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

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

Answer: The total possible outcome 23 =8 HHH,HHT,HTT,THT,TTH,HTH,THH,TTT

Number of favorable outcomes =3

P(two head and one tail)=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

Solution:   
a) Total number of outcomes = 36  
 Number of outcomes sum equal to 1=0  
 p(equal to 1) = 0/36

b) Total number of outcomes =36  
 Number of outcomes less than or equal to 4 = 6  
 p(less than or equal to 4) = 6/36 =1/6

c) Total number of outcomes =36  
 Number of outcomes sum is divisible by 2 and 3 =6  
 p(sum is divisible by 2 and 3)=1/6

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?

Answer: Number of ways of drawing 2 balls out of 7 n(S)= 7C2 = 21  
 Number of ways of drawing 2 balls out of 5 n(S) =5C2 = 10  
 Required 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

Answer: Expected number of candies for a randomly selected child

= 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2.012

= 0.015+0.8+1.95+0.025+0.06+0.24

= 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**

**Answer : For Points:**

* Mean (points) =3.578709677419355
* Median (points) = 3.69
* Mode (points) = 3.92
* Variance (points) = 0.27694755859375
* Standard deviation (points) = 0.5346787360709715
* Range (points)= 2.17

Inference:

* The mean, Median & mode of this dataset are all approximately equal values. So, the dataset is normally distributed.
* Hence, the curve is bell shaped

**For Score:**

* Mean (score) =3.2172500000000004
* Median (score)= 3.325
* Mode (score) = 3.44
* Variance (score) = 0.927460875
* Standard deviation (score) =0.9784574429896967
* Range (score) = 3.9110000000000005

Inference:

* The Mean, Median, & Mode of this dataset are all approximately equal values. So the dataset is normally distributed.
* Hence, the curve is bell shaped

**For Weight :**

* Mean (weight) = 17.8487500000000003
* Median (weight) = 17.71
* Mode (weight) = 17.02 & 18.90
* Variance (weight) = 3.1933796874999997
* Standard deviation (weight) = 1.786943236096843
* Range (weight) = 8.3999999999999999

Inference:

* The Mean, Median & Mode of this dataset are all approximately equal values. So the dataset is normally distributed.
* Hence, the curve is bell shaped.

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 : Expected value = Summation (probability \*value)

= summation(p(x).E(x))

Probability of selecting each patient =1/9

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

=1/9(1308)

=145.33

The expected value is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer: Skewness of cars speed = -0.11395477

Skewness of cars distance = 0.78248352

1. Cars speed is left skewed because of negative skewness.
2. Cars distance is right skewed because of positive skewness.

Kurtosis of cars speed = -0.57714742

Kurtosis of cars distance= 0.24801866

1.cars speed is platykurtic because of negative kutosis.

2.cars distance is leptokurtic because of positive kurtosis.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: Skewness of cars SP = 1.58145368

Skewness of cars weight = -0.60330993

1. Cars SP is right skewed because of positive skewness.
2. Cars weight is left skewed because of negative skewness.

Kurtosis of cars SP= 2.72352149

Kurtosis of cars weight = 0.819465588

1. Cars SP is leptokurtic because of positive kurtosis.

2. Cars weight is leptokurtic because of positive kurtosis.

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



Answer: 1. More data points are concentrated in the interval 50-100 with 200 frequency.  
 2. Fewer data points are concentrated in the interval 350-400 with 10 frequency.  
 3. It is right skewed because of a long tail towards right side.

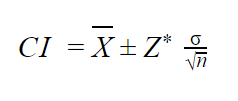


Answer : 1. There is more data points between Q3 and top point.  
 2. There is fewer data points between Q1 and bottom points.  
 3. There is outlier on the top side.

**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?

Answer: formula:

The confidence interval for data which follows a standard normal distribution is:



At 94% confidence intervals are (198.738325292158, 201.261674707842)  
 At 96% confidence intervals are (198.62230334813333, 201.377699665186667)  
 At 98% confidence intervals are (198.43943840429978, 201.56056159570022)

**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.

Answer: Mean= 41.0

Median= 40.5

Variance= 25.529411764705884

Standard deviation= 5.05266382858645

1. What can we say about the student marks?

Ans: There are 2 outliers in Students marks : 49 and 56 (inference by the box plot )

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

Answer: The nature of skewness is Symmetric.

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

Answer: The nature of skewness is positively skewed.

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

Answer: The nature of skewness is negatively skewed.

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

Answer: It indicated a distribution is peaked and possess thick tails

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

Answer: It indicates a distribution has flat and thin nails

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



What can we say about the distribution of the data?

Answer: Let’s assume above boxplot is about ages of students in a school.

50% of the people are above 10 years old and remaining are less.

And students whose age is above 15 are approximately 40%.

What is nature of skewness of the data?

Answer: Left skewed, median is greater than mean

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

Answer: Approximately = 8.

Q19) Comment on the below Boxplot visualizations?



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

Answer: By observing both the plots whisker’s level is high in boxplot 2, mean and

median is equal hence distribution is symmetrical.

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

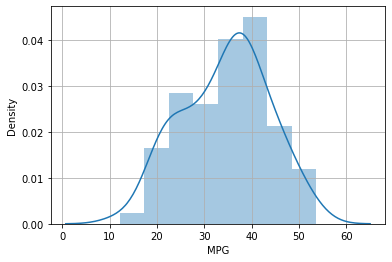
* 1. P(MPG>38)  
     1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())  
     =0.3475939251582705
  2. P(MPG<40)  
     stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())  
     =0.7293498762151616
  3. P (20<MPG<50)  
     stats.norm.cdf(0.50,cars.MPG.mean(),cars.MPG.std())-stats.norm.cdf(0.20,cars.MPG.mean(),cars.MPG.std())  
      = 1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: Mean= 34.422075728024666  
 Median= 35.15272697  
 Mode = 0 29.629936 dtype: float64



From above plot and values we can say that data is fairly symmetrical, i.e fairly normally distributed.

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

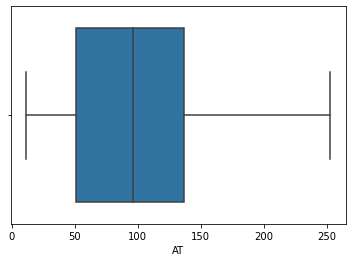
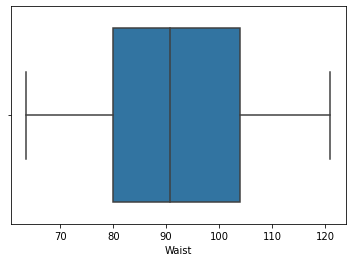
Dataset: wc-at.csv

Answer: Mean= Waist 91.901835, AT 101.894037, dtype: float64  
 Median = Waist 90.80, AT 96.54, dtype: float64  
 Mode = Waist AT

0 94.5 121.0

1 106.0 123.0

2 108.5 NaN

  
In Adipose Tissue (AT) , right whisker is larger than left whisker. So, we can say that AT is positively skewed i.e., doesn’t follow normal distribution  
  
   
In Waist Circumference (WC), both the whiskers are same size and median is fairly shifted to left-side. So, we can say that WC is fairly normally distributed i.e., fairly symmetrical.

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

Answer: Z Score of 90% confidence interval = 1.644853626951722

Z Score of 94% confidence interval = 1.8807936080512509

Z Score of 60% confidence interval = 0.8416212335729143

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

Answer: t scores of 95% confidence interval = 2.060

t scores of 96% confidence interval = 2.167

t scores of 99% confidence interval = 2.787

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

Answer :

> p <- pt((260-270)/(90/sqrt(18)), 18)

> (260-270)/(90/sqrt(18))

= 0.4714045