|  |  |
| --- | --- |
| 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 | Ordinal |
| High School Class Ranking | Internal |
| Celsius Temperature | Internal |
| Weight | Ratio |
| Hair Color | Ordinal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Internal |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Internal |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| 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?

Ans: 3/8

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

1. Equal to 1 Ans: 0
2. Less than or equal to 4 Ans:6/36=1/6
3. Sum is divisible by 2 and 3. Ans: 2/36= 1/18

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: Probability of drawing a non-blue ball on first draw: 2/7+3/7=5/7. Now left with 6 balls with 2 blues.

Probability of drawing a non-blue ball on the second draw: 4/6=2/3

The probability that neither ball drawn blue is (5/7) \* (2/3) =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 Value= (1\*0.015) +(4\*0.20) +(3\*0.65) +(5\*0.005) +(6\*0.01) +(2\*0.120)

Ev= 0.015+0.80+1.95+0.025+0.06+0.24

Ev=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 | Median | Mode | Variance | Std Dev | Range |
| Points | 3.957 | 3.695 | 3.92 | 0.286 | 0.535 | 2.17 |
| Score | 3.212 | 3.325 | 3.44 | 0.933 | 0.966 | 3.83 |
| Weigh | 17.85 | 17.17 | 17.02 | 3.193 | 1.787 | 8.40 |

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?

Total Patients – 9

Probability of choosing 1 patient – 1/9

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

EV = 1/9(1308)

EV = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.11 | -0.50 |
| **Distance** | 0.80 | 0.40 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.61 | 2.97 |
| **WT** | -0.61 | 0.95 |

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



Ans :- The Histogram is right skewed, 200 chicks are having weight from 50 - 100



Ans:- Outliers are present at upper extreme, and the box plot is right skewed.

**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:- Sample Size = 2000**

**Mean = 200**

**Standard Deviation = 30**

Confidence Interval = x̄ ± z(σ/√n)

confidence interval of 94%= 200 ± 1.262, or from 198.7 to 201.3

confidence interval of 96%= 200 ± 1.378, or from 198.6 to 201.4

confidence interval of 98%= 200 ± 1.561, or from 198.4 to 201.6

**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 |
| Mode | 41 |
| Variance | 25.52 |
| Standard Deviation | 5.05 |

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

Ans:- Normal Distribution, Bell Curve

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

Ans:- Positive Skewness

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

Ans:- Negative Skewness

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

Ans:- Positive Kurtosis means the curve is more peaked and have less kurtosis than a normal distribution.

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

Ans:- Negative Kurtosis means the curve will is more peaked and have less kurtosis than a normal distribution.

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



What can we say about the distribution of the data?

Ans:- Not normally distributed and there is skewness in data.

What is nature of skewness of the data?

Ans:- Negatively Skewed.

What will be the IQR of the data (approximately)?   
Ans:- IQR = 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.

* In both Boxes there are no outliers.
* Both are normally distributed and there is not skewness in both boxplots.
* IQR of Boxplot 1 is less compared to Boxplot 2

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

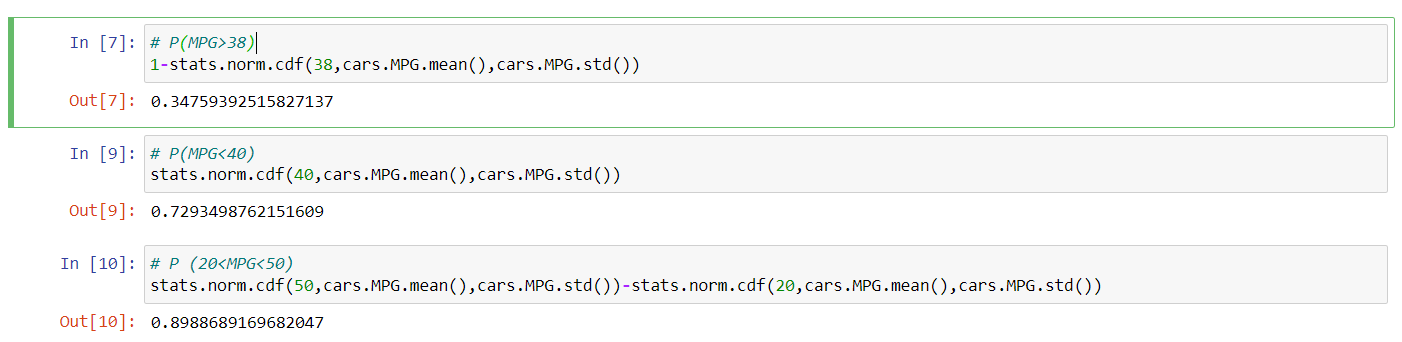
Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

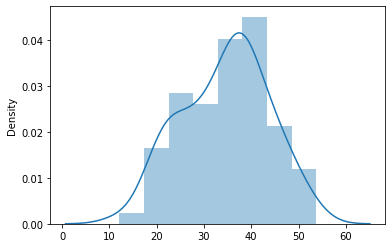


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

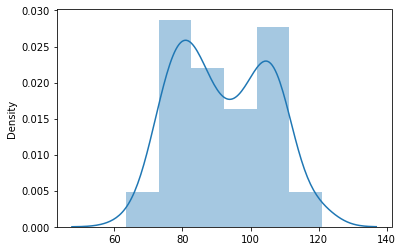
As mean, median and mode are not equal so MPG of cars follows Normal Distribution

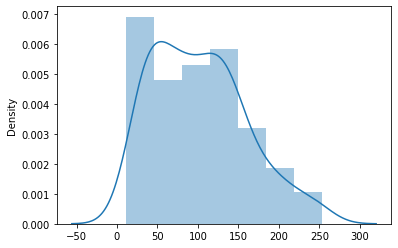


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Both do not follow Normal Distrubution.





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

Ans:- For 90% confidence interval:

Significance level at 5 % for two tailed test:

α = 5 % = 0.05, From Z table, Z value will be

z = 1.645.

For 94 % confidence interval:

Significance level at 3 % for two tailed test:

α = 3 % = 0.03, From Z table, Z value will be

z = 1.555.

For 60 % confidence interval, we get:

We have the significance level at 20 % for two tailed test:

α =20 % = 0.2, From Z table, Z value will be

z = 0.253

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

Sample size=25

Degree freedom = 25-1=24

With the help of t-score table

t scores of 95% confidence interval = 2.064

t scores of 94% confidence interval = 2.053

t scores of 99% confidence interval = 2.797

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

