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

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?

Answer: Total number of possible outcomes =[ hhh, hht, htt, ttt, tth, thh, hth, tht ]

Favourable outcome=3

Probability=3/8=37.5%

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

Answer: total number of outcomes =

(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. Equal to 1 = **0%**
2. Less than or equal to 4= 6/36 = **1/6=16.666667%**
3. Sum is divisible by 2 and 3 = 6/36= **1/6=16.666667%.**

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?

## Answer: Let S be the sample space. Then, n(S) = Number of ways of drawing 2 balls out of 7 = (7×6)/ (2×1)​ =21

## Let E = Event of drawing 2 balls, none of which is blue. n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.​ = (5×4)/ (2×1)​ =10 P(E)= n(E)/ n(S)​=10/21=47.6%.​

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

Answer:1+4+3+5+6+2=21/6=3.5.

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**

**Answer: Import pandas as pd**

**Import numpy as np**

**From pandas import DataFrame,Series**

**df=Pd.read\_csv(‘Q7.csv’)**

**df.describe(),df.mode(),df.var(),df.std()**

**Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84**

**Median for Points = 3.69, Score = 3.32 and Weigh = 17.71**

**Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02**

**Variance for Points = 0.28, Score = 0.95, Weigh = 3.19**

**Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78**

**Range [Min-Max] for Points [2.76– 4.93], Score [1.51 – 5.42] and Weigh [14.5 – 22.9]**

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?

Answer:108+110+123+134+135+145+167+187+199=1308/9=145.3.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer:**

**Import pandas as pd**

**Import numpy as np**

**From pandas import DataFrame,Series**

**df=Pd.read\_csv(‘Q9\_a.csv’)**

**df.skew()**

speed -0.117510

dist 0.806895

**df.kurt()**

speed -0.508994

dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Df1=Pd.read\_csv(‘Q9\_b.csv’)**

**Df1.skew()**

SP 1.611450

WT -0.614753

**Df1.kurt()**

SP 2.977329

WT 0.950291

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



**Asnwer: The histograms peak has right skew and tail is on right.**

**The boxplot has outliers towards upper extreme.**

**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?

Answer:

Import pandas as pd

Import numpy as np

From scipy import stats

conf\_int\_z = stats.norm.interval(0.94, loc=200,scale=30/np.sqrt(2000))

conf\_int\_z

array([198.73 – 201.26])

conf\_int\_z = stats.norm.interval(0.98, loc=200,scale=30/np.sqrt(2000))

array([198.43 – 201.56])

conf\_int\_z = stats.norm.interval(0.96, loc=200,scale=30/np.sqrt(2000))

array([198.62 – 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?

Answer: df=[34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

df=pd.DataFrame(df)

df.mean()

0 41.0

dtype: float64

df.median()

0 40.5

dtype: float64

df.var()

0 25.529412

dtype: float64

df.std()

0 5.052664

dtype: float64

Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.

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

Answer: no skewness, perfectly symmetric

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

Answer: Right skewed

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

Answer: Left skewed

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

Answer: Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Answer: Negative Kurtosis means the curve will be flatter and broader, it is Platokurtic.

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



What can we say about the distribution of the data?

Answer: It is not normally distributed.

What is nature of skewness of the data?

Answer: It is left skewed.

What will be the IQR of the data (approximately)?   
Answer: Inter Quartile Range =Upper Quartile- 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.

Answer: 1) The median of the two boxplots are same = 262.5

2) The boxplots are not skewed in +ve or –ve direction.

3) Outliers doesn’t exist in both of the boxplots.

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)

Answer: a. P(MPG>38)

from scipy import stats

1-stats.norm.cdf(38,34.422076,9.131445)

0.34759394041453007

b.P(MPG<40)

stats.norm.cdf(40,34.422076,9.131445)

0.7293498604157946

c.P (20<MPG<50)

stats.norm.cdf(50,34.422076,9.131445)

0.9559926858516099

stats.norm.cdf(20,34.422076,9.131445)

0.05712377822429007

0.9559926858516099-0.05712377822429007 =0.8988689076268

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:

df=pd.read\_csv('Cars.csv')

df

mpg=df[{'MPG'}]

mpg

import matplotlib.pyplot as plt

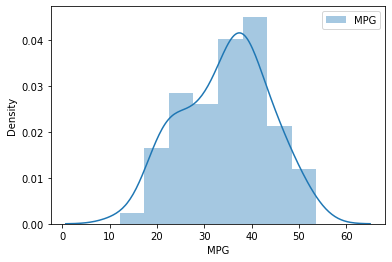
import seaborn as sn

sn.distplot(mpg.MPG,label='MPG');

plt.xlabel('MPG');

plt.ylabel('Density');

plt.legend();



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

Answer:

df1=pd.read\_csv('wc-at.csv')

df1

df\_wc=df1[{'Waist'}]

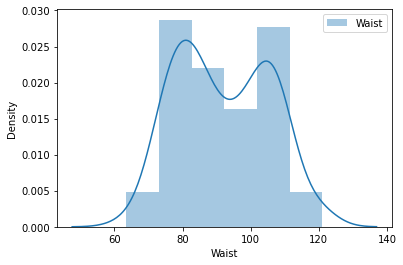
df\_wc

sn.distplot(df\_wc.Waist,label='Waist');

plt.xlabel('Waist');

plt.ylabel('Density');

plt.legend();



df\_at=df1[{'AT'}]

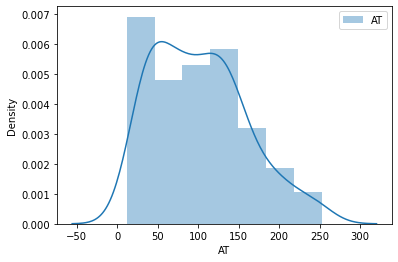
df\_at

sn.distplot(df\_at.AT,label='AT');

plt.xlabel('AT');

plt.ylabel('Density');

plt.legend();



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

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

Answer: from scipy import stats

for 90% stats.norm.ppf(.050)= -1.6448536269514729

For 94% stats.norm.ppf(.030)= -1.880793608151251

For 60% stats.norm.ppf(.200) = -0.8416212335729142

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

Answer: from scipy import stats

for 95% stats.t.ppf(0.025,df=24)= -2.063898561628021

For 96% stats.t.ppf(0.020,df=24)= -2.171544676008068

For 99% stats.t.ppf(0.005,df=24)= -2.796939504772805

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

Answer: from scipy import stats

t\_value = (260 - 270)/(90/np.sqrt(18))

t\_value

-0.4714045207910317

stats.t.cdf(-0.4714045207910317,df=17)

0.32167253567098364

=32%.

probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.32167253567098364