Q1) Identify the Data type for the Following:

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

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:**

When Three coins are tossed the total no of possible combinations are:

{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

i.e., Total no of outcomes = 8.

The number of combinations which have two heads and one tail are:

{HHT, HTH, TTH}

i.e., Number of favorable outcomes = 3.

We know that,

Probability= No of favorable outcomes **/** Total number of outcomes.

= 3/8

Probability= 0.375

Therefore, the probability of getting two heads and one tails in the toss of three coins is = 0.375

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

**Answer:**

if two dice are rolled, then total number of possible combinations = 36

1. **Equal to 1**

Total favorable outcomes (having sum =1) = 0.

As minimum sum is 2 for outcome (1,1).

Therefore, probability of having sum equal to 1 is 0.

1. **Less than or equal to 4**

Number of favorable outcomes (having sum <=4) = 6

Probability= No of favorable outcomes **/** Total number of outcomes.

= 6/36

Probability=1/6.

Therefore, probability of having sum less than or equal to 4 is 1/6.

1. **Sum is divisible by 2 and 3**

Number of favorable outcomes = 6

Probability= No of favorable outcomes **/** Total number of outcomes.

= 6/36

Probability=1/6.

Therefore, probability of having sum divisible by 2 and 3 is 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:**

Total no of balls = 7

No. of ways of drawing 2 balls out of 7 = 7C2 = (7\*6) / (2\*1) = 42/2 = 21

No of balls other than blue = 5

No of ways of drawing 2 balls out of 5 = 5C2 = (5\*4) / (2\*1) = 20/2 = 10

Probability = 5C2 / 7C2 = 10/21.

the probability that none of the balls drawn was blue is = 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:**

Expected number of candies for a randomly selected child

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

= 0.015+0.8+1.95+0.025+0.06+0.24

= 3.09

Therefore, 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**

**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** |
| **Variance** | **0.285881** | **0.957379** | **3.193166** |
| **Standard Deviation** | **0.534679** | **0.978457** | **1.786943** |
| **Max** | **4.93** | **5.424** | **22.9** |
| **Min** | **2.76** | **1.513** | **14.5** |
| **Range** | **2.17** | **3.911** | **8.4** |

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:**

Expected Value =∑ (Probability \* Value)

 ∑ P(x)\*E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Expected Value = (1/9) (108) + (1/9)(110) + (1/9)(123) + (1/9)(134) + (1/9)(135) + (1/9)(145) + (1/9)(167) + (1/9)(187) + (1/9)(199)

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

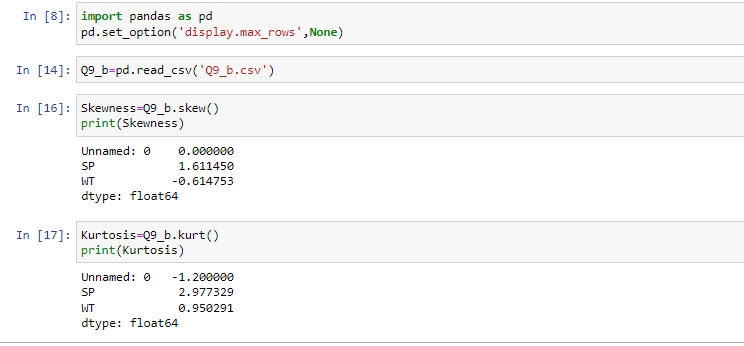
**Ans:**



**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**



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



**Ans:**

**Histogram: -**

* Data is not normally distributed.
* Chick weight data is right skewed (Positively skewed).
* The most frequent chick weight is between 50 to 100.
* More than 50% Chick weight is between 50 to 150.

  
**Boxplot: -**

* The Data is not normally distributed.
* The data is Right Skewed.
* There are outliers 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:**

Using the t-distribution,

The 94% confidence interval is (198.73, 201.27)

The 96% confidence interval is (198.61, 201.39)

The 98% confidence interval is (198.43, 201.57)

We have given standard deviation for sample (s) and we don’t know the standard deviation for population, that’s why we use t-distribution.

Given:

Sample mean = 200.

Sample standard deviation s = 30.

Sample size n = 2000.

The interval is:

In which t is critical value for the two-tailed confidence interval.

Considering 94% confidence level, with 200-1=199 df, the critical value is t=1.8916 ,

= 200-1.8916 =198.73

= 200+1.8916 =201.27

**The 94% confidence interval is (198.73, 201.27).**

Considering 96% confidence level, with 200-1=199 df, the critical value is t=2.0673 ,

= 200-2.0673 =198.61

= 200+2.0673 =201.39

**The 96% confidence interval is (198.61, 201.39).**

Considering 98% confidence level, with 200-1=199 df, the critical value is t=2.3452 ,

= 200-2.3452 =198.43

= 200+2.3452 =201.57

**The 98% confidence interval is (198.43, 201.57).**

**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 = 41.

Median = 40.5.

Variance = 25.52941.

Standard deviation = 5.052664.

1. The students marks are more consistent, so there is less variance in Sample Data. Low variance is ideal because we can better predict about students marks based on sample data.

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

**Ans:** Data is Normally Distributed and there is no skewness.

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

**Ans:** Positive Skewness (Right skewed)

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

**Ans:** Negative Skewness (Left Skewed)

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

**Ans:** Positive kurtosis indicates narrow pick and narrow tail. It indicates Minimum variance in Data.

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

**Ans:** Negative kurtosis indicates that thick peak and thick tail. It indicates Maximum variance in Data.

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



What can we say about the distribution of the data?

**Ans:** Not normally Distributed.

What is nature of skewness of the data?

**Ans:** Negative Skewness. (Left Skewed because median closer to upper quartile.)

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

**Ans:** IQR = Q3 – Q1 = 18-10 = 8.

IQR will be 8.

Q19) Comment on the below Boxplot visualizations?



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

**Ans:**

1. Median for Boxplots1 & Boxplot2 is to be appears approximately 262.5
2. Data in Boxplot2 is more Scattered as compared to Data in Boxplot1.
3. Boxplot1 range from a low of 240 to a high of 287, so their range is 287−240=4, whereas Boxplot2 range from lowest score is 180 and the highest approximately is 350, so their range is 350−180=170.
4. In Boxplot1 data points consistently hover around the center values that means less variable data in Boxplot1, whereas Boxplot2 imply more variable Data.

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)

**Ans:**

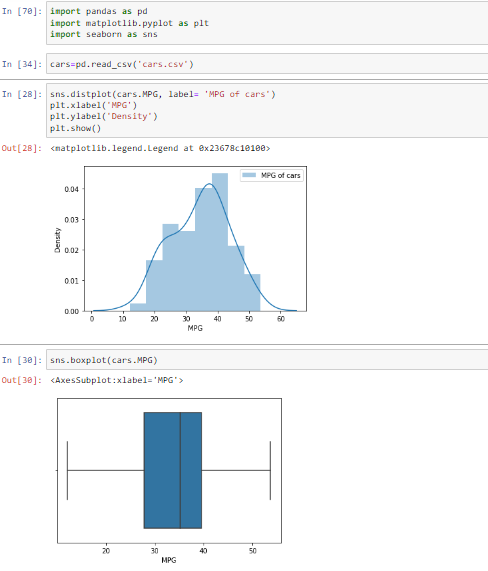


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:



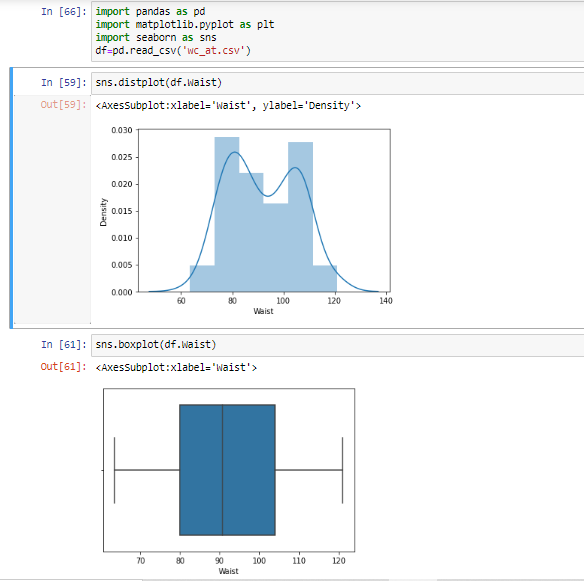
For MPG the left whisker is slightly larger than right whisker, Data is not normally distributed. Data is little bit left skewed.

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

Dataset: wc-at.csv

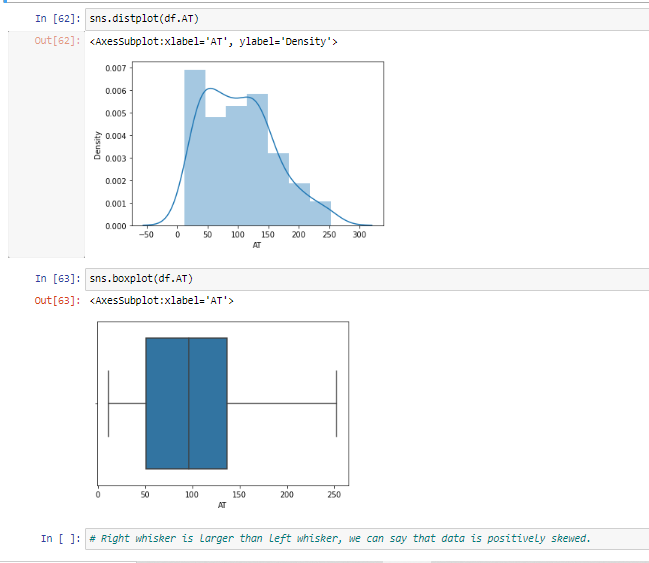
**Ans:**

For Waist Circumference(Waist)



For Waist Circumference Both the whisker are of same length, Median is slightly towards left, Data is fairly symmetric.

For Adipose Tissue (AT)



For Adipose Tissue Right whisker is larger than left whisker, we can say that data is positively skewed.

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

**Ans:**

|  |  |
| --- | --- |
| **Confidence interval** | **Z scores** |
| 60% | 0.253 |
| 90% | 1.645 |
| 94% | 1.555 |

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

**Ans:**

|  |  |
| --- | --- |
| **Confidence interval** | **t score** |
| 95% | 2.262 |
| 96% | 2.171 |
| 99% | 2.796 |

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:**

t-statistics for the data is given as follows:

x = sample mean = 260

µ=population mean=270

s=standard deviation of the sample=90

n= number of items in the sample = 18

t =

t =

t =

t =

t =

t = -0.471

For probability calculations the no of degrees of freedom is n-1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t<-0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained with 17 degrees of freedom and t score of -0.471, the probability of bulb lasting less than 260 days on average of **0.3218** assuming the mean life of the bulbs is **300 days.**