|  |  |
| --- | --- |
| 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(categorical data) |
| 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 | Nominal |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

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: c

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=0.476

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

Here they are asking expected value so that is,

μ= ∑Xp(x)

so that is,

FOR A PROBABILITY DISTRUBUION THE MEAN OF THE DISTRUBUTION IS KNOW AS EXEPCTED VALUE SO NOW

(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.



import pandas as pd

x = {"POINTS":pd.Series([1,2,3,4,5,7,8]),"SCORE":pd.Series(["A","B","C","D","E","F","G"]),"WEIGHT":pd.Series([1,2,3,4,5,7,8]),"D":pd.Series([1,2,3,4,5,7,8,11])}

new\_x = pd.DataFrame(x) ## MAKING HERE ALSO DATA FRAME---pandas.core.frame.DataFrame

type(new\_x)

new\_x

1. POINTS

new\_x['Points'].mean()

new\_x['Points'].median()

new\_x['Points'].mode()

new\_x['Points'].var()

new\_x['Points'].std()

max(new\_x['POINTS'])

min(new\_x['POINTS'])

range = max(new\_x['POINTS'])-min(new\_x['POINTS'])

range

2. SCORES

new\_x['SCORES'].mean()

new\_x['SCORES'].median()

new\_x['SCORES'].mode()

new\_x['SCORES'].var()

new\_x['SCORES'].std()

max(new\_x['SCORES'])

min(new\_x['SCORES'])

range = max(new\_x['SCORES'])-min(new\_x['SCORES'])

range

3. WEIGHTS

new\_x['WEIGHTS'].mean()

new\_x['WEIGHTS'].median()

new\_x['WEIGHTS'].mode()

new\_x['WEIGHTS'].var()

new\_x['WEIGHTS'].std()

max(new\_x['WEIGHTS'])

min(new\_x['WEIGHTS'])

range = max(new\_x['WEIGHTS'])-min(new\_x['WEIGHTS'])

range

FOR SAMPLE VALUES I HAVE CALCULATED BUT ACUAL RESULT IS

Ans:score:mean=3.23

median=3.325

mode =3.44

variance=0.957

stdeviation=0.9784

Range=3.91

weight:mean=17.84

median=17.71

mode=17.02

variance=3.129

stdeviation=1.768

range=8.2

Point:mean=3.619063

median=3.715

mode=3.92

variance=0.278086

St Dev=0.527339

Range=2.17

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?

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

**Cars speed and distance**



**SP and Weight(WT)**



**ANS:**

**import pandas as pd**

**mtcars = pd.read\_csv("C:\\EXCELR\\Cars.csv", usecols = lambda column : column not in ["HP" , "MPG", "VOL"])**

**mtcars.skew(axis=0, skipna=True)**

**SP 1.611450**

**WT -0.614753**

**dtype: float64**

**mtcars.kurt(axis=0, skipna=True)**

**SP 2.977329**

**WT 0.950291**

**dtype: float64**

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



Ans:Histogram: It is used to find frequency plot and shape of distribution.It also

find skewness

Boxplot:It is visual representation of the statistical five number summary of a

dataset

\*Minimum

\*first quartile

\*median

\*third quartile

**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 ?

**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.

import pandas as pd

df=pd.Series([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])

df

df.mean()

df.median()

df.var()

df.std()

df.mean()

Out[138]: 41.0

df.median()

Out[139]: 40.5

df.var()

Out[140]: 25.529411764705884

df.std()

Out[141]: 5.05266382858645

1. What can we say about the student marks?

range = max(df)-min(df)

range=22

The range of students marks is 22

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

Ans: Symmetric.

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

Ans: positively charged.

Positive skewness

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

Ans: negatively charged.

Negative skewness

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

Ans: The distribution has heavier tails and a sharper peak.

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

Ans: The distribution has lighter tails and flatter peak.

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



What can we say about the distribution of the data?

Ans:a) The median of boxplot is 15.5

first quartile=10

third quartile=18.2

What is nature of skewness of the data?

b) negatively skewness

What will be the IQR of the data (approximately)?   
c) 8.2

Q19) Comment on the below Boxplot visualizations?



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

Ans:IQR for boxplot 1=25

IQR for boxplot 2=87

The median for box plot 1 and 2 is 262.5

The skewness for box plot1 and 2 is Symmetric.

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)=0.68
  2. P(MPG<40)=0.04

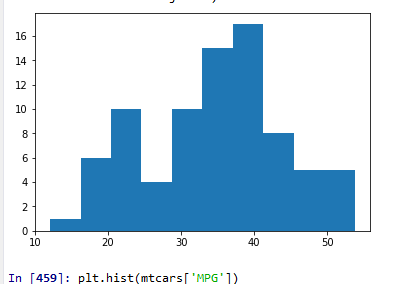
c. P (20<MPG<50)=0.79

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

import matplotlib.pyplot as plt

plt.hist(mtcars['MPG'])



Dataset: Cars.csv

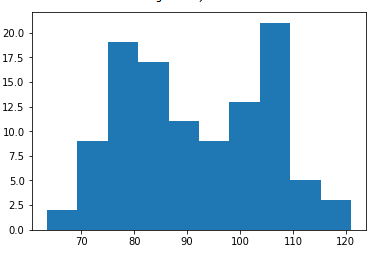
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

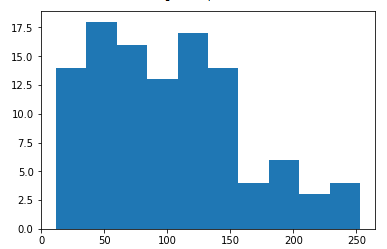
Ans:

wcat=pd.read\_csv("C:\\EXCELR\\NOTES WRITTEN\Regression\\wc-at.csv")

plt.hist(wcat.Waist)



plt.hist(wcat.AT)



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

Ans:z=90%=1.645

import scipy.stats as st

st.norm.ppf(0.95)

=1.645

Ans: z =94%=1.88

import scipy.stats as st

st.norm.ppf(0.97)

=1.88

Ans: 60%=0.841

import scipy.stats as st

st.norm.ppf(0.80)

=0.841

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

Tscore=t+/-1-alpha s/squareroot(n)

For 95%, n-1=24

2.064

For 96%, n-1=24

For 99%, n-1=24

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