30

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
CREDIT HOURS ENGINEERING PROGRAMS
SOPHOMORE (All Programs)

Fall 2022	Major Task (Part 1)	Total: 15 marks

PHM111: Probability and Statistics

1/18

Name: Mohamed Hamdy Awaad **ID:** 21P0207 **Deadline: Week 7**

Please, Solve each problem in its assigned place ONLY (the empty space below it)

Statistics

Part I

1. Name and define the two areas of statistics.
Descriptive :- Describe represent data with table, charts
Infential : Make Conclusion about Population based on specific Sample

2. Why are samples used in statistics?
less effort & time consumed than Population

3. In each of these statements, tell whether descriptive or inferential statistics have been used:
a. In the year 2030, 148 million Americans will be enrolled in an HMO. (Descriptive)
b. Nine out of ten on-the-job fatalities are men. (Infrential)
c. The median household income for people aged 25–34 is $35,888. (Descriptive)

4. Classify each variable as qualitative or quantitative:
a) Number of bicycles sold in 1 year by a large sporting goods store. (quantitative)
b) Classification of children in a day care center (infant, toddler, preschool) (qualitative)
c) Weights of fish caught in Lake George. (quantitative)
d) Marital status of faculty members in a large university. (qualitative)

5. Classify each variable as discrete or continuous:
a) Water temperatures of six swimming pools in Pittsburgh on a given day. (Continuous)
b) Weights of cats in a pet shelter. (Continuous)
c) Number of cheeseburgers sold each day by a hamburger stand on a college campus. (Discrete)
d) Number of DVDs rented each day by a video store. (Discrete)

6. Name the four basic sampling methods.

Random / Cluster / Systematic / Stratified

7. Classify each sample as random, systematic, stratified, or cluster:
   a) In a large school area, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years. (Cluster)
   b) Every seventh customer entering a shopping mall is asked to select her or his favorite store. (Systematic)

Part II

1. Name the three types of frequency distributions and explain when each should be used.

Qualitative → Categorical
Grouped
Quantitative → Ungrouped

2. The following two frequency distributions are incorrectly constructed. State the reason why.

a.

Class	Frequency
27–32	1
33–38	0
39–44	6
45–49	4
50–55	2

Class width not the same 50-45=5
→ CW=6

b.

Class	Frequency
123–127	3
128–132	7
138–142	2
143–147	19

5
10

Class width not the same
138-128=10, 128-123=5

3. Shown here are the number of inches of rain received in 1 year in 25 selected cities in the United States. Construct a grouped frequency distribution and a cumulative frequency distribution with 6 classes.

6	37	14	45	22	32	33	49	55
94	38	83	85	40	67	36	67	
71	52	52	55	63	49	44	58	

Name: **ID:**

6 - 14 - 22 - 32 - 33 - 36 - 37 - 38 - 40 - 44 - 45 - 49 - 49 , 52 , 52
55 , 58 , 58 , 63 , 67 , 67 , 71 , 83 , 85 , 94

Range = 88
C-W = 15

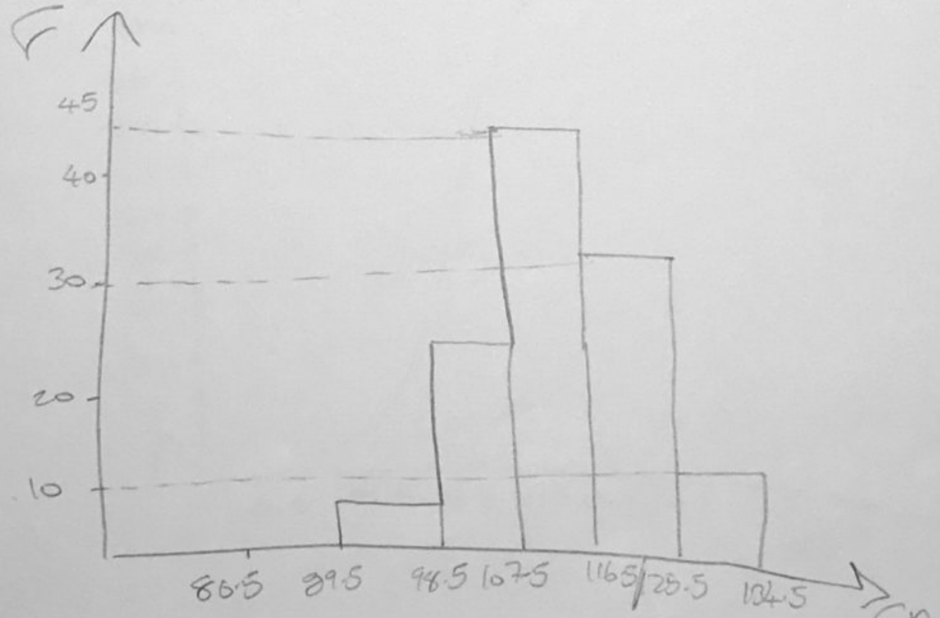
Class	Class limits	Class boundaries	Frequency	Cumulative Frequency
1	6 - 20	5.5 - 20.5	2	2
2	21 - 35	20.5 - 35.5	3	5
3	36 - 50	35.5 - 50.5	8	13
4	51 - 65	50.5 - 65.5	6	19
5	66 - 80	65.5 - 80.5	3	22
6	81 - 95	80.5 - 95.5	3	25
			25	

4. **Do Students Need Summer Development?** For 108 randomly selected college applicants, the following frequency distribution for entrance exam scores was obtained. Construct a histogram, frequency polygon, and ogive for the data.

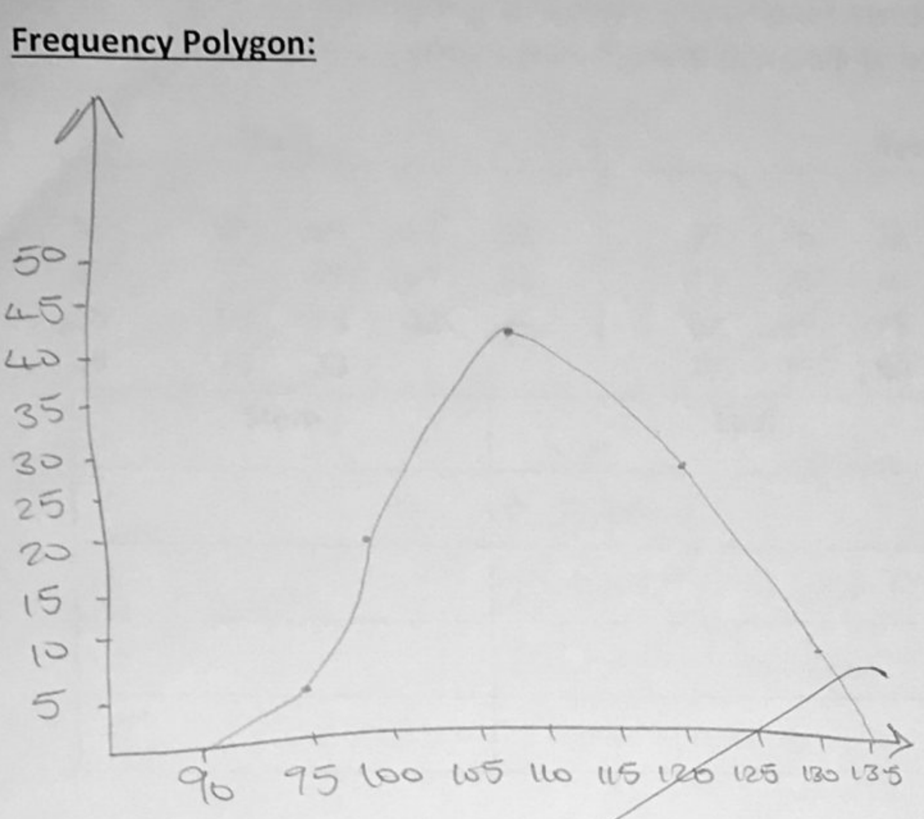
Class	Class limits	Class boundaries	Frequency	Cumulative Frequency
1	90–98	89.5 - 98.5	6	6
2	99–107	98.5 - 107.5	22	28
3	108–116	107.5 - 116.5	43	71
4	117–125	116.5 - 125.5	28	99
5	126–134	125.5 - 134.5	9	108

Applicants who score above 107 need not enroll in a summer developmental program. In this group, how many students do not have to enroll in the developmental program?

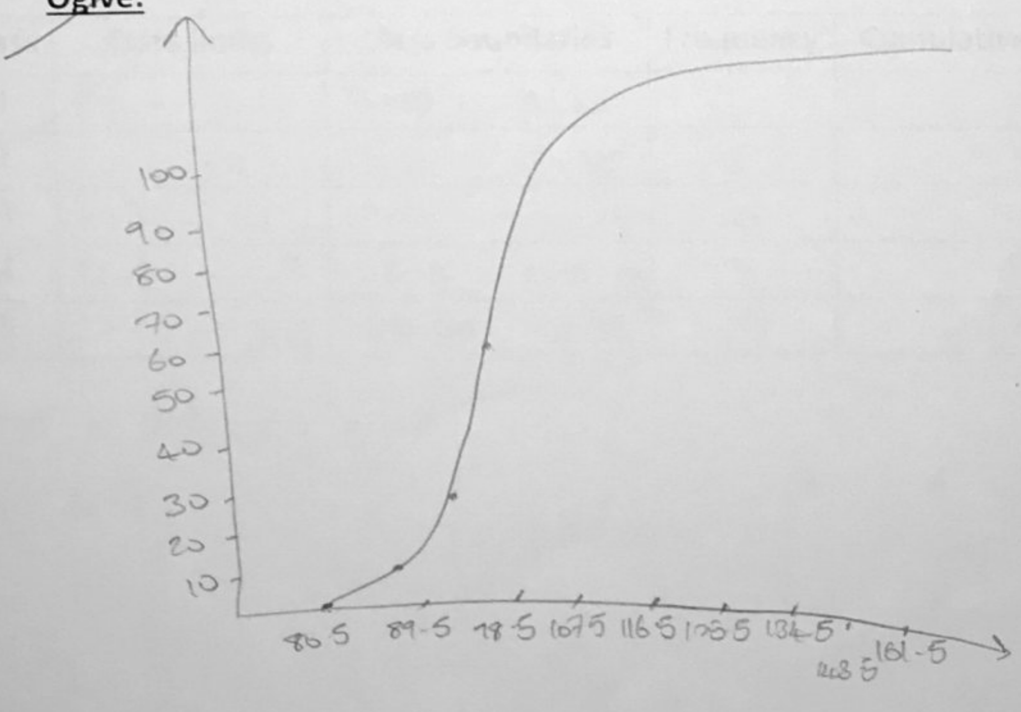
Histogram:



Frequency Polygon:



Ogive:



5. The math and reading achievement scores from the National Assessment of Educational Progress for selected states are listed below. Construct a back-to back stem and leaf plot with the data.

Math					Reading				
52	66	69	62	61	65	76	76	66	67
63	57	59	59	55	71	70	70	66	61
55	59	74	72	73	61	69	78	76	77
68	76	73			77	77	80		

Stem	Leaf Math	Leaf Reading
5	2557999	
6	123689	1156679
7	23346	0016667778
8		0

6. The state gas tax in cents per gallon for 25 states is given below. Construct a grouped frequency distribution and a cumulative frequency distribution with 5 classes.

7.5	16	23.5	17	22
21.5	19	20	27.1	20
22	20.7	17	28	20
23	18.5	25.3	24	31
14.5	25.9	18	30	31.5

Class	Class limits	Class boundaries	Frequency	Cumulative Frequency
1	7.5 – 12.3	7.45 – 12.35	1	1
2	12.4 – 17.2	12.35 – 17.25	4	5
3	17.3 – 22.1	17.25 – 22.15	10	15
4	22.2 – 27	22.15 – 27.05	5	20
5	27.1 – 31.9	27.05 – 31.95	5	25

Range = 31.5 - 7.5 = 24

C.W = 4.9

Part III

1. The average undergraduate grade point average (GPA) for the 25 top-ranked medical schools is listed below.

3.80	3.77	3.70	3.74	3.70
3.86	3.76	3.68	3.67	3.57
3.83	3.70	3.80	3.74	3.67
3.78	3.74	3.73	3.65	3.66
3.75	3.64	3.78	3.73	3.64.

Find the mean, the median, the mode, and the midrange.

Data in ascending order

3.57	3.64	3.64	3.65	3.66	3.67	3.67	3.68	3.7	3.70
3.70	3.73	3.73	3.74	3.74	3.74	3.75	3.76	3.77	3.78
3.78	3.80	3.80	3.86						

Mean =	$\frac{\Sigma x}{n} = \frac{93.09}{25} = 3.723$		
Median =	3.73		
Mode =	3.7, 3.74	**Midrange =**	3.715

2. For the following data, construct a grouped frequency distribution with six classes then find the mean and modal class.

Range = 1995
C.W = 333

1013 1867 1268 1666 2309 1231 3005 2895 2166 1136
1532 1461 1750 1069 1723 1827 1155 1714 2391 2155
1412 1688 2471 1759 3008 2511 2577 1082 1067 1062
1319 1037 2400.

Class	Class limits	Class boundaries	Frequency	Cumulative Frequency
1	1013 – 1345	1012.5 – 1345.5	11	11
2	1346 – 1678	1345.5 – 1678.5	4	15
3	1679 – 2011	1678.5 – 2011.5	7	22
4	2012 – 2344	2011.5 – 2344.5	3	25
5	2345 – 2677	2344.5 – 2677.5	5	30
6	2678 – 3010	2677.5 – 3010.5	3	33

The mean = 1804.6

The modal class = first 1216

$$1012.5 + \left(\frac{11}{18}\right)(333) = \boxed{1216}$$

3. Find the weighted mean price of three models of automobiles sold. The number and price of each model sold are shown in this list.

Model	Number	Price
A	8	$10,000
B	10	12,000
C	12	8,000.

$\frac{\Sigma w \cdot x}{\Sigma w} = 9.87$

....

4. An instructor grades exam, 20%; term paper, 30%; final exam, 50%. A student had grades of 83, 72, and 90, respectively, for exams, term paper, and final exam. Find the student's final average. Use the weighted mean.

$83(0.2) + 72(0.3) + 90(0.5) = 83.2$

....

5. Calculate the median from the following data:

Group	Frequency		Cum-Freq	
60 – 64	1	59.5 - 64.5	1	
65 – 69	5	64.5 - 69.5	6	
70 – 74	9	69.5 - 74.5	15	
75 – 79	12	74.5 - 79.5	27	
80 – 84	7	79.5 - 84.5	34	
85 – 89	2	84.5 - 89.5	36	

Median $= 74.5 + 5\left(\frac{18-15}{12}\right) = 75.6$ $n = 36$

$n/2 = 18$

6. Calculate the mode from the following data:

Group	Frequency	C.F	Class Bou	
150 - 154	5	5	149.5 - 154.5	
155 - 159	2	7	154.5 - 169.5	
160 - 164	6	13	159.5 - 164.5	
165 - 169	8	21	164.5 - 169.5	
170 - 174	9	30	169.5 - 174.5	
175 - 179	11	41	174.5 - 179.5	
180 - 184	6	47	179.5 - 184.5	
185 - 189	3	50	184.5 - 189.5	

Mode $= 174.5 + 5\left(\frac{2}{22-9-6}\right) = 175.9$

....

....

....

7. For these situations, state which measure of central tendency (mean, median, or mode) should be used.

a) The most typical case is desired. (...Mean.......)

b) The data are categorical. (...Mode.....)

c) The values are to be divided into two approximately equal groups, one group containing the larger values and one containing the smaller values. (...Median...)

8. Describe which measure of central tendency—mean, median, or mode—was probably used in each situation.

a) The average number of children per family in the Plaza Heights Complex is 1.8. (...Mean.....)

b) Most people prefer red convertibles over any other color. (...Mode.......)

c) The average age of college professors is 42.3 years. (.....Mean......)

Part IV

1. Find the range, variance, and standard deviation. Assume the data represent samples, and use the shortcut formula for the unbiased estimator to compute the variance and standard deviation:

The normal daily high temperatures (in degrees Fahrenheit) in January for 10 selected cities are as follows:

50, 37, 29, 54, 30, 61, 47, 38, 34, 61

The normal monthly precipitation (in inches) for these same 10 cities is listed here:

4.8, 2.6, 1.5, 1.8, 1.8, 3.3, 5.1, 1.1, 1.8, 2.5

Which set is more variable?

Range = 32

$Cvar = \frac{S}{\bar{X}} \times 100$

$= 27.56$

For normal daily high temperatures:

n	X	X^2	S^2	$\bar{X}$
1	29			
2	30			
3	34			
4	37			
5	38			
6	47			
7	50			
8	54			
9	61			
10	61			
	$\Sigma x = 441$	$\Sigma x^2 = 20777$	147.66	44.1

$S = 12.15$

....

....

....

....

....

For normal monthly precipitation:

n	X	X^2		
1	1-1			
2	1-5			
3	1-8			
4	1-8			
5	1-8			
6	2-5			
7	2-6			
8	3-3			
9	4-8			
10	5-1			
	26-3	86-13	s^2=1-885	$\bar{X}$=2-63

s = 1-373

Cvar = $\frac{1-373}{2-63}$ × 100 = 52-2 %

....

....

....

The more variable set is normal monthly ppt

Because larger Cvar

2. Team batting averages for major league baseball in 2005 are represented below. Find the variance and standard deviation for each league.

NL		AL	
0.252–0.256	4	0.256–0.261	2
0.257–0.261	6	0.262–0.267	5
0.262–0.266	1	0.268–0.273	4
0.267–0.271	4	0.274–0.279	2
0.272–0.276	1	0.280–0.285	1

For NL:

Class limits	Class boundaries	C.M	F	xF	x^2F
0.252 – 0.256	0.2515 – 0.2565	0.254	4		
0.257 – 0.261	0.2565 – 0.2615	0.259	6		
0.262 – 0.266	0.2615 – 0.2665	0.264	1		
0.267 – 0.271	0.2665 – 0.2715	0.269	4		
0.272 – 0.276	0.2715 – 0.2755	0.274	1		
			16	4.184	1.095

$$S^2 = \frac{n\left[\sum x^2F - (\sum xF)^2\right]}{n(n-1)} = \frac{16(1.095) - (4.184)^2}{(16)(15)} = 4.36\times10^{-5}$$

$S = 6.603\times10^{-3}$, $C_{var} = \frac{S}{\bar{x}}\times100 = 2.53\%$

For AL:

Class limits	Class boundaries	Class middle	F	xF	x^2F
0.256 – 0.261	0.2555 – 0.2615	0.2585	2		
0.262 – 0.267	0.2615 – 0.2675	0.2645	5		
0.268 – 0.273	0.2675 – 0.2735	0.2705	4		
0.271 – 0.279	0.2735 – 0.2795	0.2765	2		
0.280 – 0.285	0.2795 – 0.29	0.2825	1		
			14	3.757	1.0088

$$S^2 = \frac{n\sum x^2F - (\sum xF)^2}{n(n-1)} = 4.48\times10^{-5} \qquad S = 6.7\times10^{-3}$$

$C_{var} = \frac{S}{\bar{y}}\times100 = 2.49\%$

3. The average age of senators in the 108th Congress was 59.5 years. If the standard deviation was 11.5 years, find the *z* scores corresponding to the oldest and youngest senators: Robert C. Byrd (D, WV), 86, and John Sununu (R, NH), 40.

For Robert C. Byrd:

$$Z = \frac{X-\bar{X}}{S} = \frac{86-59.5}{11.5} = 2.3$$

For John Sununu:

$$Z = \frac{X-\bar{X}}{S} = \frac{40-59.5}{11.5} = -1.7$$

Name: **ID:** **11/18**

4. The average teacher's salary in a particular state is $54,166. If the standard deviation is $10,200, find the salaries corresponding to the following *z* scores.

a. 2 *b.* -1 *c.* 0

a- $2 = \frac{x - 54166}{10200}$ $x = 74566$ \$

b- $x = 43966$ \$

c- $x = 54166$ \$

5. Find the percentile rank for each value in the data set. The data represent the values in billions of dollars of the damage of 10 hurricanes.

1.1, 1.7, 1.9, 2.1, 2.2, 2.5, 3.3, 6.2, 6.8, 20.3.

1.1 is $\frac{0+0.5}{10} \times 100 = 5\%$

1.7 is $\frac{1+0.5}{10} \times 100 = 15\%$

1.9 is $\frac{2+0.5}{10} \times 100\% = 25\%$

2.1 is 35 %

2.2 is 45 %

2.5 is 55 %

3.3 is 65 %

6.2 is 75 %

6.8 is 85 %

20.3 is 95 %

6. check each data set for outliers.

a. 24, 32, 54, 31, 16, 18, 19, 14, 17, 20

14, 16, 17, 18, 19, 20, 24, 31, 32, 54 IQR = 14

[-4, 52], 54 is outlier

b. 321, 343, 350, 327, 200

200, 321, $\boxed{327}$, 343, 350 $Q_1 = \frac{321+200}{2} = 260.5$

No outlier $Q_3 = 346.5$ IQR = 86

7. identify the five-number summary and find the interquartile range. [131.5, 475.5]

19, 16, 48, 22, 7

$Q_2 = 19$ IQR = 23.5

$Q_1 = 11.5$

$Q_3 = 35$ [-23.75, 70.25]

8. Construct a boxplot for the following data representing the number of games pitched by major league baseball's earned run average (ERA) leaders for the past few years.

30, 34, 29, 30, 34, 29, 31, 33, 34, 27, 30, 27, 34, 32.

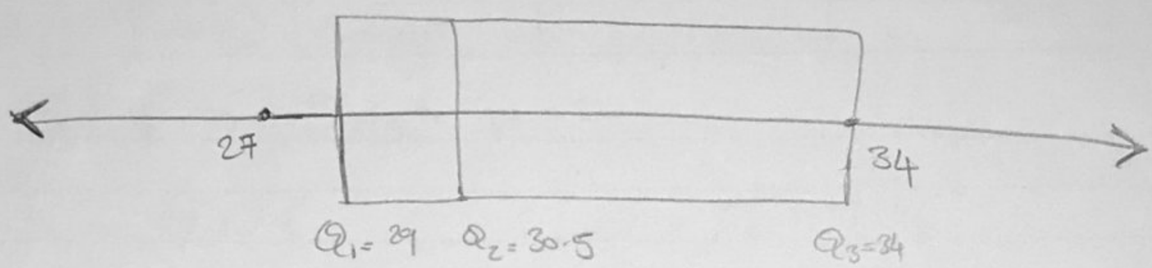
27, 27, 29, 29, 30, 30, 30 | 31, 32, 33, 34, 34, 34, 34

$Q_1 = 29$ $Q_2 = 30.5$

$Q_3 = 34$ Max = 34

Min = 27

The Box Plot:



9. Starting teacher salaries (in equivalent U.S. dollars) for upper secondary education in selected countries are listed below. Which set of data is more variable? (The U.S. average starting salary at this time was $29,641.)

Europe	
Sweden	$48,704
Germany	41,441
Spain	32,679
Finland	32,136
Denmark	30,384
Netherlands	29,326
Scotland	27,789

Asia	
Korea	$26,852
Japan	23,493
India	18,247
Malaysia	13,647
Philippines	9,857
Thailand	5,862

For Europe:

n	X			
1	48704			
2	41441			
3	32679			
4	32136			
5	30384			
6	29326			
7	27786			
	242459	8.7×10^9		

$S^2 = 57.9 \times 10^6$ $n = 7$

$S = 7609.8$, $C_{var} = 21.97\%$

For Asia:

n	X	X^2		
1	26852			
2	23493			
3	18247			
4	13647			
5	9857			
6	5862			
	97958	1923668064		

$S^2 = 64.9 \times 10^6$

$S = 8054.5$ $C_{var} = 49.3\%$

The more variable set is Asia

Because Larger coff of Variation

PHM111: Probability and Statistics

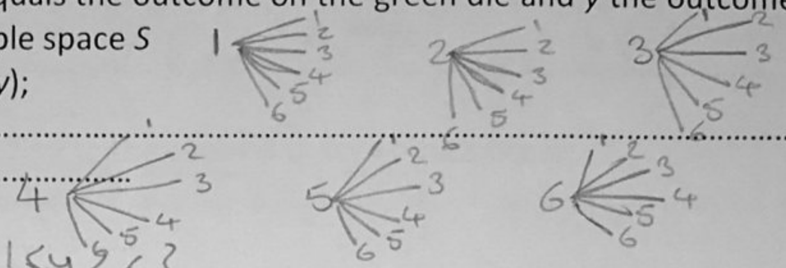
Name: **ID:**

Classical Probability, Counting Techniques, Conditional Probability

Part I

1. An experiment involves tossing a pair of dice, one green and one red, and recording the numbers that come up. If x equals the outcome on the green die and y the outcome on the red die, describe the sample space S

(a) by listing the elements (x, y);



(b) by using the rule method.

$S = \{(x,y) \mid 1 \le x \le 6, 1 \le y \le 6\}$

2. For the sample space of Exercise 1,

(a) list the elements corresponding to the event A that the sum is greater than 8;

{(3,6)(6,3)(4,5)(5,4)(4,6)(6,4)(5,5)(6,6) ~~(6,3)~~ (6,5)(5,6)}

(b) list the elements corresponding to the event B that a 2 occurs on either die;

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

(c) list the elements corresponding to the event C that a number greater than 4 comes up on the green die;

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

(d) list the elements corresponding to the event $A \cap C$;

{(5,4) (5,5) (5,6) (6,3) (6,4) (6,5) (6,6)}

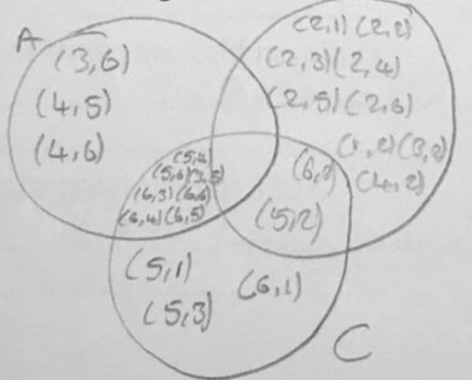
(e) list the elements corresponding to the event $A \cap B$;

$A \cap B = \emptyset$

(f) list the elements corresponding to the event $B \cap C$;

{(5,2), (6,2)}

(g) construct a Venn diagram to illustrate the intersections and unions of the events A, B and C.



3. An experiment consists of tossing a die and then flipping a coin once if the number on the die is even. If the number on the die is odd, the coin is flipped twice. Construct a tree diagram to show the 18 elements of the sample space *S*. then:

(a) list the elements corresponding to the event *A* that a number less than 3 occurs on the die;

A = {1HH, 1HT, 1TH, 1TT, 2H, 2T}

(b) list the elements corresponding to the event *B* that two tails occur;

B = {1TT, 3TT, 5TT}

(c) list the elements corresponding to the event A';

A' = {3HH, 3HT, 3TH, 3TT, 4H, 4T, 5HH, 5HT, 5TH, 5TT, 6H, 6T}

(d) list the elements corresponding to the event $A' \cap B$;

A' ∩ B = {3TT, 5TT}

(e) list the elements corresponding to the event $A \cup B$.

{1HH, 1HT, 1TH, 1TT, 3TT, 5TT, 2H, 2T}

4. If $S = \{x \mid 0 < x < 12\}$, $M = \{x \mid 1 < x < 9\}$, and $N = \{x \mid 0 < x < 5\}$, find

(a) $M \cup N$;

{x, 0<x<9}

(b) $M \cap N$;

{x, 1<x<5}

(c) $M' \cap N'$

{x, 9<x<12}

5. (a) How many three-digit numbers can be formed from the digits 0, 1, 2, 3, 4, 5, and 6 if each digit can be used only once?

6x6x5 = 180

(b) How many of these are odd numbers?

5x5x3 = 75

(c) How many are greater than 330?

(1x3x5) + (3x6x5) = 105

6. Three lottery tickets for first, second, and third prizes are drawn from a group of 40 tickets. Find the number of sample points in *S* for awarding the 3 prizes if each contestant holds only 1 ticket.

$_{40}P_3 = 59280$

7. How many ways are there to select 3 candidates from 8 equally qualified recent graduates for openings in an accounting firm?

$_8C_3 = 56$

Part II

1. A box contains 500 envelopes, of which 75 contain \$100 in cash, 150 contain \$25, and 275 contain \$10. An envelope may be purchased for \$25. What is the sample space for the different amounts of money? Assign probabilities to the sample points and then find the probability that the first envelope purchased contains less than \$100.

$P(100) = \frac{75}{500}$ $P(25) = \frac{150}{500}$ $P(10) = \frac{275}{500}$

$= \boxed{\frac{17}{25}}$

2. A pair of fair dice is tossed. Find the probability of getting

(a) a total of 8; (2,6) (3,5) (4,4), (5,3), (6,2)

$\frac{5}{36}$

(b) at most a total of 5. $1 < 1,2,3,4$ $2 < 1,2,3$ $3 < 1,2$ $4 - 1$

$\boxed{\frac{10}{36}}$

3. In a high school graduating class of 100 students, 54 studied mathematics, 69 studied history, and 35 studied both mathematics and history. If one of these students is selected at random, find the probability that:

(a) the student took mathematics or history;

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{54}{100} + \frac{69}{100} - \frac{35}{100} = \frac{88}{100}$

(b) the student did not take either of these subjects; $(A \cup B')$

$1 - \frac{88}{100} = \frac{12}{100}$

(c) the student took history but not mathematics.

$P(A \cup B') = P(A) - P(A \cap B) = \frac{54}{100} - \frac{35}{100} = \frac{19}{100}$

4. In an experiment to study the relationship of hypertension and smoking habits, the following data are collected for 180 individuals:

	Nonsmokers	Moderate Smokers	Heavy Smokers
H	21	36	30
NH	48	26	19

where *H* and *NH* in the table stand for *Hypertension* and *Non-hypertension*, respectively. If one of these individuals is selected at random, find the probability that the person is

(a) experiencing hypertension, given that the person is a heavy smoker;

$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{30}{180} / \frac{49}{180} = \frac{30}{49}$

(b) a nonsmoker, given that the person is experiencing no hypertension.

$P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{48}{93}$

5. Pollution of the rivers in the United States has been a problem for many years. Consider the following events:

A: the river is polluted,

B: a sample of water tested detects pollution,

C: fishing is permitted.

Assume

$P(A) = 0.3,\ P(B\mid A) = 0.75,\ P(B\mid A') = 0.20,\ P(C\mid A\cap B) = 0.20,$

$P(C\mid A'\cap B) = 0.15,\ P(C\mid A\cap B') = 0.80,\ P(C\mid A'\cap B') = 0.90.$

(a) Find $P(A \cap B \cap C)$.

$P(C/A\cap B) \cdot P(A\cap B) = 0.2 \times 0.225 = 0.045$

(b) Find $P(B'\cap C)$.

$= (0.8)(0.25)(0.3) + (0.9)(0.7)(0.8) = 0.564$

(c) Find $P(C)$.

$P(C\cap B) + P(B'\cap C) = 0.066 + 0.564 = 0.63$

(d) Find the probability that the river is polluted, given that fishing is permitted, and the sample tested did not detect pollution.

$P(A / B'\cap C) = \frac{0.06}{0.564} = \frac{5}{47}$

6. Police plan to enforce speed limits by using radar traps at four different locations within the city limits. The radar traps at each of the locations L_1, L_2, L_3, and L_4 will be operated 40%, 30%, 20%, and 30% of the time. If a person who is speeding on her way to work has probabilities of 0.2, 0.1, 0.5, and 0.2, respectively, of passing through these locations, what is the probability that she will receive a speeding ticket?
If the person received a speeding ticket on her way to work, what is the probability that she passed through the radar trap located at L_2?

$P(A) = 0.2(0.4) + 0.1(0.3) + 0.5(0.2) + 0.2(0.3)$

$P(A) = 0.27$

$P(B|A) = \frac{0.1}{0.27} = 0.3737$

7. A regional telephone company operates three identical relay stations at different locations. During a one-year period, the number of malfunctions reported by each station and the causes are shown below.

Station	A	B	C	
Problems with electricity supplied (ES)	2	1	1	4
Computer malfunction (CM)	4	3	2	9
Malfunctioning electrical equipment (ME)	5	4	2	11
Caused by other human errors (H)	7	7	5	19
				43

Suppose that a malfunction was reported, and it was found to be caused by other human errors. What is the probability that it came from station C?

$\frac{P(C \cap H)}{P(H)} = \frac{5/43)}{19/43)} = \boxed{\frac{5}{19}}$

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
CREDIT HOURS ENGINEERING PROGRAMS
SOPHOMORE (All Programs)

30

Total: 15 marks

Fall 2022 **Major Task (Part 2)**

PHM111: Probability and Statistics

1/15

Name: Kerolos Essam Gamal **ID:** 21P0051 **Deadline: Week 14**

Please, Solve each problem in its assigned place ONLY (the empty space below it)

Part I: Discrete Random Variables

1. Classify the following random variables as discrete or continuous:
   a) X: the number of automobile accidents per year in Virginia. (...Discrete...)
   b) Y: the length of time to play 18 holes of golf. (Continuous..)
   c) M: the amount of milk produced yearly by a particular cow. (..Continuous)
   d) N: the number of eggs laid each month by a hen. (...Discrete.)

2. Determine the value c so that the following function can serve as a probability distribution of the discrete random variable X: $f(x)=c\binom{2}{x}\binom{3}{3-x}$, for $x=0, 1, 2.$

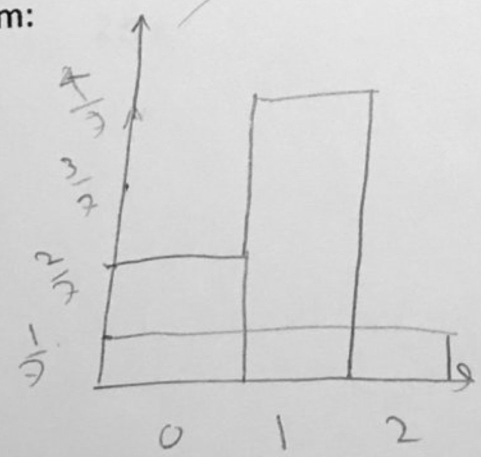
$$\sum_{x=0}^{2} f(x)=1 \rightarrow c+6c+3c=1$$

$$c=\frac{1}{10}$$

3. A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets purchased by the hotel, find the probability distribution of X.

x	0	1	2		
f(x)	2/7	4/7	1/7		

   a) Express the results graphically as a probability histogram.
   Histogram:



b) Find the cumulative distribution function of the random variable X representing the number of defectives.

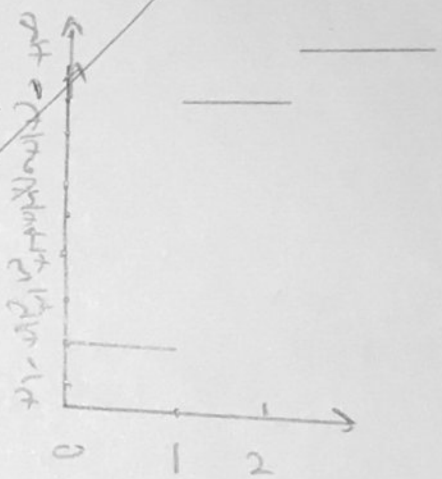
X	0	1	2
f(x)	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$
F(x)	$\frac{2}{7}$	$\frac{6}{7}$	1

c) Using F(x), find

i. P (X = 1) $= f(1) = F(1) - F(0) = \frac{6}{7} - \frac{2}{7} = \boxed{\frac{4}{7}}$

ii. P (0 < X ≤ 2) $= F(2) - F(0) = 1 - \frac{2}{7} = \boxed{\frac{5}{7}}$

d) Construct a graph of the cumulative distribution function.



4. Determine the probability mass function of X from the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \le x < 0 \\ 0.7 & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

$$f(X) \begin{cases} 0.2 & X = -2 \\ 0.5 & X = 0 \\ 0.3 & X = 2 \end{cases}$$

5. A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

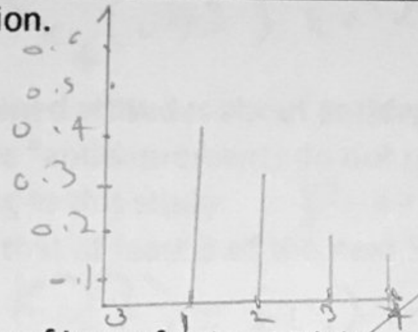
$P = \frac{1}{4}$ $E(X) = mP$ $m = 2$

$E(X) = 2\left(\frac{1}{4}\right) = \frac{1}{2}$

6. The distribution of the number of imperfections per 10 meters of synthetic fabric is given by

x	0	1	2	3	4
f(x)	0.41	0.37	0.16	0.05	0.01

a) Plot the probability function.



b) Find the expected number of imperfections, E(X) = μ.

$(X) = \sum fX = 1(0.37) + 2(0.16) + 3(0.05) + 4(0.01)$

$k = 0.88$

c) Find $E\left(x^2\right)$ $= \sum k^2 (f)(X) = 1(0.37) + 4(0.16) + 9(0.05) + 16(0.01)$

$E(X)^2 = 1.62$

d) Find the variance and standard deviation of the number of imperfections.

$var(k) = E(k^2) - E(r)^2 = 1.62 - (0.88)^2 = 0.8456$

$\sigma = \sqrt{var\,k} = 0.92$

7. If a random variable X is defined such that $E\left[(x-1)^2\right] = 10$ and $E\left[(x-2)^2\right] = 6$, find μ and σ^2.

$E(X-1)^2 = E(x^2 - 2x + 1) = E(X)^2 - 2E(X) + 1 = 10$

$E(x^2) - 2E(X) = 9 \rightarrow (1)$

$E(x^2) - 4E(k) = 2 \rightarrow (2)$

$E(X) = \frac{7}{2}$ $E(X^2) = 16$

$\mu = \frac{7}{2}$

$\sigma^2 = E(X^2) - (E(X))^2 = 16 - \left(\frac{7}{2}\right)^2 = \boxed{\frac{15}{4}}$

1.

Part II: Special Distributions of Discrete Random Variable

1. In a certain city district, the need for money to buy drugs is stated as the reason for 75% of all thefts. Find the probability that among the next 5 theft cases reported in this district

a) exactly 2 resulted from the need for money to buy drugs.

$$f(X=2) = {}^{m}C_{x}\, p^{x} q^{m-x} = {}^{5}C_{2} (0.75)^{2} (0.25)^{3}$$
$$= 0.0879$$

$m = 5$
$p = 0.75$
$q = 0.25$

b) at most 3 resulted from the need for money to buy drugs.

$$f(X \le 3) = 1 - f(X > 3) = 1 - [f(X=4) + f(X=5)]$$
$$1 - [{}^{5}C_{4} (0.75)^{4} (0.25) + {}^{5}C_{5} (0.75)^{5}] \boxed{= 0.3672}$$

2. A national study that examined attitudes about antidepressants revealed that approximately 70% of respondents believe "antidepressants do not really cure anything; they just cover up the real trouble." According to this study: $p = 0.7$ $q = 0.3$ $m = 5$

a) what is the probability that at least 3 of the next 5 people selected at random will hold this opinion?

$$P(K \ge 3) = f(3) + f(4) + f(5)$$
$$P(K \ge 3) = {}^{5}C_{3} (0.7)^{3} (0.3)^{2} + {}^{5}C_{4} (0.7)^{4} (0.3) + {}^{5}C_{5} (0.7)^{5}$$
$$\boxed{= 0.8369}$$

b) If X represents the number of people who believe that antidepressants do not cure but only cover up the real problem, find the mean and variance of X when 5 people are selected at random.

$$\mu = mp = 5(0.7) = 3.5$$
$$var(K) = mpq = 5(0.7)(0.3) = 1.05$$

3. In a batch of 2000 calculators, there are, on average, 8 defective ones. If a random sample of 150 is selected, find the probability of 5 defective ones. $m = 150$ $p = \frac{8}{2000}$

$$P(K=5) = {}^{150}C_{5} \left(\frac{8}{2000}\right)^{5} \left(\frac{1992}{2000}\right)^{145}$$

$q = \frac{1992}{2000}$

$$\boxed{= 3.388 \times 10^{-4}}$$

Name: **ID:**

4. A mail-order company receives an average of 5 orders per 500 solicitations. If it sends out 100 advertisements, find the probability of receiving at least 2 orders.

$m = 100 \quad P = \frac{5}{500}$

$\lambda = mP = \boxed{1}$

$P(k \geq 2) = 1 - [P(k=0) + P(X=1)]$

$1 - \left[\frac{e^{-1}}{0!} + \frac{(1)^1 e^{-1}}{1!}\right] = 1 - [0.368 + 0.368]$

$= 0.264$

5. A bookstore owner examines 5 books from each lot of 25 to check for missing pages. If he finds at least 2 books with missing pages, the entire lot is returned. If, indeed, there are 5 books with missing pages, find the probability that the lot will be returned.

$n = 5 \quad p = 0.2 \quad q = 0.8$

$P(X \geq 2) = 1 - [P(X=0) + P(k=1)$

$1 - [{}^5C_0 (0.8)^5 + {}^5C_1 (0.2)(0.8)^4]$

$P(X \geq 2) = 0.2627$

Part III: Continuous Random Variables

1. The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \dfrac{20{,}000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

a) at least 200 days. $P(X \geq 200) = 1 - P(X < 200) = 1 - \int^{200} \frac{20000}{(X+100)^3} =$

$1 - \left[\frac{20000}{-2}(X+100)^{-2}\right]_0^{200}$

$P(k \geq 200) = 1 - \left[\frac{-10000}{(300)^2} + \frac{10000}{(100)^2}\right] = \boxed{\frac{1}{9}}$

b) anywhere from 80 to 120 days.

$P(80 \leq X \leq 120) = \int_{80}^{120} F(X)dx = \left[\frac{-10000}{(X+100)^2}\right]_{80}^{120}$

$= \boxed{0.102}$

2. Consider the density function

$$f(x)=\begin{cases} k\sqrt{x}, & 0<x<1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Evaluate k.

$$k\int_0^1 \sqrt{x}\,dx = 1 \quad k\left[\frac{2}{3}(x)^{\frac{3}{2}}\right]_0^1 = 1 \quad k=\frac{3}{2}$$

b) Find F(x) and use it to evaluate P(0.3 < X < 0.6).

$$F(x) \text{ from } 0<x<1 = \frac{2x^{\frac{3}{2}}}{3}\cdot\frac{3}{2} = x^{\frac{3}{2}}$$

$$P(0.3<x<0.6) = F(0.6)-F(0.3) = (0.6)^{\frac{3}{2}}-(0.3)^{\frac{3}{2}} = 0.3004$$

3. Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function:

$$f(x)=\begin{cases} k(3-x^2), & -1\le x\le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Determine k that renders f(x) a valid density function.

$$\int_{-\infty}^{\infty} F(k)=1 \quad k\int_{-1}^{1} 3-\cancel{x}^2\,dk = 1 = k\left[3k-\frac{x^3}{3}\right]_{-1}^{1} = 1$$

$$k\left[\frac{8}{3}-\frac{-8}{3}\right]=1 \quad k=\frac{3}{16}$$

b) Find the probability that a random error in measurement is less than 1/2.

$$P(k<\tfrac{1}{2}) = P(-1\le \cancel{x} \le \tfrac{1}{2}) = \frac{3}{16}\int_{-1}^{\frac{1}{2}} (3-x)^2\,dk = \frac{3}{16}\left[\frac{83}{24}-\left[\frac{-8}{3}\right]\right]$$

$$P(k<\tfrac{1}{2}) = \frac{99}{128} = \boxed{0.7734}$$

c) For this particular measurement, it is undesirable if the magnitude of the error (i.e., |x|) exceeds 0.8. What is the probability that this occurs?

$$P(|x|>0.08) = 1-P(-0.8\le \cancel{x} \le 0.8) = 1-\int_{-0.8}^{0.8}\frac{3}{16}(3-k^2)\,dk$$

$$= 1-\frac{3}{16}\left[3x-\frac{\cancel{x}^3}{3}\right]_{-0.8}^{0.8} = 0.164$$

$(x)=\begin{cases}0, & \text{elsewhere.}\end{cases}$

a) Calculate $P(X \le 1/3)$. $= \int_0^{1/3} 2(1-x)\,dx$

$2\left[x-\frac{x^2}{2}\right]_0^{1/3} = \boxed{\frac{5}{9}}$

b) What is the probability that X will exceed 0.5?

$P(0.5 \le X < 1) = 2\int_{0.5}^{1} 1-x\,dx = 2\left[x-\frac{x^2}{2}\right]_{0.5}^{1} = 2\cdot\frac{1}{8} = \boxed{\frac{1}{4}}$

c) Given that $X \ge 0.5$, what is the probability that X will be less than 0.75?

$P(X<0.75 \mid X \ge 0.5) = \frac{P(0.5 \le X < 0.75)}{P(X \ge 0.25)} = \frac{\int_{0.5}^{0.75} 2(1-x)\,dx}{0.25}$

$= \frac{3/16}{0.25} = \frac{3}{4}$

5. The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$f(x)=\begin{cases}0, & x<0,\\ 1-e^{-0.01x}, & 0\le x.\end{cases}$$

Determine the probability density function of X. What proportion of reactions is complete within 200 milliseconds?

$f(x) = \frac{d}{dx} F(x) = 0.01e^{-0.01x}$ $x \ge 0$

$P(X<200) = \int_0^{200} 0.001e^{-0.01x}\,dx = \left[-e^{-0.01x}\right]_0^{200}$

$= 1 - \frac{1}{e^2} = \boxed{0.8647}$

Part IV: Special Distributions of Continuous Random Variables

1. Given a standard normal distribution,
   a) find the area under the curve that lies between z = −0.48 and z = 1.74.

   $P(z_1 \le Z < z_2) = \Phi(1.74) - \Phi(-0.48) = 0.6435$

   b) Find the value of z if the area enclosed between −z and z, is 0.9500.

   $P(-z \le Z < z) = \Phi(z) - \Phi(-z) = \Phi(z) - 1 + \Phi(z)$

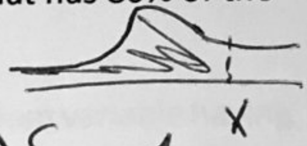
   $\Phi(z) = 0.975 \quad z = 1.96$

   c) find the value of k such that P(−0.93 < Z < k) = 0.7235.

   $\Phi k - \Phi(-0.93) = 0.7235 \quad \Phi k = 0.8997$

   $k = \boxed{1.28}$

2. Given a normal distribution with μ = 30 and σ = 6, find the value of x that has 80% of the normal curve area to the left. $P(Z \le \frac{k-30}{6}) = 0.8$



$\Phi(\frac{x-30}{6}) = 0.8 \quad \frac{x-30}{6} = 0.84 \quad x = 35.04$

3. Given the normally distributed variable X with mean 18 and standard deviation 2.5, find the value of k such that P(X >k) = 0.1814.

   $P(Z > \frac{k-18}{2.5}) = 0.1814 \quad \Phi\frac{k-18}{2.5} = 0.8186$

   $P(Z < \frac{k-18}{2.5}) = 0.8186 \quad \frac{k-18}{2.5} = 0.91 \quad k = 20.275$

4. The heights of 1000 students are normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Assuming that the heights are recorded to the nearest half-centimeter, how many of these students would you expect to have heights
   a) less than 160.0 centimeters? $P(X < 160) = P(Z < \frac{160-174.5}{6.9})$

   $\Phi(-2.1) = 0.0179 \quad 0.018 \times 1000 = 18$ students

   b) between 171.5 and 182.0 centimeters inclusive?

   $P(\frac{171.5-174.5}{6.9} \le Z < \frac{182-174.5}{6.9})$

   $\boxed{0.532}$

   c) equal to 175.0 centimeters?

   $P(0 < Z < 0.15) = \Phi 0.15 - \Phi 0 = 0.5596 - 0.5 = 0.0596$

   $= 60$ students

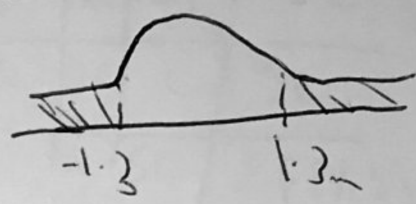
d) greater than or equal to 188.0 centimeters?

$P(k \geq 188) = 1 - P(k < 188) = 1 - P(z < \frac{188 - 174.5}{6.9}) \; 1 - \Phi(1.96)$

$1 - 0.975 = 0.025 = \boxed{25 \text{ students}}$

5. If a set of observations is normally distributed, what percent of these differ from the mean by

a) more than 1.3σ? $1 - P(-1.3 < z < 1.3)$



$1 - [0.9032 - 0.0968] = 0.1936$

19.36 %

b) less than 0.52σ?

$= 0.6985 - 0.3015 = 0.397$

$\% = 39.7\%$

6. The length of time for one individual to be served at a cafeteria is a random variable having an exponential distribution with a mean of 4 minutes. What is the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days?

$M = 4$ mins $\quad \frac{1}{M} = 0.25 \quad p = 0.5277 \quad q = 0.4724 \quad n = 6$

$F(3) = 1 - e^{-x/4} = 1 - e^{-3/4} = 0.5277$ By using Binomial dist.

$P(X \geq 4) = f(4) + f(5) + f(6)$

$= {}^6C_4 (0.5276)^4 (0.4724)^2 + {}^6C_5 (0.5276)^5 (0.4724) + {}^6C_6 (0.5276)^6$

$= \boxed{0.397}$

7. Suppose that a study of a certain computer system reveals that the response time, in seconds, has an exponential distribution with a mean of 3 seconds.

a) What is the probability that response time exceeds 5 seconds? $\frac{1}{M} = \boxed{\frac{1}{3}}$

$F(X) = 1 - e^{-x/3}$

$P(X > 5) = 1 - P(X \leq 5) = 1 - 1 + e^{-\frac{5}{3}} = 0.189$

b) What is the probability that response time exceeds 10 seconds?

$P(X > 10) = 1 - P(X \leq 10) = 1 - 1 + e^{-\frac{10}{3}}$

$= 0.0357$

PHM111: Probability and Statistics

Name: **ID:** **10/15**

Part V: Joint Distributions

1. Determine the values of c so that the following function represents joint probability distribution of the random variables X and Y:
$f(x, y) = c|x - y|$, for $x = -2, 0, 2$; $y = -2, 3$.

$\sum f(x,y) = 15c = 1$

$c = \frac{1}{15}$

$x \backslash y$	-2	0	2	$f_2(y)$
-2	0	2c	4c	6c
3	5c	3c	c	9c
$f_1(x)$	5c	5c	5c	$\sum f(x,y) = 15c$

2. If the joint probability distribution of X and Y is given by

$f(x,y)=\dfrac{x+y}{30}$, for $x=0,1,2,3;\ y=0,1,2,$, Find

a) $P(X \le 2, Y = 1)$

$\frac{1}{30} + \frac{2}{30} + \frac{3}{30} = \frac{1}{5} = \boxed{0.2}$

$x \backslash y$	0	1	2	3	$f_2(y)$
0	0	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{1}{5}$
1	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{1}{3}$
2	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{7}{15}$
$f(x)$	$\frac{3}{30}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$	$\sum f(x,y) = \boxed{1}$

b) $P(X + Y = 4)$

$\frac{4}{30} + \frac{4}{30} + \frac{4}{15} = 0.2$

c) the marginal distribution of X

X	0	1	2	3
f(Y)	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

d) the marginal distribution of Y

y	0	1	2
f(y)	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{7}{15}$

PHM111: Probability and Statistics

Name: **ID:** **11/15**

3. A fast-food restaurant operates both a drive through facility and a walk-in facility. On a randomly selected day, let X and Y , respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y)=\begin{cases}\frac{2}{3}(x+2y), & 0\le x\le 1, 0\le y\le 1,\\ 0, & \text{elsewhere.}\end{cases}$$

a) Find the marginal density of X

$f(x) = \frac{2}{3}\int_0^1 x+2y\,dy = \frac{2}{3}\left[xy+y^2\right]_0^1 = \boxed{\frac{2}{3}(x+1)}$

b) Find the marginal density of Y

$f_y = \frac{2}{3}\int_0^1 x+2y\,dx = \frac{2}{3}\left[\frac{x^2}{2}+2yx\right]_0^1 = \boxed{\frac{1}{3}(1+4y)}$

c) Find the probability that the drive-through facility is busy less than one-half of the time

$P(x<0.5) = P(0\le x<0.5,\ 0\le y\le 1)$

$= \frac{2}{3}\left[\frac{y^2}{2}+\frac{y}{8}\right]_0^1 = \frac{2}{3}\left[\frac{1}{2}+\frac{1}{8}\right] = \frac{5}{12}$

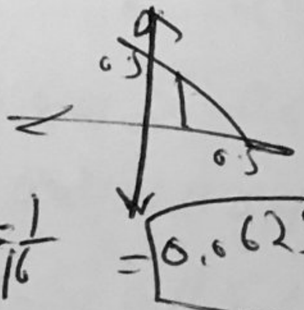
4. A candy company distributes boxes of chocolates with a mixture of creams, toffees, and cordials. Suppose that the weight of each box is 1 kilogram, but the individual weights of the creams, toffees, and cordials vary from box to box. For a randomly selected box, let X and Y represent the weights of the creams and the toffees, respectively, and suppose that the joint density function of these variables is

$$f(x,y)=\begin{cases}24xy, & 0\le x\le 1, 0\le y\le 1, x+y=1,\\ 0, & \text{elsewhere.}\end{cases}$$

a) Find the probability that in a given box the cordials account for more than 1/2 of the weight.

$P(1-x-y>\frac{1}{2}) = P(x+y<0.5)$

$= \int_0^{0.5}\int_0^{0.5-x} 24xy\,dy\,dx$



$= \frac{3}{8}-\frac{1}{2}+\frac{3}{16} = \frac{1}{16} = \boxed{0.0625}$

b) Find the marginal density for the weight of the creams.

$f(x) = 24\int_0^{1-x} xy\,dy = 12[xy^2]_0^{1-x} = 12(x(1-x^2))$

5. Suppose that X and Y have the following joint probability distribution:

f(x, y)		x: 2	x: 4
y	1	0.10	0.15
	3	0.20	0.30
	5	0.10	0.15

a) Find the marginal distribution of X

x	2	4
f(x)	.4	0.6

b) Find the marginal distribution of Y

y	1	3	5
f(y)	0.25	.5	0.25

c) Determine whether the two random variables X and Y are dependent or independent.

$E(x) = \Sigma x f(x) = 3.2 \quad E(xy) = 9.6$

$E(y) = 3$

$Cov(xy) = E(xy - E(x)E(y)) = 0$

x, y are independent

6. A coin is tossed twice. Let Z denote the number of heads on the first toss and W the total number of heads on the 2 tosses. If the coin is unbalanced and a head has a 40% chance of occurring, find

a) the joint probability distribution of W and Z

$6(H) = 0.4 \quad p(T) = 0.6$

$= 0.16 \quad p(H,T) = 0.24 \quad p(T,T) = 0.36$

Z \ W	0	1	2	f(z)
0	0.36	0.24	0	0.6
1	0	0.24	0.16	0.4
f(w)	0.36	0.48	0.16	$\Sigma = 1$

b) the marginal distribution of W

W	0	1	2
F(W)	0.36	0.48	0.16

c) the marginal distribution of Z

Z	0	1
F(Z)	0.6	0.4

d) the probability that at least 1 head occurs.

$= P(H,H) + P(H,T) + P(T,H) = 0.16 + 0.24 + 0.24 = \boxed{0.64}$

7. The joint probability density function of the random variables X, Y, and Z is

$$f(x,y,z)=\begin{cases}\frac{4xyz^2}{9}, & 0<x,y<1, 0<z<3,\\ 0, & \text{elsewhere.}\end{cases}$$, Find

a) the joint marginal density function of Y and Z;

$F(y,z) = \int_0^1 \frac{4}{9} xyz^2 dx = \frac{2}{9}\left[x^2 yz^2\right]_0^1 = \boxed{\frac{2yz^2}{9}}$

b) the marginal density of Y

$F(y) \int_0^1\int_0^2 \frac{4}{9} yz^2 dz\, dx = \frac{4}{9}\int_0^1 x \frac{yz^3}{3}\Big|_0^3 dx = 4\int_0^1 xy\, dx$

$\left[x^2\right]_0 \boxed{2y}$

c) $P(\frac{1}{4}<X<\frac{1}{2}, Y>\frac{1}{3}, 1<Z<2)$;

$= \cancel{\frac{4}{9}\left[\frac{x^2}{2}\right]^{1/2}} \frac{4}{9}\int_{1/4}^{1/2}\int_{1/3}^{1}\int_{1}^{2} xyz^2\, dz\, dy\, dx = \frac{4}{9}\cdot\int_{1/4}^{1/2} x\, dx \cdot \int_{1/3}^{1} y\, d$

$= \frac{4}{9}\left[\frac{x^2}{2}\right]_{1/4}^{1/2} \cdot \left[\frac{y^2}{2}\right]_{1/3}^{1} \cdot \int_1^2 \frac{z^3}{3} dz$

$= \frac{4}{9} \cdot \frac{3}{32} \cdot \frac{4}{9} \cdot \frac{7}{3} = \frac{7}{162} = 0.0432$

8. Two tire-quality experts examine stacks of tires and assign a quality rating to each tire on a 3-point scale. Let X denote the rating given by expert A and Y denote the rating given by B. The following table gives the joint distribution for X and Y.

f(x, y)		y		
		1	2	3
x	1	0.10	0.05	0.02
	2	0.10	0.35	0.05
	3	0.03	0.10	0.20

Find μ_x and μ_y.

$E(x) = \Sigma x f(x) = 1(0.17) + 2(0.5) + 0.33(3) = \boxed{2.16}$

$E(y) = \Sigma y f(y) = 1(0.23) + 2(0.5) + 3(0.27)$

$= 2.04$

9. Random variables X and Y follow a joint distribution

$$f(x,y) = \begin{cases} 2, & 0 < x \leq y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the correlation coefficient between X and Y.

$E(x) = \int_0^1 x \cdot 2(1-x)\,dx = 2\int_0^1 x - x^2\,dx$

$E(x) = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right] = 2\left(\frac{1}{6}\right) = \boxed{\frac{1}{3}}$

$E(x^2) = 2\int_0^1 x^2 - x^3\,dx = 2\left[\frac{x^3}{3} - \frac{x^4}{4}\right] = 2\left(\frac{1}{12}\right) = \boxed{\frac{1}{6}}$

$E(y) = 2\int_0^1 y^2\,dy = 2\left[\frac{y^3}{3}\right] = \boxed{\frac{2}{3}}$

$E(y^2) = 2\int_0^1 y^3\,dy = 2\left[\frac{y^4}{4}\right] = \boxed{\frac{1}{2}}$

10. If X and Y are independent random variables with variances $\sigma_X^2 = 5$ and $\sigma_Y^2 = 3$, find the variance of the random variable Z = −2X + 4Y − 3.

$Var(x) = E(x^2) - (E(x))^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \boxed{\frac{1}{18}}$

$Var(y) = E(y^2)(E(y))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \boxed{\frac{1}{18}}$ →

10 $Var z = Var(-2x+4y-3) = 4 Var x + 16 Var y$

$Var y = 4(5) + 16(3) = \boxed{68}$

PHM111: Probability and Statistics

Name: **ID:** **15/15**

11. There are two service lines. The random variables X and Y are the proportions of time that line 1 and line 2 are in use, respectively. The joint probability density function for (X, Y) is given by

$$f(x,y)=\begin{cases}\frac{3}{2}(x^2+y^2), & 0\le x,y\le 1,\\ 0, & \text{otherwise.}\end{cases}$$

a) Determine whether or not X and Y are independent.

$f_1(x) = \int_0^1 \frac{3}{2}(x^2+y^2)\,dy = \frac{3}{2}\left[x^2y+\frac{y^3}{3}\right]_0^1 = \frac{3}{2}\left[x^2+\frac{1}{3}\right]$

$f_2(y) = \int_0^1 \frac{3}{2}(x^2+y^2)\,dx = \frac{3}{2}\left[\frac{x^3}{3}+y^2x\right]_0^1 = \frac{3}{2}\left[\frac{1}{3}+y^2\right]$

$f(x)\cdot f(y) = \frac{3}{2}\left[y^2+\frac{1}{3}\right]\cdot\frac{3}{2}\left[x^2+\frac{1}{3}\right] \neq \frac{3}{2}(x^2+y^2)$ $\boxed{\text{X, Y are not independent}}$

b) It is of interest to know something about the proportion of Z = X + Y, the sum of the two proportions. Find E(X + Y). Also find E(XY).

$E(X+Y) = \frac{3}{2}\left[\frac{y}{4}+\frac{y^3}{6}+\frac{y^2}{6}+\frac{y^4}{4}\right]_0^1 = \frac{3}{2}\left[\frac{5}{6}\right] = \boxed{1.25}$

$E(XY) = \frac{3}{2}\int_0^1 \frac{x^4y}{4}+\frac{y^2x^2}{2}\Big|_0^1 dy = \frac{3}{2}\int_0^1 \frac{y}{4}+\frac{y^3}{2}\,dy$

$E(XY) = \frac{3}{2}\left[\frac{y^2}{8}+\frac{y^4}{8}\right]_0^1 = 0.375$

c) Find Var(X), Var(Y), and Cov(X,Y).

$E(X) = \frac{3}{2}\left[\frac{5}{12}\right] = \frac{5}{8}$

$E X^2 = \frac{3}{2}\left[\frac{14}{45}\right] = \frac{7}{15}$

$E Y = \frac{5}{8}$

$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 0.076$

$\mathrm{Var}\,Y = [E(Y^2)) - E(Y))^2 = 0.076$

$\mathrm{Cov}(X,Y) = \boxed{-\frac{1}{64}}$

d) Find Var(X + Y).

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X,Y)$

$= \boxed{0.12}$