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

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

Two heads and on tail combinations are (H,H,T) , (H,T,H) , (T,H,H) . So total 3 Combinations.

Probability = Obtained combinations / total number of combinations

P= 3 / (2\*2\*2)

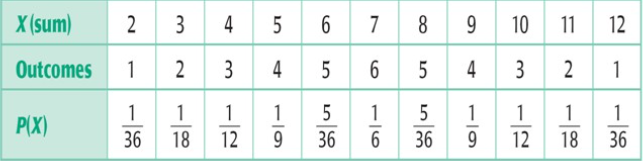
= 3/8

Probability = 0.375

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

Two dices are rolled the below are the Outcome's with sum:



1. P(X=1) = 0

Because the sum of the least two dices will start from 2 only. So The probability of having sum is zero

1. P(X<=4) = P(X=1) + P(X=2) + P(X=3) +P(X=4)

= 0 + 1/36 + 1/18 + 1/12

=0+0.0277+0.055+0.0833

=0.166

1. Sum is divisible by 2 and 3

Here sums I.e 2,4,6,8,10,12 which are divisible by 2

3,6,9,12 which are divisible by 3.

So in the common we have only two sum which is 6, 12 which can be divisible by 2 and 3 .

So p(x)= p(x=6)+p(x=12)

=5/36 + 1/36

=6/36

=0.166

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?

Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

=7C2 = (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.

=5C2 = (5×4)/(2×1)

=10

∴P(E) = n(E)/n(S)​= 10/21 = 0.4761

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 Value = sigma x\*p(x)

= x1.p(x1) + x2.p(x2) + ….+ xn.p(xn)

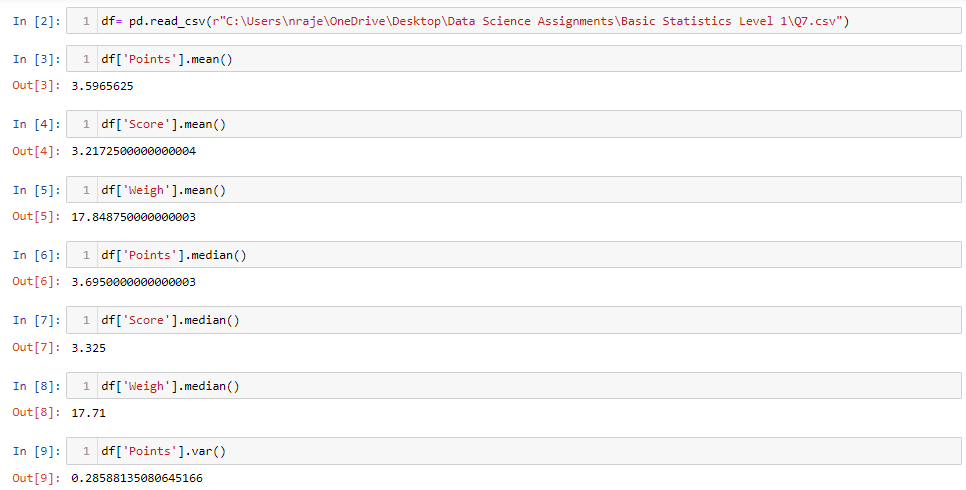
=(1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120)

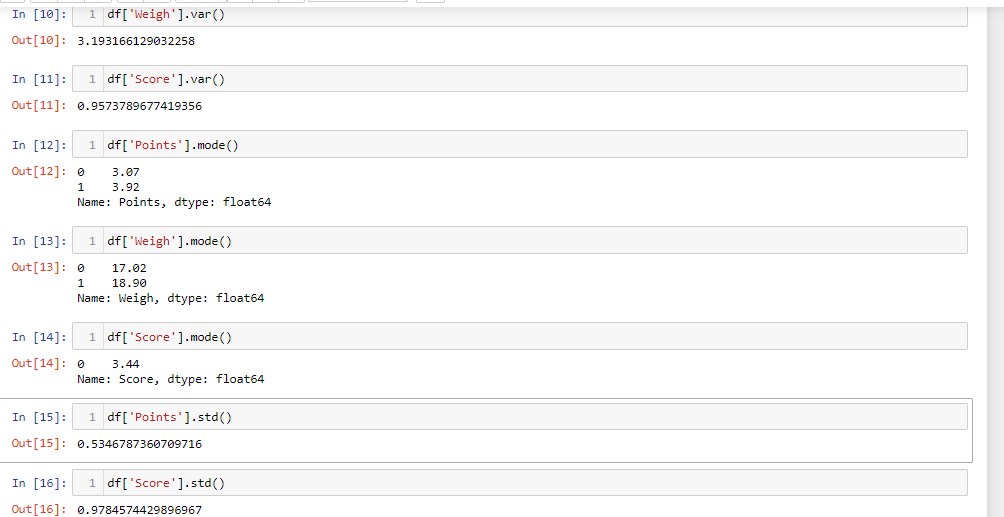
Expected Value = 3.09

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**

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:

Expected value=sigma x.p(x) =x1.p(x1) + x2.p(x2) + ….+xn.p(xn)

=1/9(108+110+123+134+135+145+167+187+199)

=1308/9

Expected value =145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answers attached in MAIL**

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



ANS: Right Side skewed or positively skewed



Ans : Positively skewed

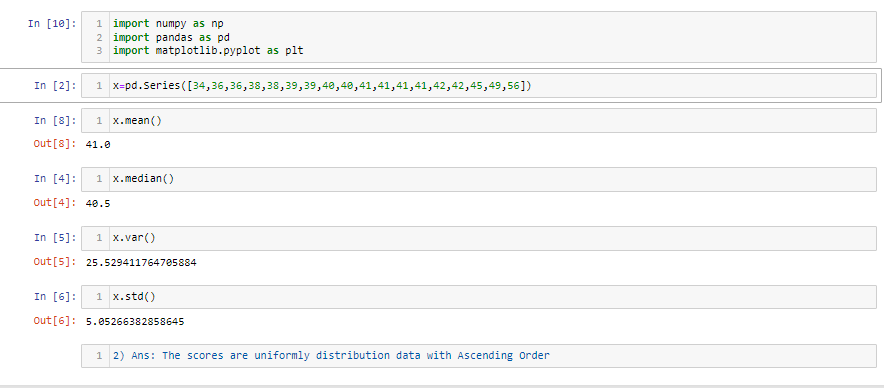
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Answer: python files are attached in mail

**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?



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

Ans: Perfect Symmetrical distribution

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

Ans: Right Skewness

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

Ans: Left skewness

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

Ans: LeptoKurtic Data or postive kurtosis.

Sharp peak in the plot. less gap between tails to x-axis

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

Ans : Flatter Data or Negative Kurtosis

Border peak under the curve and more gap between tails and x-axis

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



What can we say about the distribution of the data?

ANS: The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

ANS: The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

ANS: The Inter Quantile Range = Q3 Upper quartile – Q1 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.

Ans: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

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

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

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

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