Activity Sheet

Please go through 'Central Tendency' slides of today's lecture and spend time in understanding the concepts, examples problems explained and then solve the problems given below.

Activity - 1

1. Two people work in a factory making parts for cars. The table shows how many complete parts they make in one week.

Worker	Mon	Tue	Wed	Thu	Fri
Philip	20	21	22	20	21
Mathews	30	15	12	36	28

(a) Find the mean, median and range for Philip and Mathews.
(b) Who is more consistent?

(a) Philip: Mean= (20+21+22+20+21)/5=20.8, Median=(20 20 21 21 22) = , Range=22-20=2

Matthews: Mean=(30+15+12+36+28)/5=24.2, Median=12 15 28 30 36, Range=36-12=24

(b) Look for least standard deviation/variance; hence Philip is more consistent.

2. Find the mode for 8,6,2,4,6,8,10,8

The frequency for 8 is 3, and all other values occur less frequently. Therefore the mode is 8

3. Analyze the performance of your class in the first WUQ taken at INSOFE

Scores: 11, 7.5, 8.5, 10, 10, 10.5, 5.5, 10, 9, 9.5, 5.25, 8, 6.5, 10.5, 8.75, 0, 6, 6, 6.75, 8.75, 0, 9.5, 7.5, 8.5, 7

a) How is the spread of the scores? Compute range, variance & standard deviation
b) Find the 25th percentile, 50th percentile and 75 percentile for this data.

a) Range = Max-Min;

Variance = $\sum(x - \text{mean})^2/n$;

Stdev = sqrt(Variance)

b)

Percentile	Formula	Value
25	[(n+1)/4] th value	6.5
50	[2(n+1)/4] th value	8.5
75	[3(n+1)/4] th value	9.5

4. Temperatures in 5 cities measured on 12 days is given below. The weather department says that two cities have similar weather. Use central tendencies to identify those two cities

City 1	29	32	36	40	43	37	36	33	32	37	31	29
City 2	20	24	31	37	40	38	37	34	34	33	28	23
City 3	23	26	32	38	41	40	35	33	35	37	30	25
City 4	20	24	29	34	37	36	32	30	33	32	27	23
City 5	19	24	29	38	43	38	33	34	36	34	29	23

	Mean	Median	Mode	Stdev
City 1	34.58333	34.5	29	4.33712
City 2	31.58333	33.5	37	6.487167
City 3	32.91667	34	35	5.915439
City 4	29.75	31	32	5.327885
City 5	31.66667	33.5	29	7.062492

Using mean, median and stdev, we conclude City 2 and 5 have similar weather.

5. What is the probability that we get a 5th Tuesday in a 30-day month?

30 days => 4 weeks +2 days;
Last two days can be any one of the following:{ Sun-Mon, Mon-Tues, Tues-Wed, Wed-Thurs, Thurs-Fri, Fri-Sat, Sat-Sun}
Of them Tuesday occurs in two cases; Hence probability = 2/7

6. 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)

Joint Probability. P(Male and Graduate)= 19/100

2. What is the probability that a randomly selected individual is a male?

Marginal Probability: P(Male) = 60/100

3. What is the probability of a randomly selected individual being a graduate? What kind of probability is this?

   Marginal Probability. P(Graduate)=31/100

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?

   Conditional Probability. P(Female|Post Graduate)=28/69

7. 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 atleast 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

   Ans: Let H=the event that atleast one of parents of the randomly selected man die of cause related to renal failure.
   D= event that the randomly selected man died of renal failure.

D/H	Parent had RF	Parent ! have RF	Total
People died of RF	115	206	321
People !died of RF	345	334	679
Total	460	540	1000

$$P(D|H') = \frac{P(D \cap H')}{P(H')}$$
$$= 206/540=38\%$$

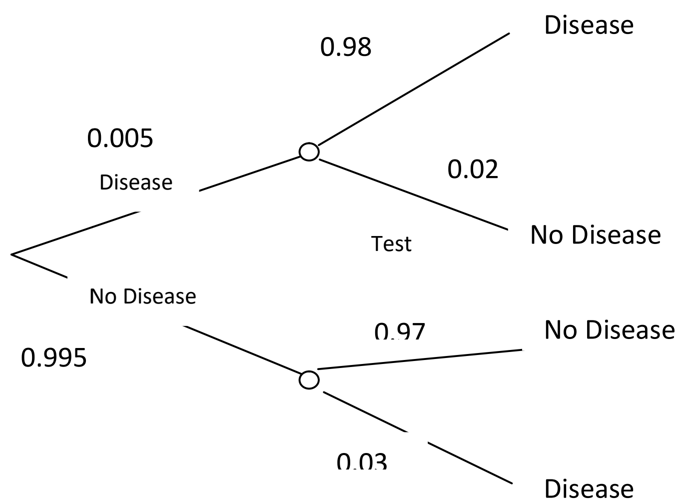
8. 0.5 percent of the population of an area is affected by a particular disease. A test is developed to detect the disease. This test gives a false positive 3% of the time and false negative 2% of the time.

   a) Draw the tree diagram for this problem.

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

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

Ans: a)



b) We want to compute P(T). We do so by conditioning on whether or not Joe has the disease:
P(T) = P(T|D)P(D) + P(T|Dc)P(Dc) = (.98)(.005) + (.03)(.995)
By Law of total probability

c) We want to compute P(D|T) = P(D ∩ T)/P(T)
= P(T|D)P(D)/(P(T|D)P(D) + P(T|Dc)P(Dc))
= (.98)(.005)/((.98)(.005) + (.03)(.995)) ≈ .14

9. Let three fair coins be tossed. Let Event A = {all heads or all tails}, Event B = {at least two heads}, and Event C = {at most two tails}. Of the pairs of events, (A,B), (A,C), and (B,C), which are independent and which are dependent? (Justify).

If A and B are independent, then P(A ∩ B) = P(A) * P(B).
If A & B are dependent, P(A ∩ B) = P(A)*P(B|A) = P(B)*P(A|B) OR P(A ∩ B) != P(A)*P(B)

We write the event space for each of A, B and C.
A = {HHH,TTT},
B = {HHH,HHT,HTH,THH},
C = { HHH, HHT, HTH, THH, HTT, THT, TTH}.

P(A ∩ B) = 1/8 and P(A) · P(B)= (2/8)(4/8) =1/8 so A and B are independent.
P(A ∩ C) = 1/8 and P(A) · P(C)= (2/8)(7/8), so A and C are dependent.
P(B ∩ C) =4/8 and P(B) · P(C)= (4/8)(7/8), so B and C are dependent

10. 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

a. 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.)

TP= 8000, TN= 1000, FP= 100 FN=900

Accuracy= (8000+1000)/(8000+900+100+1000)

Precision= 8000/(8000+100)

Recall= 8000/(8000+900)

F1 Statistic= 2*Precision*Recall/(Precision + Recall)

b. In this analysis, will you be more worried about false positives or false negatives?
In this case, we would be more worried about false positives (i.e. predicting a non-credit worthy person as credit worthy)

11. Let us suppose, you tossed two two-sided fair coins:

a. Compute the PMF for heads in this experiment

Outcome space = {HH, HT, TH, TT}. Let X denote random variable i.e. the number of Heads we get after tossing two 2-sided fair coin. Let r = {0, 1, 2}

PMF = P(X=r) i.e. probability that the random variable X can take a specific value r.
P (Heads = 0) = 1/4 =0.25
P (Heads = 1) = 2/4=1/2 = 0.5
P (Heads = 2) = 1/4 =0.25

r	0	1	2
P(X=r)	1/4	1/2	1/4

b. Compute Expectation of heads

E(X) = ∑ X*P(X) = (0*1/4) + (1*1/2) + (2*1/4) = 1

Expected value = [no. of times the event happened *probability of the event]
= [X * P(X)]

12. For a given probability density function

$$f(x) = \begin{cases} 3x^{-4}, & x > 1 \\ 0, & elsewhere \end{cases}$$

Find the following:

c. (X = 2)
d. P(X <= 4)
e. P(X < 1)
f. P(2 <= X <= 3)

a. P(X = 2) By definition of PDF, it is 0

b. P(X<= 4)

$$P(X \leq 4) = \int_1^4 3x^{-4} \, dx$$

$$P(X \leq 4) = \left[-x^{-3}\right]_1^4$$

$$P(X \leq 4) = -(4)^{-3} - -(1)^{-3}$$

$$P(X \leq 4) = \frac{63}{64}$$

c. P(X<1) = 0

d. P(2<=X<=3) $= \int_2^3 3x^{-4}\, dx$

$$= \left[-x^{-3}\right]_2^3$$
$$= -\frac{1}{27} - -\frac{1}{8} = \frac{19}{216}$$