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

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

**ANS:** P : { HHH, TTT, HTT, THH, HHT, TTH, THT, HTH}

Conclusion The probability that two heads and one tail is 3/8.

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

1. Equal to 1 **= 0**
2. Less than or equal to 4 = **1/6**
3. Sum is divisible by 2 and 3 = **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:** 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 is

=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, 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 | Weigh |
| Mean | 3.596563 | 3.217250 | 17.848750 |
| Median | 3.695 | 3.325 | 17.710 |
| Mode | 3.92 | 3.44 | 18.90 |
| Standard Deviation | 0.534679 | 0.978457 | 1.786943 |

**Inference:** From the data, we can observe that mean, median and mode are not equal. Hence, we can say that data is skewed.

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:** By using R-Studio:

x <- c(108, 110, 123, 134, 135, 145, 167, 187, 199)

mean(x)

**OUTPUT: 145.3333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.117510 | 0.806895 |
| Kurtosis | -0.508994 | 0.405053 |

**Inference:** From the skewness of speed we can say that the data of speed is

symmetric and from the distance we can say that the data is positively skewed.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | SP | Weight (WT) |
| Skewness | 1.611450 | -0.614753 |
| Kurtosis | 2.977329 | 0.950291 |

**Inference:** From the skewness of SP we can say that the data of SP is positively

skewed and from the skewness of weight we can say that the data is

negatively skewed.

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



**ANS:** From the above Histogram we can say that the data is Positively Symmetric i.e. data is skewed towards right.



**ANS:** From the above Boxplot, we can say that few outliers are present in the data, and data is skewed towards right.

**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:**

**By using python:**

import numpy as np

from scipy import stats

stats.norm.interval(alpha = 0.94,loc = 200,scale = 30/np.sqrt(2000)) #for 94%

stats.norm.interval(alpha = 0.98,loc = 200,scale = 30/np.sqrt(2000)) #for 98%

stats.norm.interval(alpha = 0.96,loc = 200,scale = 30/np.sqrt(2000)) #for 98%

**Output:**

|  |  |  |
| --- | --- | --- |
| Confidence Interval | Lower Limit | Upper Limit |
| 94% | 198.738325292158 | 201.261674707842 |
| 96% | 198.62230334813333 | 201.37769665186667 |
| 98% | 198.43943840429978 | 201.56056159570022 |

**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:**

**BY using python:**

Import numpy as np

marks = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

np.mean(marks)

np.median(marks)

np.var(marks)

np.std(marks)

**Output:**

Mean= 41.0

Median= 40.5

Variance = 24.1111

Standard deviation = 4.91030

Students average marks is 41.

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

**ANS:** When mean, median are equal then we can say that data is perfectly symmetric.

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

**ANS:** The nature of skewness is positively skewed.

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

**ANS:** The nature of skewness is negatively skewed.

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

**ANS:** A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

**ANS:** A distribution with 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 is skewed distribution. The data is skewed towards left.

What is nature of skewness of the data?

**ANS:**  The nature of skewness of the data is negatively skewed.

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: Inference:**  1. No outliers present in boxplot1 and boxplot2.

2. Both follows normal distribution.

3. Difference is boxplot1 has lesser range compared to boxplot2.

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)
  3. P (20<MPG<50)

**ANS: Python code:**

**import pandas as pd**

**from scipy import stats**

data = pd.read\_csv("cars.csv")

average = data["MPG"].mean()

std = data["MPG"].std()

stats.norm.cdf(38,loc=average,scale=std) #p(MPG>38)

stats.norm.cdf(40,loc=average,scale=std) #p(MPG>40)

a = stats.norm.cdf(50,loc=average,scale=std) #p(MPG<50)

b = stats.norm.cdf(20,loc=average,scale=std) #p(20<MPG)

a-b

**Output: (MPG>38) = 0.6524060748417295**

**(MPG>40) = 0.7293498762151616**

**(20<MPG<50) = 0.11424755264231871**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS: Python code:**

**import pandas as pd**

**import matplotlib.pyplot as plt**

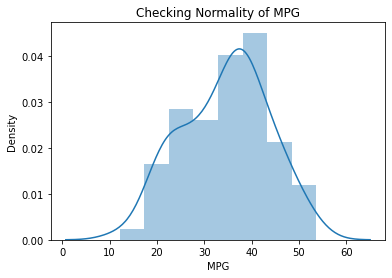
**import seaborn as sns**

data = pd.read\_csv("cars.csv")

sns.distplot(data.MPG)

plt.title("Checking Normality of MPG")

plt.show()



**Conclusion:** From the above histogram plot, we can say that 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: Python code:**

**import pandas as pd**

**import matplotlib.pyplot as plt**

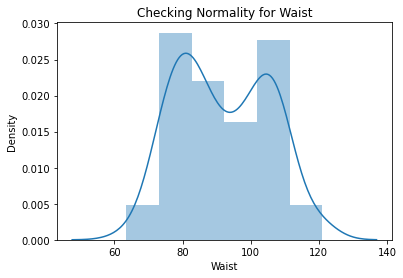
**import seaborn as sns**

df = pd.read\_csv("wc-at.csv")

sns.distplot(df.Waist)

plt.title("Checking Normality for Waist")

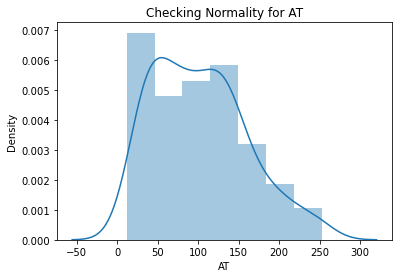
plt.show()



sns.distplot(df["AT"])

plt.title("Checking Normality for AT")

plt.show()



**Conclusion:** Data is normally distributed for “WAIST”.

Data is skewed towards right side for “AT”.

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

**ANS:** The Z scores of 90% =  1.645

The Z scores of 94% = 1.880

The Z scores of 60% = 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- score of 95% = 2.0638985616280205

t- score of 96% = 2.1715446760080677

t- score of 99% = 2.796939504772804

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:** µ=270, =260, SD=90, n=18, df=n-1=18-1= 17

tscore= = = -10/21.23= -0.47

pt(-0.47,17)

Required probability = 0.32=32%

***The probability of the bulbs lasting less than 260 days on average of 0.32***

**p(0.3216)**