Q1) Identify the Data type for the Following:

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

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 | Nominal |
| 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 & Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer:

Sample Space: {HHH, HHT, HTH,THH,

TTT, TTH, THT, HTT}

Two heads and one tail: {HHT, HTH, HHT}

P(2heads & 1tail): Favorable outcomes / Total Outcomes

= 3/8

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:

SS: {(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)}

a) **0**

b) sum <=4: (1,1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)

Total number of outcomes = 36

P (sum <=4): 6/36

: **1/6**

C) divisible by 2&3: (1,5), (2, 4), (3, 3), (4, 2), (5, 1), (6, 6)

P (divisible by 2&3): 6/36

: **1/12**

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:

Total Num of balls in a bag: 2+3+2=7

P(red)= 2/7

P(green)= 3/7

P(blue)= 2/7

Number of ways of drawing 2 balls out of 7 =

7C2= (7 x 6)/(2x1) = 42/2 = 21

Number of balls other than blue = 5

Number of ways of drawing 2 balls out of 5 = 5C₂

= (5 x 4)/ (2 × 1) = 20/2 = 10

**Required Probability = 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

Answer:

Expected number of candies:

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = **3.09**

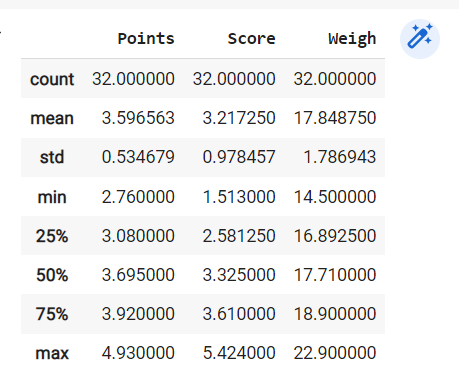
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

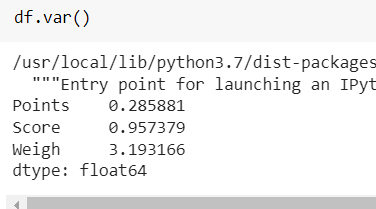
* For Points,Score,Weigh>

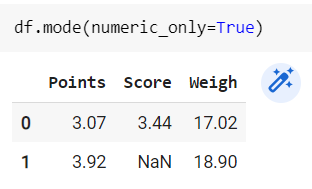
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

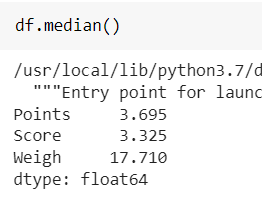
**Use Q7.csv file**

Answer:









In Points COLUMN:

We can see that the variance is low and mean,median,mode also have very minor difference. Hence,we can tell that the points are less deviated from its mean.

In Score COLUMN:

We can see that the variance is high when compared to Points column and mean,median,mode also have very minor difference. Hence,we can tell that the points are little deviated from its mean.

In Weigh COLUMN:

We can see that the variance is high and mean,median,mode have very minor difference. Hence,we can tell that the points are dispersed from its mean.

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:

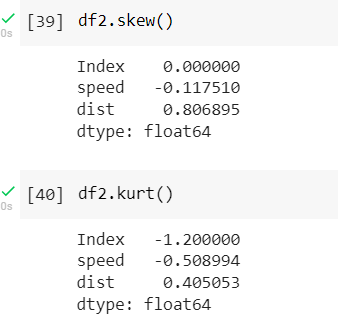
**145.33** (Solution provided in .ipynb File)

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer:



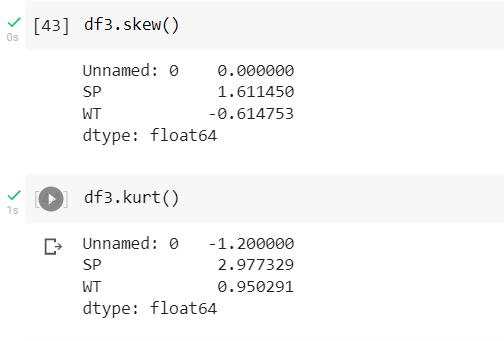
In **SPEED** the skewness is negative, that means the data points are more concentrated towards the right-hand side of the distribution. This makes the mean, median, and mode bend towards the right. Hence these values are always negative. In this distribution, Mode > Median > Mean. Where as the kurtosis is also negative, that means the distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

In **DISTANCE** the skewness is positive , that means the values are more concentrated towards the right side, and the left tail is spread out. Hence, the statistical results are bent towards the left-hand side. Hence, that the mean, median, and mode are always positive. In this distribution, Mean > Median > Mode. Where the kurtosis is also positive, that means distribution is peaked and possesses thick tails.

**SP and Weight (WT)**

**Use Q9\_b.csv**

Answer:



In **WT** the skewness is negative, that means the data points are more concentrated towards the right-hand side of the distribution. This makes the mean, median, and mode bend towards the right. Hence these values are always negative. In this distribution, Mode > Median > Mean. Whereas the kurtosis is positive, that means distribution is peaked and possesses thick tails.

In **SP** the skewness is positive , that means the values are more concentrated towards the right side, and the left tail is spread out. Hence, the statistical results are bent towards the left-hand side. Hence, that the mean, median, and mode are always positive. In this distribution, Mean > Median > Mode. Whereas the kurtosis is less than 3 (beta values is <3) so it is Platykurtic. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

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





Answer:

In the above HISTOGRAM, we see that it is positively skewed where the mean>median>mode. that means the values are more concentrated towards the right side, and the left tail is spread out. And the SECOND BIN of the histogram is the mode.

In the above BOX-PLOT, we see that the whisker is more in the upper side therefore it is a positively skewed i.e mean>median>mode. After the upper extreme we see there are some outliers.

**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:

94% : (143.54415570565965, 256.45584429434035)

98%: (130.15355671679083, 269.84644328320917)

96%: (138.34730111522666, 261.6526988847733)

**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.

Answer:

Mean = 41.0

Median = 40.5

Mode= 41

Standard Deviation = 5.05

Variance = 25.52

1. What can we say about the student marks?

Answer:

We can say that the student’s performance in all the subjects is the same and has scored same score which is 41 in few subjects.

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

Answer: The Distribution is Symmetric.

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

Answer: Positively Skewed

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

Answer: Negatively Skewed

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

Answer: It indicates that the distribution is peaked and possesses thick tails.

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

Answer: Distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

Answer: The data is assigned in De-assigned format

What is nature of skewness of the data?

Answer:

Left Skewed

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

Answer:

Q3-Q1

= 18 – 10

= **8 Is the IQR**

Q19) Comment on the below Boxplot visualizations?



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

Answer:

Here we see that the both boxplot 1 & 2 have the same median. But the data in boxplot1 is more dense as compared to the boxplot2 since the area of the boxplot 2 is more dispersed(IQR is more for Boxplot2 than boxplot1). Boxplot 2 has a normal distribution since the median is exactly in the middle of the box.

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)

ANSWER:

a. 0.34759392515827137

b. 0.7293498762151609

c. 0.8988689169682047

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANSWER:

Since the mean,median and mode aren’t equal the MPG column of the cars dataset doesn’t follow 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:

Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data, do not follow Normal Distribution as the mean,median and mode for both the columns aren’t equal.

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

ANSWER:

90%: 1.6448536269514722

94%: 1.8807936081512509

60%: 0.8416212335729143

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

ANSWER:

95%: 2.0638985616280205

96%: 2.1715446760080677

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 à pt(tscore,df)

df à degrees of freedom

ANSWER:

Probability: **0.32167**