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

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

Sol: The probability of two heads and one tail is =

=3/8 = 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

Sol: a) Probability = 0

b) Probability =

= 6/36 = 0.167

1. Probability =

= 6/36 = 0.167



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?

Sol: Total no. of balls =(2+3+2)=7

n(S) = No. of ways of drawing 2 balls out of 7 = 7C2 = =21

n(E) = No. of ways of drawing 2 balls out of (2+3) balls

= 5C2 = = 10

The probability that none of the balls drawn is blue = n(E)/n(S)

=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

Sol: Expected no. of candies for 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

=3.090



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

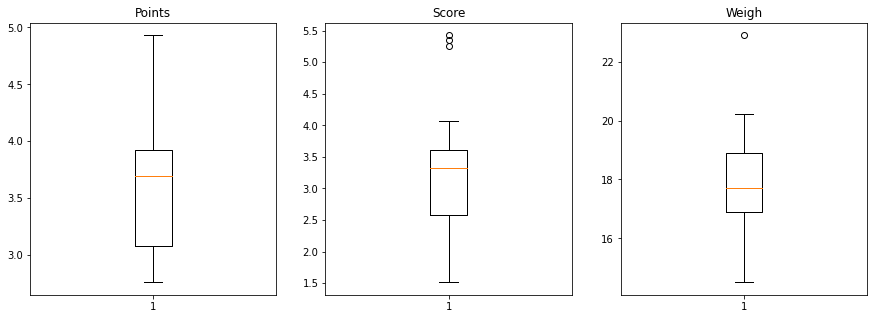
* For Points,Score,Weigh>

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

**Use Q7.csv file**

Sol:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.217250 | 17.848750 |
| Median | 3.695000 | 3.325000 | 17.710000 |
| Mode | 3.07  3.92 | 3.44 | 17.02  18.90 |
| Standard  Deviation | 0.534679 | 0.978457 | 1.786943 |
| Variance | 0.28588 | 0.95737 | 3.19316 |
| Range | 2.17 | 3.911 | 8.3999 |

****

# Mean value are closer for both points and 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?

Sol: Probability of selecting each patient (P(x)) = 1/9

Total weight of patients at the clinic (E(x)) = 1308 pounds

When one of the patient is chosen at random ,

The expected value of the weight of that person = P(x)\*E(x)

= (1/9)\*1308

=145.33



**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**

**Sol: Q9\_a.csv**

Skewness Kurtosis

Speed = -0.117510 speed = -0.508994

Dist = 0.806895 Dist = 0.405053

For speed skewness of data is negative, so it indicates that the data are skewed towards left and kurtosis of data is also negative, so it indicates that the distribution has lighter tails than the normal distribution.

For distance skewness of data is positive , so it indicates that the data are skewed towards right and kurtosis of data is positive ,so it indicates that the distribution is peaked and possesses thick tail.

**Q9\_b.csv**

Skewness Kurtosis

SP = 1.611450 SP = 2.977329

WT = -0.614753 WT = 0.950291 

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



**Sol:**

* In the above histogram the tail is on right. Mean is greater than median and its peak has right skew and have outliers on the higher side.
* The boxplot has outliers on the maximum 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?

**Sol:** Mean sample weighs(x) = 200 pounds

Random sample(n) = 2000 men

Samplestandard deviation(s) = 30 pounds

Interval = x±

(1-α)=probability of average weight of samples

94% confidence interval =[198.74 – 201.26]

98% confidence interval = [198.44 – 201.56]

96% confidence interval = [198.62 – 201.38]

****

**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?

**Sol: 1)**.



MEAN = 41

MEDIAN = 40.5

VARIANCE = 25.529

STANDARD DEVIATION = 5.053

**2)**. We can say that there is no outlier and the marks are slightly skewed

towards right because mean is greater than median.

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

**Sol:** When the mean and median of a data are equal , then the skewness is zero.

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

**Sol:** If the mean is greater than median , then it is positively skewed.

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

**Sol:** If the median is greater than mean ,then it is negatively skewed.

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

**Sol:** Positive kurtosis value indicate that distribution is peaked and possesses thick tail . An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

**Sol:** The negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**Sol:**

* In the above boxplot the median is towards the higher value , so it is not normally distributed and there might be influence of outlier.
* The data is skewed towards left .The whisker range of minimum value is greater than maximum.
* The IQR =Q3(Upper quartile)-Q1(Lower quartile)  
   =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.

**Sol:** In this distribution both the boxplot shares the same median which is in between the range of 250 to 275 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range and there are no outliers.

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)

**SOL:** a. P(MPG>38)= 0.348

b. P(MPG<40)= 0.729

c. P(20<MPG<50)= 0.013000000000000012

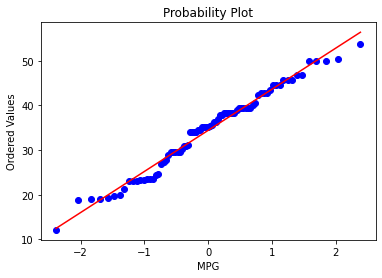


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

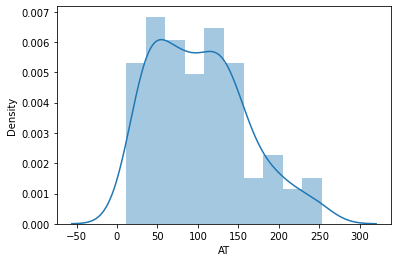
**Sol: a)** MPG of cars follow normal distribution.

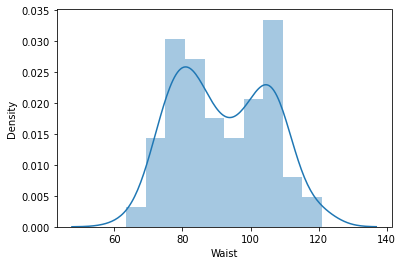
****

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

Dataset: wc-at.csv

**Sol: b)** The adipose tissue and circumference from the data set does not follow normal distribution.

****



****

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

**Sol :** For 90% confidence interval

Confidence level (cl) = 90/100 =0.90

Area = = = 0.95

stats.norm.ppf(0.95)

stats.norm.ppf(0.05)

For 94% confidence interval

Confidence level (cl) = 94/100 = 0.94

Area = = = 0.97

stats.norm.ppf(0.97)

stats.norm.ppf(0.03)

For 60% confidence interval

Confidence level(cl) = 60/100 = 0.60

Area = = = 0.8

stats.norm.ppf(0.8)

stats.norm.ppf(0.2)



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

**Sol:** For 95% confidence interval

Confidence level (cl) = 95% = 0.95

Significance level (α) = 1-cl = 1-0.95 = 0.05

Degree of freedom (df) =25-1 = 24

T(α/2) = T(0.05/2) =T(0.025) for a t-distribution with 24 degree of freedom

=-2.064

For 96% confidence interval

Confidence level (cl) = 96% = 0.96

Significance level (α) = 1-cl = 1-0.96 = 0.04

Degree of freedom (df) =25-1 = 24

T(α/2) = T(0.04/2)=T(0.02) for a t-distribution with 24 degree of freedom

=-2.172

For 99% confidence interval

Confidence level (cl) = 99% = 0.99

Significance level (α) = 1-cl = 1-0.99 = 0.01

Degree of freedom (df) =25-1 = 24

T(α/2) = T(0.01/2)=T(0.005) for a t-distribution with 24 degree of freedom

=-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

**Sol:** Sample size (n) = 18

Sample average (x) =260

Standard deviation(std)= 90

Average(µ) = 270

T-Score = (x-µ)/(std/)

= = -0.471

Degree of freedom(df) = 18-1 = 17

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days

= stats.t.cdf(T-score,df)

=0.32 or 32 %

