Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | **Discrete (integer)** |
| Results of rolling a dice | **Discrete(integer)** |
| Weight of a person | **Continuous (float)** |
| Weight of Gold | **Continuous (float)** |
| Distance between two places | **Continuous (float)** |
| Length of a leaf | **Continuous (float)** |
| Dog's weight | **Continuous (float)** |
| Blue Color | **Discrete(Boolean)** |
| Number of kids | **Discrete (integer)** |
| Number of tickets in Indian railways | **Discrete (integer)** |
| Number of times married | **Discrete (integer)** |
| Gender (Male or Female) | **Discrete (Boolean)** |

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 | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| 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?

**Probability formula=No. of interested events/no. of total events**

= **HHH, HHT, HTH, HTT,THH,THT,TTH,TTT =8**

**P= 3/8**

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

**When 2 dice rolled probability comes:**

**(1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6)**

**(2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6)**

**(3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6)**

**(4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6)**

**(5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6)**

**(6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6)**

a) **0%**

b) **1/6 (1,1)(1,2)(1,3)(2,1)(2,2)(3,1)**

c) **1/6** (**4 times 6, 1 time 12**)

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?

**Total event = 7 =7! / 2!\*5!**

**Interested event= 5 = 5! /2!\*3!**

**= 10/21 = 0.47**

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

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| **MEAN** | **3.596563** | **3.21725** | **17.84875** |  | **TOTAL MEAN** | **8.220854** |
| **MEDIAN** | **3.695** | **3.325** | **17.71** |  | **TOTAL MEADIAN** | **3.875** |
| **MODE** | **3.92** | **3.44** | **17.02** |  | **TOTAL MODE** | **3.15** |
| **STD DEVIATION** | **0.534679** | **0.978457** | **1.786943** |  | **TOTAL STD DEV** | **6.950394** |
| **VARIANCE** | **0.285881** | **0.957379** | **3.193166** |  | **TOTAL VAR** | **48.30797** |
| **MAX** | **4.93** | **5.424** | **22.9** |  | **TOTAL MAX** | **22.9** |
| **MIN** | **2.76** | **1.513** | **14.5** |  | **TOTAL MIN** | **1.513** |
| **RANGE** | **2.17** | **3.911** | **8.4** |  |  |  |

**Inference: Follow normal distribution**

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?

**=108+110+ 123+ 134+135+ 145+ 167+187+ 199/9=145.3333 (mean)**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Q9-A** |  | **Q9-B** |  |
|  | **Car speed** | **dist** | **sp** | **wt** |
| **skewness** | -0.11750986144663393  **-ve** | -0.5089944204057617  **-ve** | 1.6114501961773555  **+ve** | -0.6147533255357768  **-ve** |
| **kurtosis** | -0.11750986144663393  **-ve** | 0.4050525816795765  **+ve** | 2.9773289437871764  **+ve** | 0.9502914910300326  **+ve** |

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



**HISTOGRAM= +ve skewed and 350-400 data can be considered as outlier**

**BOXPLOT = +ve skewed , showing 7 outliers at U.E**

**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?

**94%= (198.7376089443071, 201.2623910556929)**

**96%= (198.6214037429732, 201.3785962570268)**

**98%= (198.4381860483216, 201.5618139516784)**

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.

|  |  |
| --- | --- |
| **41** | **mean** |
| **40.5** | **median** |
| **25.52941** | **variance** |
| **5.052664** | **std. dev** |

1. What can we say about the student marks?

= **Average marks of student lies in between 41 marks**

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

= **zero skewness , symmetrical distribution**

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

= **+ve skewed (right)**

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

**= -ve skewed (left)**

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

= **thin pick wide tail**

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

= **wide peak thin tail**

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



What can we say about the distribution of the data?

**= 50% of the data is in between 10-18 (IQR), 25% data 2-10(UQ), 25% data 18 above (LQ)**

What is nature of skewness of the data?

**= -ve skewed (left)**

What will be the IQR of the data (approximately)?   
  
= **IQR= (U.Q-L.Q) = 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.

|  |  |  |
| --- | --- | --- |
|  | **B1** | **B2** |
| **UQ** | **287.5** | **337.5** |
| **MEDIAN** | **212.5** | **212.5** |
| **LQ** | **250** | **200** |

**NO. OUTLIERS= 0**

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

**P(MPG>38)=** 0.34824060748417295

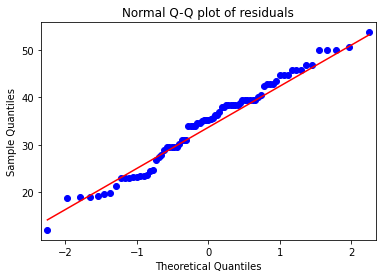
**P(MPG<40)=** 0.7293498762151616

**P (20<MPG<50) =** 0.8988689169682046

Q 21) Check whether the data follows normal distribution

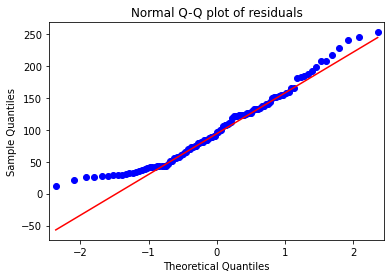
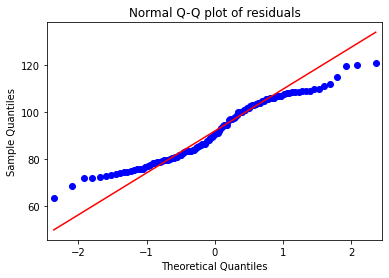
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

= **IT 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

= **IT FOLLOWS NORMAL DISTRIBUTION**

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

**= 1.6448536269514722, 1.8807936081512509, 0.8416212335729143**

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

= **2.0638985616280205, 2.1715446760080677, 2.796939504772804**

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

**probability**=**0.32167253567098364= 32%**

Rcode → pt (tscore, df)

df → degrees of freedom

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

print('critical value = ', np.round(t\_value, 2))

print('probabilty for average life of no more than 260 days is', np.round(stats.t.cdf(t\_value, df=17), 2))

**critical value = -0.47**

**probabilty for average life of no more than 260 days is 0.32**

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom