|  |  |
| --- | --- |
| 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 | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Nominal |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |
|  |  |

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

= The probalbility of getting two head 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 = 5/6

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?

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

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.

Ans.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | median | mode | varience | STD DEVIATION | range |
| point | 0.08625 | 3.695 | 3.92 | 0.276948 | 0.526258072 | 2.17 |
| SCORE | 0.111563 | 3.325 | 3.44 | 0.927461 | 0.963047701 | 3.911 |
| WEIGH | 0.625 | 17.71 | 17.02 | 3.09338 | 1.758800639 | 8.4 |

**Use Q7.csv file**

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?

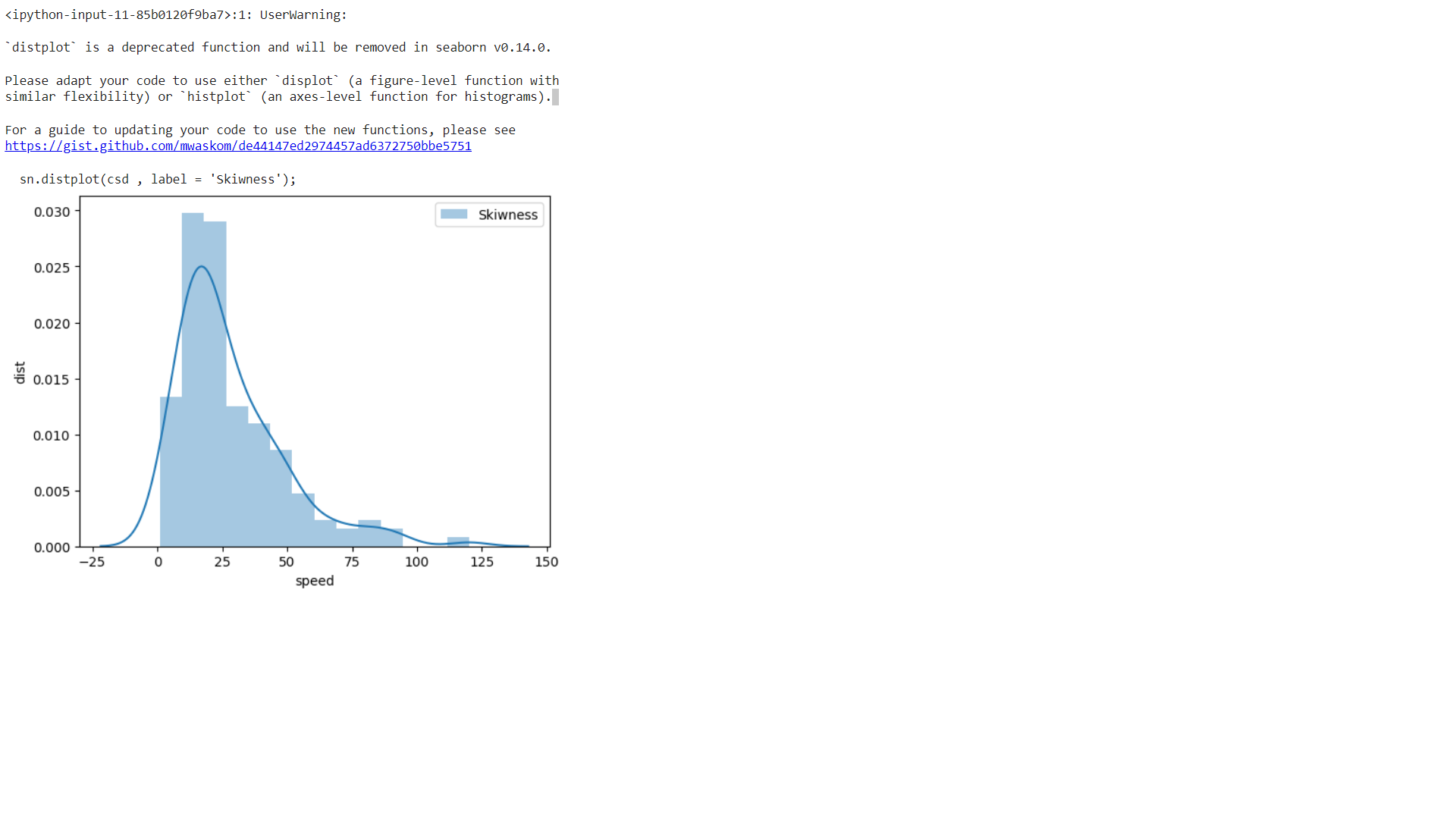
Ans . Expected value = (probability \* values) = 145

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS .**

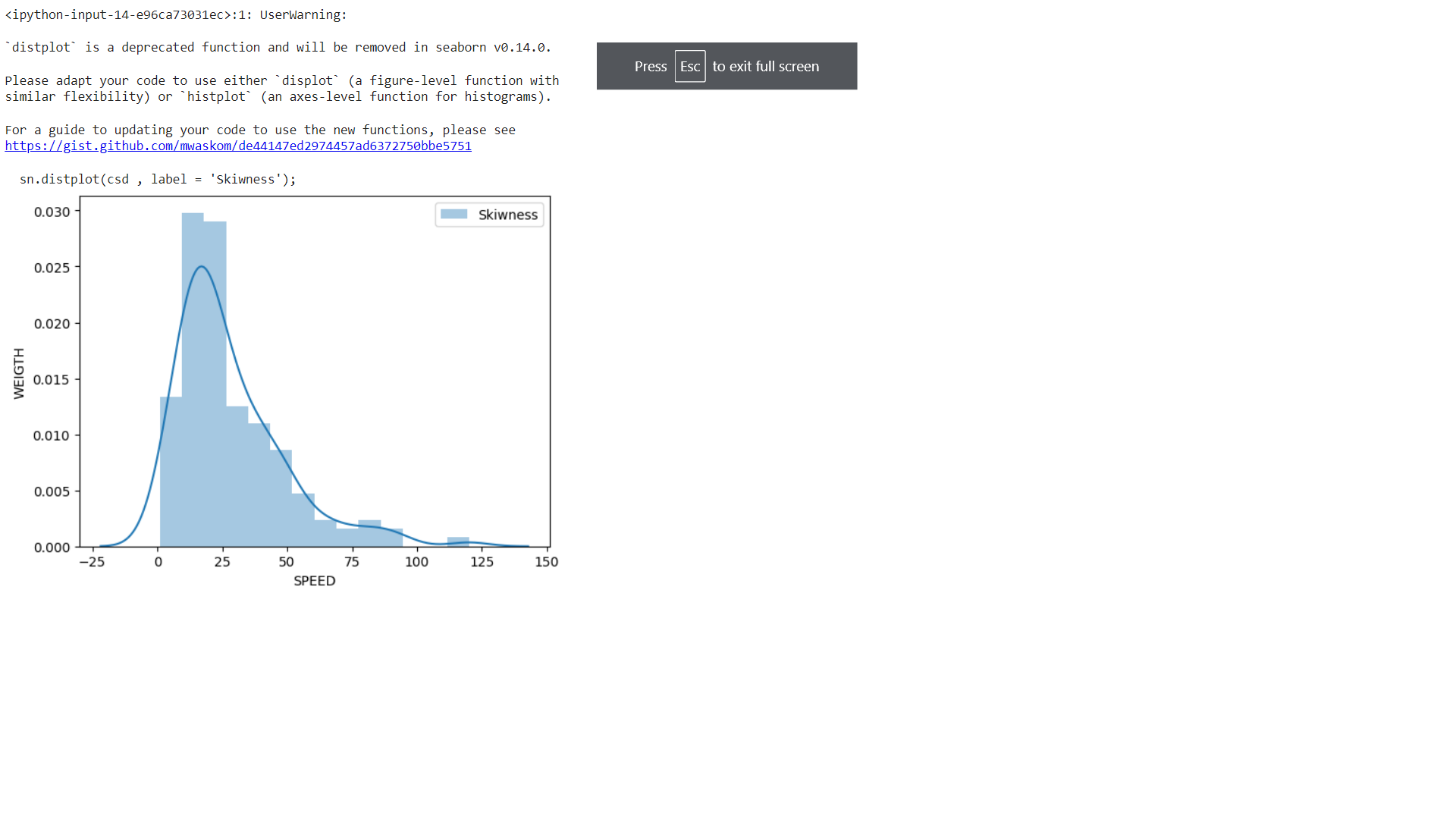


**is positive Skewness**

**and the peak defind as its positive kurtosis**

**SP and Weight(WT)**

**Use Q9\_b.csv**



**is positive Skewness**

**and the peak define as its positive kurtosis**

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



ANS. Histogram: - Chick weight data is positively skewed AND 50% Chick Weight is between 50 to 150.

The data is right skewed And There are outliers at upper 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 .

Using the t-distribution, it is found that:

Sample mean of 200

Sample standard deviation of 30

Sample size of =2000.

By given formula :

The 94% confidence interval is (198.73, 201.27).

the 96% confidence interval is (198.61, 201.39).

98% confidence interval is (198.43, 201.57).

OR

**import** numpy **as** np

**import** pandas **as** pd

**from** scipy **import** stats

**from** scipy.stats **import** norm

In [13]:

*# Avg. weight of Adult in Mexico with 94% CI*

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

Out[13]:

**(198.738325292158, 201.261674707842)**

In [14]:

*# Avg. weight of Adult in Mexico with 98% CI*

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

Out[14]:

**(198.43943840429978, 201.56056159570022)**

In [15]:

*# Avg. weight of Adult in Mexico with 96% CI*

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

Out[15]:

**(198.62230334813333, 201.37769665186667)**

**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 .

Mean =41

Median = 40.5

Variance= 25.51

Standard Deviation =5.05

The score data is arrange as ascending order

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

ANS

No Skewness or Normal Skewness

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

ANS

Positive Skewness

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

ANS

Negative Skewness

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

ANS

It’s a sharp Peak of plot

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

ANS

Don’t have a peak it thinner tailed .

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



What can we say about the distribution of the data?

ANS

Data Distribution in assigned format

What is nature of skewness of the data?

ANS

Negatively Skewed

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

IQR=Q3-Q1 = 18-10 = 8 is IQR

Q19) Comment on the below Boxplot visualizations?



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

ANS.

First there are no out layers,

Both box plot shares the same median that as between 250-275 range

And they are normally distributed .

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)

Prob\_MPG\_greater\_than\_38 **=** np**.**round(1 **-** stats**.**norm**.**cdf(38, loc**=** q20**.**MPG**.**mean(), scale**=** q20**.**MPG**.**std()),3)

print('P(MPG>38)=',Prob\_MPG\_greater\_than\_38)

**P(MPG>38)= 0.348**

In [27]:

prob\_MPG\_less\_than\_40 **=** np**.**round(stats**.**norm**.**cdf(40, loc **=** q20**.**MPG**.**mean(), scale **=** q20**.**MPG**.**std()),3)

print('P(MPG<40)=',prob\_MPG\_less\_than\_40)

**P(MPG<40)= 0.729**

prob\_MPG\_greater\_than\_20 **=** np**.**round(1**-**stats**.**norm**.**cdf(20, loc **=** q20**.**MPG**.**mean(), scale **=** q20**.**MPG**.**std()),3)

print('p(MPG>20)=',(prob\_MPG\_greater\_than\_20))

**p(MPG>20)= 0.943**

In [31]:

prob\_MPG\_less\_than\_50 **=** np**.**round(stats**.**norm**.**cdf(50, loc **=** q20**.**MPG**.**mean(), scale **=** q20**.**MPG**.**std()),3)

print('P(MPG<50)=',(prob\_MPG\_less\_than\_50))

**P(MPG<50)= 0.956**

In [43]:

prob\_MPG\_greaterthan20\_and\_lessthan50**=** (prob\_MPG\_less\_than\_50) **-** (prob\_MPG\_greater\_than\_20)

print('P(20<MPG<50)=',(prob\_MPG\_greaterthan20\_and\_lessthan50))

**P(20<MPG<50)= 0.013000000000000012**

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 Distribition.



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

Dataset: wc-at.csv

ANS . AT and waist not followed normal Distribution.



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

Ans.

z value for 90% confidence interval

print('Z score for 90% Confidence Interval =',np.round(stats.norm.ppf(.05),4))

Z score for 90% Confidence Interval = -1.6449

z value for 94% confidence interval

print('Z score for 94% Confidence Interval =',np.round(stats.norm.ppf(.03),4))

Z score for 94% Confidence Interval = -1.8808

z value for 60% confidence interval

print('Z score for 60% Confidence Interval =',np.round(stats.norm.ppf(.2),4))

Z score for 60% Confidence Interval = -0.8416

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

Ans:

t score for 95% confidence interval

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))

T score for 95% Confidence Interval = -2.0639

t value for 94% confidence interval

print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4))

T score for 94% Confidence Inteval = -1.974

t value for 99% Confidence Interval

print('T score for 99% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))

T score for 99% Confidence Interval = -2.7969

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: import numpy as np

Import scipy as stats

t\_score = (x - pop mean) / (sample standard daviation / square root of sample size)

(260-270)/90/np.sqrt(18))

t\_score = -0.471

stats.t.cdf(t\_score, df = 17)

0.32 = 32%