MATH 3305

Topic 1. Descriptive Statistics

1.1 Exploring Numerical Data: Frequency Distribution Table and Histogram.

1. A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature:

   24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27.

   a. Sort raw data in ascending order:

   b. Find range. Range =

   c. Select number of classes: (usually between 5 and 15)

   d. Compute interval width: $w = \text{interval width} = \dfrac{Range}{\text{number of desired intervals}} =$

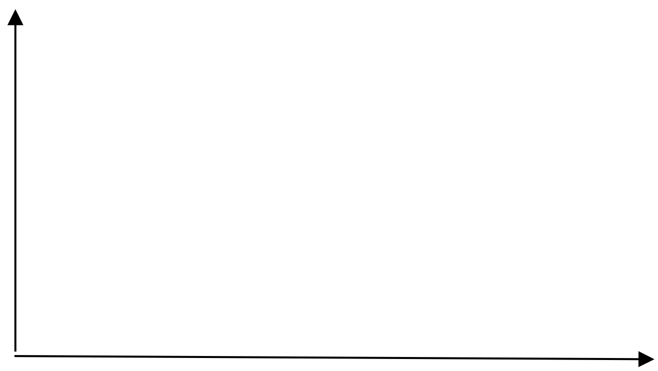
   e. Determine interval boundaries

   f. Count observations (frequencies)

   g. Feel out the table and build the histogram

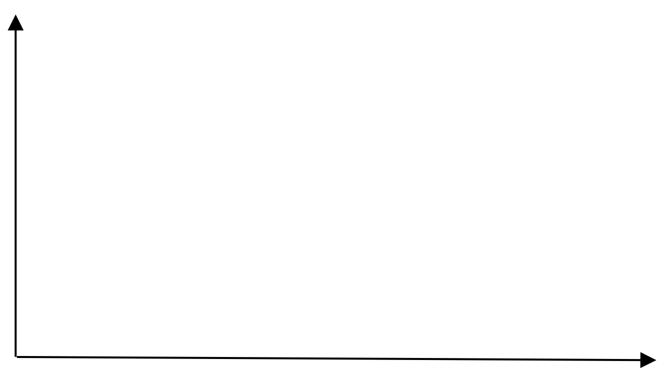
Interval	Frequency	Relative Frequency	%

Histogram: (Frequencies)



Histogram (Relative Frequencies)

Interval	Relative Frequency (**AREA**)	Base = 10 Height



Total area of any RF Histogram always = 1

2. US Presidents (age at inauguration).

President	Age
Washington	57
J.Adams	61
Jefferson	57
Madison	57
Monroe	58
J.Q Adams	57
Jackson	61
Van Buren	54
W.H.Harrison	68
Tyler	51
Polk	49
Taylor	64
Fillmore	50
Pierce	48
Buchanan	65
Lincoln	52
A.Johnson	56
Grant	46
Hayes	54
Garfield	49
Arthur	51
Cleveland	47
B.Harrison	55
Cleveland	55
McKinley	54
T.Roosevelt	42
Taft	51
Wilson	56
Harding	55
Coolidge	51
Hoover	54
F.D.Roosevelt	51
Truman	60
Eisenhower	61
Kennedy	43
L.B.Johnson	55
Nixon	56
Ford	61
Carter	52
Reagan	69
G.H.W.Bush	64
Clinton	46
G.W.Bush	54
Obama	47

President	Age
1	42
2	43
3	46
4	46
5	47
6	47
7	48
8	49
9	49
10	50
11	51
12	51
13	51
14	51
15	51
16	52
17	52
18	54
19	54
20	54
21	54
22	54
23	55
24	55
25	55
26	55
27	56
28	56
29	56
30	57
31	57
32	57
33	57
34	58
35	60
36	61
37	61
38	61
39	61
40	64
41	64
42	65
43	68
44	69

Range =

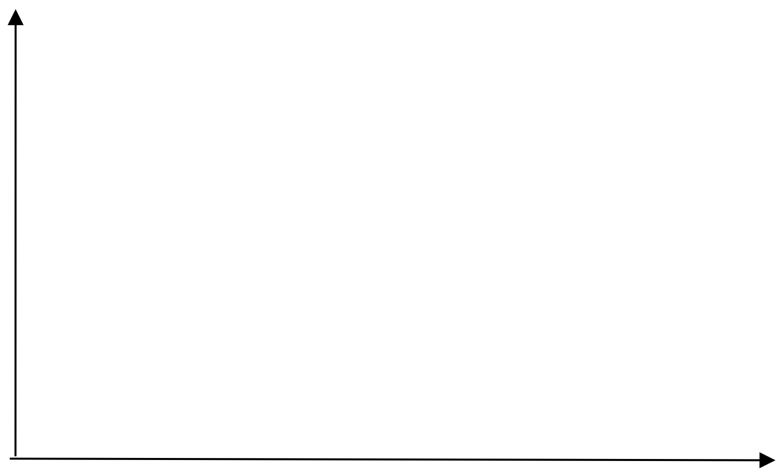
Number of intervals =

Interval width =

Frequency Distribution Table:

Interval	Freq

Histogram: Age



1.2 Numerical Summaries. Measures of Center: Mean, Median, Mode $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

3. Find the Mean and Median for two samples:

Sample #1 11 12 13 14 15

Sample #2 11 12 13 14 25

4. Find the Median for two samples:

Sample #1 15 12 14 13 11

Sample #2 14 12 11 15 16 13

5. Find the Mean, Median and Mode

X – number of cloudy days per year in Chicago

year	X (days)	
1	201	
2	220	
3	207	
4	227	
5	247	
6	247	
7	213	
8	214	
9	235	
10	209	
Sum =	2220	

X - City gas mileage for 2-seater cars.

	Model	City mpg ↓
1	Lamborghini Aventador	11
2	Aston Martin V8	13
3	Jaguar XK	16
4	Nisan GT	16
5	Chevrolet Camaro	17
6	Hyundai Genesis	17
7	Lotus Evora	18
8	Nissan 370Z	18
9	BMW	19
10	Ford Mustang	19
11	Infinity G	19
12	Porsche 911	19
13	Porsche Cayman	20
14	Mercedes-Benz C	21
15	Audi TT	22
16	Scion FR-S	22
17	Subaru BRZ	22
18	Honda Accord	24
19	Kia Forte	24
20	Honda Civic	28
21	Honda CR-Z	36
	Sum =	421

Quartiles, 5-number summary and Boxplot.

Four steps to build a boxplot:
1. Interquartile range IQR= Q3 – Q1, shows the spread of the middle 50% of the data.
2. Fences: Lower Fence LF = Q1 – 1.5(IQR) Upper Fence UF = Q3 + 1.5(IQR)
3. Whiskers - extend from Q1 and Q3 to the smallest and largest observations within the fences.
4. Outliers (if any) - extreme observations that fall outside the fence.

6. Gas mileage for 2-seater cars. Find the five numbers summary, build the boxplot. Compare with the histogram.

Max	
Q3	
Median Q2	
Q1	
Min	

1.3 Measures of Variation: Variance and Standard Deviation.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{n}{n-1}(\bar{x})^2$$

Example:

	x	$(x-\bar{x})$	$(x-\bar{x})^2$
	4	4 – 3 = 1	1
	1	1 – 3 = -2	4
	3	3 – 3 = 0	0
	4	4 – 3 = 1	1
Sum	12		6
Mean	3	Variance	2

	x	x^2
	4	16
	1	1
	3	9
	4	16
Sum		42

$$Var = s^2 = \frac{1}{3}\cdot 42 - \frac{4}{3}\cdot 3^2 = 2 \qquad s = \sqrt{Var} = \sqrt{2}$$

7. X – number of cloudy days per year in Chicago. Find the Variance and Standard Deviation

year	x	$(x-\bar{x})$	$(x-\bar{x})^2$
1	201		
2	220		
3	207		
4	227		
5	246		
6	247		
7	213		
8	214		
9	235		
10	210		
Sum			
Mean			

year	x	x^2
1	201	
2	220	
3	207	
4	227	
5	246	
6	247	
7	213	
8	214	
9	235	
10	210	
Sum		

1.4 Standard Deviation as a Ruler of Variation. Chebyshev's Theorem and Empirical Rule.

Chebyshev's Theorem: Regardless of how the data are distributed, at least $\left(1-\frac{1}{k^2}\right)\times 100\%$ of observations will fall within **k** standard deviations of the mean (for k > 1)

8. According to Chebyshev's Theorem, what percent of the observations lie within 2.25 standard deviations of the mean?

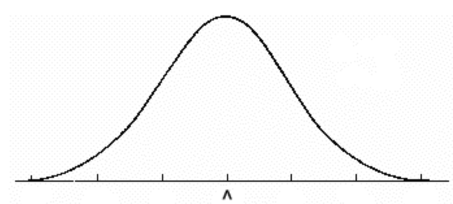
Empirical Rule: for any symmetrical and unimodal distribution, 68% - 95% - 99.7% of observations will fall within one, two and 3 standard deviations of the mean.

9. How the Empirical Rule works: US Presidents (age at inauguration).

Interval	Rule	Real data	
	%	Frequency	%
1 st dev			
2 st. dev			
3 st.dev			

Find three intervals, associated with the Empirical Rule, determine the percentage of observations within each interval and compare with the theoretical statement. Empirical Rule: (68% - 95% - 99.7%)

10. Suppose that the variable Math SAT scores is bell-shaped with a mean of 490 and a standard deviation of 100. Then,
    a. 68% of all test takers scored between
    b. 95% of all test takers scored between
    c. 99.7% of all test takers scored between



Comparing the relative standing of observations: Standardizing. $z = \frac{x - \bar{x}}{s}$

11. A town's January high temperatures average is 36°F with the standard deviation of 10°F, while in July the mean high temperature is 74°F and the standard deviation is 8°F.
    In which month it is more unusual to have a day with a high temperature of 55°F?

Topic 2. Counting. Sets.

2.1 Counting

Arrangements. Basic principle of Counting: $n_1 \times n_2 \times n_3 \times .... \times n_r$

How many different ordered arrangements of the *n* objects are possible? $n \times (n-1) \times (n-2) \times .... \times 3 \times 2 \times 1 = n!$

12. How many different 7-place license plates are possible if the first three places are to be occupied by letters and the final four by numbers?

13. How many different ordered arrangements of the letters A,B,C are possible?

14. How many different batting orders are possible for a baseball team consisting of 9 players?

15. An auto manufacturer produces 7 models, each available in 6 different colors, with 4 different upholstery fabrics and 5 interior colors. How many varieties of the auto are available?

16. A social security number has 9 digits. How many social security numbers are there? The US population in 2012 was about 314 million. Is it possible for every US resident to have a unique SSN?

Permutations (order is relevant!)

How many different ordered arrangements of the *k* objects chosen from *n* objects are possible?

$${}_nP_k = n \times (n-1) \times (n-2) \times ... \times (n-k+1) \quad or \quad {}_nP_k = \frac{n \times (n-1) \times ... \times (n-k+1) \times (n-k) \times ... \times 2 \times 1}{(n-k) \times ... \times 2 \times 1} \qquad {}_nP_k = \frac{n!}{(n-k)!}$$

17. A baseball team has 20 players. How many 9-player batting <u>orders</u> are possible?

Combinations (order isn't relevant!)

How many different groups of *k* objects chosen from *n* objects are possible? ${}_nC_k = \dfrac{n!}{(n-k)!k!}$

18. A baseball team has 20 players. How many 9-player groups are possible?

19. A lottery game requires that you pick 6 different numbers from 1 to 99.
    a. How many ways are there to choose 6 numbers if order matter?

    b. How many ways are there to choose 6 numbers if order is not important?

20. A bridge hand consists of 13 cards from a desk of 52 cards.
    a. How many hands exist?

    b. How many ways are to get 6 face cards?

21. From a group of 16 smokers and 20 nonsmokers, a researcher wants to randomly select 8 smokers and 8 nonsmokers for d study. In how many ways can the study group be selected?

2.2 Sets and Venn diagrams.

22. A marketing survey of 1,000 commuters found that 600 answered listen to the news, 500 listen to music, and 270 listen to both. Let *N* = set of commuters in the sample who listen to news and *M* = set of commuters in the sample who listen to music. Find the number of commuters in the set $N \cap M'$

23. In a certain class, there are 23 majors in Psychology, 16 majors in English and 7 students who are majoring in both Psychology and English.
    a. How many students are majoring in Psychology or English?
    b. If there are 50 students in the class, how many students are majoring in neither of these subjects?
    c. How many students are majoring in Psychology alone?

24. A survey of 100 college faculty who exercise regularly found that 45 jog, 30 swim, 20 cycle, 6 jog and swim, 1 jogs and cycles, 5 swim and cycle, and 1 does all three.
    a. How many of the faculty members do at least one of these three activities?
    b. How many of the faculty members do not do any of these three activities?

25. 50 families were surveyed about their pets:
    17 families have dogs
    24 families have cats
    9 families have parakeets
    10 families have cats and dogs
    6 families have cats and parakeets
    7 families have dogs and parakeets
    5 families have cats, dogs and parakeets

    How many families do not have a pet?
    How many families have dogs (only)?
    How many families have cats or dogs but not parakeets?

Topic 3. Probability Basic

3.1 Theoretical Definition of Probability

26. A bag contains 6 red marbles, 5 blue marbles, and 3 green marbles. What is the probability of choosing a blue marble?

27. A bag contains four chips, one of each of different colors: red, blue, green, and yellow. A chip is selected at random from the bag and then replaced in the bag. A second chip is then selected at random. Determine the set of possible outcomes. What is the probability that the two chips selected are the same color?

28. A card is selected randomly from a deck of 52. The events A, B, and C are defined as follows.
    A = event the card selected is a black card
    B = event the card selected is a club
    C = event the card selected is an ace
    Determine the corresponding probabilities.

29. If a person is randomly selected, find the probability that his or her birthday is in May. Ignore leap years.

30. Two 6-sided dice are rolled. Determine the set of the possible outcomes. What is the probability that
    a. the sum of the numbers on the dice is 6?

    b. the sum of the numbers on the dice is 10?

31. A lottery game requires that you pick 6 different numbers from 1 to 99.
    a. Assuming order is unimportant, what is the probability of picking all six numbers correctly to win the big prize? If order matter?

    b. What is the probability of picking 5 out of 6 numbers correctly?

32. A bag contains 5 black, 1 red and 3 yellow jelly beans. You take three at random.
    a. How many samples are possible?

    Find the probability that selected jelly beans are:

    b. All black

    c. All red

    d. All yellow

    e. 2 black, 1 red

    f. 2 yellow, 1 black

33. Two cards are drawn at random from the desk of 52 cards.
    a. How many 2-card hands are possible?

    b. Find the probability that the 2-card hand contains two queens.

    c. Find the probability that the 2-card hand contains no aces.

    d. Find the probability that the 2-card hand contains both cards of the same suite.

    e. Find the probability that the 2-card hand contains both cards of the different suites.

3.2 Empirical Definition of Probability

34. A polling firm, hired to estimate the likelihood of the passage of an up-coming referendum, obtained the set of survey responses to make its estimate.
The encoding system for the data is: 1 = FOR, 2 = AGAINST.
The survey responses are shown below:
1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 1, 1, 2, 1, 2, 1, 1, 1, 2, 1, 1, 2, 2, 1, 1, 2, 1, 2, 2, 2, 1, 2, 1, 1, 1
If the referendum were held today, find the probability that it would pass.

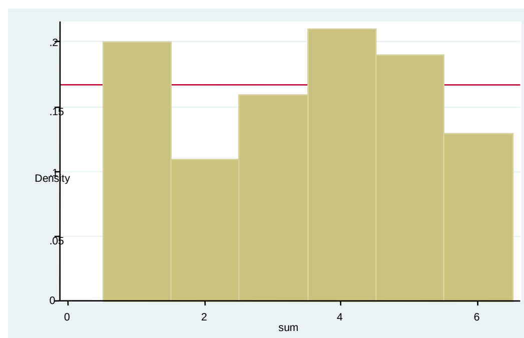
35. The distribution of B.A. degrees conferred by a local college is listed below, by major.
What is the probability that a randomly selected degree is in Mathematics?

Major	**Frequency**
English	2073
Mathematics	2164
Chemistry	318
Physics	856
Liberal Arts	1358
Business	1676
Engineering	868
Total	**9313**

36. From the statistical simulation of rolling a die 100 times we see that the number 4 was rolled 21 times.

a. What is the empirical probability that the number 4 was rolled?

b. What is the probability that the number 4 was rolled under the equally likely assumption?



Law of Large Numbers and Logical Thinking

Question 1. If a fair coin is tossed many times and the last seven tosses are all heads, then the chance that the next toss will be heads is somewhat less than 50%. True or false and justify.

Question 2. A coin is tossed and you win a prize if there are more than 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

Question 3. A coin is tossed and you win a prize if there are more than 40% heads. Which is better: 10 tosses or 100 tosses? Explain.

Question 4. A coin is tossed and you win a prize if there are between 40% and 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

Question 5. A coin is tossed and you win a prize if there are exactly 50% heads. Which is better: 10 tosses or 100 tosses? Explain.

Topic 4: Probability Rules.

4.1 Complement Rule $P(A) = 1 - P(A')$

37. A bag contains 6 red marbles, 5 blue marbles, and 3 green marbles. What is the probability that a marble chosen at random is not a blue marble?

38. A percentage distribution is given below for the size of families in one U.S. city.

Size	**Percentage**
2	45.6
3	22.3
4	20.0
5	7.5
6	2.8
7+	1.8

A family is selected at random.

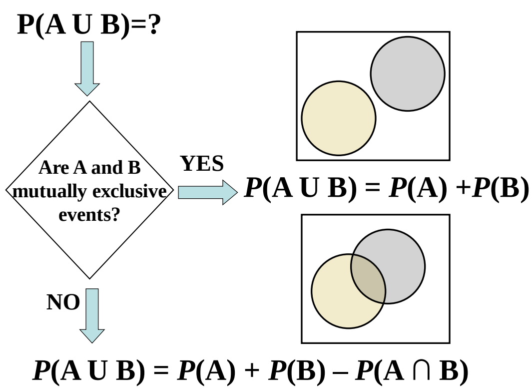
a. Find the probability that the size of the family is less than 7.

b. Find the probability that the size of the family is at least 3.

39. You are dealt a hand of three cards, one at a time. Find the probability that you have at least one red card.

40. Two cards are drawn at random from the desk of 52 cards.
Find the probability that the 2-card hand contains both cards of the different suites.

4.2 **Addition Rule** (Probability of a Union of Two or more Events)



41. Two 6-sided dice are rolled. What is the probability that the sum of the numbers on the dice is 6 or 10?

42. Two balanced dice are rolled. Find the probability that either doubles are rolled or the sum of the dice is 8.

43. For a person selected randomly from a certain population, events A and B are defined as follows.
A = event the person is male
B = event the person is a smoker
For this particular population, it is found that $P(A) = 0.50$, $P(B) = 0.28$, and $P(A \cap B) = 0.15$.
Find $P(A \cup B)$.

44. Let A and B be events such that $P(A) = 7/36$, $P(B) = 2/9$, and $P(A \cup B) = 29/72$. Determine $P(A \cap B)$.

45. Of the coffee makers sold in an appliance store, 6.0% have either a faulty switch or a defective cord, 2.8% have a faulty switch, and 0.5% have both defects. What percent of the coffee makers will have a defective cord?

46. The table below describes the smoking habits of a group of asthma sufferers.

	Nonsmoker	Occasional smoker	Regular smoker	Heavy smoker	Total
Men	363	33	89	46	531
Women	443	40	66	44	593
Total	806	73	155	90	1124

If one of the 1124 people is randomly selected, find the probability that

a. the person is a man.

b. the person is a heavy smoker.

c. the person is a man or a heavy smoker.

4.3 Conditional Probability

47. If a single fair die is rolled,

a. find the probability of a 5 given that the number rolled is odd.

b. find the probability to get an odd number given that the number rolled is 5.

48. At a California college, 23% of students speak Spanish, 5% speak French, and 3% speak both languages. A student is chosen at random from the college. What is the probability that the student speaks Spanish if she speaks French?

49. A group of volunteers for a clinical trial consists of 81 women and 77 men. 18 of the women and 19 of the men have high blood pressure.

a. If one of the volunteers is selected at random find the probability that the person has high blood pressure given that it is a woman.

	High blood pressure	Normal blood pressure	Total
Men			
Women			
Total			

b. If one of the volunteers is selected at random find the probability that it is a woman given the person has high blood pressure.

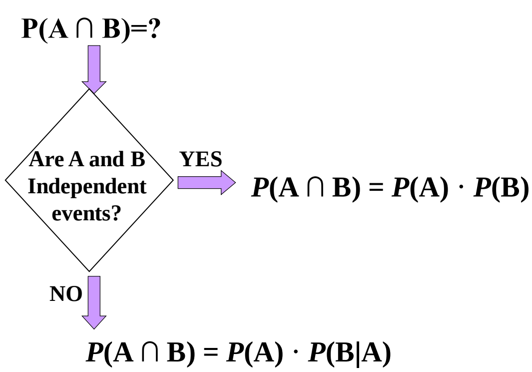
50. An auto insurance company was interested in investigating accident rates for drivers in different age groups. The following table was based on a random sample of drivers and classifies drivers by age group and accident rate.

	Under 25	25 - 45	Over 45	Total
0	0.121	0.194	0.363	0.678
1	0.080	0.075	0.103	0.258
2+	0.031	0.015	0.018	0.064
Total	0.232	0.284	0.484	1

a. What is the probability that a person, who had no accidents in the past three years, is over 45?

b. What is the probability that a person, who is over 45, had no accidents in the past three years

4.4 Multiplication Rule (Probability of an Intersection of Two Events)



51. Two cards are drawn with replacement from an ordinary deck of cards. Find the probability that two clubs are drawn in succession.

52. Two cards are drawn without replacement from an ordinary deck of cards. Find the probability that two clubs are drawn in succession.

53. Two marbles are drawn with replacement from a bag containing 7 blue and 3 red marbles. What is the probability of getting a blue on the first draw and a red on the second draw?

54. Two marbles are drawn with replacement from a bag containing 7 blue and 3 red marbles. What is the probability of getting one blue and one red marbles?

55. Two marbles are drawn without replacement from a bag containing 7 blue and 3 red marbles. What is the probability of getting one blue and one red marbles?

56. You are drawing three cards from a desk of 52, one at a time, with replacement. Find the probability that your cards are all diamonds.

57. You are dealt a hand of three cards, one at a time. Find the probability that your cards are all diamonds.

58. You are dealt a hand of three cards, one at a time. Find the probability that you have at least one red card.

59. 60% of students at one college drink coffee, and 16% of people who drink coffee suffer from insomnia. What is the probability that a randomly selected student drinks coffee and suffers from insomnia?.

4.5 Independent Events

60. Are Blood Pressure (BP) and Gender statistically independent?

	High BP	Normal BP	Total
Men	0.097	0.400	0.497
Women	0.114	0.389	0.503
Total	0.211	0.789	1

61. Are the accidents rate and age independent?

	Under 25	25 - 45	Over 45	Total
0	0.121	0.194	0.363	0.678
1	0.080	0.075	0.103	0.258
2+	0.031	0.015	0.018	0.064
Total	0.232	0.284	0.484	1.000

62. According to a survey, 8% of students at a college are left handed, 53% are female, and 4.24% are both female and left handed. Is being left handed independent of gender? Explain.
    a. No, 4.24% of students are both female and left handed
    b. No, because $P(L \text{ and } F) \neq P(L) \cdot P(F)$
    c. Yes, 8% of all students are left handed and 8% of female students are left handed.
    d. Yes, P(left handed) = P(left handed and female)
    e. No, 8% of all students are left handed but 4.24% of female students are left handed. These are not equal

63. Prove the following Proposition: If events A and B are independent, so are A' and B'.

64. Prove the following Proposition: If events A and B are independent, so are A and B'.

4.6 Bayes' Formula

65. Samuel has a flight to catch on Monday morning. His father will give him a ride to the airport. If it rains, the traffic will be bad and the probability that he will miss his flight is 0.05. If it doesn't rain, the probability that he will miss his flight is 0.01. The probability that it will rain on Monday is 0.28.
    a. What is the probability that Samuel misses his flight?

    b. Samuel missed his flight. Find the probability that it was a rainy day.

66. 58% of drivers in Texas are males. The probability to have an accident for male driver is 0.12, for female driver is 0.09.
    a. Find the probability that one randomly selected driver had an accident.

    b. If one randomly selected driver has an accident in the record, what is the probability that this driver is a man?

67. A manufacturing firm orders computer chips from three different companies: 20% from Company A; 30% from Company B; and 50% from Company C. Some of the computer chips that are ordered are defective: 3% of chips from Company A are defective; 1% of chips from Company B are defective; and 2% of chips from Company C are defective.

    a. Find the probability that a worker at the manufacturing firm discovers that a randomly selected computer chip is defective.

    b. What is the probability that this defective computer chip came from Company A?

Topic 5. Discrete Random Variables (DRV)

5.1 Probability Distribution of a Discrete Random Variable

Events 	Zero Heads	One Head	Two Heads	
Outcomes				
Probability				

68. Two coins are flipped. Random Variable X - Number of heads; x – possible values
Probability Model:

Values ***x***	0	1	2	
Probability ***p***				Sum = 1

Write the probability Distribution Function:

Write the Cumulative Distribution Function:

Sketch both graphs:

69. You roll a pair of fair dice. If you get a sum greater than 10 you win $60. If you get a double you win $20. If you get a double and a sum greater than 10 you win $80. Otherwise you win nothing.
Create a probability model for the amount you win at this game.

70. Determine if the following is a probability distribution.

x	$P(x)$
0	0.243
1	0.167
2	0.213
3	0.149
4	0.232
5	0.164

5.2 DRV: Center of the Distribution (Expected Value) $\mu = E(X) = \sum x_i \cdot p_i$

The value we expect a random variable to take on, notated µ***(x)*** for population mean or E***(X)*** for Expected Value.
The expected value of a discrete random variable can be found by summing the products of each possible value by the probability that it occurs.

Note: Be sure that every possible outcome is included in the sum and verify that you have a valid probability model to start with.

71. A wheel comes up green 50% of the time and red 50% of the time. If it comes up green, you win $100, if it comes up red you win nothing. Intuitively, how much do you expect to win on one spin, on average?

72. Two coins are flipped. Find the expected number of Heads.

Values ***x***	0	1	2	
Probability ***p***				Sum = 1
$x_i \cdot p_i$				$\sum x_i \cdot p_i = E(X)$

73. Find the expected value of the random variable X, using the chart below:

x	10	20	30	
p	0.3	0.5	0.2	1

74. Suppose you pay $2.00 to roll a fair die with the understanding that you will get back $4.00 for rolling a 2 or a 3, nothing otherwise. What is your expected value?

75. A bowl contains five chips, which cannot be distinguished by a sense of touch alone. Three of the chips are marked $1 each and the remaining two are marked $4 each.
A player is blindfolded and draws, at random and without replacement, two chips from the bowl. The player is paid an amount equal to the sum of the values of the two chips that he draws and the game is over.
If it costs $4.75 to play this game, would we care to participate for any protracted period of time?

76. You pick a card from a deck. If you get a face card, you win $15. If you get an ace, you win $30 plus an extra $50 for the ace of hearts. For any other card you win nothing.

a. Create a probability distribution table for variable X - the amount you win at this game.

X					
P(***X*** = *x*)					

b. How much you expect to win playing this game for a long time?

c. Assume you loose $10 for any card other than face card or an ace. Find the expected value of X.

d. How to make this game fair?

77. Let X be a random variable – a number of girls in a tree-child family.

a. Build the probability distribution table: show all possible values and corresponding probabilities.

Values ***x***					
Probability ***p***					

b. Sketch the graph of the distribution.

c. Find the mean – the expected number of girls.

5.3 Discrete Random Variables: Solving problems using Simulations

Coin Simulation.

Suppose we want to find the probability that a three-child family contains exactly one girl. We can find the theoretical answer by applying the rules of probability. We can also estimate this probability if we could observe a large number of three-child families and count the number that contain exactly one girl. We can also simulate the outcomes for three-child families.

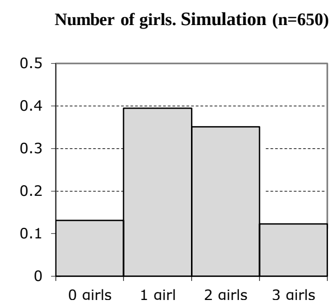
Simulation is a procedure developed for answering questions by running experiments that closely resemble the real situation. We use a **randomizing device** (fair coin, die, deck of cards, spinner, or random number table) with probability of having a girl = 0.50. We let "coin lands heads up" or "even digit" or "red card" to represent a female birth.

Coin Simulation: Number of Girls in the Family of Three

If we know that the chance of having a girl is 0.50 we can use a "fair" coin, a deck of cards, a spinner, or any device that generates P(girl) = 0.5. Since we want three-child families, we need to toss 3 coins or one coin 3 times; select one card, return it, shuffle, select a second card, return and shuffle and pick the third card; use the spinner 3 times to represent each birth.

150 students used the coin and repeated the three tosses many, many times (650). We got the following results:

Number of Heads	0 girl	1girl	2girls	3girls	Total
Observed Frequencies	85	257	228	80	650
Relative Frequencies	0.131	0.395	0.351	0.123	1
Theoretical Probabilities	0.125	0.375	0.375	0.125	1



Compare with the theoretical distribution obtained in Problem 76b.

Compare the theoretical and empirical probabilities that one randomly selected family has at least two girls.

5.4 DRV: Variation (spread, dispersion)

The Variance for a discrete random variable is: $\sigma^2 = Var(X) = \sum (x_i - \mu)^2 \cdot p_i$

The Standard Deviation for a random variable is: $\sigma = SD(X) = \sqrt{Var(X)}$

78. Two coins are flipped. Find the variance and standard deviation for the random variable X – number of Heads.

x_i	p_i	$x_i \cdot p_i$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 \cdot p_i$
		$E(X) = \mu =$			$Var(X) = \sigma^2 =$

How to use the shortcut formula for the Variance: $Var(X) = \sigma^2 = E(X^2) - [E(X)]^2$

x_i	p_i	$x_i \cdot p_i$	$(x_i)^2$	$(x_i)^2 \cdot p_i$
		$E(X) =$		$E(X^2) =$

$\sigma^2 = E(X^2) - [E(X)]^2 =$

79. The probability model for a particular life insurance policy is shown.

X - cost	Death \$100,000	Disability \$50,000	Neither \$0	
Probability	1/1,000	2/1,000	997/1,000	1

a. Find the expected annual payout on a policy.

b. Find the standard deviation of the annual payout.

5.5 Properties of the Expected Value, Variance and Standard Deviation:

1. $E(X+b) = E(X) + b$	1. $Var(X+b) = Var(X)$
2. $E(X \cdot a) = a \cdot E(X)$	2. $Var(X \cdot a) = a^2 \cdot Var(X)$
3. $E(aX+b) = aE(X) + b$	3. $Var(aX+b) = a^2 Var(X)$
4. $E(X \pm Y) = E(X) \pm E(Y)$	4. $Var(X \pm Y) = Var(X) + Var(Y)$ $-$ *for Independent RV*

80. X - employee's monthly salary Mean = \$5,000 Var =\$1,000,000 St.Dev = \$1,000

a. Everyone in a company receiving a \$500 increase in salary. New random variable (X + 500).
Find Expected Value, Variance and Standard Deviation

b. Salary was doubled. New random variable is 2X. Find the Expected Value, Variance and Standard Deviation

81. Two independent random variables X and Y are determined as follow:

Find Expected Value, Variance and Standard Deviation for the following random variables S, D and V:

	Mean	Var	St. Dev
X	20	25	5
Y	10	4	2

a. $S = X + Y$

b. $D = X - Y$

c. $V = 2X + 3Y$

82. Consider the return per $1000 for two types of investments. Find the Mean and Standard Deviation of Investment Returns.

Probability	Economic Condition	Investment	
		Passive Fund X	Aggressive Fund Y
0.2	Recession	- $25	- $200
0.5	Stable Economy	+ $50	+ $60
0.3	Expanding Economy	+ $100	+ $350

Topic 6. Discrete Probability Distributions

6.1 Bernoulli Random Variable: result of a single trial

	success	*failure*	
Values X	1	0	
Probabilities	*p*	*1 - p*	**1**

Expected Value and Variance:

$$\mu = E(X) = p$$
$$\sigma^2 = Var(X) = p(1-p)$$
$$\sigma = \sqrt{p(1-p)}$$

Let $X_1, X_2, ...X_n$ be n independent Bernoulli random variables.

Then $X = X_1 + X_2 + ... + X_n$ - the number of successes in n independent trials

6.2 Binomial Distribution: result of n independent Bernoulli trials

Random Variable X -the number of successes in n independent trials

Values X	0	1	2	...	n	
Probabilities	$p_0 = P(X=0)$	$p_1 = P(X=1)$	$p_2 = P(X=2)$	...	$p_n = P(X=n)$	$\sum_{i=1}^{n} p_i = 1$

Notations:

n - number of trials
p - probability of success
(1 – p)= q - probability of failure

Binomial distribution appears when:
there are n independent trials;
there are two possible outcomes for each trial (success, failure)
the probability of success, *p*, is constant for all n trial

$$Bin(n, p)$$

$$P(X=k) = {}_nC_k \cdot p^k \cdot (1-p)^{n-k} \quad where \quad {}_nC_k = \frac{n!}{k!(n-k)!} \quad and \quad n! = n \times (n-1) \times ... \times 2 \times 1$$

83. 3 fair coins are tossed. What is the probability to get exactly 2 heads? No heads? At least one?

84. Toss 1 coin 5 times in a row. Note number of tails. What's the probability of 3 tails?

85. Multiple-choice questions:

Choosing 8 marbles from a box of 40 marbles (20 purple, 12 red, and 8 green) one at a time without replacement, keeping track of the number of red marbles chosen.

a. Not binomial: there are too many trials.
b. Not binomial: there are more than two outcomes for each trial.
c. Procedure results in a binomial distribution.
d. Not binomial: the trials are not independent.

Choosing 8 marbles from a box of 40 marbles (20 purple, 12 red, and 8 green) one at a time with replacement, keeping track of the number of red marbles chosen.

a. Not binomial: there are too many trials.
b. Not binomial: there are more than two outcomes for each trial.
c. Procedure results in a binomial distribution.
d. Not binomial: the trials are not independent.

Choosing 10 marbles from a box of 40 marbles (20 purple, 12 red, and 8 green) one at a time with replacement, keeping track of the colors of the marbles chosen.

a. Not binomial: there are too many trials.
b. Procedure results in a binomial distribution.
c. Not binomial: there are more than two outcomes for each trial.
d. Not binomial: the trials are not independent.

86. You're a telemarketer selling service contracts for Macy's. You've sold 20 in your last 100 calls. If you call 12 people tonight, what's the probability of

a. No sales?

b. Exactly 2 sales?

c. At least 2 sales?

87. Assume that 13% of people are left-handed. If we select 10 people at random, what is the probability that exactly 3 of them are left-handed?

88. A coin is tossed and you win a prize if there are exactly 50% heads. Which is better: 10 tosses or 100 tosses?

Expected Value and Variance of Binomial Random Variable

Note: Binomial variable can be considered as the sum of n independent Bernoulli random variables with the mean *p* and variance *p(1-p).* Therefore the mean and variance of this sum is the sum of means and variances.

$E(X) = np$ $\quad\quad Var(X) = \sigma^2 = np(1-p) = npq$ $\quad\quad St.Dev(X) = \sigma = \sqrt{npq}$

89. What is the expected number of heads if you toss the coin two times? Three times? Five times?

90. . What is the mean and standard deviation of Binomial distribution with parameters
    a. $n = 5$ $p = 0.1$
    b. $n = 5$ $p = 0.5$

91. Assume that 13% of people are left-handed. If we select 10 people at random. Find the expected value (mean), variance and standard deviation of the number of left handed people.

92. Suppose that only 25% of all drivers come to a complete stop at an intersection having flashing red lights in all directions when no other cars are visible. What is the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions?
    a. Exactly 6 will come to a complete stop?
    b. Between 5 and 7 (inclusive) will come to a complete stop?
    c. At least 3 will come to a complete stop?
    d. How many of the next 20 drivers do you expect to come to a complete stop?

6.3 Poisson Distribution $P(X=k)=\dfrac{\lambda^k \cdot e^{-\lambda}}{k!}$

93. Customers arrive at a rate of 72 per hour. What is the probability of 4 customers arriving in 3 minutes?

94. If approximately 2% of the people in a room of 200 people are left-handed, find the probability that exactly five people there are left-handed. Compare with the binomial probability.
    Poisson Model: Binomial Mod

Does the Poisson distribution show good approximation to the true value when n is large, and p is small?

95. A scientific article reports that 1 in 200 people carry the defective gene that causes inherited colon cancer. In a sample of 1,000 individuals, what is distribution of the random variable – the number who carries this gene?
    a. Find the probability that between 2 and 4 (inclusive) carry this gene.

    b. At least 3 carry the gene?

96. A classic example of the Poisson distribution involves the number of deaths caused by horse kicks to men in the Prussian Army between 1875 and 1894. Data for 14 corps were combined for the 20-year period, and the 280 corps-years included a total of 196 deaths. After finding the mean number of deaths per corps-year, find the probability that a randomly selected corps-year has the following number of deaths: 0, 1, 2, 3 , 4 .
The actual results consisted of these frequencies: 0 deaths in 144 corps-years; 1 death in 91 corps-years; 2 deaths in 32 corps-years; 3 deaths in 11 corps-years; 4 deaths in 2 corps-years.

Compare the actual results to those expected from the Poisson probabilities. Does the Poisson distribution serve as a good device for predicting the actual results?

The mean number of deaths per corps-year is lambda – the parameter of Poisson distribution $\lambda = \frac{196}{280} = 0.7$

X	$P(X=k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$	Expected frequencies	Actual frequencies
0	P(X = 0) =		144
1	P(X = 1) =		91
2	P(X = 2) =		32
3	P(X = 3) =		11
4	P(X = 4) =		2

The comparison of frequencies in the last two columns shows that a Poisson distribution is a good model for our data.

6.4 Geometric Distribution $\quad Geom(p) \quad P(X=k)=(1-p)^{k-1}\cdot p$

97. Three fair dice are tossed until a sum of 18 appears for the first time.
What is the probability that less than 6 rolls will be required for that to happen?

98. People with O-negative blood are called “universal donors” because O-negative blood can be given to anyone else, regardless of the recipient’s blood type. Only about 6% of people have O-negative blood. If donors line up at random for a blood drive, how many do you expect to examine before you find someone who has O-negative blood? What’s the probability that the first O-negative donor found is one of the first four people in line?

6.5 Negative Binomial Distribution $\quad NegBin(r,p) \quad P(X=k)={}_{k-1}C_{r-1}\cdot p^r\cdot(1-p)^{k-r}$

99. From the past experience, it is known that a telemarketer makes a sale with probability 0.20. Assuming the results from one call to next are independent, find the probability that the telemarketer makes the second sale after 10^{th} call. What is the expected number of calls before the 2nd sale?

6.6 Hypergeometric Distribution $\quad Hyp(n,m,r) \quad P(X=k)=\dfrac{{}_rC_k\cdot{}_{n-r}C_{m-k}}{{}_nC_m}$

100. A shipment contains 100 items and a sample of size m = 2 is inspected. Let a RV ***X*** denote the number of defectives in the sample. If the manufacturer decides to return all 100 items if $X\geq 1$, what is the probability that 10% defective shipment will be accepted?

101. 3 different computers are checked out from 10 in the department. 4 of the 10 computers have illegal software loaded.
What is the probability that 2 of the 3 selected computers have illegal software loaded

Topic 7. Continuous Random Variables (CRV)
7.1 CRV: Probability Distribution

$$pdf \quad f(x): P(a \le X \le b) = \int_a^b f(x)dx$$

$f(x) \ge 0 \ for \ \forall x$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$P(X = c) = 0$

$P(a \le X \le b) = P(a < X < b)$

$P(a \le X \le b) = F(b) - F(a)$

102. Random variable X has the pdf $f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$

Find the following probabilities:

a. $P(X \le 1)$

b. $P(0.5 \le X \le 1.5)$

c. $P(X > 1.5)$

103. For the given pdf $f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$, find

a. Cumulative distribution function cdf and sketch its graph

$$F(x) = \begin{cases} & x < 0 \\ & 0 \le x \le 2 \\ & x > 2 \end{cases}$$

b. $P(X \le 1)$

c. $P(0.5 \le X \le 1.5)$

d. $P(X > 1.5)$

7.2 Numerical Characteristics of CRV:

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx$$

$$Var(X)=\int_{-\infty}^{+\infty}\left[x-E(X)\right]^2\cdot f(x)dx=\sigma^2$$

$$St.Dev(X)=\sqrt{Var(X)}=\sqrt{\sigma^2}=\sigma$$

Shortcut Formula for Variance

$$Var(X)=E(X^2)-\left[E(X)\right]^2=$$

$$=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx-\left[\int_{-\infty}^{+\infty} x\cdot f(x)dx\right]^2$$

104. Find the Expected Value, Variance and Standard Deviation

$$f(x)=\begin{cases}0.5x & 0\le x\le 2\\ 0 & otherwise\end{cases}$$

7.3 Chebyshev's Theorem

$P(|X-\mu|<\varepsilon)\ge 1-\dfrac{\sigma^2}{\varepsilon^2}$ or $P(|X-\mu|<k\cdot\sigma)\ge 1-\dfrac{1}{k^2}$ when $\varepsilon=k\cdot\sigma$

105. A machine used to fill out cereal boxes dispenses, on the average, **μ** ounces per box. The manufacturer wants the actual ounces dispensed Y to be within 1 ounce of **μ** at least 75% of the time.
What is the largest value of **σ**, the standard deviation of Y that can be tolerated if the manufacturer's objectives are to be met?

Topic 8. Continuous Probability Distributions

8.1 Uniform Distribution

$$pdf \quad f(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & otherwise\end{cases} \qquad cdf \quad F(x)=\begin{cases}0 & x\langle\, a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & x\,\rangle\, b\end{cases} \qquad E(X)=\frac{a+b}{2} \quad Var(X)=\frac{(b-a)^2}{12}$$

106. A Random Variable X is uniformly distributed between -5 and +5.

a. Find pdf. Sketch the graph

b. $P(X\le 0)$

c. $P(0\le X\le 1)$

d. Find Expected Value and Variance of random variable X.

107. The time required to prepare a dry cappuccino using whole milk at the Coffee House is uniformly distributed between 25 and 35 seconds. Assuming a customer has just ordered a whole - milk dry cappuccino.

a. Determine the probability density function of the preparation time.

b. What is the probability that the preparation time will be more than 29 seconds?

c. What is the probability that the preparation time will be between 28 and 33 seconds?

d. What percentage of a whole - milk dry cappuccinos will be prepared within 31 seconds?

8.2 Exponential Distribution

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & otherwise\end{cases} \qquad E(X)=\frac{1}{\lambda} \quad Var(X)=\sigma^2=\left(\frac{1}{\lambda}\right)^2 \quad \sigma=\frac{1}{\lambda}$$

The exponential distribution occurs naturally when describing the lengths of the inter-arrival times in a Poisson process.

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(y)dy=1-\mathrm{e}^{-\lambda x} \qquad P(a\le X\le b)=F(b)-F(a)=\mathrm{e}^{-\lambda a}-\mathrm{e}^{-\lambda b}$$

108. Customers arrive at the service counter according Poisson distribution with the rate of 15 per hour.

a. What is the probability that the arrival time between consecutive customers is less than three minutes?

b. What is the average wait time between two consecutive customers?

109. The average time between consecutive arrivals at the car wash station is 6 minutes. What is the probability that the arrival time between two cars

a. doesn't exceed 15 minutes?

b. is between 5 and 15 minutes?

110. A manufacturing plant uses a specific bulk product. The amount of product used in one day can be modelled by an exponential distribution with an average of 4 tons.

a. Find the probability that the plant will use more than 4 tons on a given day.

b. How much of the bulk product should be stocked so that the plant's chance of running out of the product is 5%?

111. Historical evidence indicates that times between fatal accidents on scheduled American domestic passenger flights have an approximately exponential distribution. The mean time between accidents is 44 days.

a. If one of the accidents occurred on July 1 of a randomly selected year, what is the probability that another accident occurred that the same month?

b. What is the standard deviation of the time between accidents?

Topic 9. Normal Distribution and its Applications

9.1 Normal Distribution X **Standard Normal Distribution** $z = \dfrac{x-\mu}{\sigma}$

$pdf \quad f(x;\mu,\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$P(a \le X \le b) = \int_a^b \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

$E(X) = \mu \quad Var(X) = \sigma^2 \quad St.Dev(X) = \sigma$

$f(z) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$

$P(Z \le z) = \int_{-\infty}^{z} \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt = F(z)$

$E(Z) = \mu = 0 \quad Var(Z) = 1 \quad St.Dev(Z) = \sigma = 1$

For all problems sketch the Normal curve and shade the areas.

112. If Z is a standard normal variable, find the probabilities:

a. The probability that Z is less than 1.13

b. The probability that Z lies between 0.7 and 1.98

c. The probability that Z lies between -1.10 and -0.36

d. P(Z > 0.59)

113. For a standard normal distribution, find the percentage of data that are more than 1 standard deviation away from the mean.

114. If P(Z > c) = 0.1093, find ***c***.

115. Find the 95th percentile of the standard Normal Distribution.

116. Assume that random variable ***X*** is normally distributed.

a. The mean is μ = 60.0 and the standard deviation is σ = 4.0. Find the probability that ***X*** is less than 53.0.

b. The mean is μ = 15.2 and the standard deviation is σ = 0.9. Find the probability that ***X*** is greater than 15.2.

c. The mean is μ = 137.0 and the standard deviation is σ = 5.3. Find the probability that ***X*** is between 134.4 and 140.1.

117. For a recent English exam, use the Normal model N(73, 9.2) to find the score that

a. represents the 90th percentile.

b. Cuts off bottom 5% of the distribution

118. The volumes of soda in quart soda bottles are normally distributed with a mean of 32.3 oz and a standard deviation of 1.2 oz. What is the probability that the volume of soda in a randomly selected bottle will be

a. less than 32 oz?

b. More than 34 oz?

119. Assume that the weights of quarters are normally distributed with a mean of 5.67 g and a standard deviation 0.070 g. A vending machine will only accept coins weighing between 5.48 g and 5.82 g. What percentage of legal quarters will be rejected?

9.2 Accessing Normality

How to see whether a Normal model is reasonable? (Statistical software FATHOM)

1. Login to TACS/Resources/Data Files and open the file Normal Distributionю
2. On the **FATHOM** window double click on the **Collection** US Presidents Age.
   Open a new **Graph** and make a histogram. You should try to vary the bin-widths to see how this affects the shape.
   Double click on graph and change the value of **binWidth** and **binAllignment Position** to make the histogram better.
   Describe the shape and the characteristic features of the distribution. Do you think the Normal Model will fit these data?
3. A more reliable way of checking to see whether the distribution is "nearly normal" is with a Normal Quantile Plot. If the distribution is normal, this plot is a straight line. If it is nearly normal, the plot will be nearly straight.
   Open a new **Graph** for presidents **Age**. From the Graph pull-down menu, select **Normal Quantile Plot.**
   Analyze the graph and make your conclusion: is the distribution of the data approximately Normal or the Normal distribution is not reasonable.

In addition, we can see graphically how the Normal curve fits the histogram:

4. Open the **Summary Table** and add **Basic Statistics**. Record the values of the mean and standard deviation.
5. Change the scale for the histogram by selecting **Graph _ Scale _ Density.** The graph now shows the histogram using densities (relative frequency equals area of corresponding bar).
6. Add the graph of Normal density curve: **Graph _ Plot Function**. In a new window select
   Functions /Distributions /Norma/l normalDensity(Age, *μ, σ*) – use the values of the mean and standard deviation.
   Make your conclusion how the Normal curve fits the data.

9.3 Normal Approximation to Binomial Distribution

$$Bin(n,p) \to N\left(\mu = np; \sigma = \sqrt{np(1-p)}\right), \; when \quad n \to \infty$$

Continuity correction: $P(a \le X \le b) \to \quad P(a - 0.5 \le X \le b + 0.5)$

120. A tennis player makes a successful first serve 78% of the time. If she serves 80 times, is it appropriate to use a normal model to approximate the distribution of the number of good first serves? Assume that each serve is independent of the others.
    a. Yes; normal model with μ = 62.4 and σ = 3.71 can be used to approximate the distribution
    b. Yes; normal model with μ = 62.4 and σ = 13.73 can be used to approximate the distribution
    c. Yes; normal model with μ = 17.6 and σ = 13.73 can be used to approximate the distribution
    d. No; normal model cannot be used to approximate the distribution because nq < 10 and np < 10

121. Assume that 13% of people are left-handed. If we select 100 people at random,
    a. What is the probability that at least 15 of them are left-handed?

    b. What is the probability that exactly 15 of them are left-handed?

122. Bill claims that he has a coin which is biased and which comes up heads more than tails. His claim is based on a trial in which he flipped the coin 200 times and got 110 heads. Is his claim fair?

Standard Normal Table

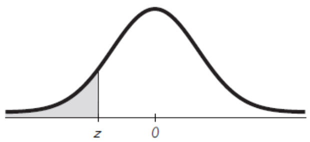


TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051 *	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505 *	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

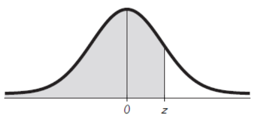


TABLE A-2 (continued) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495 *	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949 *	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 and up	.9999									

Topic 10 : Chi-square Distribution and Goodness – of – Fit Test

Goodness – of – Fit Test

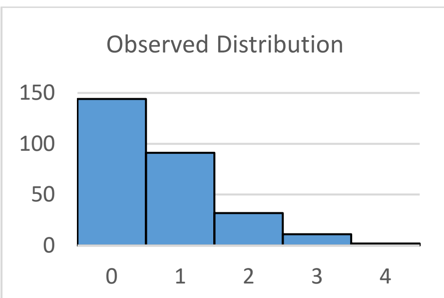
The Goodness-of-Fit tests measure the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the distribution you selected fits to your data.

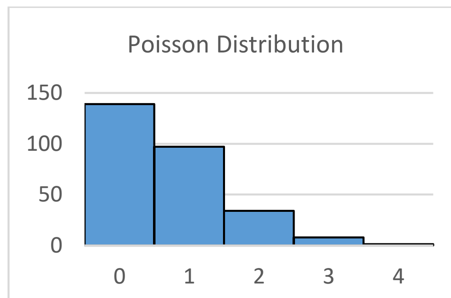
123. A classic example of the **Poisson distribution** involves the number of deaths caused by horse kicks to men in the Prussian Army between 1875 and 1894. Data for 14 corps were combined for the 20-year period, and the 280 corps-years included a total of 196 deaths. After finding the mean number of deaths per corps-year, find the probability that a randomly selected corps-year has the following number of deaths: 0, 1, 2, 3, 4.
The actual results consisted of these frequencies: 0 deaths in 144 corps-years; 1 death in 91 corps-years; 2 deaths in 32 corps-years; 3 deaths in 11 corps-years; 4 deaths in 2 corps-years.
Compare the actual results to those expected from the Poisson probabilities. Does the Poisson distribution serve as a good device for predicting the actual results?

The mean number of deaths per corps-year is lambda – the parameter of Poisson distribution $\lambda = \dfrac{196}{280} = 0.7$ Test:

$H_0 : f(x) = Poisson(\lambda = 0.7)$ $\qquad H_0 : f(x) = \dfrac{0.7^k \cdot e^{-0.7}}{k!}$

$H_0 : f(x) \neq Poisson(\lambda = 0.7)$ $\qquad H_0 : f(x) \neq \dfrac{0.7^k \cdot e^{-0.7}}{k!}$





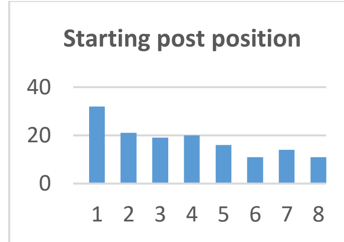
	Actual frequencies x_i	$P(X=k) = \dfrac{0.7^k \cdot e^{-0.7}}{k!}$	**Expected Frequencies** np_i	$(x_i - np_i)$	$(x_i - np_i)^2$	$\dfrac{(x_i - np_i)^2}{np_i} \sim \chi^2_{k-1}$
0	144	0.497	$0.497 \cdot 280 = 139.16$			
1	91	0.348	$0.348 \cdot 280 = 97.44$			
2	32	0.122	$0.122 \cdot 280 = 34.16$			
3	11	0.0284	$0.0284 \cdot 280 = 7.95$			
4	2	0.00497	$0.00497 \cdot 280 = 1.39$			
						Sum =

df = $\qquad$ Critical value (table) = $\qquad$ Observed χ^2 =

Sketch the Chi-square distribution and make your decision:

124. Is the distribution of racetrack winners a function of their starting-post positions? Test hypotheses at the level of significance $\alpha = 0.01$

Starting post	1	2	3	4	5	6	7	8	
Number of	32	21	19	20	16	11	14	11	Sum =



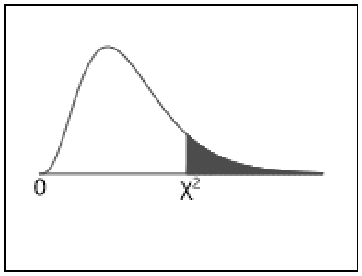
	Observed frequencies x_i	Probabilities p_i	Expected frequencies np_i	$(x_i - np_i)$	$(x_i - np_i)^2$	$\frac{(x_i - np_i)^2}{np_i} \sim \chi^2_{k-1}$
1	32					
2	21					
3	19					
4	20					
5	16					
6	11					
7	14					
8	11					
n	144					Sum =

125. At the significance level $\alpha = 0.01$ test the hypothesis that the distribution of IQ scores is Normal: $N(\mu = 100, \sigma = 15)$
An observed frequency distribution of sample of n = 200 IQ scores is as follow:

IQ score	Less than 80	80 - 95	95 - 110	110 - 125	125 - 140	More than 140	n
Frequency	14	26	80	40	31	9	200

Interval	Observed frequencies x_i	Probabilities p_i	Expected frequencies np_i	$(x_i - np_i)$	$(x_i - np_i)^2$	$\frac{(x_i - np_i)^2}{np_i} \sim \chi^2_{k-1}$
Less than 79.5	14					
79.5 – 94.5	26					
94.5 – 109.5	80					
109.5 - 124.5	40					
124.5 – 139.5	31					
More than 139.5	9					

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Topic 11. Joint Distributions

11.1 Discrete Joint Distributions

126. A supermarket has two express lines. Let X and Y denote the number of customers in the first and in the second, respectively, at any given time. During non-rush hours, the joint pdf of X and Y is summarized by the following table:

a. Find marginal probabilities of X and Y

		X				
		0	1	2	3	
Y	0	0.10	0.20	0	0	
	1	0.20	0.25	0.05	0	
	2	0	0.05	0.05	0.025	
	3	0	0	0.025	0.05	

b. Find the expected value and variance of the number of customers for each line.

c. What is the probability that there is exactly the same number of customers in both lines?

d. Find $P(|X-Y|=1)$, the probability that X and Y differ by exactly 1.

e. Are the numbers of customers in the two lines independent?

127. Five cards are dealt from a standard poker deck. Let X be the number of aces received, and Y, the number of kings. Compute $P(X=2\,|Y=2)$.

128. When an automobile is stopped by a roving safety patrol, each tire is checked for tire wear, and each headlight is checked to see whether it is properly aimed. Let X denotes the number of headlights that need adjustment, and let Y denote the number of defective tires. The joint probability distribution table is shown below:

		X					
Y		0	1	2	3	4	
	0	0.30	0.05	0.025	0.025	0.10	
	1	0.18	0.03	0.015	0.015	0.06	
	2	0.12	0.02	0.010	0.010	0.04	

Are X and Y independent?

129. Suppose that X and Y are discrete random variables with the joint distribution given as the following table:

(x, y)	$p(x, y)$
(1,2)	0.5
(1,3)	0.25
(2,1)	0.125
(2,4)	0.125

a. Find the covariance and correlation coefficient

b. Find the variance of the sum and the difference of X and Y

11.2 Continuous Joint Distributions

130.Joint pdf for X and Y is given as $f_{XY}(x,y)=\dfrac{c}{x}, \quad 0<y<x, \quad 0<x<1$

a. Determine the value of c

b. Find marginal distributions $f_X(x)$ and $f_Y(y)$

c. Find $P(X\ge 0.5, 0.1\le Y\le 0.3)$

131.Joint pdf for X and Y is given as $f_{XY}(x,y)=x+y, \quad 0\le x\le 1, \quad 0\le y\le 1$

a. Find $P(0\le X\le 0.5, 0\le Y\le 0.5)$

b. Find the conditional pdf of Y given *x*

c. Find $P(0\le Y\le 0.2|_{X=0.5})$

132. Are random variables X and Y independent if

a. $$f_{XY}(x,y)=\begin{cases}2 & 0\le x\le y\le 1\\ 0 & elsewhere\end{cases}$$

b. $$f_{XY}(x,y)=\begin{cases}e^{-x-y} & x\ge 0 \quad y\ge 0\\ 0 & x<0 \quad y<0\end{cases}$$

133. Suppose that random variables X and Y are independent with marginal pdf

$$f_X(x)=2x,\ 0\le x\le 1 \qquad f_Y(Y)=3y^2,\ 0\le y\le 1$$

Find $P(Y<X)$

134. Suppose that X and Y have the joint pdf $f_{XY}(x,y)=x+y,\quad 0\le x\le 1,\quad 0\le y\le 1$.

Find the covariance $Cov(X,Y)$ and correlation coefficient $\rho(X,Y)$.

Topic 12. Moments and Moment Generating Functions

12.1 Moments of a Random Variable

The *r*th noncentral moment of X is determined as $\mu_r = E(X^r)$

The *r*th central moment of X is determined as $\mu_r' = E\left(X - E(X)\right)^r$

The mean is the 1st noncentral moment of X: $E(X) = \mu_1$

The variance is the 2nd central moment of X: $Var(X) = E\left[X - E(X)\right]^2 = \mu_2'$

The skewness is the 3rd central moment of X: $\gamma = E\left[X - E(X)\right]^3 = \mu_3'$

The kurtosis is the 4th central moment of X: $k = E\left[X - E(X)\right]^4 = \mu_4'$

135. Find the expression for 2nd, 3rd and 4th central moments in terms of noncentral moments.

136. Find the coefficient of skewness for Bernoulli distribution.

$$\gamma_1 = \frac{E\left[X - E(X)\right]^3}{\sigma^3} \qquad \gamma_1 = \frac{\mu_3'}{\sigma^3}$$

137. Find the coefficient of skewness for an exponential random variable having pdf $f(x) = e^{-x}, \quad x \ge 0$

138. Find the center, variation, skewness and excess kurtosis for Uniform distribution (0, 1).

$$\gamma_2 = \frac{E\left[X - E(X)\right]^4}{\sigma^4} - 3 \qquad \gamma_2 = \frac{\mu_4'}{\sigma^4} - 3$$

139. The probability density function of the random variable X is given by $f(x) = \begin{cases} \dfrac{1}{2\sqrt{x}} & if\ 0 < x < 1 \\ 0 & elsewhere \end{cases}$

Find the mean, variance, coefficient of skewness and excess kurtosis.

.

140. Suppose the W is a random variable for which $E\left[W - \mu\right]^3 = 10$ and $E(W^3) = 4$. Is it possible that $\mu = 2$?

141. Suppose that the random variable X is described by the pdf $f(x) = cx^{-6}, \quad x > 1$.

a. Find c.

b. What is the highest moment of X that exists?

12.2 Moment Generating Functions

142. Two chips are drawn at random without replacement from a box that contains five chips numbered 1 through 5. If the sum of chips drawn is even, the random variable X equals 5; if the sum of chips drawn is odd, $X = -3$.
   a. Find the moment-generating function (mgf) for X.
   b. Use mgf to find the first and second non-central moments.
   c. Find the expected value and variance of X.

143. Find the moment-generating function for *X*, if X is a Bernoulli random variable with the pdf
$$f_X(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0 \text{ and } x = 1 \\ 0 & \text{otherwise} \end{cases}$$
Use mgf to find the expected value and variance of *X*.

144. Find the moment-generating function for X, if X is a Poisson random variable with the pdf $f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}$.

145. Find the moment-generating function for X, of X is an exponential RV: $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

146. Find the moment-generating function for the standard normal random variable Z with the pdf $f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$.
Use your result to find the moment-generating function for normally distributed random variable $X \sim N(\mu, \sigma)$.
Hint: use the property of moment-generating function $Y = a + bx \Rightarrow M_Y(t) = e^{at}M_X(bt)$.

147. Let X have pdf $f_X(x) = \frac{1}{x!}e^{-3} \cdot 3^x$
   a. Find the moment-generating function (mgf) $M_X(t)$ for X.
   b. Use mgf to find the first and second non-central moments.
   c. Find the expected value and variance of X.

148. Let *X* and *Y* be independent random variables and let α and β be scalars. Find an expression for the mgf of $Z = \alpha X + \beta Y$ in terms of mgf's of *X* and *Y.*

149. Use the mgf to show that if *X* follows the exponential distribution, *cX* ($c > 0$) does also.

150. Show that:
   a. If X_i follows a binomial distribution with n_i trials and probability of success $p_i = p \quad where\ i = 1, 2, \ldots n$ and the X_i are independent, then $\sum_{i=1}^{n} X_i$ follows a binomial distribution.
   b. Referring to part a, show that if p_i are unequal, the sum $\sum_{i=1}^{n} X_i$ does not follows a binomial distribution.