**Activity Sheet:**

1. Below is a table of graduates and post graduates

|  |  |  |  |
| --- | --- | --- | --- |
|  | Graduate | Post Graduate | Total |
| Male | 19 | 41 | 60 |
| Female | 12 | 28 | 40 |
| Total | 31 | 69 | **100** |

* 1. What is the probability that a randomly selected individual is a male and a graduate? What kind of probability is it (Marginal/ Joint/Conditional)
     1. P(Male and Graduate) = 19/100 = 0.19
     2. It’s joint probability
  2. What is the probability that a randomly selected individual is a male? What kind of probability is it (Marginal/ Joint/Conditional)
     1. Probability of male is: 60/100 = 0.6
     2. Marginal
  3. What is the probability of a randomly selected individual being a graduate? What kind of probability is this?
     1. Probability of Graduate is: 31/100 = 0.31
     2. Marginal
  4. What is the probability that a randomly selected person is a female given that the selected person is a post graduate? What kind of probability is this?
     1. P(Female | Postgraduate) – Female given she is postgraduate
     2. (P(Female and Postgraduate))/P(PostGraduate)
     3. 0.28/0.69 = 0.405
     4. Probability is Conditional Probability

1. In a particular region during a 1year period, there were 1000 deaths. It was observed that 321 people died of a renal failure and 460 people had at least one parent with renal failure. Of these 460 people, 115 died of renal failure. Calculate the probability of a person that he dies of renal failure if neither of his parents had a renal failure

Parents RF - People with parents dying with RF

~Parents RF - opp parents RF

|  |  |  |  |
| --- | --- | --- | --- |
|  | **RF** | **~RF** | **Deaths** |
|  |  |  |  |
| **Parents RF** | 115 | 345 (460-115) | 460 |
| **~Parents RF** | 206 (321-115) | 334 (540-206) | 540 |
|  | 321 | 679 | 1000 |

P( RF | ~Parents RF) = 206/540

1. 0.5 percent of the population of an area is affected by a disease. A test is developed to detect the disease. This test gives a false positive 3% of the time and false negative 2% of the time.

Draw the tree diagram for this problem.

What is the probability that the test gives a positive result?

If a person's test turns out to be positive, what is the probability that he actually has the disease

Positive (0.98)

(0.5)

Disease

Negative

(0.02)

Population

Negative(0.97)

No Disease

(99.5) Positive (0.03)

ii. (0.5% \* 98% ) +(99.5%\*3%)

=0.035

iii.

P( Disease | given test is positive) =

(p( test is positive |D) \* p(Disease) )/p(testPositive )

Numerator = (0.005 \* 0.98)

Denominator = (0.005\*0.98) + (0.995 \* 0.03)

= numerator/denominator = 0.014

1. Let us suppose, you tossed two two-sided fair coins:
   1. Compute the PMF (Probability Mass function) for heads in this experiment
      1. HH, HT,TH,TT

P( 0 heads) - 1 / 4

P(1 heads) – 2/4

P(2 heads) – 1/4

* 1. Compute Expectation of heads

|  |  |  |  |
| --- | --- | --- | --- |
| No. of heads | 0 | 1 | 2 |
| Frequency | 1 | 2 | 1 |
| Probability | 1/ 4 | 2/4 | ¼ |

Expectation is mean = (1/4 \* 0) + (1/2 \* 1) \* (1/4 \* 2) = 1

1. A bank has developed an analytical model that helps them assess the credit worthiness of individuals and offer loans accordingly. To validate the performance of the model, they constructed a classification matrix on historical data.

|  |  |  |
| --- | --- | --- |
|  | **Predicted as credit worthy** | **Predicted as not credit worthy** |
| **Truly credit worthy** | 8000 | 900 |
| **Truly not credit worthy** | 100 | 1000 |

Identify “True Positives, True Negatives, False positives and False Negatives” from the table and compute “Accuracy, Precision, Recall and F1 statistic”. (Please write the formula used to calculate each metric and substitute appropriate values to score.)

True Positives = 8000

True Negatives = 1000

False Positive = 100

False Negative = 900

Recall = 8000/8900 ( True positive / Total +ve) = 0.89

Precision = 8000/8100 (True Positive / Predicted +ve ) = 0.988

Accuracy = ( T +ve + T –ve)/total = 9000/10000 = 0.9  
F1 = 2PR/(P+R) = (2\*0.988\*0.89)/(0.988+ 0.89)

= (1.758)/(1.878) = 0.94

In this analysis, will you be more worried about false positives or false negatives?

I will be worried about False Positivies (100) as we will end up giving loan to them.

1. Here are the two stock options given to participants in an experiment.

Stock A : 80% Chance of gaining $ 4000

20% Chance of gaining $ 0

stock B : 100% Chance of gaining $ 3000

Intuitively, which stock do you think they will prefer?

Now , Can you compute the Expected values in each case. Do you stick to earlier choice or would like to switch?

Stock A : 80% Chance of loosing $ 4000

20% Chance of loosing $ 0

stock B : 100% Chance of loosing $ 3000

Intuitively, which stock do you think they will prefer?

Now , Can you compute the Expected values in each case. Do you stick to earlier choice or would like to switch?

What did you learn from the above exercises? Did you see the application of expected values in real world?

* + 1. Intuitively I will prefer to go with Stock B as we don’t have loss and always $3000 gaining is guaranteed.

P(X=x) => 40,000 \* 0.8 + 0 \*0.2

* 3200
* If we do this event ‘n’ number of times, Stock A is better.
* I would not stick to my decision of going with Stock B.
  + 1. Intuitively I will prefer to go with Stock B as loss is only 3000$ when compared to stock A where loss is 4000$.

P(X=x) => -40,000 \* 0.8 + 0\*0.2

* -3200
* I would stick to my decision of going with Stock B as I will have less loss there.

Intuition can be wrong and its always better to check what happens when ‘n’ number of events are performed.

1. Jeevansh Dad gave him a die for his birthday . He wanted to make sure that it was fair so he took the die to school and rolled it 500 times and kept track of how many times the die rolled each number on a paper.

In another experiment, he calculated the expected value of the sum of of 20 rolls to be 67.4. On his way home from school it was raining and 2 values were washed away from her table.

|  |  |
| --- | --- |
| **Die Value** | **Frequency** |
| 1 | A |
| 2 | 110 |
| 3 | 95 |
| 4 | 70 |
| 5 | 75 |
| 6 | B |
| **Total** | **500** |

Can you help Jeevansh to find the 2 missing frequencies from the above table?

E(20 rolls) => 67.4

E(1 roll) -> 67.4/20

Total number of events = 500

A + 110 + 95 + 70+ 75+B = 500

A + B = 150

Mean and average will not change with number of experiments.

* (1\*A/500) + (2\*110/500) + (3 \*95/500) + (4\*70/500) + (5 \*75/500) + (B\*6/500) = 67.4/20
* (A + 6B + 1160)/500 = 67.4/20
* A+6B = 525

By computing A+ B = 150 and A + 6B = 525

We get to know that A = 75 and B = 75

1. Consider the favorite coin toss experiment. If you toss a biased coin, the probability of obtaining heads is 0.6. If you toss the coin 10 times, what is the probability of getting heads exactly 4 times?

P(H) = 0.6

P(T) = 0.4

N = 10, r = 4

P(H=4) = 10C4 \* (0.6)4  \* (0.4)6