|  |  |
| --- | --- |
| 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 |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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

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

Total possible outcomes = result^No of event= 2^3=8

Total possible outcomes=(HHH, TTT, HTT, THT,TTH,HHT,HTH,THH)

Favorable outcomes =3

Probability= 3/8=37.5%

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

Total possible outcomes = result^No of event= 6^2=36

a)0

b)6/36

c)6/36

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 outcomes=7c2=7\*6/2\*1=21

Favorable outcomes=5c2=5\*4/2\*1=10

Probability=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

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

=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.

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 |
| Range | 2.17 | 3.911 | 8.4 |

**Use Q7.csv file**

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?

Expected Value = ∑ (probability \* Value)

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

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Speed dist

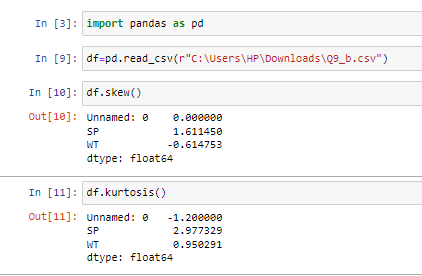
|  |  |  |
| --- | --- | --- |
| Skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

The value of skewness is more in distance, means data has more spread.

Value of kurtosis is more for speed means it has heavy tails, or outliers.

**SP and Weight(WT)**

**Use Q9\_b.csv**



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



Ans:The graph is positively skewed. It has outliers on 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?

Range: x-(z\*u)/n^(0.5)- x+(z\*u)/n^(0.5)

Stats.norm.interval(alpha=0.94,loc=200,scale=30/np.sqrt(2000))assign

|  |  |
| --- | --- |
| 198.7385 | 201.2615 |
| 198.4433 | 201.5567 |
| 198.6245 | 201.3755 |

**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?

|  |  |
| --- | --- |
| Mean | 41 |
| median | 40.5 |
| variance | 25.52941 |
| Std. dev | 5.052664 |
|  |  |

This plot is positively skewed.

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

It means that data is normalized. Also it is symmetrical.

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

Positively skewed distribution.

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

Negatively skewed distribution.

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

Leptokurtic: It means it has a sharper peak and heavier tails compared to normal distribution. That means it has less data near mean and more towards tail.

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

Platykurtic: It means it has a flatter peak and thinner tails compared to normal distribution. That means it has more data near mean and less towards tail.

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



What can we say about the distribution of the data?

Not Normally distributed

What is nature of skewness of the data?

Negatively skewed

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

10-18

Q19) Comment on the below Boxplot visualizations?



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

Both the plots looks symmetrical as mean and median are same. There is more range of data in box plot 2 than in 1.

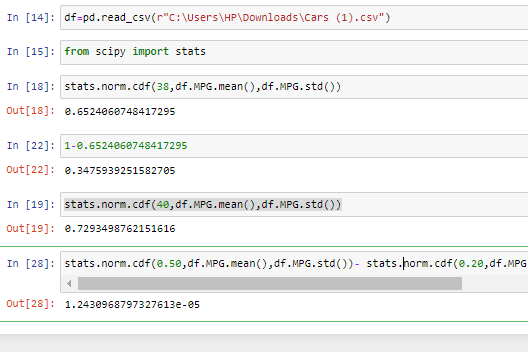
Q 20) Calculate probability from the given dataset for the below cases

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)



Q 21) Check whether the data follows normal distribution

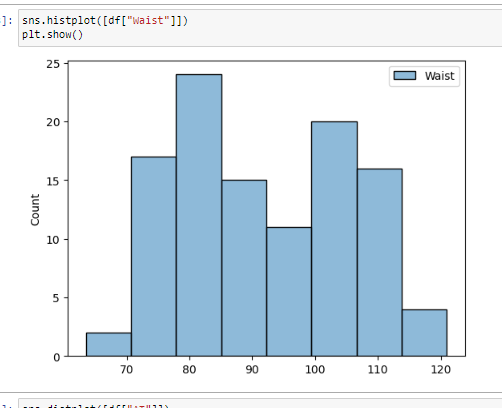
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

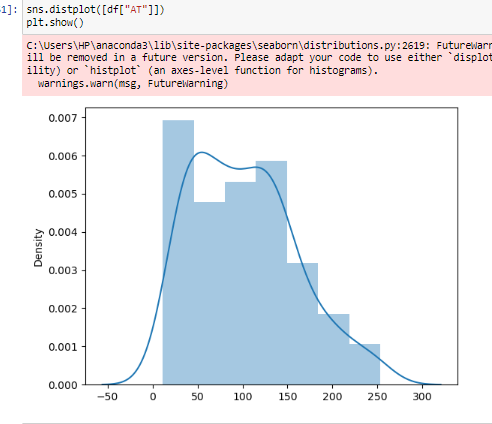
It does follow normal distribution as mean=median =mode for MPG.

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

Dataset: wc-at.csv

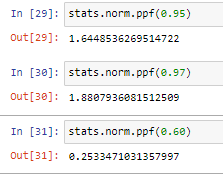


It does follow normal distribution.

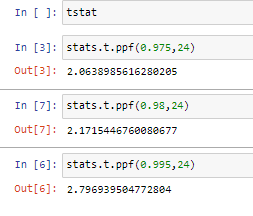


It is positively skewed.

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

x-u/6/n^0.5

(270-260)/(90/(18^0.5))

0.471

