|  |  |
| --- | --- |
| 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 | CATEGORICAL/NOMINAL |
| Number of kids | DISCRETE |
| Number of tickets in Indian railways | DISCRETE |
| Number of times married | DISCRETE |
| Gender (Male or Female) | CATEGORICAL/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 | 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer: Total possible events: 8

No of desired events : 3

Probability : 3/8 or 0.375

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

1. Equal to 1 = 0
2. Less than or equal to 4 = 3/36 or 1/12
3. Sum is divisible by 2 and 3 = 6 /36 or 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?

**Answer** : 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

**Answer**: Sum of (candies count + probability of given child) gives us expected number of candies for a randomly selected child

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

=**4.125**

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**

**Answer**

**Mean:**

* Points 3.596563
* Score 3.217250
* Weigh 17.848750
* dtype: float64

**Mode:**

* Points 3.92
* Score 3.44
* Weigh 17.02

**Median:**

* Points 3.695
* Score 3.325
* Weigh 17.710
* dtype: float64

**Range:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | MAX | MIN | RANGE |
| POINTS | 4.93 | 2.76 | 2.17 |
| SCORE | 5.424 | 1.513 | 3.911 |
| WEIGHT | 22.9 | 14.5 | 8.4 |
|  |  |  |  |

**Standard Deviation**

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

**Variance**

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

|  |
| --- |
|  |
|  |
|  |
|  |

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?

**Answers:**

Expected value of the weight of a patient: 145.33333333333334

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answers:**

Skewness:

speed -0.117510

dist 0.806895

dtype: float64

kurtosis:

speed -0.508994

dist 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answers:**

Skewness:

SP 1.611450

WT -0.614753

kurtosis:

SP 2.977329

WT 0.950291

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



**Histogram**

A histogram is a graphical representation of data that displays the distribution of values in different intervals or bins. It consists of a series of bars, where the height of each bar corresponds to the frequency or relative frequency of the data within that interval. Here the histogram is right skewed and the mean of the data is around 200-220 . median of the data is 180-200 . mode of the data is 50-100.



**BOXPLOT**

A boxplot, also known as a box-and-whisker plot, is a visual representation of the distribution of a dataset. It displays the median, quartiles, and potential outliers, providing a summary of the central tendency, spread, and skewness of the data. Here the data is negatively skew . Negative skew in a boxplot indicates that the tail of the distribution is elongated towards the lower values.

**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?

**Answers:**

For CI for 94%

stats.norm.interval(0.94,200,30/np.sqrt(2000))

(198.738325292158, 201.261674707842)

For CI for 98%

stats.norm.interval(0.98,200,30/np.sqrt(2000))

For CI for 96%

(198.43943840429978, 201.56056159570022)

stats.norm.interval(0.96,200,30/np.sqrt(2000))

(198.62230334813333, 201.37769665186667)

**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?

**Answers:**

Mean:

41.0

Median:

40.5

Variance:

25.529

STD:

5.052

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

ANSWER: Symentrically Distributed

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

ANSWER: Positive skewness

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

ANSWER: Negative skewness

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

ANSWER: positive kurtosis indicate the nature of a graph (i.e) in positive kurtosis the graph has thick tails (fat tails)

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

ANSWER: negative kurtosis indicate the nature of a graph (i.e) in negative kurtosis the graph has thin tails (lighter tails)

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



What can we say about the distribution of the data?

Answer: Here the data is left skewed

What is nature of skewness of the data?

Answer: Negative skewness

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

Answer: 8(Eight)

Q19) Comment on the below Boxplot visualizations?



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

1. Data is Normally Distributed. No Outliers. Center around 262.5. Comparatively, first graph has less range
2. Data is Normally Distributed. No Outliers. Center around 262.5

Comparatively, second graph has more 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)
  3. P (20<MPG<50)

ANSWERS:

A:0.3475939251582705

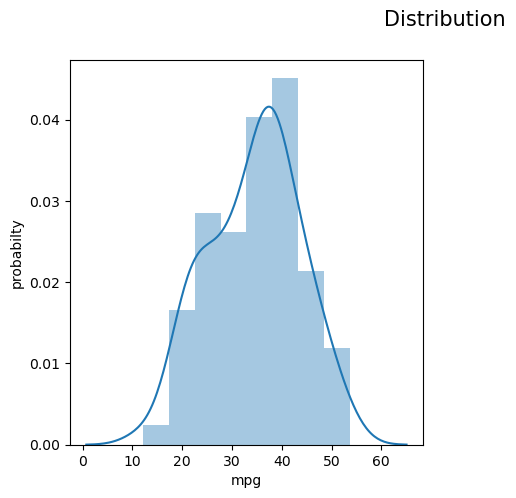
B: 0.7293498762151616

C: 0.8988689169682046

Q 21) Check whether the data follows normal distribution

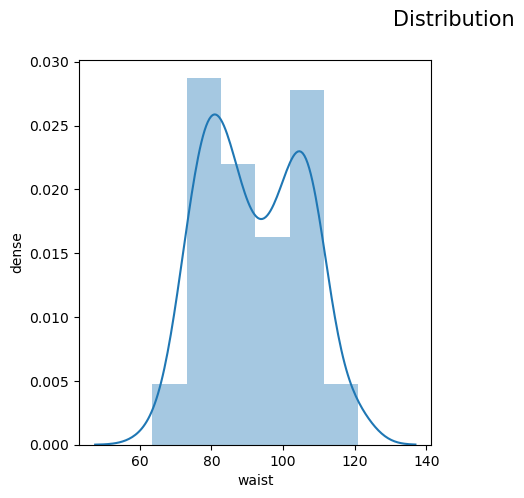
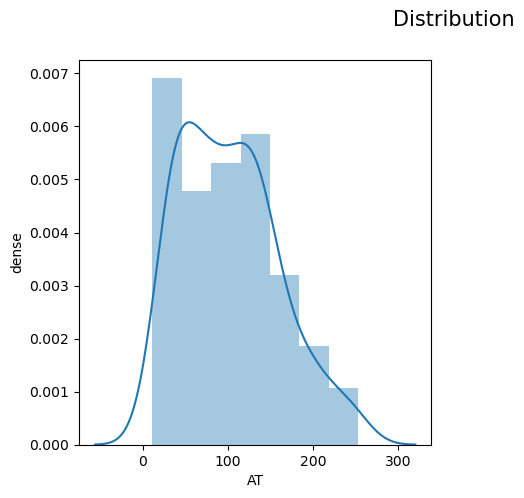
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

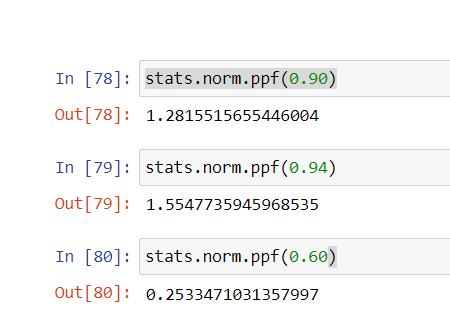


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

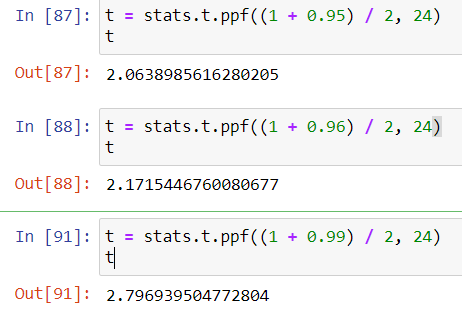
Dataset: wc-at.csv

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

(By using hypothesis testing)

