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

**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 | 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:

three coins are Tossed.

(HHH),(TTT),(TTH),(THT),(THH),(HHT),(HTH),(HTT)

Therefore Total No Of Events (N)=8

Condition Is Two Heads And One Tail

(THH),(HHT),(HTH)

Interested Events = 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 Dice Are Rolled.

Possible Cases Are

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

* A: Sum Is Equal To 1

Is There is no Case That’s Sum IS 1

So Probability Is 0

* B: Sum Is Less Than OR Equal To 4

Possible cases are (1,1)(1,2)(1,3)(2,1)(2,2)(3,1)

Total Interesed Events 6

Probability : 6/36=1/6=0.16666

* C: Sum Is Divisible by 2 and 3

Possible cases are (1,5)(2,4)(3.3)(4,2)(5,1)(6,6)

Total Interested Events = 6

Probability : 6/36=1/6=0.1666

**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

No. of ways of drawing 2 balls out of 7 = 2C7

n(E)= 2C7 =(7\*6\*5\*4\*3\*2\*1)/(2\*1) =21

No. of balls drawing which is not blue i.e.(2+3) = 5C2

n(S) **=**5C2 = (5\*4\*3\*2\*1)/(2\*1) =10

Probability = No. of possible events / Total no. of events

= 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**

Ans :

Expected number of candies = (1\*0.015) + (4\*0.2) + (3\*0.65) + (5\*0.005) +

(6\*0.01) + (2\*0.12)

***Expected number of candies = 3***

**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.**

Ans:

**For Importing File**

**R Studio**: q7 <- read.csv(file.choose())

View(q7)

**Python**:

q7 = pd.read\_csv('Q7.csv')

# Points:

**Mean**:

**R Code:**

> mean(q7$Points)

[1] 3.596563

**Python Code:**

np.mean(q7.Points)

Out[15]: 3.5965625000000006

**Median:**

**R Code:**

> median(q7$Points)

[1] 3.695

**Python Code:**

np.median(q7.Points)

Out[19]: 3.6950000000000003

**Mode :**

**R Code:**

> nummode <- function(x){

+ uniquv <- unique(x)

+ uniquv[which.max(tabulate(match(x,uniquv)))]

+ }

> nummode(q7$Points)

[1] 3.92

**Python Code:**

**Varience:**

**R Code:**

> var(q7$Points)

[1] 0.2858814

**Python Code:**

np.var(q7.Points)

Out[11]: 0.27694755859375003

**Standarad Daviation:**

**R Code:**

> sd(q7$Points)

[1] 0.5346787

**Python Code:**

np.std(q7.Points)

Out[12]: 0.5262580722361891

**Range:**

**R Code:**

> range(q7$Points)

[1] 2.76 4.93

> rangevalue <- function(x){max(x)-min(x)}

> rangevalue(q7$Points)

[1] 2.17

**Python Code**

# range =max(q7['Points'])- min(q7['Points'])

# range

# Out[25]: 2.17

# Score:

**Mean**:

**R Code:**

> mean(q7$Score)

[1] 3.21725

**Python Code:**

np.mean(q7['Score'])

Out[21]: 3.2172499999999995**:**

**Median:**

**R Code:**

median(q7$Score)

[1] 3.325

**Python Code:**

np.median(q7['Score'])

Out[26]: 3.325

**Mode**:

**R Code:**

> nummode(q7$Points)

[1] 3.92

**Python Code:**

**Varience:**

**R Code:**

> var(q7$Score)

[1] 0.957379

**Python Code:**

np.var(q7.Score)

Out[29]: 0.927460875

**Standarad Daviation:**

**R Code:**

> sd(q7$Score)

[1] 0.9784574

**Python Code:**

np.std(q7.Score)

Out[30]: 0.9630477013107918

**Range:**

**R Code:**

> range(q7$Score)

[1] 1.513 5.424

> rangevalue(q7$Score)

[1] 3.911

**Python Code**

range =max(q7['Score'])- min(q7['Score'])

range

Out[23]: 3.9109999999999996

# Weigh:

**Mean**:

**R Code:**

> mean(q7$Weigh)

[1] 17.84875

**Python Code**:

np.mean(q7['Weigh'])

Out[31]: 17.848750000000003

**Median:**

**R Code:**

> median(q7$Weigh)

[1] 17.71

**Python Code:**

np.median(q7['Weigh'])

Out[32]: 17.71

**Mode**:

**R Code:**

> nummode(q7$Weigh)

[1] 17.02

**Python Code:**

**Varience:**

**R Code:**

> var(q7$Weigh)

[1] 3.193166

**Python Code:**

np.var(q7['Weigh'])

Out[33]: 3.0933796874999997

**Standarad Daviation:**

**R Code:**

> sd(q7$Weigh)

[1] 1.786943

**Python Code:**

np.std(q7['Weigh'])

Out[34]: 1.758800638929836

**Range:**

**R Code:**

> range(q7$Weigh)

[1] 14.5 22.9

> rangevalue(q7$Weigh)

[1] 8.4

**Python Code:**

range =max(q7['Weigh'])- min(q7['Weigh'])

range

Out[35]: 8.399999999999999

**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:

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

=145.333

It is not necessary for an expected value lies between given sample

space. In above case expected value is 145.333 which is not in sample space.

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

**Cars speed and distance**

**Ans:**

> q9 <- read.csv(file.choose())

Python

q9 = pd.read\_csv('Q9\_a.csv')

**Speed**

**Skewness**:

**R Code:**

> skewness(q9$speed)

[1] -0.1139548

**Python Code:**

q9['speed'].skew()

Out[9]: -0.11750986144663393

**Kurtosis**:

**R Code:**

> kurtosis(q9$speed)

[1] 2.422853

**Python Code:**

q9['speed'].kurt()

Out[12]: -0.5089944204057617

**Distance**

**Skewness**:

**R Code:**

> skewness(q9$dist)

[1] 0.7824835

**Python Code:**

q9['dist'].skew()

Out[13]: 0.8068949601674215

**Kurtosis**:

**R Code:**

> kurtosis(q9$dist)

[1] 3.248019

**Python Code:**

q9['dist'].kurt()

Out[14]: 0.4050525816795765

Inferences:



**Q.9 b : SP and Weight(WT)**

**Sp**

**Skewness**:

**R Code:**

> skewness(Q9$SP)

[1] 1.581454

**Python Code:**

q9['SP'].skew()

Out[17]: 1.6114501961773555

**Kurtosis**:

**R Code:**

> kurtosis(Q9$SP)

[1] 5.723521

**Python Code:**

q9['SP'].kurt()

Out[18]: 2.9773289437871764

**Weight:**

**Skewness**:

**R Code:**

> skewness(Q9$WT)

[1] -0.6033099

**Python Code:**

q9['WT'].skew()

Out[19]: -0.6147533255357768

**Kurtosis**:

**R Code:**

> kurtosis(Q9$WT)

[1] 3.819466

**Python Code:**

q9['WT'].kurt()

Out[20]: 0.9502914910300326



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



**Ans :**  1) The most of the points lies between 50-100 range having maximum frequency 200.

2) Above Histogram is positively Skewed because long tail is towards right side .

3) Expected Value for above graph is 75



Ans:

1. Negatively Skewed data ,because long tail is towards left side.
2. Median < Mean , and outlier is on upper side of box plot.
3. Less no. of points lies between Q1 and bottom point

**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:**

**Formula= X+/- Z(1-α) \*(σ/√n)**

Degrees of freedom= 2000-1=1999

Confidence Interval= **94% = 1.89**

Confidence Interval= **98% = 2.33**

Confidence Interval= **96% = 2.06**

**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 :

In Python

p = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

**Mean**:

np.mean(p)

Out[26]: 41.0

**Median:**

np.median(p)

Out[27]: 40.5

**Varience:**

np.var(p)

Out[28]: 24.11111111111111

**Standard Deviation :**

np.std(p)

Out[29]: 4.910306620885412

Students gets marks between 34 and 56.

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

Ans: symmetrical Skewness

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

**Ans:**

If mean > median

Then Skewness Is Positive.

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

**Ans :**

If median > mean

Then Skewness Is Negetive.

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

**Ans :**

Data is Normally distributed having Kurtosis value is Zero (0).

Positive Kurtosis Value Indicates

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

**Ans:**

The distribution of the data has lighter tails and flatter peak than Normal distribution.

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



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

**Ans:**

The above box plot represents the information about Ages of students in School.

50% of students are above 10 yrs of old and remaining are less. And the students having age above 15yrs are approximately 40%.

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

**Ans:**

Negatively Skewed , as Long Tail is Left side.

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

**Ans:**

Lower Quartile(Q1)=10, Upper Quartile(Q3)=18  
  
 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.**

Ans:

In both Boxplots, Mean, Median and Mode are same and hence the distribution is

Symmetrical the only difference is that Boxplot 2 having high Wisker level as compare toBoxplot 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)**

**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**

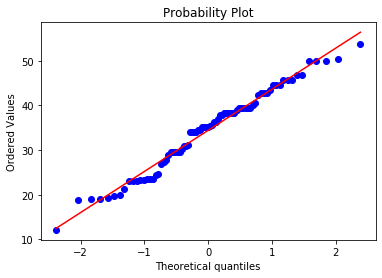
Ans:

Cars = pd.read\_csv('Cars.csv')

import pylab

import scipy.stats as st

st.probplot(Cars['MPG'],dist="norm",plot=pylab)



**Conclusion:**

Yes, MPG of Cars follows Normal Distrubution Because most Of The Points Lies On The Red Line.

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

**Dataset: wc-at.csv**

Ans:

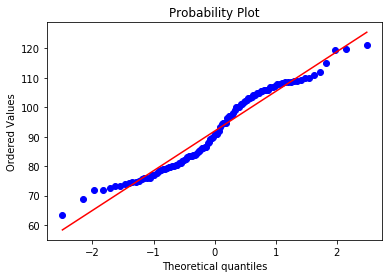
wc = pd.read\_csv('wc-at.csv')

import pylab

import scipy.stats as st

**Waist:**

st.probplot(wc['Waist'],dist='norm',plot=pylab)



**Conclusion:**

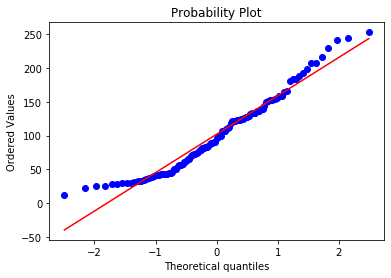
Yes, Circumference(Waist) Follows Normal Distribution Because most Of The Points Lies On The Red Line.

**AT:**

st.probplot(wc['AT'],dist='norm',plot=pylab)

**Conclusion:**

Yes, Adipose Tissue (AT) Follows Normal Distribution Because most Of The Points Lies On The Red Line.

Adipose Tissue (AT 

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

**Ans:**

**Z Score for 90% confidence Interval:**

**In R : In Python:**

|  |
| --- |
| > qnorm(0.950) stats.norm.ppf(0.950,0,1)  [1] 1.644854 1.6448536269514722 |
|  |
| |  | | --- | |  | |

**Z Score for 94% confidence Interval:**

**In R : In Python:**

> qnorm(0.970) stats.norm.ppf(0.970,0,1)

[1] 1.880794 1.8807936081512509

**Z Score for 60% confidence Interval:**

**In R : In Python:**

> qnorm(0.800) stats.norm.ppf(0.800,0,1)

[1] 0.8416212 0.8416212335729143

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

Ans:

**T Score for 95% confidence Interval:**

**In R : In Python:**

> qt(0.975,25) stats.t.ppf(0.975,25)

[1] 2.059539 2.059538552753294

**T Score for 96% confidence Interval:**

**In R : In Python:**

> qt(0.98,25) stats.t.ppf(0.980,25)

[1] 2.166587 2.1665866344527562

**T Score for 99% confidence Interval:**

**In R : In Python:**

> qt(0.995,25) stats.t.ppf(0.995,25)

[1] 2.787436 2.787435813675851

**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:**