|  |  |
| --- | --- |
| 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 | CATEGORICAL |
| Number of kids | DISCRETE |
| Number of tickets in Indian railways | DISCRETE |
| Number of times married | DISCRETE |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

P(2H AND 1T) = 3/8 = .375 = 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

A . 0

B. .166 = 16.6%

C. .166 = 16.6%

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?

= 5 C 2 / 7C 2 = 10 / 21 =.476 = 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.

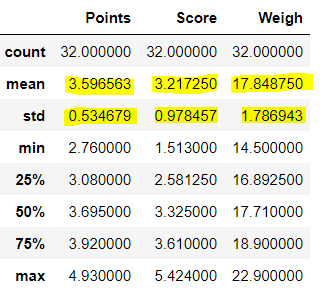
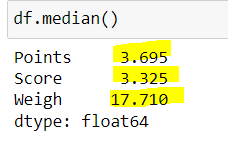
Child B – probability of having 4 candies = 0.20

Expected number of candies = 3.09 candies

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | POINTS | SCORE | WEIGHT |
| RANGE | 2.17 | 3.911 | 8.39 |
| Variance | 0.285 | 0.95 | 3.19 |

MODE = 0

**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?

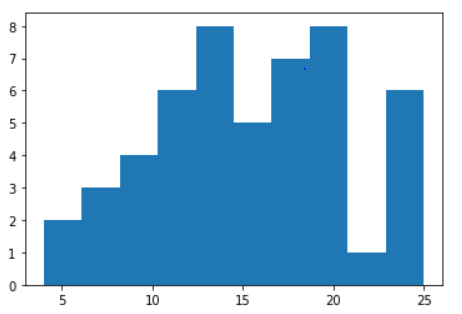
Expected Value of the Weight of that patient = 144.098 = 144

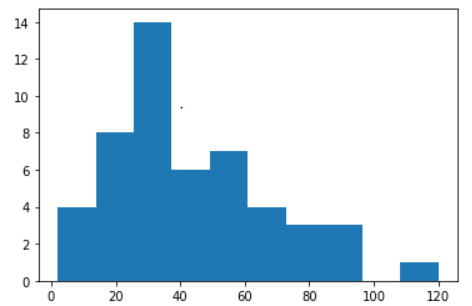
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

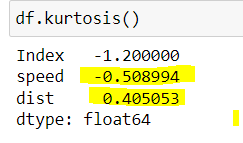
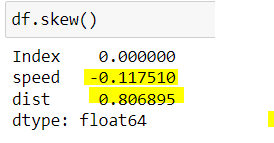
**Cars speed and distance Use Q9\_a.csv**

**SPEED - DISTANCE -**

1. **NEGATIVELY SKEWED POSITIVELY SKEWED**
2. **NEGATIVE KURTOSIS POSITIVE KURTOSIS**



**-**



**SP and Weight(WT)**

**Use Q9\_b.csv**

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



* MOST OF THE DATA LIES IN THE RANGE OF 0 TO 200
* DATA IS POSITIVELY SKEWED
* DATA HAS OUTLIERS



\* DATA HAS MANY OULIERS (POINTS ABOVE THE UPPER EXTREME).

\* OUTLIERS ABOVE UPPER EXTREME INDICATES DATA IS POSITIVEY 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?

Avg weight of Adult in Mexico with 94% CI – (198.74 , 201.26)

Avg weight of Adult in Mexico with 98% CI – (198.44 , 201.56)

Avg weight of Adult in Mexico with 96% CI – (198.62 , 201.38)

**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.

* MEAN - 41
* MEDIAN - 40.5
* VARIANCE - 25.52
* STANDARD DEVIATION - 5.05

1. What can we say about the student marks?

* Student marks has 2 outliers – 49 , 56
* It is positively skewed.

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

* Skewness will be zero (symmetric distribution of data)

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

* Skewness will be positive

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

* Skewness will be negative

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

* It indicates that distribution is peak and has thick tails

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

* It indicates that 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?



* Presence of higher values has dragged the mean towards upper whisker.
* Data is **not** normally distributed

What is nature of skewness of the data?

* Negatively skewed data (left Skewed)

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

IQR = Q3 – Q1

= 18 – 10

= 8 IQR

Q19) Comment on the below Boxplot visualizations?



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

* Both plots are normally distributed , where as Boxplot 2 is plotted for larger number of data’s 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)

33/81 = .407

* 1. P(MPG<40)

61/81 = .753

* 1. P (20<MPG<50)

69/81 = .851

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

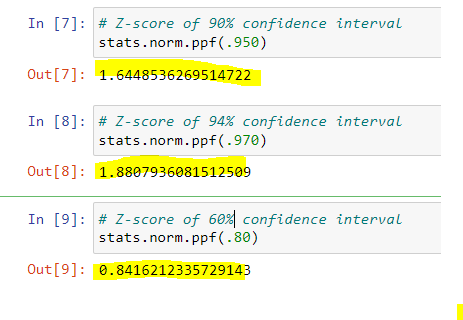
* MPG of cars doesn’t follow Normal Distribution ( No bell shaped curve)

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

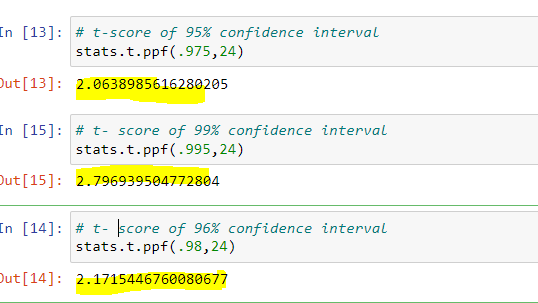
Dataset: wc-at.csv

* Both AT and WC doesn’t follow Normal Distribution

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

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

On calculation t score = -0.471 (formula) , df = n-1 = 18 -1 = 17

