|  |  |
| --- | --- |
| 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 | Ordinal |
| 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 |  |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans. 3/8 or .375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Ans. 0/36 probability is 0 because the lowest number of sum is 2.

1. Less than or equal to 4

Ans. 1/6

1. Sum is divisible by 2 and 3

Ans. 2/36 = 1/18

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. 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

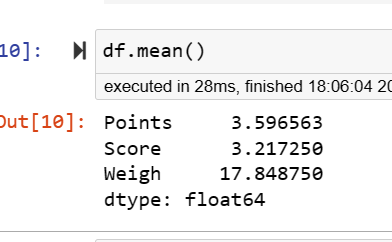
Ans. 3.14

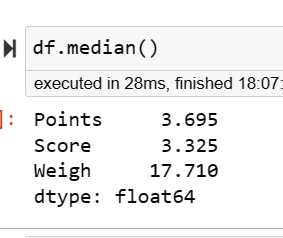
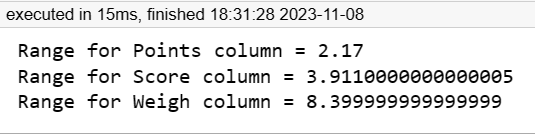
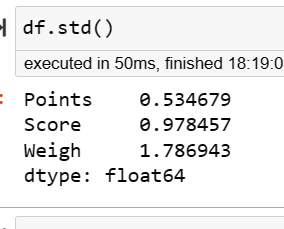
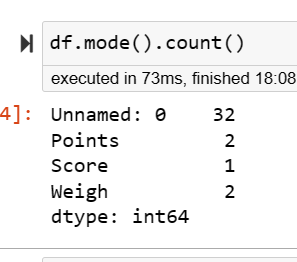
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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

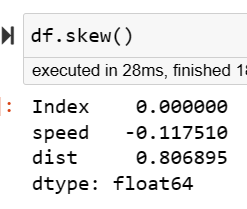
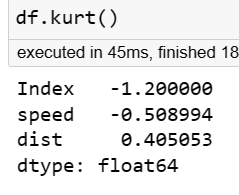
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans. 132.78

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

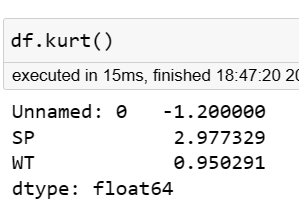
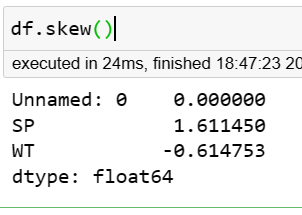
**Cars speed and distance**

**Use Q9\_a.csv**



**SP and Weight(WT)**

**Use Q9\_b.csv**

** **

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





Ans 10. The histogram is ‘Right skewed histogram’(positive skewed). It has some lesser observation which we called ‘outliers’.

And in the boxplot we have 5 points to notice which is lower limit(L.L),quartile 1 (q1),median,qurtile3(q3),upper limit(U.L) AND the outliers.

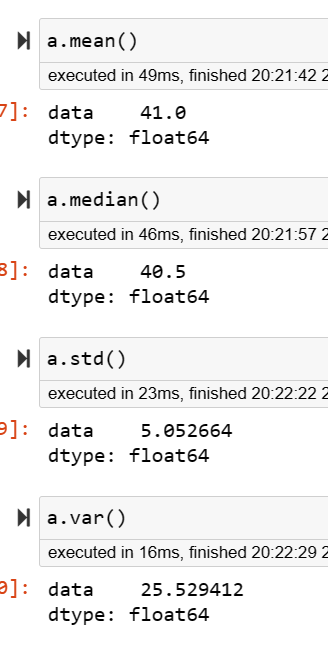
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

Ans. 98% = 2.3282 , 96% = 2.0551 , 94% = 1.8819

**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. 

* We can say that the average marks students have obtained is 41.0.
* we have variance is 25 there is some variability in our score. variety in our data (the more variety I have more information I get)
* The median we have is 40.5 that indicates oiur data is symmetrical.
* And the last but not least our Standard Deviation is 5.052664 it indicates dispersion of our score.

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

Ans. The nature of skewness is symmetric.

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

Ans. It indicates that the distribution is positively skewed.

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

Ans. It indicates that the distribution is negatively skewed.

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

Ans. The distribution has fatter tails and more peaked centre.

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

Ans. The distribution has lighter tails and flatter peak.

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



What can we say about the distribution of the data?

Ans. The data is skewed.

What is nature of skewness of the data?

Ans. The data is left skewed (neagative skewness).

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

Ans. IQR = Q3-Q1

= 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. The IQR, represent by the length of box, a longer box indicated greater variability so Boxplot 2 has greater variability compare to boxplot 1. Boxplot2 has higher median than boxplot1.

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. 0.3476

* 1. P(MPG<40)

Ans. 0.7293

* 1. P (20<MPG<50)

Ans. 0.8989

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. Yes, MPG of cars follows 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

Ans. No, the Adipose Tissue (AT) is right skewed not follows normal distribution. and Waist Circumference(Waist) is also not follows normal distribution.

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

Ans. 90% = 1.64 , 94% = 1.88 , 60% = 0.84

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

Ans. 95% = 2.063 , 96% = 2.17 , 99% = 2.79

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. Mean = 260 , population mean = 270 , N = 18 , std = 90 , df = 17

So there in this data the average bulb life may be no more than 270 days. Because t-statistical value(-0.47) is less than p-value(0.05) so we will have to reject null-hypothesis and accept alternate hypotyhesis area.