|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete & Continuous |
| 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 | Interval |
| Celsius Temperature | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ordinal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Answer = 3/8 or 0.375.

(3 coins are tossed the total number of possible combinations, that are HHH,TTT,HTH,THT,HHT,TTH,THH,HTT. Probability combination for 2 Heads and 1 tails are - HHT,HTH,THH)

P(2Head, 1Tail) = 3(HHT,HTH,THH) / 8(are HHH, TTT,HTH,THT,HHT,TTH,THH,HTT)

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

1. Equal to 1 – None p=0/n(36)=0
2. Less than or equal to 4 –1/12 event sum of >=4(1,3),(2,2),(3,1) P=3/n36=1/12
3. Sum is divisible by 2 and 3 = 1/6 or 0.166

Total Outcome is 36(n)

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 = 10/21 or 0.47(Probability=5C2/7C2)

Total no balls(7), Total combinations for drawn 2 ball(6)

7 x 6/2 x 1= 21

Number of other balls not blue=5, Total combinations for drawn 2 balls in 5(4)

5 x 4 / 2 x 1= 10

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 = 3.09

(1 x 0.015+4 x 0.20…..)

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset.

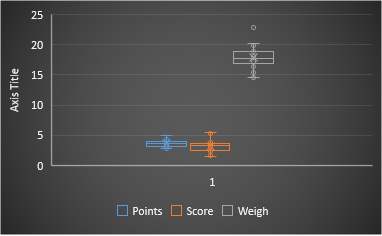
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Ans - For Points, Score, Weight>

Used excel formulas for the output. i.e for mean = avg formula. And for range get the Max & min value, used (Max-Min)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.597 | 3.217 | 17.849 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| 3.07 |  | 18.9 |
| Range | 2.17 | 3.911 | 8.4 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Std Deviation | 0.534679 | 0.978457 | 1.786943 |



Conclusion – Data is normally distributed., Std Dev of Weight is > point & Score

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 = 145.33(mean value)

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

**Cars speed and distance**

**Use Q9\_a.csv ANS=**

|  |  |  |
| --- | --- | --- |
|  | speed | dist. |
| **Skewness** | -0.11751 | 0.806895 |
| **kurtosis** | -0.50899 | 0.405053 |

* Using python-df.skew(axis=0,skipna = True).dropna() & df.kurtosis(axis=0,skipna = True).dropna()
* & With excel using skew & kurt formula.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans -**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| **Skewness** | 1.61145 | -0.61475 |
| **Kurtosis** | 2.977329 | 0.950291 |

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



**Ans =** Right Skewed (Positive), Highest frequency is 180 in the bin (Weight) 50-100. There are some outliers which you can identify in box plot.

Majority of the data is in 50-150(weight) with 100 -180 Frequency.

**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** | **Z Value** | **Range** |
| **94%** | 1.889 | 198.7383, 201.2616 |
| **98%** | 2.32 | 198.4394, 201.5605 |
| **96%** | 2.056 | 198.6223, 201.3776 |

(Using excel sheet for Calculations & Method used in python,scipy- stats.norm.interval(0.94,200,30/(2000\*\*0.5)))

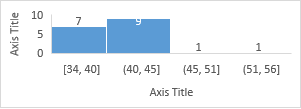
**Q12)** Below are the scores obtained by a student in tests

**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 =

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52941 |
| Std Deviation | 5.052664 |



* Majority of Students marks obtained between 38 to 43
* Skewness is 1.68 which is positive & right side

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

Ans= Zero Skew(Data Normally distributed)

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

Ans = Positive Skew, right sided.

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

Ans= Negative Skew, left sided.

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

Ans = Positive kurtosis implies closer peak & thick tails.

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

Ans = Negative kurtosis implies wider peak & thin tails

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



What can we say about the distribution of the data?

Ans= Not Normally distributed. Negative Skewed.

What is nature of skewness of the data?

Ans = Negative Skewed.

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

Ans = 8.1

IQR= Q3(18.3)-Q1(10.2)

Q19) Comment on the below Boxplot visualizations?



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

Ans = Box plot 1 data max distributed in range between 260-280 & Box plot 2 data max distributed in range between 225-315. Both box plot seen as a 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) Ans – 0.65240(38(x),34.422(m),9.1314(stdv))
  2. P(MPG<40) Ans – 0.72935

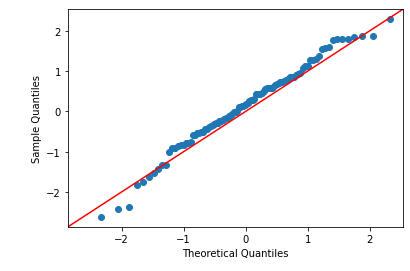
c. P (20<MPG<50) Ans – 1.24309

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans – Distributed normally.

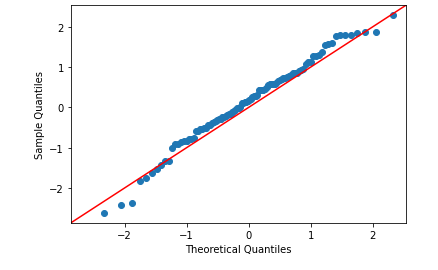


1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

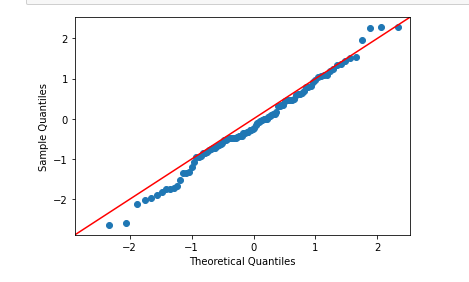
Dataset: wc-at.csv

Ans – Both are Normally distributed

Waist->



AT->



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

Ans- stats.norm.ppf(0.90)

|  |  |
| --- | --- |
| **Confidence Interval** | **Z score** |
| 90% | 1.281552 |
| 94% | 1.554774 |
| 60% | 0.253347 |

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

|  |  |
| --- | --- |
| **Confidence Interval** | **T score** |
| 95% | 2.063 |
| 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 = Probability of 18 bulbs life no more than 260 days is **0.0321**