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

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

**Answer -:** P (Two heads and one tail) = 37.5%

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

1. Equal to 1

**Answer -: P(A) = 0**

1. Less than or equal to 4

**Answer -: P(A) = 0.166 =16.66%**

1. Sum is divisible by 2 and 3

**Answer -: P(A) = 0.166 = 16.66%**

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 -:**

N (Event (2 balls are drawn randomly from bag) = 7! / 2! \* 5!

= (7\*6\*5\*4\*3\*2\*1) /

(2\*1) \* (5\*4\*3\*2\*1)

N (Event (2 balls are drawn randomly from bag) = (7\*6)/ (2\*1) = 21

If none of them drawn 2 balls are blue = 7 – 2 = 5

N (Event (None of the balls drawn is blue) = 5! / 2! \* 3! = (5\*4) / (2\*1)

= 10

P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) /

N (Event (2 balls are drawn randomly from

bag)

= 10 / 21

**Answer = 0.4716**

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 -: 0.015+0.8+1.95+0.025+0.06+0.24 = 3.09**

**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.

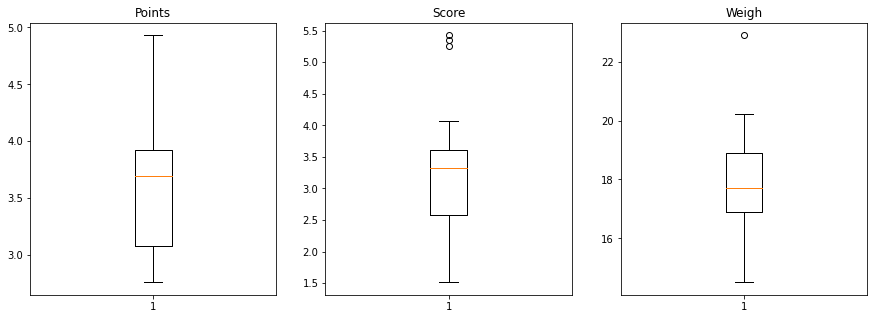
**Use Q7.csv file**

**Answer -:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Points | weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Mode | 3.695 | 3.325 | 17.71 |
| Median | 3.92 | 3.44 | 17.02 |
| Std | 0.53468 | 0.97846 | 1.78694 |

Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]

**Draw Inferences -:**

****

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?

**Answer -:** Expected value = Sum (X \* Probability of X)

= (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+

(1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

**Expected value = 145.33**

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

Cars speed and distance

**Use Q9\_a.csv**

**Answer -: Skewness - np.round(data.skew(),2)**

**Cars speed = -0.12**

**Cars distance = 0.81**

**Kurtosis -: np.round(data.kurt (),2)**

**Cars speed = -0.51**

**Cars distance = 0.41**

SP and Weight(WT)

Use Q9\_b.csv

**Answer -: Skewness - np.round(data.skew(),2)**

**Cars speed = 1.61**

**Cars weight = -0.61**

**Kurtosis -: np.round(data.kurt (),2)**

**Cars speed = 2.98**

**Cars weight= 0.95**

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

**HISTOGRAM -:**



-chick Weight data is right skewed or positively skewed

-more than 50% chick weight is between 50 to 150.

-most of the chick weight is between 50 and 100

-highest frequecy is near to 200.

-The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



**BOXPLOT**

-The data is right skewed .

-There are outliers at upperside

**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?

**Answer -:**

**Average weight of adult in mexico with 94% CI is =**np.round(stats.norm.interval(0.94,200,30/(2000\*\*0.5) ),4)

([198.7383, 201.2617])

**Average weight of adult in mexico with 98% CI is =**

np.round(stats.norm.interval(0.98,200,30/(2000\*\*0.5) ),4)

([198.4394, 201.5606])

**Average weight of adult in mexico with 96% CI is =**

np.round(stats.norm.interval(0.96,200,30/(2000\*\*0.5) ),4)

([198.6223, 201.3777])

**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.

**Answer -:**

**Mean -: 41.0**

**Median -: 40.5**

**Variance -: 25.5294117647**

**Standard deviation -: 5.052663828**

1. What can we say about the student marks?

**Answer -:**

Most of the student marks in between 36 to 42.

Maximum marks is 56.

Minimum marks is 34.

The students marks are good

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

**Answer -:** distribution is symmetric and it has zero skewness.

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

**Answer -:** positive/right nonparametric skewness.

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

**Answer-:** negative/left nonparametric skew

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

**Answer -:**  A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution

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

**Answer-:** A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

**Answer-:** This distribution say all data present between 10 to 18

Median value is 15 and maximum value is 18 and minimum value is

10. these distribution show left skewness because median is closer to

The upper quartile

What is nature of skewness of the data?

**Answer-:** distribution is negatively skewed

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

**Answer -: IQR = 8**

Q19) Comment on the below Boxplot visualizations?



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

**Answer -:**  First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range. second boxplot is more spread

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)

**Answer-: = 1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())**

**= 0.3475**

* 1. P(MPG<40)

**Answer-: = stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())**

**= 0.7293**

* 1. P (20<MPG<50)

**Answer-:**

**= stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())-**

**stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())**

**= 1.2430**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer-:**

**cars.MPG.median() = 35.1527**

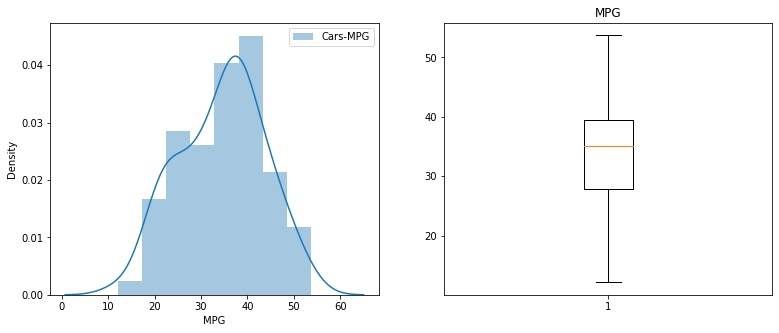
**cars.MPG.mean () = 34.4220**

**cars.MPG.std() = 9.1314**

**cars.MPG.var () = 83.3832**

**cars.MPG.min () = 12.10126289**

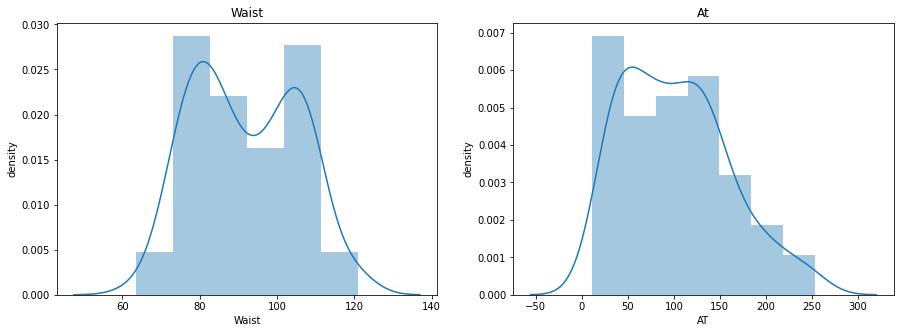
**cars.MPG.max () = 53.70068138**

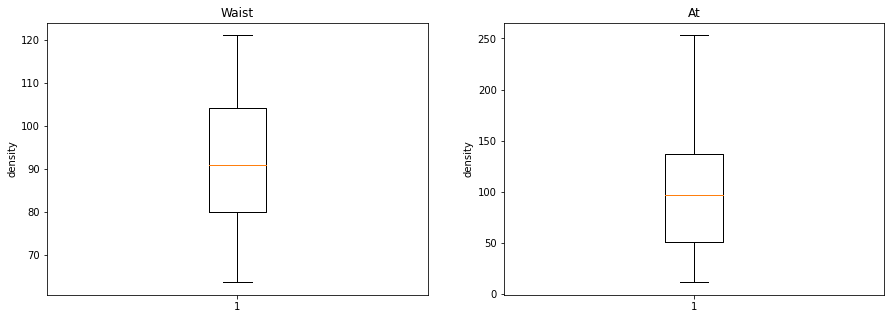
****

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

Dataset: wc-at.csv

|  |  |  |
| --- | --- | --- |
|  | Adipose Tissue (AT) | Waist Circumference(Waist) |
| Mean | 101.8940 | 91.9018 |
| Median | 96.54 | 90.8 |
| Std | 57.2947 | 13.5591 |
| Var | 3282.6898 | 183.8496 |
| Max | 253.0 | 121.0 |
| Min | 11.44ss | 63.5 |
|  |  |  |





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

**Answer-:**

**Z-score of 90% confidence interval**

**=stats.norm.ppf(0.95)**

**=1.6448**

**Z-score of 94% confidence interval**

**=stats.norm.ppf(0.97)**

**=1.8807**

**Z-score of 60% confidence interval**

**=stats.norm.ppf(0.8)**

**=0.8416**

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

**Answer-:**

1. **t scores of 95% confidence interval for sample size of 25**

**=stats.t.ppf(0.975,24)**

**=2.0638**

1. **t scores of 98% confidence interval for sample size of 25**

**=stats.t.ppf(0.98,24)**

**=2.1715**

1. **t scores of 99% confidence interval for sample size of 25**

**=stats.t.ppf(0.995,24)**

**=2.7969**

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

**Answer-:**

**import numpy as np**

**Import scipy as stats**

**t\_score = (x - pop mean) / (sample standard daviation / square root of sample size)**

**(260-270)/90/np.sqrt(18))**

**t\_score = -0.471**

**stats.t.cdf(t\_score, df = 17)**

**0.32 = 32%**