|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 Data |
| 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 | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans. S={HHH, HHT, HTT, HTH,THH, THT, TTH, TTT}

P(S)= 3/8 = 0.375 = 37.5%

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

1. Equal to 1

Ans: S= {}

P(X) = 0

1. Less than or equal to 4

Ans: S= {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

P(X) = 6/36=1/6

1. Sum is divisible by 2 and 3

Ans: S= {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

P(X) = 6/36=1/6

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?

Ans. P(blue) =2/7

P(no blue) = 1- p(blue)

=1-2/7

=5/7

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. E(X) = 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.12

= 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**

**Ans:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Points |  | Score |  | Weigh |
|  |  |  |  |  |  |
| Mean | 3.596563 |  | 3.21725 |  | 17.84875 |
| Median | 3.695 |  | 3.325 |  | 17.71 |
| Mode | 3.92 |  | 3.44 |  | 17.02 |
| Standard Deviation | 0.534679 |  | 0.978457 |  | 1.786943 |
| Sample Variance | 0.285881 |  | 0.957379 |  | 3.193166 |
| Range | 2.17 |  | 3.911 |  | 8.4 |

**Inferences:**

1. The distribution of the point and score is negatively skewed.(mean < median)

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. Probability = 1308/ 9

= 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**

**Ans.**

**a)**

|  |  |  |
| --- | --- | --- |
|  | speed | dist |
| Skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

**b)**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| skewness | 1.61145 | -0.61475 |
| kurtosis | 2.977329 | 0.950291 |

**Q10) Draw inferences about the following box plot & histogram**





Ans. Histogram:

* Chick weight data is right skewed or positively skewed.
* Most of the chick weight is between 50 to 100.
* More than 50% chick weight is between 50 to 150.

Box plot:

* There are outlier at upper side.

**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:

* Confidence interval at 94%

stats.norm.interval(0.94,200,30) = array([143.576, 256.424])

* Confidence interval at 98%

stats.norm.interval(0.98,200,30) = array([130.21, 269.79])

* Confidence interval at 96%

stats.norm.interval(0.96,200,30) = array([138.388, 261.612])

**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. 1)

* Mean

np.mean(x) = 41

* Median

np.median(x) = 40.5

* Variance

np.var(x)= 24.1111111

* Standard deviation

np.std(x)= 4.9103066

* Most of the students get **41** marks in their test. We don’t have outliers .

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

Ans. The distribution is symmetric.

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

Ans. The distribution is positively skewed.

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

Ans. The distribution is negatively skewed.

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

Ans. The distribution is peaked and possesses thick tail.

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

Ans. The distribution is less peaked and possesses thin tail.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans. 1) The data is not normally distributed the median is toward the higher value.

* The data is right skewed . The whisker range of minimum value is greater than maximum.
* The Inter Quartile Range = Upper quartile – lower quartile = 18 – 10 = 8

Q19) Comment on the below Box plot visualizations?



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

Ans. There are no outliers. Both the box plot has same median. They are normally distributed.

Q 20) Calculate probability from the given dataset for the below cases

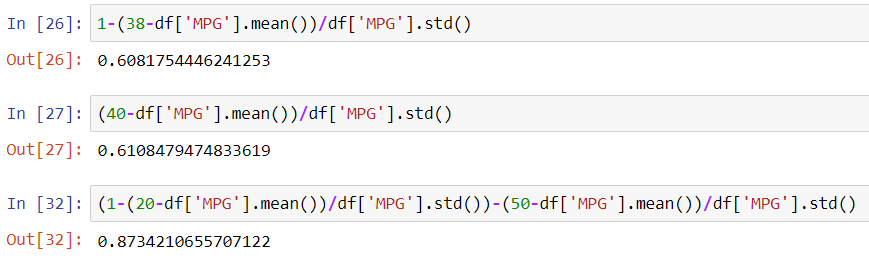
Data \_set: Cars.csv

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)

Ans.

a. P(MPG> 38) = 0.6081

b. P(MPG< 40) = 0.610

c. P(20<MPG<50) = 0.873

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. MPG of cars is not 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. Adipose Tissue(AT) and Waist Circumference are not follow normal distribution.

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

Ans. Z score of 90% = 1.28155 ( stats.norm.ppf(0.9))

Z score of 94% = 1.55477 (stats.norm.ppf(0.94))

Z score of 60% = 0.2533 (stats.norm.ppf(0.60))

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

Ans. t score of 95% = 1.71088 ( stats.t.ppf(0.95,24))

T score of 96% = 1.8280 (stats.t.ppf(0.96,24))

T score of 99% = 2.4921 (stats.t.ppf(0.99,24))

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(t score, df)

df 🡪 degrees of freedom

Ans.

Ho = Avg life of Bulb >= 260 days

H1 = Avg life of Bulb < 260 days

type 3

t\_val=(260-270)/(90/np.sqrt(18))

t\_val

= -0.4714

P\_val=stats.t.cdf(t\_val,17)

P\_val

= 0.3216

