

from scipy.stats import norm

cars=pd.read\_csv('C:/Assignments/Assignment 1 Basic stats level 1/Cars.csv')

cars

Graphical user interface, text, application, email

Description automatically generated

stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())

stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std()) -stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())

1. P(MPG>38)

=0.3475

1. P(MPG<40)

=0.729

1. P (20<MPG<50)

=1.2430

Graphical user interface, text, application, email

Description automatically generated

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** MPG of cars follows normal distribution.

cars=pd.read\_csv('C:/Assignments/Assignment 1 Basic stats level 1/Cars.csv')

Graphical user interface, text, application, email

Description automatically generated

Chart, line chart

Description automatically generated

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

Dataset: wc-at.csv

**Ans:**

Adipose Tissue (AT) and Waist does not follow Normal Distribution

df=pd.read\_csv('C:/Assignments/Assignment 1 Basic stats level 1/wc-at.csv')

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

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

**Ans:**

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

**Ans:**

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 follows:



x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

find t-scores









t = - 0.471

 degrees of freedom are n - 1,

The probability that **t < - 0.471 with 17 degrees of freedom**

