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

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

Ans: The Possible Outcomes:- TTT,HHH,HTT,THT,TTH,THH,HTH,HHT

Possible Outcomes = 2^3=8

The Probability of two heads and one tail are = HHT,HTH,THH

No. of outcomes = 3

Probability = 3/8

**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**

Ans: Two dices are rolled possible outcomes n= 36

1. The sum is equal to 1 = 0 because dice starts from 1
2. N=36

Probability that both the event occurs=2/36

Sum of less than or equal to 4

(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)

We have 6 favourable outcomes

So probability of the sum of both the dice is lesser than or equal to 4=6/36

(2/36)/(6/36)=1/3

1. Sum is divisible by 2 and 3

N=36

Divisible by 2 and 3

(1,5) (3,3) (4,2) (5,1) (6,6)

Favourable Outcomes=5

Probability = 5/36

**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?**

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

2 balls drawn out of 7 = 7c2

7\*6/2\*1 =21

None of the balls drawn is blue is = 5 balls

5c2 = 5\*4/2 = 10

P(e)=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 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

=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**

**Ans:**

**Points**

Mean =3.59 Variance =0.28

Median =3.69 Std.deviation =0.53

Mode =3.92 Range =2.17

**Score**

Mean =3.21 Variance =0.95

Median =3.32 Std.deviation =0.97

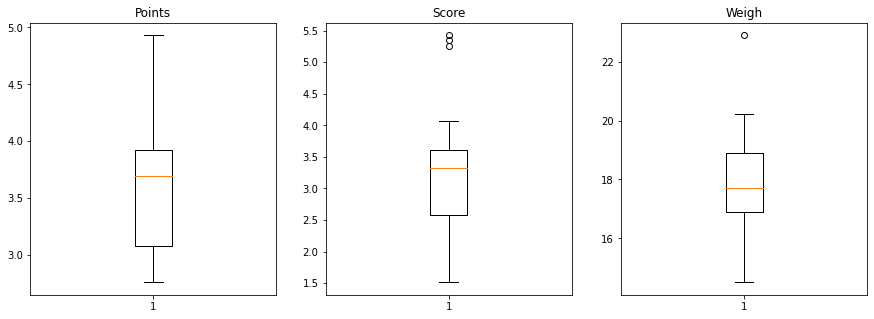
Mode =3.44 Range =3.911

**Weighs**

Mean =17.84 Variance =3.19

Median =17.71 Std.deviation =1.78

Mode =17.02 Range =8.4

****

**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?**

Ans: 108+110+123+134+135+145+167+187+199

=1308/9

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans: Speed Distance

Skewness= -0.11 Skewness= 0.80

Kurtosis= -0.50 Kurtosis= 0.40

Skewness and Kurtosis values are negative in speed data.

SP and Weight(WT)

**Use Q9\_b.csv**

Speed Weight

Skewness=1.61 Skewness= -0.614

Kurtosis= 2.977 Kurtosis= 0.95

Negative Skewness is present in Weight Data.

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



This Histogram diagram shows the measure of asymmetry in the distribution.

It is a positive right Skew.



The Median is closer to the bottom of the box and the whisker is shorter on the lower end of the box, hence the distribution 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:

94% of confidence interval

Sample weighs = 200

Std Dev = 30

N=2000

Confidence interval **c=94%**=0.94

α = 1-0.94 = 0.06

z= α/2=0.06/2=0.03=0.97

z=1.88

200+1.88(30/√2000)=201.259

200-1.88(30/√2000)=198.740

Confidence interval **c=96%**=0.96

α = 1-0.96 = 0.04

z= α/2=0.04/2=0.02=0.98

z=2.05

200+2.05(30/√2000)=201.37

200-2.05(30/√2000)=198.627

Confidence interval **c=98%**=0.98

α = 1-0.98 = 0.02

z= α/2=0.02/2=0.01=0.99

z=2.32

200+2.32(30/√2000)=201.55

200-2.32(30/√2000)=198.44

**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: 1) mean= 41

Median=40.5

Variance=25.5

Std dev=5.05

2) mean of marks of student is 41 which is slightly greater than median.

Most of the students got marks in between 41-42.

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

**Ans:** If the mean is equal to the median as well as the mode, hence **the skewness is zero**. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.

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

**Ans:** If the mean is greater than the median, the distribution is positively skewed

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

**Ans:** If the mean is less than the median, the distribution is negatively skewed.

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

**Ans:** Positive values of kurtosis indicate that **distribution is peaked and possesses thick tails**. 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?**

**Ans:** Negative values of kurtosis indicate that **a distribution is flat and has thin tails**.

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



**What can we say about the distribution of the data?**

When the median is closer to the top of the box, and if the whisker is shorter on the upper end of the box, then the data is distributed negatively.

**What is nature of skewness of the data?**

The distribution is negatively skewed (skewed left).

**What will be the IQR of the data (approximately)?**Q3=18

Q1=10

IQR= Q3-Q1

=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.**

The medians is in the same level of data for Boxplot 1 and Boxplot 2. However the box plots in these examples show very different distributions of views.  
It always important to consider the pattern of the whole distribution of responses in a box plot.

**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: 1-stats.norm.cdf(38,cars.MPG.mean(),cars.MPG.std())

0.347593

* 1. **P(MPG<40)**

Ans: stats.norm.cdf(40,cars.MPG.mean(),cars.MPG.std())

0.729349

**c. P (20<MPG<50)**

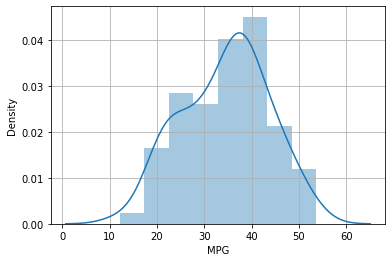
Ans: stats.norm.cdf(50,cars.MPG.mean(),cars.MPG.std())-(1-stats.norm.cdf(20,cars.MPG.mean(),cars.MPG.std()))

0.0131

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

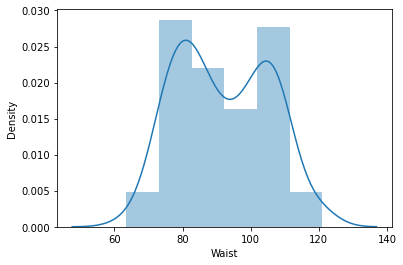
**Dataset: Cars.csv**



Data of Cars MPG 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**



-Waist Circumference(Waist) normally distributed.

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

Ans: Z scores of 90% = 1.644

Z scores of 94% = 1.880

Z scores of 60% = 0.8416

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

Ans: t scores of 95% = 2.0638

t scores of 96% = 2.171

t scores of 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: 0.3216