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

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Ans:

Events: HHH, HHT, HTH, THH, HTT, TTH, THT, TTT

Total number of events: 8

Required Events: HHT, HTH, THH

Probability that two heads and one tail are obtained: 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

Ans:

1. Probability that sum is 1: 0
2. Probability that sum is less than or equal to 4: 6/36 = 1/6
3. Probability that sum is divisible by 2 and 3: 6/36

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:

Total number of events: 21

Probability of not getting blue ball: 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

Ans:

Expected value =

Expected number of candies for a randomly selected child: 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.

Ans:

.ipynb file is attatched

* Count of data is same in Points, Score, Weigh
* There is no similarity between mean, median, mode of Points, Score, Weigh
* There is no case which mean=median=mode

**Use Q7.csv file**

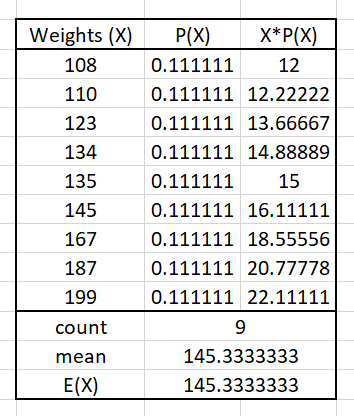
Q8) Calculate Expected Value for the problem below

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:

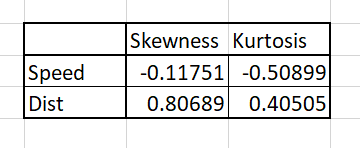


Expected Value of the Weight of that patient: 145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans:

* Both ‘dist’ and ‘speed’ are negatively skewed.

ie. ‘dist’ and ‘speed’ have more weight in left tail of the distribution

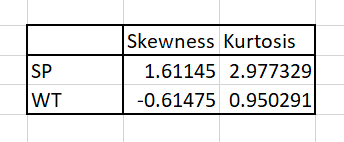
* Both ‘dist’ and ‘speed’ are positive kurtosis.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans:

.ipynb file is attatched

****

* SP is positively skewed where WT is negatively skewed.

ie. SP have more weight in the left tail of the distribution whereas WT have more weight in the right tail of the distribution.

* Both SP and WT have positive kurtosis

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



Ans:

* Histogram shows the distribution of weight of chicks.
* Most of the chicks have the weight between 50-100 followed by 100-150 and then 150-200
* Data is positively skewed as it has more weight on the left tail of the distribution.
* Hence the average weight is 50-200, chicks of weight <50 can be classified as under-weight and chicks of weight >200 can be classified as over-weight.



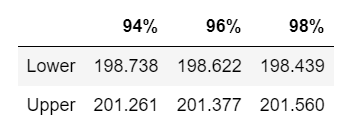
Ans:

* Data have upper long whisker
* Data is positively skewed
* Data have 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?

Ans:

.ipynb file is attatched



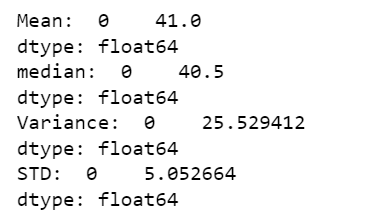
**Q12)** Below are the scores obtained by a student in tests

**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.

Ans:

.ipynb file is attatched



|  |  |
| --- | --- |
| Mean | 40.1 |
| Median | 40.5 |
| Variance | 25.53 |
| Standard Deviation | 5.052 |

1. What can we say about the student marks?

Ans:

* The students have an average mark of 40.1
* Have a maximum mark of 56 and a minimum mark of 34

ie. Marks range between 56 and 34

* Data has outliers

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

Ans:

When the mean and median of the data are equal then the distribution will be symmetrical. That is the distribution will be in a symmetrical bell shaped curve.

Skewness will be equal to zero.

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

Ans:

When the mean is greater than the median of the data, then the distribution will be positively skewed.

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

Ans:

When the median is greater than the mean of the data, then the distribution will be negatively skewed.

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

Ans:

Positive values of kurtosis indicate that the distribution is having long peaked center.

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

Ans:

Negative values of kurtosis indicate that the distribution is wider peaked in the center.

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



What can we say about the distribution of the data?

Ans:

Data is not symmetric.

Data is concentrated towards right side.

Negatively skewd data.

What is nature of skewness of the data?

Ans:

Negatively skewed data

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

IQR = 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:

Boxplot 1

* Data is normally distributed.
* Have no outliers.
* Center is at around 262.5
* Ranges between 237.5 and 287.5 approx

Boxplot 2

* Data is normally distributed.
* Have no outliers.
* Center is at around 262.5

Boxplot 1 and Boxplot 2 have almost same center

Boxplot 2 have a higher range than Boxplot 1

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)

Ans:

0.3475

* 1. P(MPG<40)

Ans:

0.7293

c. P (20<MPG<50)

Ans:

0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

The data follows normal distribution

.ipynb file is attatched

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

Dataset: wc-at.csv

Ans:

Both are normally distributed

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

Ans:

.ipynb file is attatched

|  |  |  |  |
| --- | --- | --- | --- |
|  | 90 % | 94% | 60% |
| Z-Score | -1.6448  1.6448 | -1.8807  1.8807 | -0.8416  0.8416 |

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

Ans:

.ipynb file is attatched

|  |  |  |  |
| --- | --- | --- | --- |
|  | 90% | 96% | 99% |
| t-Score | 1.3178 | 1.6121 | 2.4921 |

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:

* n=18
* df=18-1 =17
* mean=260
* std=90

t = -0.471404

P(X>=260 days) = 0.3216

.ipynb file is attatched