Supplementary Exercises for Math 218
Version I

written (except where indicated) and compiled by Cymra Haskell

Please note: These exercises are meant to supplement rather than replace the exercises in the text.

Sheet I
Probability

1. Consider the experiment of choosing one person at random from all students at USC. Consider the events:

   A = the student is female
   B = the student is a junior
   C = the student is a senior
   D = the student has taken a math course
   E = the student is in the business school

   Describe the following events in words.

   (a) A^c
   (b) $B \cup C$
   (c) $(B \cup C)^c$
   (d) $B \cap C$
   (e) $A \cap E$
   (f) $C \cap D^c$

2. Consider the experiment of drawing two cards from a deck of cards. Consider the events:

   A = 1st card is a face card
   B = 1st card is an ace
   C = 2nd card is a face card
   D = 2nd card is an ace

   In parts (a)-(f) write the event described in terms of A, B, C, and D.

   (a) Both cards are aces.
   (b) Neither card is an ace.
   (c) Exactly one of the cards is an ace.
   (d) The 1st card is either a face card or an ace.
   (e) At least one of the cards is a face card.
   (f) The 1st card is neither a face card nor is it an ace.
   (g) Consider all the pairs of events: AB, AC, AD, BC, BD, CD. Which of them, if any, are mutually exclusive?

3. For each experiment either draw a picture of the sample space that clearly indicates all its members, or list every element in the sample space, or describe clearly what the elements of the sample space are in such a way that it is apparent how many elements there are total. In each case choose the sample space so that all the outcomes are equally likely.

   (a) Roll two dice.
   (b) Select an entree at random from the three fish, three meat, and four vegetarian entrees offered at a restaurant.
   (c) Select two marbles at random (without replacement) from a bag that contains a red marble, an orange marble, a yellow marble, a green marble, and a blue marble.
   (d) Select two marbles at random (without replacement) from a bag that contains one blue marble and three green marbles.
   (e) Two siblings, Aisha and Bob, are randomly assigned to the chores of cleaning the bathroom, washing the dishes, and mopping the floor in such a way that each chore gets done and each sibling could end up doing none of the chores, exactly one of the chores, two of the chores or all three of the chores.
   (f) Select three people at random (without replacement) from Andrew, Betty, Chuck, David and Elizabeth.
   (g) Select three people at random (without replacement) from a group of two girls and three boys.

4. Consider the experiment of rolling a pair of dice.

   (a) Are the events "sum is 6" and "at least one of the dice shows a 3" mutually exclusive?
   (b) Are the events "sum is 6" and "at least one of the dice shows a 6" mutually exclusive?
   (c) Are the events "at least one of the dice shows a 3" and "at least one of the dice shows a 6" mutually exclusive?
   (d) Are the events "exactly one of the dice is 6" and "exactly one of the dice is 3" mutually exclusive?
   (e) Are the events "sum is less than 7" and "sum is greater than 7" mutually exclusive?

5. Consider the experiment of selecting 2 people from a group consisting of 8 women and 12 men.

   (a) Are the events "1st person selected is a man" and "2nd person selected is a man" mutually exclusive?
   (b) Are the events "1st person is a man" and "1st person is a woman" mutually exclusive?
   (c) Are the events "1st person is a man and second is a woman" and "1st person is a woman and second is a man" mutually exclusive?

(d) Are the events "exactly one woman is selected" and "exactly two women are selected" mutually exclusive?

6. Many businesses use TQM (Total Quality Management) but they profess different reasons for doing so. One hundred and twenty businesses that use TQM were surveyed and asked why they use TQM. The results are shown in the table below.

Reason	Number
increase productivity	61
stay competitive	50
comply with policies	5
other	4

One of these businesses is selected at random.

(a) What is the probability that the selected business uses TQM in order to comply with policies?

(b) What is the probability that the selected business uses TQM either to increase productivity or to stay competitive?

7. The contingency table below shows the ages and incomes of residents of a small town.

Income	Age			
($1000)	<18	$18-35$	>35	Total
<15	190	50	40	280
$15-50$	10	120	480	610
>50	0	30	80	110
Total	200	200	600	1000

A resident is selected at random.

(a) What is the probability that the resident selected is older than 35 and earns more than $50,000?

(b) What is the probability that the resident selected earns at least $15,000?

(c) What is the probability that the resident selected is less than 18 years old?

(d) What is the probability that the resident selected is at least 18 years old?

(e) Are the events "resident is less than 18 years old" and "resident is between 18 and 35 years old" mutually exclusive?

(f) Are the events "resident is more than 35 years old" and "resident earns less than $15,000" mutually exclusive?

(g) Are the events "resident is less than 18 years old" and "resident earns more than $50,000" mutually exclusive?

8. 62%of all students at a certain school have access to a computer at home. 53% have easy access to a library in their neighborhood that has a computer. 30% have access to a computer both at home and at a library in their neighborhood. A student is selected at random. What is the probability that the student has access to a computer either at home or at a library in their neighborhood?

9. Mary is particularly partial to raspberries and hazelnuts in chocolate candies. 80% of the candies in a box of assorted chocolates contain either raspberries or hazelnuts (or both). 40% contain hazelnuts and 60% contain raspberries. Mary selects a chocolate at random from the box. What is the probability that it contains both raspberries and hazelnuts?

10. Five candies are drawn from a bag full of candies. The probability that none of them are chocolate is .19. The probability that exactly one of them is a chocolate candy is .44.

(a) Find the probability that at most one of them is a chocolate candy.

(b) Find the probability that at least one of them is a chocolate candy.

11. Every time a particular circuit is turned on there is a chance that one or more of its components will fail. The probability that none of the components fail is .35. The probability that exactly one of the components fail is .39, and the probability that all the components fail is 10^{-10}.

(a) Find the probability that at most one of the components fails.

(b) Find the probability that at least one of the components fails.

(c) Find the probability that at least one of the components functions properly (does not fail).

12. One card is drawn at random from a deck of cards. Let A be the event that the card is a king, let B be the event that the card is a heart and let C be the event that the card is a diamond.

(a) Find $P(A/B)$

(b) Find $P(B/A)$

(c) Find $P(B/C)$

13. Two cards are dealt from a deck of cards.

    (a) Find the probability that the first card is a spade.
    (b) Find the probability that the second card is a spade given that the first card is a spade.
    (c) Find the probability that the second card is a spade.

14. A card is drawn from a deck of cards. It is replaced, the deck is reshuffled, and a second card is drawn (i.e. two cards are drawn with replacement).

    (a) Find the probability that the first card is a king.
    (b) Find the probability that the second card is a king given that the first card is a king.
    (c) Find the probability that the second card is a king.

15. A coin is tossed twice. Let A be the event that the first toss is a head, let B be the event that both tosses are heads, let C be the event that the second toss is tails.

    (a) Find $P(A)$
    (b) Find $P(B/A)$
    (c) Find $P(C/A)$
    (d) Find $P(A/B)$
    (e) Find $P(A/C)$

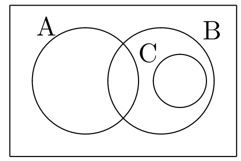
16. Two marbles are drawn from a bag (without replacement) that contains 2 white marbles, 3 blue marbles and 1 green marble. Let A be the event that the marbles are different colors, let B be the event that one of the marbles is green, and let C be the event that at least one of the marbles is blue.

    (a) Draw a picture of the sample space.
    (b) Find $P(A)$
    (c) Find $P(A/C)$
    (d) Find $P(C/A)$
    (e) Find $P(A/B)$. Look at your answer. Could you have worked this out without the picture of the sample space?
    (f) Find $P(B/A)$

17. A pair of dice is rolled.

    (a) Find the probability that one of the dice shows 4 given that the sum of the dice is 7.
    (b) Find the probability that the sum is 7 given that one of the dice shows 4.
    (c) Find the probability that the dice show the same number given that one of the dice shows 6.

18. Consider the Venn diagram below. $P(A) = .5$, $P(B) = .4$, $P(A \cap B) = .2$, and $P(C) = .1$.



    (a) Find $P(A/B)$
    (b) Find $P(B/A)$
    (c) Find $P(B/C)$
    (d) Find $P(C/A)$
    (e) Which pairs of events: AB, AC, BC, if any, are independent?
    (f) Which pairs of events, if any, are mutually exclusive?

19. Consider the contingency table in question 7 above showing the ages and incomes of residents of a small town. A resident of the town is selected at random.

    (a) A resident that earns less than \$15,000 is selected at random. What is the probability that the person selected is less than 18 years old?
    (b) An adult resident (at least 18 years old) is selected at random. What is the probability that the person selected earns more than \$50,000?
    (c) A resident of the town is selected at random. Are the events {earns more than \$50,000} and {is more than 35 years old} independent?

20. A class of 30 school children were asked to rate how much they liked spinach and how much they liked school. They gave each one a rating of L (like), O (okay), or H (hate). The results are shown in the contingency table below. In addition it is known that the events {like spincah} and {hate school} are mutually exclusive, the events {spinach is okay} and {school is okay} are independent and one sixth of the kids that think school is okay like spinach.

	school			
spinach	L	O	H	Total
L				6
O				8
H			4	
Total		30	5	40

(a) Fill in the gaps in the contingency table.

(b) What is the probability that a randomly selected child from this class hates school?

(c) A child that is selected at random hates spinach. What is the probability that they also hate school?

(d) Are the events {hate spinach} and {like school} independent? Explain.

21. In a certain company 83% of the female employees have one or more dependents and 53% of all the employees are women. An employee is selected at random. What is the probability that a woman with one or more dependents is selected?

22. In a certain school district 24% of all the schools are high schools, and 63% of the high schools have been rated "above average" based on the performance of their students on a standardized test. A school from the district is selected at random. What is the probability that it is a high school that has been rated above average?

23. A coin is tossed 3 times. Consider the following events.

$A = \{$1st toss is a head$\}$

$B = \{$2nd toss is a head$\}$

$C = \{$1st toss is a head and the second is not$\}$

$D = \{$2nd toss is a head and the first is not$\}$

(a) Are A and B independent? Are they mutually exclusive?

(b) Are C and D independent? Are they mutually exclusive?

(c) List all the outcomes that lie in A. List all the outcomes that lie in C. How are A and C related?

(d) Draw a Venn diagram that shows the relationship between the events A, B, C and D.

24. Two people are selected at random from a group of 8 women and 12 men. Consider the following events.

$A = \{$1st person is a woman$\}$

$B = \{$2nd person is a woman$\}$

$C = \{$1st person is a woman and the second is a man$\}$

$D = \{$1st person is a man and the second is a woman$\}$

$E = \{$Exactly one of the two people selected is a woman$\}$

(a) Find $P(A)$, $P(B)$, and $P(B|A)$.

(b) Are A and B independent?

(c) Are A and B mutually exclusive?

(d) Find $P(A \cap B)$.

(e) Are C and D independent? Are they mutually exclusive?

(f) Find $P(C)$ and $P(D)$.

(g) How is E related to C and D?

(h) Find $P(E)$.

25. Two cards are dealt from a deck of cards.

(a) Find the probability that both cards are aces.

(b) Find the probability that neither card is an ace.

(c) Find the probability that at least one card is an ace.

26. Three cards are dealt from a deck of cards.

(a) Find the probability that all three cards are spades.

(b) Find the probability that the first two cards are spades and the third card is a heart.

(c) Find the probability that none of the cards is a spade.

(d) Find the probability that at least one of the cards is a spade.

(e) Find the probability that exactly one card is a spade.

(f) Find the probability that at most one card is a spade.

27. Three marbles are selected from a bag that contains four red marbles, three black marbles and three green marbles.

(a) Find the probability that all three marbles are red.

(b) Find the probability that none of the marbles are red.

(c) Find the probability that at least one of the marbles is red.

(d) Find the probability that exactly one marble is red.

(e) Find the probability that at most one marble is red.

28. A class of 30 people was asked to name their favorite soft drink. The responses are shown in the table below.

Drink	Number of responses
Coca-cola	5
Pepsi	11
Dr. Pepper	8
7-up	6

Three people in the class are selected at random.

(a) What is the probability that they all prefer Pepsi?

(b) What is the probability that they all prefer something other than Pepsi?

(c) What is the probability that at least one of them prefers Pepsi?

29. The winning number of a lottery is a three digit number between 000 and 999. Each of the three digits is chosen randomly from the numbers between 0 and 9, independently of the other two digits.

(a) What is the probability that the winning number is 666?

(b) What is the probability of winning the lottery?

(c) What is the probability that the winning number contains no 6's?

(d) What is the probability that the winning number contains at least one 6?

30. A carrier has a 1 in 4 chance of passing a particular genetic disease to each child. The event that any one child acquires the disease is independent of the event that any other child acquires the disease. A carrier has four children.

(a) Find the probability that all of them acquire the disease.

(b) Find the probability that none of them acquire the disease.

(c) Find the probability that exactly one of them acquires the disease.

(d) Find the probability that at least one acquires the disease.

31. *(Adapted from an exercise in "A First Course in Statistics" by McClave and Sincich.)* Recall a few years ago when a flaw was discovered in the Pentium chip that was installed in many PC's. The flaw caused the computer to sometimes produce an incorrect result when dividing two numbers. This occurred in about one in every 9 billion (9,000,000,000) divides. At first Intel did not offer to replace the chip since that seemed like such a small probability. However, it is not uncommon for many statistical software packages to perform 1 billion divides over short period of time.

(a) Find the probability that in one billion divides all of them are performed correctly.

(b) Find the probability that in one billion divides at least one of them is performed incorrectly.

32. There are three power lines from an island to the mainland. In a fierce storm each power line has a 10% chance of going out. What is the probability that at least one power line is functional in a fierce storm?

33. Consider the contingency table below that shows the percentage of residents of a hypothetical country that live in an urban or rural environment and have the income level indicated.

	Place of Residence		
Income	urban	rural	Total
$< 10K$	15%	20%	35%
$10K - 30K$	30%	14%	44%
$30K - 50K$	6%	3%	9%
$50K - 70K$	6%	2%	8%
$> 70K$	3%	1%	4%
Total	60%	40%	100%

Four residents are selected at random.

(a) Find the probability that they all earn less than 10K.

(b) Find the probability that at least one earns more than 10K.

(c) Find the probability that at least one of the residents lives in an urban environment.

Sheet II
Combinatorics

1. A committee consisting of 5 people is to be chosen from a group of 20 people.
   (a) How many different committees are possible?
   (b) How many committees are there that include both Aaron and Bernadette?

2. *(Adapted from an exercise in "An Introduction to Probability and Its Applications" by Larsen and Marx.)* A woman is attacked by two assailants. A day later the police show her a line-up of seven suspects. The line-up includes the two guilty people.
   (a) How many ways can the woman choose two people out of the line-up?
   (b) How many ways can the woman choose two people out of the line-up exactly one of whom is guilty?

3. A committee consisting of 5 people is to be chosen from a group of 8 men and 12 women. The committee must consist of 2 men and 3 women. How many such committees are possible?

4. A bag contains 25 marbles, 10 of which are white and 15 of which are black. How many samples of size 7 are there that consist of 3 white marbles and 4 black marbles?

5. 100 people participate in a raffle that will have three winners. How many different combinations of three winners are possible?

6. 100 people participate in a raffle. The raffle will have three grand-prize winners but everybody else will receive a small prize for their participation. How many different ways could the small and grand prizes be distributed among the participants?

7. Six people sit down to dinner. Two people order a cocktail but when the waiter brings them to the table he can't remember who ordered them. How many ways could the waiter distribute the cocktails to the dinner guests?

8. Six people sit down to dinner. Two of the guests order lamb and the other four order salmon but the waiter is having a bad day and can't remember who ordered what. How many ways could he distribute the meals to the dinner guests?

9. Three indistinguishable red marbles and five indistinguishable black marbles are to be lined up in a row. How many patterns are possible?

10. How many code-words can be made out of the letters SSSSSSFF?

Problems 11-14 were adapted from exercises in "Understandable Statistics," by Brase and Brase.

11. You have two decks of cards and draw one card from each deck.
    (a) How many different possible pairs of cards are there that you could draw?
    (b) How many of these pairs consist of a king and a king?
    (c) What is the probability that you draw a pair of kings?

12. A professor assigns 12 homework problems and randomly chooses 5 to be graded.
    (a) How many different possible sets of five problems are there that she might grade?
    (b) Jerry only did five problems. What is the probability that the professor chooses exactly those five problems as the ones that she grades?
    (c) Silvia did seven problems. Of all the sets of five problems that the professor has to choose among to grade, how many of them did Silvia complete? What is the probability that she completed all five of the problems that the professor grades?

13. The qualified applicant pool for six management trainee positions consists of seven women and five men.
    (a) How many different groups of applicants can be selected for the positions?
    (b) How many of those groups consist entirely of women?
    (c) If the trainee positions are selected by drawing names at random from the qualified applicant pool, what is the probability that the trainee class will consist entirely of women?

14. In the Colorado State Lotto game there are 42 numbers of which players choose any 6. The state then selects 6 numbers at random. The winning tickets are those on which the player's 6 numbers match the state's 6 numbers.

(a) From 42 numbers how many different sets of 6 numbers are possible?

(b) If you buy one lottery ticket what is your chance of winning the grand prize?

(c) If you buy ten lottery tickets (with different choices on each ticket) what is your chance of winning the grand prize?

15. A bag of marbles contains 6 red marbles and 2 black marbles.

    (a) How many samples of size 4 are there?

    (b) How many samples of size 4 are there that consist of 3 red marbles and 1 black marble?

    (c) If a simple random sample of size 4 is selected from the bag, what is the probability that it will consist of 3 red marbles and 1 black marble?

16. Consider a population consisting of 20 women and 10 men.

    (a) How many samples of size 3 are there?

    (b) How many samples of size 3 are there that consist of 2 women and 1 man?

    (c) If a simple random sample of size three is selected from this population, what is the probability that it will consist of 2 women and 1 man?

17. *(Adapted from an exercise in "An Introduction to Probability and its Applications," by Larsen and Marx.)* A woman is attacked by two assailants. A day later the police show her a line-up of seven suspects. The line-up includes the two assailants. Suppose the woman never saw her assailants and chooses two people from the line-up completely at random.

    (a) What is the probability that among the two people she chooses exactly one of them is guilty?

    (b) What is the probability that neither of the people she chooses is guilty?

18. An IRS auditor randomly chooses two returns out of 50 to audit. Unknown to the auditor, 15 of the 50 returns contain mathematical errors. Find the probability that both returns selected contain mathematical errors.

19. Widgets coming off an assembly line have a 10% defective rate. Five widgets are selected at random and tested.

    (a) What is the probability that all five of them are defective?

    (b) What is the probability that the first two are defective and the other three are not?

    (c) What is the probability that the second and fifth ones are defective and the other three are not?

    (d) How many different outcomes result in exactly two of the five being defective?

    (e) What is the probability that exactly two of the five are defective?

20. When a fish is hooked and then released there is a 5% chance that it will die. Suppose a group of anglers goes fishing and releases a total of 20 fish.

    (a) What is the probability that the 1st, 2nd, and 3rd fish released all die and the rest survive?

    (b) What is the probability that the 5th, 16th, and 17th fish released all die and the rest survive?

    (c) How many outcomes result in exactly three of the fish dying and the rest surviving?

    (d) What is the probability that exactly three of the fish die?

21. A quiz contains 10 multiple choice questions each with three possible answers only one of which is correct. A student guesses the answer to every question.

    (a) What is the probability that the student will answer the first six questions correctly and the rest incorrectly?

    (b) What is the probability that the student will answer the 1st, 2nd, 4th, 6th, 7th, and 10th questions correctly and the rest incorrectly?

    (c) How many ways can the student get exactly 6 correct answers?

    (d) What is the probability that the student answers exactly 6 questions correctly?

22. Suppose that, when properly trained, people have an 80% chance of passing a lie detector test even when lying. Eight members of a terrorist group that were involved in a bombing are detained by the police and given a lie detector test. All are well-trained in passing the lie-detector test. What is the probability that exactly two of them will be detected as lying?

Sheet III
Population variables and distributions

1. In each case determine if the variable described is qualitative, quantitative and discrete, or quantitative and continuous.

   (a) the presidential voting preferences of voters in the U.S.

   (b) the number of books bought by customers of a bookstore

   (c) the weight of professional football players

   (d) the marital status of all adults in the U.S.

   (e) the amount spent by customers of a bookstore

   (f) the strength in pounds per square inch of beams produced on an assembly line

   (g) the age of U.S. citizens

2. Many businesses use TQM (Total Quality Management) but they profess different reasons for doing so. A study revealed that the primary reason businesses use TQM has the distribution shown below.

Reason	Percentage
increase productivity	51%
stay competitive	42%
comply with policies	4%
other	3%

   (a) What percentage of businesses use TQM to comply with policies?

   (b) What percentage of businesses use TQM either to increase productivity or to stay competitive?

   (c) Of all the businesses that use TQM in order to either increase productivity or to stay competitive what percentage of them use it to increase productivity?

   (d) If one of the businesses is chosen at random what is the probability that it uses TQM in order to increase productivity?

3. The hair color of students in a class has the distribution below. There are 200 students total in the class.

Hair color	Fraction
red	0.01
black	0.55
brown	0.37
blond	

   (a) How many students have red hair?

   (b) What fraction of the students have blond hair?

   (c) What is the most commonly occurring hair color?

   (d) What fraction of the students have dark hair (black or brown)?

   (e) Of all the students in the class that have dark hair (brown or black) what percentage have black hair?

   (f) A student is chosen at random. What is the probability that the student has brown hair?

   (g) A student is chosen at random that has dark hair. What is the probability that the student has brown hair?

   (h) Two students are chosen at random. What is the probability that they both have brown hair?

4. A company makes packages of colored candies by randomly selecting a dozen candies from a pool of candies whose colors are distributed as below.

Color	Percent
red	30%
orange	10%
yellow	20%
green	25%
blue	15%

   You buy a bag of candies.

   (a) What is the probability that your bag doesn't contain any orange candies?

   (b) What is the probability that your bag contains exactly 4 green candies?

5. The age (in years) of students in a class has the distribution below. Students that are at least 21 years old are classified as 'mature' students by the university.

age	number of students
16	1
18	30
19	80
20	22
21	7
22	3
23	2
24	2
30	1
33	1
42	1

(a) What fraction of the students are mature?

(b) If a student is chosen at random from the class what is the probability that they are mature?

(c) If a mature student is chosen at random from the class what is the probability that they are at least 30 years old?

(d) Two students are chosen at random from the class. What is the probability that one of them is mature and the other is not?

6. It is of interest to employers to know how many dependents their employees have since this has an effect on how much they might expect to pay out in health care and other benefits. A certain company employs many young people with few dependents. The distribution of the number of dependents is given below. No employee has more than 5 dependents.

number of dependents	percentage
0	45%
1	27%
2	13%
3	7%
4	4%
5	

(a) What percentage of employees have 5 dependents?

(b) Of all employees that have at least one dependent, what percentage have more than one?

(c) What is the most common number of dependents that employees have?

(d) If an employee is chosen at random what is the probability that they have at least three dependents?

(e) What is the median number of dependents?

(f) What is the mean number of dependents, μ?

(g) Find the standard deviation, σ, of the number of dependents.

(h) Calculate the two numbers $\mu - 2\sigma$ and $\mu + 2\sigma$. What percentage of employees have between $\mu - 2\sigma$ and $\mu + 2\sigma$ dependents?

7. Explain why each of the following could or could not be the distribution of some variable.

(a)

x	-2	0	1	4
fraction	.1	.3	.4	.2

(b)

x	0	1	2	3
fraction	-.1	.5	.4	.2

(c)

x	-2	-1	1	2
fraction	.2	.3	.4	.2

8. Consider the distribution below.

x	0	10	20	30	40	50
%	5%	5%	25%	20%	30%	15%

(a) Find μ, σ^2, and σ.

(b) What percentage of the population has values of the variable that lie between $\mu - 2\sigma$ and $\mu + 2\sigma$?

(c) If one member of the population is chosen at random what is the probability that the x-value of the individual chosen is greater than 20?

9. Consider the distributions below.

x	4	5	6
fraction	.1	.8	.1

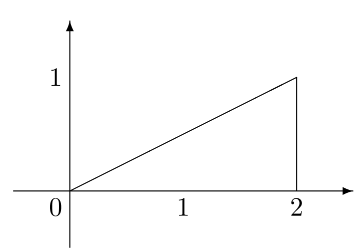
y	0	1	2
fraction	.3	.4	.3

z	4	5	6
fraction	.3	.4	.3

(a) Use your intuition to find the means of x, y and z.

(b) Without doing any calculations, determine which variable has the largest standard deviation and which has the smallest standard deviation. Why?

10. A continuous variable, x, has the distribution shown below.



(a) Find the proportion of the population for which the value of x lies between 0 and 1.

(b) Find the median of x.

(c) Without doing a calculation, is the mean of x

i) between 0 and 1, or

ii) equal to 1, or

iii) between 1 and 2?

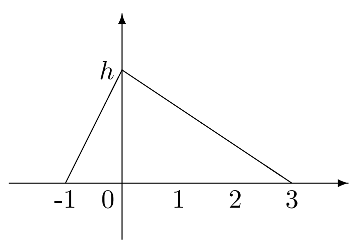
(d) Now calculate the mean of x.

(e) Find the standard deviation of x.

(f) A member of the population is selected at random. What is the probability that their x-value is greater than 1?

(g) Two members of the population are selected at random. What is the probability that both of them have x-values that are greater than 1?

11. A continuous variable, x, has the distribution shown below.



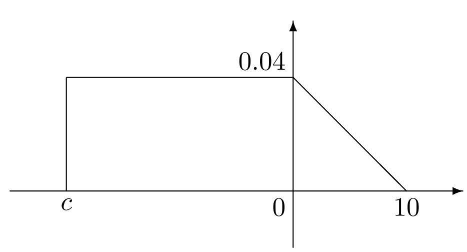
(a) Find the value of h.

(b) Find the equations of the two lines that make up the distribution of x.

(c) Find a value, c, such that 70% of the population has an x-value that is less than c.

(d) A member of the population is selected at random. Their x-value is greater than 1. What is the probability that their x-value is less than 2?

12. A continuous variable, x, has the distribution shown below.



(a) Find the proportion of the population that has a negative value of x.

(b) Find the number c.

(c) Find the mean of x.

(d) Two members of the population are chosen at random. The first one has a negative value of x. What is the probability that the second one also has a negative value of x?

13. Let x denote the lifetimes of lightbulbs manufactured by Lightning Co. The distribution of x is $f(x) = 0.001e^{-0.001x}$, $x \geq 0$.

(a) What proportion of lightbulbs last less than 500 hours?

(b) What proportion of lightbulbs last longer than 500 hours?

(c) If we buy two lightbulbs what is the probability that they will both last longer than 500 hours?

(d) If we buy two lightbulbs what is the probability that at least one of them lasts longer than 500 hours?

14. Let x be a variable whose distribution is given by $f(x) = c(1 - x^2)$, $-1 < x < 1$.

(a) What is the value of c?

(b) What proportion of the population has an x-value that is between 0 and 0.5?

(c) Find the mean of x.

(d) Find the standard deviation of x.

15. Let x be a variable whose distribution is given by $f(x) = 4x(1 - x^2)$, $0 < x < 1$.

(a) What proportion of the population has an x-value that is greater than 0.5?

(b) Find the mean of x.

(c) Find the standard deviation of x.

Sheet IV
Random variables and distributions

General distributions

1. *(Adapted from a problem in "A First Course in Probability" by Sheldon Ross)* A filling station is supplied with gasoline once a week. Its weekly volume of sales in thousands of gallons, X, has the distribution $f(x) = 5(1-x)^4$, $0 < x < 1$.

   (a) What is the probability that it will sell more than 750 gallons of gasoline in the week beginning March 2nd?

   (b) What should the capacity of the tank be so that the probability of the supply's being exhausted in a given week is 0.01?

   (c) What is its expected weekly sales of gasoline?

2. A random variable, X, has the distribution $f(x) = \frac{1}{3} - \frac{1}{18}x$, $0 < x < 6$.

   (a) Find $P[X > 3]$

   (b) Find c so that $P[X < c] = 0.5$

   (c) Find $E[X]$

   (d) Find $\mathrm{Var}[X]$ and $\mathrm{SD}[X]$.

Constructing probability distributions

3. A bag contains 3 blue marbles, 2 red marbles and a white marble. Two marbles are drawn at random from the bag. Let X denote the number of blue marbles that are drawn.

   (a) Draw a picture of the sample space and next to each outcome label the value of X associated with that outcome.

   (b) What outcomes lie in the event $\{X = 2\}$?

   (c) Find $P[X = 2]$.

   (d) What outcomes lie in the event $\{X < 2\}$?

   (e) Find $P[X < 2]$.

   (f) Write out the probability distribution of X.

   (g) Calculate $E[X]$. Might you have guessed this without calculating it?

4. You roll a pair of dice. If both numbers are greater than two then you win a dollar. Otherwise you lose a dollar. Let X denote your winnings.

   (a) Draw a picture of the sample space and next to each outcome write down the value of X associated with that outcome.

   (b) Write out the probability distribution of X.

   (c) Find $E[X]$. Is this a fair game?

5. A couple have six kids. Alex, Beth and Charles are all 5 years old (they were triplets), Deborah is 8 years old, and Ethan and Fran are both 10 years old (they were twins). Two children are selected at random to visit their grandparents. Let X denote the average age of the two children that are selected.

   (a) If Beth and Charles are chosen what is the value of X?

   (b) Write out all the outcomes in the event $\{X = 9\}$.

   (c) Draw a picture of the sample space of this experiment and next to each outcome write down the value of X.

   (d) Write out the probability distribution of X.

   (e) Find $E[X]$.

6. A bag contains ten tickets numbered 1 through 10. Three tickets are drawn at random (without replacement). Let Y denote the largest number that is drawn.

   (a) There are many outcomes to this experiment. Give some examples. Be sure to choose your outcomes in such a way that it is easy to find their probability of occurring. How many outcomes are there in total?

   (b) List all the outcomes in the event $\{Y = 3\}$.

   (c) Find $P[Y = 3]$.

   (d) List all the outcomes in the event $\{Y = 4\}$.

   (e) Find $P[Y = 4]$.

   (f) Find $P[Y = 7]$. *Hint:* The event $\{Y = 7\}$ is the same as the event that 7 is one of the three cards that is picked and the other two cards that are picked are both less than 7.

   (g) Find $P[Y = 6]$.

   (h) Write out the probability distribution of X.

   (i) Find $E[X]$, $\mathrm{Var}[X]$, and $\mathrm{SD}[X]$.

7. A machine forms butter into bars. All of the bars are labeled as containing 4 oz of butter, but the actual weights of the bars it produces are normally distributed with a mean of 4.1 oz and a standard deviation of 0.05 oz. An inspector

selects five bars at random. If all five of them weigh at least 4 oz then the company passes the inspection. If one or more of them weighs less than 4 oz then the company fails the inspection. Let X denote the random variable that assumes the value 'P' if the company passes and 'F' if the company fails. Find the probability distribution of X.

8. Goodies Inc. produces bags of assorted candies. Each bag contains either 5, 6, or 7 chocolate candies. Let x denote the number of chocolate candies in the bags produced by Goodies Inc. The distribution of x is given in the table below.

x	fraction
5	.3
6	.5
7	.2

Beth buys two bags of candies. Let X denote the total number of chocolate candies in both bags combined.

(a) Suppose the first bag that Beth opens has six chocolate candies in it and the second one has seven. What is the value of X? What is the probability of this occurring?

(b) Suppose the first bag that Beth opens has seven chocolate candies in it and the second one has six. What is the value of X? What is the probability of this occurring?

(c) Find $P[X = 13]$.

(d) Suppose Beth ends up with 12 chocolate candies total. One way this might have happened is if the first bag contained 5 chocolate candies and the second bag contained 7 chocolate candies. What are the other ways in which this may have happened?

(e) Find $P[X = 12]$.

(f) What are all the different possible values that X can assume?

(g) Write out the probability distribution of X.

(h) Find $E[X]$.

9. A coin is tossed three times. Let X take on the value 1 if the second toss is heads and take on the value 0 if the second toss is tails.

(a) Write out all the outcomes of this experiment and next to each outcome write down the corresponding value of X.

(b) Write out the probability distribution of X.

Sheet V
Commonly occurring families of distributions

Recognising and using the hypergeometric distribution

1. A five card hand is dealt from a deck of cards. Let X denote the number of spades in the hand.
   (a) What kind of distribution does X have? Be as spcific as you can.
   (b) Find $P[X = 3]$
   (c) Find $P[X \geq 3]$
   (d) Find $E[X]$.

2. A class consists of 25 girls and 30 boys. Seven students are selected at random for a school outing. Let X denote the number of girls that are selected.
   (a) What kind of distribution does X have? Be as specific as you can.
   (b) Find the probability that less than half of the students selected are girls.
   (c) What is the expected number of girls selected?

3. A city council has 15 members; 7 are Democrats, 5 are Republicans, and the other 3 are Independent. A committee of 5 people is selected at random from among all the members.
   (a) Find the probability that at most three of the committee members are Democrats.
   (b) Find the probability that at least 2 of the committee members are Independent.

4. The hair color of students in a class has the distribution below.

Hair color	Number of students
red	1
black	4
brown	23
blond	3

   Four students are selected at random to present problems at the board.
   (a) What is the probability that at least three of the four students selected have blond hair?
   (b) How many of the students are expected to have brown hair?

5. A professor assigns twenty problems for homework and chooses 5 of them at random to grade. Jason did 7 of the twenty problems thoroughly, 5 of the problems only partially and didn't attempt the other 8 problems.
   (a) What is the probability that all 5 of the problems that the professor grades are ones that Jason did thoroughly?
   (b) What is the probability that at least 3 of the problems that the professor grades are ones that Jason either did thoroughly or partially?

Recognising and using the binomial distribution

6. Widgets coming off an assembly line have a 5% chance of being defective. A quality control inspector selects eight widgets at random. Let X denote the number of defective widgets among those selected.
   (a) What is the distribution of X? Be as specific as you can.
   (b) What is the probability that two or more of the widgets selected are defective?
   (c) How many widgets are expected to be defective?
   (d) What is the standard deviation of the number of defective widgets selected?

7. As a marketing strategy Goodies Inc. is putting award tokens in its bags of mixed candies. Bags have either no award tokens in them, an award token that can be redeemed for a free bag of mixed candies, or an award token that can be redeemed for $100. The percentages of bags with the various types of tokens is shown in the table below.

Type	Percentage
No token	60%
Candies token	39%
$100 token	1%

   The Smith family buys seven bags of mixed candies in preparation for Easter.
   (a) What is the probability that at least three of them contain tokens for a free bag of candies?
   (b) What is the probability that at least one of them contains a $100 token?

8. Every week the professor selects a student at random to present a problem at the board. The selections from one day to the next are independent of each other and throughout the semester a student can be selected once, more than once, or not at all. There are 15 weeks in the semester. There are 12 students in the class of which 7 are girls and 5 are boys.

   (a) What is the probability that a girl is selected at least 9 times?

   (b) John is a student in the class. What is the probability that John gets selected more than once?

   (c) What is the probability that at least one student gets selected more than once?

9. The lifetimes of bulbs produced by Lightning Co. have the continuous distribution $f(t) = 0.001e^{-0.001t}$, $t \geq 0$. You buy a package of eight Lightning Co. lightbulbs. What is the probability that more than 2 of them blow out in less than 500 hours?

Binomial approximations of hypergeometric distributions

10. (a) A class with 60 students is voting on their representative to the PTA. There are two candidates, Andy and Beth. In fact 20 of the students favor Andy and 40 favor Beth but of course nobody knows these figures for sure. A student is curious how the election is going to turn out and decides to do a small poll to find out how her classmates plan to vote. Using the class enrollment list she is able to select a simple random sample of 5 people. Let X be the number of people in the sample that plan to vote for Andy. What kind of distribution does X have? Write out the distribution explicitly.

    (b) A party game involves rolling a die. If the side facing up is 1 or 2 then you get a chocolate. You play the game five times. Let Y be the number of chocolates you win. What kind of distribution does Y have? Write out the distribution of Y explicitly.

    (c) Compare the distributions of X and Y.

11. (a) You choose three cards from a deck of cards (without replacement). Let X denote the number of hearts in your sample. What kind of distribution does X have? Write out the distribution explicitly.

    (b) Components coming off an assembly line have a 25% chance of being defective. Let Y denote the number of defective components in a random sample of size three. What kind of distribution does Y have? Write out the distribution explicitly.

    (c) Compare the distributions of X and Y.

12. The Denver Post reported that a recent audit of the Los Angeles 911 calls showed that 85% were not emergencies. A sample of 4 of these calls is selected at random.

    (a) What is the probability that three of the four calls are not emergencies?

    (b) Is your answer to (b) exact? If not, how accurate do you think it is?

13. Suppose about 40% of all drivers flash their headlights to warn oncoming traffic of a speed trap ahead. Suppose seven cars pass a speed trap. What is the probability that at least one of the drivers will warn the oncoming traffic of the speed trap?

Uniform distribution

14. The time it takes a mechanic to change the oil in a car is uniformly distributed between 10 minutes and 25 minutes.

    (a) What is the average or mean time it takes for her to change the oil?

    (b) What is the standard deviation of this time?

    (c) What proportion of the time does it take her more than 20 minutes?

    (d) You go to her to get your oil changed. What is the probability that it takes her less than 15 minutes?

    (e) You go to her to get your oil changed but you have to wait for her to finish working on the previous customer's car. She has already been working on the previous customer's car for 15 minutes. What is the probability that it will take her at least an additional 5 minutes?

    (f) You send two of your friends to her to get their oil changed. What is the probability that it takes her longer than 15 minutes to change the oil in at least one of your two friends' cars?

15. In London because of unpredictable traffic, the buses don't run on a fixed schedule. They do run at fixed frequencies however. On a certain line the buses run every fifteen minutes. This means that the amount of time passengers have to wait for the next bus is uniformly distributed between 0 and 15 minutes.

    (a) What is the probability that a passenger arriving at the bus stop has to wait more than 10 minutes?

    (b) Ten passengers are selected at random. What is the probability that at least 4 of them have to wait more than 10 minutes?

    (c) Among the ten passengers selected how many are expected to have to wait more than 10 minutes?

Normal distribution

16. A variable, z, has a standard normal distribution. What proportion of the population has a z-value that lies in the indicated range?

    (a) Between 0 and 2.35.

    (b) Between -3.12 and 0.

    (c) Greater than 2.14.

    (d) Less than 1.63.

    (e) Between 1.90 and 2.55.

17. A random variable, Z, has a standard normal distribution. Find the following probabilities.

    (a) $P[-2.05 < Z \leq -1.82]$.

    (b) $P[-2.30 \leq Z \leq 1.65]$.

    (c) $P[Z < -3.75]$.

    (d) $P[Z \geq -3.45]$.

    (e) $P[Z > 4.30]$.

18. A random variable, Z, has a standard normal distribution. Find the value, c, where

    (a) $P[0 < Z < c] = 0.3$

    (b) $P[Z < c] = 0.65$

    (c) $P[Z \geq c] = .25$

19. A variable, z, in a population has a standard normal distribution. Find the value of c where

    (a) 25% of the population has a z-value that lies between 0 and c.

    (b) 45% of the population has a z-value that is less than c.

    (c) 70% of the population has a z-value that is greater than c.

    (d) 80% of the population has a z-value that lies between $-c$ and c.

20. A random variable, X is normally distributed with $E[X] = 30$ and $SD[X] = 5$. Find the probabilities indicated.

    (a) $P[30 < X < 40]$.

    (b) $P[X < 45]$.

    (c) $P[36 < X < 42]$.

    (d) $P[22 < X < 27]$.

21. A variable, x, is normally distributed with $\mu = 30$ and $\sigma = 5$. Find the proportion of the population that has x-values in the range given.

    (a) Greater than 32.

    (b) Between 25 and 40.

    (c) Less than 24.

    (d) Greater than 27.

22. A variable, x, is normally distributed with $\mu = 12.2$ and $\sigma = 0.1$. Find the proportion of the population that has x-values in the range indicated.

    (a) Greater than 12.5.

    (b) Less than 12.

    (c) Between 11.9 and 12.25.

23. A certain brand of tire manufactured by Run Wild Inc. has a lifetime that is normally distributed with a mean of 81,200 miles and a standard deviation of 4,300 miles. For how many miles should they warranty their tires if they want 95% of them to last longer than the warranty?

24. Scores on a national test are normally distributed with a mean of 650 and a standard deviation of 75. The top 15% get an A. What scores receive A grades?

25. The pistons in a certain car are supposed to have a diameter of 5.3 inches with a tolerance of ± 0.02 inches. Under normal operations the machine used in the manufacturing process produces pistons whose diameters are normally distributed with a mean of 5.3 inches and a standard deviation of 0.01 inches.

(a) Under normal operations what proportion of the pistons produced are within the tolerance levels?

(b) The machine used in the manufacturing process is malfunctioning. The result is that the pistons presently being produced have diameters that are normally distributed with a mean of 5.321 inches and a standard deviation of 0.017 inches. What proportion of these pistons have diameters that are within tolerance levels?

(c) The piston manufacturers are investigating new machines to buy.Specifications of machines on the market indicate the mean and standard deviation of the pistons that they produce. The manufacturers need to have a mean of 5.3 inches. The smaller the standard deviation the more costly the machine. What standard deviation do the manufacturers need if they want 99% of all pistons produced by the machine to be within tolerance levels?

26. In a hypothetical country the heights of adult males is normally distributed with a mean of 70 inches and a standard deviation of 3 inches. Three adult males are selected at random.

    (a) What is the probability that all three of them are taller than 72 inches?

    (b) What is the probability that the shortest one is taller than 72 inches?

    (c) What is the probability that the shortest one is shorter than 72 inches?

    (d) What is the probability that exactly two of the three are taller than 72 inches?

27. A machine forms butter into bars. The weights of the bars are normally distributed with a mean equal to 4.1 oz and a standard deviation equal to 0.05 oz.

    (a) A bar is selected at random. It weighs more than 4 oz. What is the probability that it weighs less than 4.1 oz?

    (b) The bars are advertised as weighing at least 4 oz. An inspector randomly selects 8 bars and weighs them. If two or more of the bars weigh less than 4 oz the company is fined. What is the probability that the company will be fined upon inspection?

Sheet VI
Joint distributions and independence

1. The table below shows the joint distribution of the age and income of residents of a small town.

Income (*in $1000*)	*Age* < 18	18 − 35	> 35	Total
< 15	.19	.05	.04	.28
15 − 50	.01	.12	.48	.61
> 50	0	.03	.08	.11
Total	.2	.2	.6	1

   (a) What proportion of the residents earn less than $15,000?

   (b) What proportion of the residents that are less than 18 years old earn less than $ 15,000?

   (c) Is the set of individuals that earn less than $ 15,000 independent of the set of individuals that are less than 18 years old?

   (d) Are income and age independent?

2. 30% of the members of a club are married and 45% are managers. 35% are neither married nor are they managers.

   (a) Create a table showing the joint distribution of marital status (married/not married) and profession (manager/not manager) of the members of the club.

   (b) What percentage of the married members are managers?

3. The table below shows the joint distribution of sex and income level of employees of an hypothetical company.

Income	*Sex* female	male	Total
< 40K	246	62	308
40K - 80K	67	48	115
80K - 120K	13	9	22
> 120K	4	4	8
Total	330	123	453

   (a) What proportion of the employees earn less than 40 K?

   (b) What proportion of the female workers earn less than 40 K?

   (c) Are income and sex independent among the employees of this company?

4. To plan for a banquet the potential attendees are questioned by the caterers on whether or not they like shrimp and whether or not they like pasta. It turns out that their preferences for shrimp and pasta are independent. Complete the contingency table below.

Shrimp	*Pasta* like	dislike	Total
like			81
dislike			
Total	96		108

5. The joint distribution of random variables X and Y is given below.

X	Y = 0	1	
-1	.12	.08	.2
0	.06	.04	.1
1	.42	.28	.7
	.6	.4	

   (a) Write down the distribution of X alone.

   (b) Write down the distribution of Y alone.

   (c) Are the events $\{X = -1\}$ and $\{Y = 1\}$ independent?

   (d) Are X and Y independent?

6. The joint distribution of random variables X and Y is given below.

X	Y = 0	1	
-1	.12	.08	.2
0	.08	.02	.1
1	.4	.3	.7
	.6	.4	

   (a) Write down the distribution of X alone.

   (b) Write down the distribution of Y alone.

   (c) Are the events $\{X = -1\}$ and $\{Y = 1\}$ independent?

   (d) Are X and Y independent?

7. The distributions of the random variables X and and Y are given below.

r	-1	1
$P[X = r]$	0.4	0.6

r	0	1
$P[Y = r]$	0.7	0.3

(a) Write out the joint distribution if X and Y are independent.

(b) Create a possible joint distribution in which X and Y are dependent.

8. An experiment has two random variables, X and Y, that have the same distribution shown below.

r	0	1	2
Probability	.6	.3	.1

   (a) Write out the joint distribution of X and Y if they are independent.

   (b) Write out a possible joint distribution in which they are dependent.

9. A coin is tossed three times. Let X denote the number of heads that are tossed and let Y denote the number of tails that are tossed.

   (a) Write down all the outcomes of the experiment with their corresponding values of X and of Y.

   (b) Write out the probability distribution of X alone.

   (c) Write out the probability distribution of Y alone.

   (d) Write out the joint probability distribution of X and Y.

   (e) Are X and Y independent?

10. An experiment consists of tossing a coin and rolling a die. Let X denote what is showing on the coin (Heads/Tails) and let Y denote what is showing on the die (1/2/3/4/5/6).

   (a) Write out all the outcomes of the experiment and next to each outcome write down the value of X and the value of Y.

   (b) Write out the distribution of X alone.

   (c) Write out the distribution of Y alone.

   (d) Write out the joint distribution of X and Y.

   (e) Are X and Y independent?

11. A coin is tossed twice. Let X take the value 1 if the first toss is heads and 0 otherwise. Let Y take the value 1 if the second toss is heads and 0 otherwise.

   (a) Write out all the outcomes of the experiment and next to each outcome write the values of X and Y.

   (b) Write out the distribution of X alone.

   (c) Write out the distribution of Y alone.

   (d) Write out the joint distribution of X and Y.

   (e) Are X and Y independent?

12. A bag contains three blue marbles, two white marbles and one green marble. Two marbles are drawn from the bag (without replacement). Let X be the color of the first marble that is drawn and let Y be the color of the second marble.

   (a) Write out all the outcomes of the experiment and next to each outcome write the values of X and Y.

   (b) Write out the distribution of X alone.

   (c) Write out the distribution of Y alone.

   (d) Write out the joint distribution of X and Y.

   (e) Are X and Y independent?

13. In each case consider the experiment and random variables, X and Y, described. Determine whether or not X and Y have the same distribution and whether or not they are independent.

   (a) Three percent of the screws produced by a certain machine do not satisfy certain tolerances and are considered to be defective. A quality control inspector selects five screws at random. X is equal to 1 if the first screw selected is defective and 0 otherwise. Y is equal to 1 if the second screw selected is defective and 0 otherwise.

   (b) A school club has 15 members; 9 girls and 6 boys. Two students are selected at random to serve as a delegation to the president of the school. Let X be the sex of the first person selected and let Y be the sex of the second person selected.

   (c) The experiment is deal two cards from a deck of cards. X is the suit of the first card. Y is the suit of the second card.

   (d) The experiment is draw a card from a deck of cards. X is the suit of the card drawn (spades, hearts, diamonds, clubs) and Y is the number showing on the card (Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King).

   (e) A machine fills orange juice containers. It is set to fill each container with 12 oz of orange juice but the actual amount in each container varies. A container is selected at random. Let X denote its weight in ounces and let Y denote its weight in grams.

Sheet VII
Linear combinations of variables and random variables

1. A bag contains 3 blue marbles, 2 white marbles, and a green marble. Two marbles are drawn at random (without replacement). Let X denote the number of blue marbles that are drawn.
   (a) Draw a picture of the sample space of this experiment writing the value of X next to each outcome. List all the outcomes in the event $\{X = 1\}$.
   (b) Write out the distribution of X.
   (c) You win \$5 for every blue marble that is drawn. Let Y denote your total winnings in dollars. Notice that $Y = 5X$. Include the value of Y next to each outcome in your picture of the sample space.
   (d) Write out the distribution of Y.

2. The distribution of a variable x is shown below. $y = 3x - 2$.

value of x	fraction
1	.1
2	.3
3	.4
4	.2

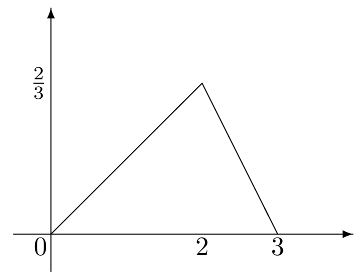
   (a) One member of the population has a y-value equal to 4. What is its x-value?
   (b) What fraction of the population has their y-value equal to 4?
   (c) What fraction of the population has their y-value equal to 10?
   (d) Write down the distribution of y.

3. A random variable, X, has the distribution below. $X = 2Y - 6$.

r	$P[X = r]$
0	.5
1	.3
2	.1
3	.1

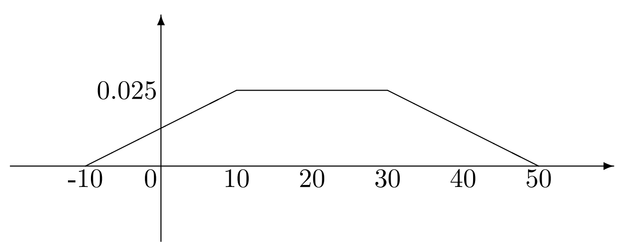
   (a) Find $P[Y = 4]$.
   (b) Write out the distribution of Y.

4. A random variable, X, has the distribution below.



The random variable, Y is given by $Y = 2X$.
   (a) If we get an outcome where $Y = 4$ what is the value of X?
   (b) Consider all the outcomes that lie in the event $\{2 < Y < 4\}$. What are the X-values of these outcomes?
   (c) Find $P[2 < Y < 4]$.
   (d) Find $P[4 < Y < 6]$.

5. Some antipsychotics have the side effect that the person taking them gains or loses weight. Let x be the weight, in pounds, gained by people that take a certain hypothetical antipsychotic. The distribution of x is shown below.



   (a) What proportion of people lose weight?
   (b) What proportion of people gain more than 10 lb?
   (c) Let y be the weight in kilograms gained by people that take this antipsychotic. (1 kg = 2.2 lb) What proportion of people gain more than 10 kg?
   (d) It can be shown that $\mu_x = 20$ and ${\sigma^2}_x = 55$. Find μ_y and ${\sigma^2}_y$.

6. The temperature at noon on 4th April in a hypothetical city in the northern hemisphere, Frigid, has been recorded for the last 100 years. It has a mean of 27^oF and a standard deviation of 8^oF. Find the mean and standard deviation of the temperature measured in ^{o}C. (Note: to get temperatures in ^{o}C from temperatures in ^{o}F subtract 32 and multiply by $\frac{5}{9}$.)

7. Sally has been practising various tricks among which is throwing a peanut in the air and catching it in her mouth. She misses the peanut 10% of the time. She is planning a talent show for her parents. As part of the show she will perform the act of catching a peanut in her mouth six times.

   (a) What is the probability that Sally misses the peanut one third of those six times?

   (b) What is the probability that Sally misses the peanut at least half of those six times?

   (c) What fraction of times out of six is she expected to miss the peanut?

   (d) What is the standard deviation of the fraction of times out of six that she misses the peanut?

8. A machine forms butter into bars. The weights of the bars are normally distributed with a mean of 4.1 oz and a standard deviation of 0.05 oz. Any bar that weighs less than 4 oz is considered defective. An inspector randomly selects twenty bars of butter.

   (a) What is the probability that exactly 10% of the bars inspected weigh less than 4 oz?

   (b) What is the probability that less than 20% of the bars weigh less than 4 oz?

   (c) What percentage of bars in the inspector's sample is expected to weigh less than 4 oz?

   (d) What is the standard deviation of the percentage of bars in such a sample that weigh less than 4 oz?

9. A continuous random variable, X, is uniformly distributed on the interval $[5, 10]$. The random variable, Y, is given by $Y = 3X + 5$.

   (a) Find $P[23 < Y < 26]$ by writing the event $\{23 < Y < 26\}$ in terms of X.

   (b) What is the distribution of Y? Use this to find $P[23 < Y < 26]$.

10. A continuous random variable, X, is uniformly distributed on the interval $[0, 10]$. The random variable, Y, is given by $Y = 2X + 5$. Sketch the distribution of Y.

11. A continuous random variable, X, has a normal distribution with expected value 5 and standard deviation 1. The random variable, Y, is given by $Y = 3X - 5$. Find $P[10 < Y < 13]$.

12. A random variable X is normally distributed with expected value equal to 32 and standard deviation equal to 2. The random variable Y is given by $Y = \frac{X-32}{2}$. What is the distribution of Y? Be as specific as you can.

13. A random variable X is normally distributed with expected value equal to 3000 and standard deviation equal to 100. Use the "Z-method" to find the following probabilities.

   (a) $P[3035 < X < 3230]$

   (b) $P[2940 < X < 3010]$

   (c) $P[X < 2850]$

   (d) $P[X < 3325]$

14. Suppose the heights of adult women in a certain city are approximately normally distributed with a mean of 63.5 inches and a standard deviation of 2.5 inches. Use the "z-method" to find the following quantities.

   (a) The proportion of adult women that are taller than 6 feet.

   (b) The proportion of adult women that are between 5 feet and 6 feet tall.

   (c) The heights of the tallest 10% of the population.

15. Consider the joint distribution of the variables x and y below. Let $w = x + y$.

	y			
x	-1	0	1	
1	.2	.3	.2	.7
2	0	.2	.1	.3
	.2	.5	.3	

   (a) Find the proportion of the population that has a w-value that is equal to 2.

   (b) Find the proportion of the population that has a w-value that is less than 2.

   (c) Write out the distribution of w.

16. The random variables X and Y are independent and have the distributions below. $W = X + Y$.

r	0	5	10
$P[X = r]$	.1	.7	.2

r	-5	5
$P[Y = r]$	.4	.6

   (a) Find $P[W = 5]$

(b) Find $P[W < 10]$

(c) Write out the distribution of W.

17. When a slot machine is played the amount won, W (in cents), has the distribution below.

r	$P[W = r]$
0	.6
25	.3
50	.05
100	.05

You play the machine twice.

(a) What is the probability that you win \$1.50 total?

(b) What is the probability that you win \$0.75 total?

(c) What is the probability that you win \$1.00 total?

(d) Find the probability distribution of your total winnings.

18. A coin is tossed three times. Let X denote the number of heads that are tossed and let Y denote the number of tails that are tossed. Let $W = X + Y$.

(a) Find $E[X]$, $E[Y]$, $\text{Var}[X]$, and $\text{Var}[Y]$. (*Hint:* write out the distributions of X and of Y.)

(b) Write out the distribution of W.

(c) Find $E[W]$. Is it equal to $E[X] + E[Y]$? Should it be?

(d) Find $\text{Var}[W]$. Is it equal to $\text{Var}[X] + \text{Var}[Y]$? Should it be?

19. A coin is tossed twice. Let X take on the value 1 if the first toss is a head and 0 otherwise. Let Y take on the value 1 if the second toss is a head and 0 otherwise. Let $V = X + Y$ be the number of heads that are tossed.

(a) Write out all the outcomes of the experiment with the corresponding values of X, Y, and V.

(b) Write out the joint distribution of X and Y and the distribution of V.

(c) Find the expected values and variances of X, and Y.

(d) Find $E[V]$. Is it equal to $E[X] + E[Y]$?

(e) Find $\text{Var}[V]$. Is it equal to $\text{Var}[X]+\text{Var}[Y]$? Should it be?

20. Suppose that X and Y are independent and that they both have the same distribution shown below.

r	0	1	2
Probability	.2	.5	.3

Notice that $E[X] = 1.1$ and $\text{SD}[X] = .7$. Let $W = X + Y$.

(a) Write out the distribution of W.

(b) Use your answer to (a) to calculate $E[W]$ and $\text{Var}[W]$.

(c) Is it true that $E[W]$ is equal to $E[X]+E[Y]$? Should it be?

(d) Is it true that $\text{Var}[W]$ is equal to $\text{Var}[X] + \text{Var}[Y]$? Should it be?

21. Suppose that X and Y both have the same distribution as in question 20 above. This time, however, X and Y are *not* independent but have the joint distribution below.

	Y			
X	0	1	2	
0	.1	.1	0	.2
1	.1	.3	.1	.5
2	0	.1	.2	.3
	.2	.5	.3	

$W = X + Y$.

(a) Write out the distribution of W.

(b) Use your answer to (a) to find $E[W]$ and $\text{Var}[W]$.

(c) Is it true that $E[W]$ is equal to $E[X]+E[Y]$? Should it be?

(d) Is it true that $\text{Var}[W]$ is equal to $\text{Var}[X] + \text{Var}[Y]$? Should it be?

22. A coin is tossed twice. Let X take on the value 1 if the first toss is a head and 0 otherwise. Let $W = X + X = 2X$. Alex and Bert are having an argument about the value of $\text{Var}[W]$. Alex says that $W = X + X$ so by the addition rule for variances $\text{Var}[W]$ should be equal to $\text{Var}[X] + \text{Var}[X] = 2\text{Var}[X]$. Bert says that $W = 2X$ so by the constant multiple rule for variances $\text{Var}[W]$ should be equal to $4\text{Var}[X]$. Who is right? What is wrong with the other person's argument?

23. Suppose X and Y are independent random variables with $E[X] = 75$, $\text{SD}[X] = 8$, $E[Y] = 42$, and $\text{SD}[Y] = 10$. Let $W = 2X - 30$ and $V = 2X + 6Y$.

(a) Find $E[W]$, $\text{Var}[W]$ and $\text{SD}[W]$.

(b) Find $E[V]$, $\text{Var}[V]$ and $\text{SD}[V]$.

24. Suppose X and Y are independent random variables with $E[X] = 15$, $\text{SD}[X] = 3$, $E[Y] = 25$, and $\text{SD}[Y] = 3$. Let $W = X + Y - 10$ and $V = 5X - 2Y + 1$.

    (a) Find $E[W]$, $\text{Var}[W]$ and $\text{SD}[W]$.

    (b) Find $E[V]$, $\text{Var}[V]$ and $\text{SD}[V]$.

25. A young entrepreneur holds a bake sale outside of his house every Saturday. He sells glasses of fresh squeezed lemonade for \$0.75 and brownies for \$1.25. He has been doing this for about a year and has kept careful records of his sales. The number of glasses of lemonade he sells averages out to 25 and has a standard deviation of 5. The number of brownies he sells averages out to 18 and has a standard deviation of 6. Find his average revenue. Assuming that sales of lemonade and sales of brownies are independent, find the standard deviation in his revenue. Does this assumption seem reasonable?

26. A small bakery owner sells both bakery goods that are not taxable, and material items such as party plates and napkins, that are taxable. The owner's monthly income from sales of bakery goods has a mean of \$5500 and a standard deviation of \$1000. Her monthly income from sales of material items has a mean of \$500 and a standard deviation of \$100. She pays 10% tax on these items. Her monthly expenses amount to a fixed \$2000 for rent and supplies. Find the mean and standard deviation of her net income after tax and after paying her expenses. You may assume that her income from sales of bakery goods and her income from sales of material goods are independent. Is this assumption, in fact, reasonable?

27. A machine shapes butter into bars. The weights of the bars it forms are normally distributed with a mean of 4.1 oz and a standard deviation of 0.05 oz. Two bars are selected at random.

    (a) What is the distribution of the combined weight of the two bars? Be as specific as you can.

    (b) What is the probability that the combined weight of the two bars is at least 8 oz?

28. A machine regulates how much dye is dispensed when mixing paint. When the machine is set to make cans of lightning blue colored paint the amount of dye it dispenses is normally distributed with a mean of 5.5 ml and a standard deviation of 0.3 ml. If a can has more than 6.0 ml or less than 5.0 ml of dye then the eye can tell that it is not the right shade and the can is considered unacceptable.

    (a) What proportion of all cans of lightning blue paint produced by this machine are unacceptable?

    (b) You pour together the contents of two cans of lightning blue paint, mix them, and then put them back again in their cans. The two cans are then either both acceptable or both unacceptable. Find the probability that they are both unacceptable.

Sheet VIII
Focus on sums and averages of iid variables

1. Goodies Inc. produces bags of assorted candies. Each bag contains somewhere between 2 and 8 chocolate candies. Let x denote the number of chocolate candies in the bags produced by Goodies Inc. The distribution of x is given in the table below.

x	fraction
2	0.15
3	0.25
4	0.35
5	0.1
6	0.05
7	0.05
8	0.05

Adrian buys 10 bags of candies. Let ΣX be the total number of chocolate candies that he gets.

(a) Find $E[\Sigma X]$, $\mathrm{Var}[\Sigma X]$ and $\mathrm{SD}[\Sigma X]$.

(b) Recall Chebychev's theorem. Would it be surprising if Adrian got 60 or more chocolate candies total?

2. The weight of a certain type of nail has an average of 3.22 grams and a standard deviation of 0.14 grams. The nails come in boxes of two thousand. What is the average weight and standard deviation of the weight of these boxes?

3. Let W denote the amount won (in cents) when a slot machine is played. The distribution of W is given in the table below.

r	$P[W = r]$
0	.6
25	.3
50	.05
100	.05

You play the machine 25 times. Let $\bar{X}$ denote your average winning per play.

(a) What is the probability that you win 50 cents or more on at least one of your plays?

(b) Find $E[\bar{X}]$, $\mathrm{Var}[\bar{X}]$, and $\mathrm{SD}[\bar{X}]$.

(c) Recall Chebychev's theorem. Would you be surprised if your average winnings came to 50 cents or more?

4. Scores on a national test are normally distributed with a mean equal to 500 and a standard deviation equal to 100. A random sample of 1000 students taking the test is selected. Let $\bar{X}$ be their average score.

(a) What is the probability that at least one of the students scores 600 or more?

(b) Find $E[\bar{X}]$, $\mathrm{Var}[\bar{X}]$ and $\mathrm{SD}[\bar{X}]$.

(c) Would you be surprised if the average score of the 1000 selected students was as large as 600?

5. A machine fills orange juice containers. The amount of juice in each container is normally distributed with a mean of 12.1 oz and a standard deviation of 0.05 oz. You buy a container of orange juice every day at school for lunch. Let X denote the total amount of orange juice that you drink in one week (Monday to Friday inclusive). What is the probability that you drink more than 60 oz of orange juice in one week?

6. Suppose the lifetimes of lightbulbs manufactured by Electrifying Delight Inc. is normally distributed with mean 600 hours and standard deviation 100 hours. You buy a package of a dozen lightbulbs which you plan to use one after the other in your desk lamp.

(a) What is the expected length of time that you will be able to study at night at your desk?

(b) What is the probability that you will be able to study at night at your desk for more than 7000 hours?

7. *(Adapted from a problem in "Probability and Statistical Inference" by Hogg and Tanis)* Assume the weights of fat-free Fig Newton cookies are normally distributed with mean 14.22 grams and standard deviation 0.29 grams. These cookies are sold in packages that have a label weight of 340 grams. How many cookies should be put into a package in order to be 95% certain that the total weight of the cookies in the package is at least 340 grams?

8. *(Adapted from a problem in "Probability and Statistical Inference" by Hogg and Tanis)* A student played the board game "Pinocchio" with his friend's nephews. In this game, to try to escape the great white whale, the player rolls a fair six-sided die 30 times adding together these 30 outcomes. To escape the whale the player must

receive a score of at least 110. For a score of less than 110 the player is "eaten by the whale." Find, approximately, the probability of escaping the whale.

9. Suppose the average distance spanned from fingertip to fingertip when a person holds out their hands from the side of their bodies is 64 inches with a standard deviation of 4 inches. Twelve hundred protestors assemble at a nuclear power plant and seek to form a human chain around the plant - a distance of 77,000 inches. What is the probability that they have enough people to be able to reach?

10. Bags of mixed candies produced by Goodies Inc. have 5, 6, or 7 chocolate candies in them. The distribution of x, the number of chocolate candies in each bag, is given below.

x	5	6	7
fraction	.3	.4	.3

   (a) You buy two bags of candies. Let $\Sigma_2 X$ denote the total number of chocolate candies that you get. Find $P[\Sigma_2 X > 12]$. Is your answer exact or is it approximate?

   (b) You buy two hundred bags for a party. Let $\Sigma_{200} X$ denote the total number of chocolate candies that you get. Find $P[\Sigma_{200} X > 1200]$. Is your answer exact or is it approximate?

11. A variable x of an infinite population has the distribution shown below.

x	0	1	2
fraction	.6	.3	.1

   (a) Two members of the population are selected at random. Let $\bar{X}_2$ be the average of their x-values. Find $P[\bar{X}_2 < 0.5]$. Is your answer exact or approximate?

   (b) One hundred members of the population are selected at random. Let $\bar{X}_{100}$ be the average of their x-values. Find $P[\bar{X}_{100} < 0.5]$. Is your answer exact or approximate?

12. (a) You toss a coin 100 times. What is the probability that more than 60% of those are heads?

   (b) You toss a coin 1000 times. What is the probability that more than 60% of those are heads?

13. A manufacturer of computer chips sends the chips to the wholesaler in packages of 300. Ten percent of the chips have defects. What is the probability that more than fifteen percent of the chips in a package are defective?

Sheet IX
Using random variables to find probabilities

1. A mid-size company has 50 employees, 15 of which are men. The company is having financial difficulties and plans to downsize. They need to let 8 people go and plan to choose them by drawing lots. What is the probability that at least 2 of the employees that they let go are men?

2. Based on long experience an airline knows that 6% of people that make a reservation on a certain flight from Los Angeles to New York do not show up for the flight. The airline overbooks the flight by selling reservations to 267 people when the airplane only has 255 seats. What is the probability that the number of passengers that show up with reservations exceeds the number of seats available?

3. A physical fitness association is including the mile run in its secondary school fitness test for boys. Running times for boys in this age group are normally distributed with $\mu = 450$ seconds and $\sigma = 40$ seconds.
   (a) What percentage of boys in this age group can run one mile in less than 7 min (420 seconds)?
   (b) If the asociation wants to designate the fastest 10% of boys as "excellent," what time should the association set for this criterion?
   (c) If 25 boys are selected at random from this age group what is the probability their average run time would be 428 seconds or less?
   (d) King Henry is a small private school with 25 boys in its graduating class. Its graduating class did the one mile run in an average time of 428 seconds. Considering your answer to (c) do you think that this quick average run time is just due to chance variation or might it be due to the King Henry's excellent athletics program?

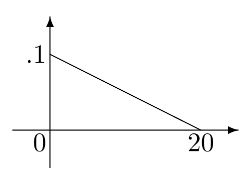
4. *(Adapted from a problem in "Understandable Statistics" by Brase and Brase)* Coal is carried from a mine in West Virginia to a power plant in New York in hopper cars on a long train. The automatic hopper car loader can be set to put any weight of coal in the hopper cars. The actual weights it will then load into the cars will be normally distributed with a mean, μ, equal to the setting, and a standard deviation, σ, of 1.5 tons.
   (a) The power plant in New York has ordered 1500 tons of coal to be shipped from West Virginia. The mine will ship the coal in 20 hopper cars. What do they need to set μ to be in order to be 99% certain that the complete load comes to at least 1500 tons?
   (b) Suppose they end up setting μ to be 74 tons. A hopper car is considered to be dangerously overfull if it weighs more than 77 tons. What is the probability that at least one of the hopper cars is dangerously overfull?

5. When a particular slot machine is played the winnings have the probability distribution below.

winnings	probability
0	.6
10	.3
20	.1

   (a) Find the mean and standard deviation of the winnings from a single play of the machine.
   (b) A person plays the machine twice. Find the probability distribution of their winnings.
   (c) The machine is played 500 times each day. How much money must be put into the machine each morning to be 99% certain that it will have enough money to provide winnings for the whole day?

6. Good Health Inc. manufactures a medical device whose lifetime (in years) has the distribution shown below.



   (a) Mary buys one of these devices. What is the probability that it lasts longer than ten years?
   (b) A clinic buys twenty of these devices. What is the probability that at least four of them last longer than ten years?
   (c) A hospital buys two hundred of these devices. What is the probability that their average lifetime is at least seven years?

7. A shipment of fifty garlic presses contains six defective items. A quality control inspector that receives the shipment selects ten at random from the shipment and tests them.

(a) What is the expected number of defective presses in the sample?

(b) What is the probability that there is more than one defective press in the sample?

8. A certain type of seed has an 80% chance of germinating.

   (a) In a package of 100 seeds what is the probability that at least 70 of them will germinate?

   (b) A particular package of 100 seeds actually contains 85 seeds that will germinate. You take 10 seeds and plant them in a small planter. What is the probability that at least 7 of the ones you planted germinate?

9. The manufacturer of a sports drink for athletes ships the drink to the bottler in barrels. The amount of drink in each barrel is normally distributed with a mean of 50.1 gallons and a standard deviation of 0.5 gallons.

   (a) If 100 barrels are shipped to the bottler what is the probability that the bottler receives at least 5000 gallons of drink?

   (b) The machine that fills the barrels is adjustable. The mean amount of drink that goes into each barrel can be set to any value but, whatever the setting, the standard deviation is 0.5 gallons. Usually the manufacturer ships 20 barrels at a time to the bottler. What should the manufacturer set the mean fill volume to be if they want to be 99% certain that the bottler receives at least 1000 gallons of drink in each shipment?

10. Goodies Inc. produces bags of candies by randomly selecting fifty candies from a pool of candies whose colors have the distribution below.

Color	proportion
red	.3
orange	.05
yellow	.25
green	.25
blue	.15

   (a) You buy a bag of candies. What is the probability that between 12 and 15 of them (inclusive) are red?

   (b) Your friend buys two bags of candies. What is the probability that your friend doesn't get any orange candies?

   (c) You buy a bag of candies. It turns out to consist of 18 red, 3 orange, 11 yellow, 12 green, and 6 blue candies. You randomly select 5 candies to give to your friend. What is the probability that your friend gets at least 3 red candies?

11. Five percent of all telephones produced by Easy Talk Inc. are defective. Easy Talk ships the phones to the retailer in boxes of 20. The retailer randomly selects five telephones from the shipment and inspects them. If more than one of the telephones inspected is defective the retailer rejects the shipment.

   (a) What proportion of all shipments end up being rejected?

   (b) Unknown to the retailer (of course) a particular shipment actually contains 2 defective phones. What is the probability that, upon inspection, the retailer will reject this shipment?

   (c) A shipment is received and inspected. All five of the telephones that are selected for inspection are in good working order. What is the probability that all the telephones in the shipment are in good working order?

Sheet X
Properties of estimators

1. The machine used to manufacture pistons for a certain car produces pistons whose diameters are normally distributed with $\sigma = 0.01$ inches. The mean, μ, can be adjusted but the gauge is broken so the only way to ascertain the value of μ is to estimate it from a sample. It is planned to take a random sample of 5 pistons and use the data from the sample to estimate the value of μ.

   (a) What is the probability that the average diameter of the 5 pistons sampled, $\bar{X}$, will be more than 0.005 inches away from μ?

   (b) What is the probability that the median diameter of the 5 pistons sampled, M, will be more than 0.005 inches away from μ? *(Hint: the median is more than 0.005 inches away from μ if and only if three or more of the pistons sampled have diameters that are less than $\mu - 0.005$ or three or more have diameters that are greater than $\mu + 0.005$.)*

   (c) In light of your answers to (a) and (b) which statistic, $\bar{X}$ or M, would appear to be a better estimator of μ?

2. A variable, x, in a population is known to be uniformly distributed over the interval $[0, T]$ but the value of T is not known. To estimate the value of T a random sample of 100 individuals is taken.

   (a) Frederic reasons that the mean of x is $\frac{0+T}{2} = \frac{1}{2}T$ so a good way to estimate T would be to use twice the sample mean, $2\bar{X}$. Is this an unbiased estimator of T? Find the probability that the value Frederic obtains for T lies within 10% of the true value of T (i.e. find the probability that $2\bar{X}$ lies within $T - (.1)T$ and $T + (.1)T$).

   (b) Lisette reasons that since T is the largest value that x can be, a good way to estimate it would be to use the largest value of x that appears in the sample, $X_{\max}$. Is $X_{\max}$ an unbiased estimator of T? Find the probability that the value Lisette obtains for T lies within 10% of the true value of T.

   (c) Which estimator is better? What information would you need in addition to your calculations to answer this question fully? Can you think of a way in which you might improve one or the other of these estimators?

3. Recall that the median of a (continuous) distribution is a value, m, where half of the population has a value greater than m and half the population has a value that is less than m. We take a random sample of size 9 and use the median, M, of the sample as an estimate for m.

   (a) Find $P[M > m]$. *(Hint : $M > m$ if and only if 5 or more of the values in the sample are greater than m).*

   (b) Find $P[M < m]$.

   (c) Does it necessarily follow that M is an unbiased estimator for m?

Sheet XI
Hypothesis Testing

Fundamental idea behind hypothesis tests

1. Suppose one mile run times of secondary school boys are normally distributed with $\mu = 450$ seconds and $\sigma = 40$ seconds.
   (a) If 25 secondary school boys are selected at random what is the probability that their average run time would be 428 seconds or less?
   (b) King Henry is a small private school with 25 boys in its graduating class. Its graduating class did the one mile run in an average time of 428 seconds. Considering your answer to (a) do you think that this quick average run time is due to chance variation or is it more likely to be due to the fact that the students at King Henry tend to do the one mile run faster than students nationwide (perhaps because of King Henry's excellent athletics program or its ability to be selective in the students it admits)?

2. Suppose it is known that, if left untreated, 20% of patients diagnosed with a particular disorder become paralysed within 12 months.
   (a) If 100 patients diagnosed with this disorder are selected at random, find the probability that 10 or fewer become paralysed within 12 months.
   (b) 100 patients diagnosed with this disorder are given vitamin supplements. They are followed for 12 months and only 10 of them become paralysed. Do you think that this low rate of paralysis is due to chance variation or is it more likely that the patients that received the vitamin supplements have a lesser tendency towards paralysis than other patients (perhaps because the vitamin supplements reduce the tendency towards paralysis or because of some other factor such as the placebo effect or the fact that the patients were self selecting)?

Basic concepts in hypothesis testing

3. *(Written by K. Alexander)* Allnighter Test Prep Services claims in its advertising that its SAT prep course will raise a student's total score by more than 80 points on average.

   (a) A government agency tells Allnighter that Allnighter must establish its claim with a high degree of certainty if they want to continue making the claim. As part of an hypothesis test, Allnighter will choose some students who each took the test twice, before and after the prep course, and determine the score increase for each student. Formulate the null and alternative hypotheses.

   (b) Allnighter's competitor, Cramwell Test Prep Services, wonders whether Allnighter's claim is false, and will do its own hypothesis test to see if there is convincing evidence that the average increase is less than 80. Formulate the null and alternative hypotheses.

4. *(Written by K. Alexander)* An auto repair shop proclaims in its advertising, "Our average repair bill is well under \$100!"

   (a) If the government can find strong evidence that the advertising claim is false, the repair shop will be fined. A hypothesis test is to be done using a randomly chosen sample of repair bills. Formulate the null and alternative hypotheses.

   (b) Instead of (a), suppose a radio station refuses to air the ads without very strong evidence that the advertising claim is true. Again, a hypothesis test is to be done using a randomly chosen sample of repair bills. Formulate the null and alternative hypotheses.

5. Consider the hypothesis test conducted in 1(b) above.
   (a) What are the null and alternative hypotheses?
   (b) What is the p-value?

6. Consider the test conducted in question 2(b) above.
   (a) What are the null and alternative hypotheses?
   (b) What is the p-value?

7. A test of H_0: $\mu \leq 5.3$ versus H_a: $\mu > 5.3$ is conducted. The population variable is normally distributed with $\sigma = .1$. Find the p-value if the data yield
   (a) $n = 100$; $\bar{X} = 5.314$
   (b) $n = 15$; $\bar{X} = 5.371$
   (c) $n = 40$; $\bar{X} = 5.291$

8. A test is conducted of H_0: $\mu = 30$ versus H_a: $\mu \neq 30$. The population variable has $\sigma = 2.5$. Find the p-value if the data yield

   (a) $n = 1200$; $\bar{X} = 29.9$

   (b) $n = 340$; $\bar{X} = 29.9$

   (c) $n = 211$; $\bar{X} = 30.7$

9. A random sample of 100 observations from a population with standard deviation 60 yielded a sample mean of 110.

   (a) Calculate the p-value if the test is of H_0: $\mu \leq 100$ versus H_a: $\mu > 100$. Interpret the results of the test if a 5% significance level is used.

   (b) Calculate the p-value if the test is of H_0: $\mu = 100$ versus H_a: $\mu \neq 100$. Interpret the results of the test if a 5% significance level is used.

10. For which of the given p-values would the null hypothesis be rejected when performing a test at the 5% significance level?

    (a) .001 (b) .021 (c) .047 (d) .078 (e) .148

11. For which of the given p-values would the null hypothesis be rejected when performing a test at the 1% significance level?

    (a) .001 (b) .021 (c) .047 (d) .078 (e) .148

12. Suppose we conduct an hypothesis test and get a p-value of 0.023. At which of the following significance levels can we reject the null hypothesis?

    (a) 10% (b) 5% (c) 2% (d) 1% (e) 0.5%

13. Suppose we conduct an hypothesis test and get a p-value of 0.35. At which of the following significance levels can we reject the null hypothesis?

    (a) 10% (b) 8% (c) 5% (d) 2% (e) 1%

14. An hypothesis test is conducted. It is decided to reject the null hypothesis if the p-value is less than 0.05. What is the significance level of the test?

15. A statistical test of $H_0 : \mu \leq 200$ versus $H_a : \mu > 200$ is conducted. The variable in the population is normally distributed with a standard deviation equal to 40. It is decided to reject H_0 if the sample mean of a random sample of 10 items is more than 215.

    (a) What is the test statistic that is being used in this test?

    (b) What is the rejection region?

    (c) What is the significance level?

16. A statistical test is conducted of $H_0 : p = .3$ versus $H_a : p \neq .3$. A random sample of 200 items is taken and it is decided to reject H_0 if $\hat{p}$ is either greater than .35 or less than .25.

    (a) What is the test statistic that is being used in this test?

    (b) What is the rejection region?

    (c) What is the significance level?

17. Suppose the number of miles driven annually by people in the United States is normally distributed with a mean of 10.4 thousand and a standard deviation of 3.1 thousand. In order to qualify for extra subsidies the Board of Supervisors needs to convince the federal government that people in LA tend to drive more than other people in the United States. They plan to take a random sample of 1000 LA residents and find the average distance these people drove last year, $\bar{X}$.

    (a) Formulate the null and alternative hypotheses.

    (b) The federal government will be convinced if the evidence is significant at the 5% level. What values of $\bar{X}$ will produce evidence that is convincing at this significance level?

18. Deborah thinks she has ESP. Her sister, Elizabeth, is skeptical. They decide to do a test whereby Elizabeth holds up between 1 and 5 fingers on her hand behind a screen and Deborah has to "sense" how many she is holding up. They will do this 100 times.

    (a) Formulate the null and alternative hypotheses.

    (b) Elizabeth is skeptical but willing to believe. She will find the evidence convincing if it is significant at the 10% level. How many times out of 100 will Deborah need to correctly "sense" the number of fingers that Elizabeth is holding up in order to convince her sister that she has ESP?

19. When the machine that makes pistons at Smooth and Slick Inc. is properly adjusted the diameters of the pistons are normally distributed with a mean of 5.3 inches and a standard deviation of 0.01 inches. Every day the machine is tested for proper adjustment by taking a random sample of

15 pistons and measuring their diameters. If the average diameter of the 15 pistons is too small or too large the machine is shut down and inspected.

(a) Formulate the null and alternative hypotheses of the daily tests.

(b) On Monday, the average diameter of the 15 pistons was too large and the machine was shut down for inspection. Upon inspection, however, the machine was found to be properly adjusted. What kind of error was made - a type I error or a type II error?

(c) On Tuesday, the average diameter of the 15 pistons was neither too large nor too small and the machine was not shut down. However, throughout the day the machine was consistently making pistons whose diameters were less than 5.3 inches and it was found that the machine actually was out of adjustment. What kind of error was made this time - a type I error or a type II error?

Tests on the value of μ when σ is known

20. A family run business makes wine. They leave the wine to age in 50 gallon oak barrels. The actual amount of wine in the barrels is normally distributed with a mean of 48 gallons and a standard deviation of 2 gallons. The family suspects that local youths entered the cellar one night and siphoned off some wine from each barrel. To test their theory they carefully measure the amount of wine they get from the next 5 barrels.

(a) Formulate the null and alternative hypotheses.

(b) The amount of wine in the next five barrels has an average of 46.4 gallons. Calculate the p-value.

(c) At which of the following significance levels would the family conclude that they are correct about the local youths?

10% 8% 5% 2% 1%

21. A test of $H_0 : \mu \geq 500$ versus $H_a : \mu < 500$ is conducted. The population variable is normally distributed with $\sigma = 10$. Find the rejection region in each case.

(a) Sample size is 20; significance level is 5%; test statistic is $\bar{X}$.

(b) Sample size is 45; significance level is 10%; test statistic is $\bar{X}$.

(c) Significance level is 1%; test statistic is $Z = \frac{\bar{X}-\mu}{(\sigma/\sqrt{n})} = \frac{\bar{X}-500}{(10/\sqrt{n})}$.

(d) Significance level is 5%; test statistic is $Z = \frac{\bar{X}-\mu}{(\sigma/\sqrt{n})} = \frac{\bar{X}-500}{(10/\sqrt{n})}$.

22. A test of $H_0 : \mu = 32$ versus $H_a : \mu \neq 32$ is conducted. The population variable is normally distributed with $\sigma = 4$. Find the rejection region in each case.

(a) Sample size is 18; significance level is 10%; test statistic is $\bar{X}$.

(b) Sample size is 55; significance level is 5%; test statistic is $\bar{X}$.

(c) Significance level is 8%; test statistic is $Z = \frac{\bar{X}-\mu}{(\sigma/\sqrt{n})} = \frac{\bar{X}-32}{(4/\sqrt{n})}$.

(d) Sample size is 25; significance level is 1%; test statistic is $Z = \frac{\bar{X}-\mu}{(\sigma/\sqrt{n})} = \frac{\bar{X}-32}{(4/\sqrt{n})}$.

23. A machine forms butter into bars. Under normal operations the amount of butter in each bar is normally distributed with a mean of 4.1 oz and a standard deviation of 0.05 oz. Sometimes, however, the machine gets out of adjustment. In this case the mean amount of butter in each bar may increase or decrease. The standard deviation remains the same. Every day quality control tests the machine to see if it is properly adjusted. This is done by taking a sample of 10 bars and weighing the amount of butter in each bar.

(a) Formulate the null and alternative hypotheses associated with the daily tests.

(b) Choose an appropriate test statistic and find the rejection region if the test is carried out at a 5% significance level.

(c) One day the ten bars tested had an average weight of 4.07 oz. What would quality control conclude that day?

Tests on the value of μ when σ is not known

24. Mary and Nathan are independently conducting statistical tests of $H_0 : \mu = 15$ versus $H_a : \mu \neq 15$. The population variable is normally distributed and its standard deviation is not known. They each take a random sample of size 25.

(a) Mary gets $\bar{X} = 15.2$ and $s = 0.5$. Calculate the value of $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$.

(b) Nathan gets $\bar{X} = 15.1$ and $s = 0.2$. Calculate the value of $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$.

(c) Who has more compelling evidence for the alternative hypothesis?

25. *(Adapted from a problem written by L. Heyer)* A researcher conducts a test of H_0: $\mu \leq 17$ versus H_a: $\mu > 17$. The variable in the population is normally distributed and its standard deviation is not known. Estimate the p-value in each case where $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$ is calculated from the data and n is the sample size used.

    (a) $T = 1.84$; $n = 15$

    (b) $T = 3.74$; $n = 26$

    (c) $T = 2.42$; $n = 14$

    (d) $T = 1.32$; $n = 150$

26. *(Adapted from a problem written by L. Heyer)* A researcher conducts a test of H_0: $\mu = 30$ versus H_a: $\mu \neq 30$. The variable in the population is normally distributed and its standard deviation is not known. Estimate the p-value in each case where $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$ is calculated from the data and n is the sample size used.

    (a) $T = 2.3$; $n = 15$

    (b) $T = -3.0$; $n = 26$

    (c) $T = 3.0$; $n = 14$

    (d) $T = -1.32$; $n = 150$

27. A car manufacturer believes that its newly developed electric and gas powered vehicle gets an average of more than 80 miles per gallon on the highway. However, to advertise this they need to produce evidence that it is the case. They test drive 25 vehicles on the highway and measure the gas mileage of each vehicle.

    (a) Formulate the null and alternative hypotheses.

    (b) The gas mileages of the 25 test drives had an average of 81.7 miles per gallon and a sample standard deviation of 5.86 miles per gallon. Estimate the p-value. You may assume that the gas mileages obtained by these vehicles are normally distributed.

    (c) At which of the following significance levels can the manufacturer claim that there is compelling evidence that their vehicle gets an average of more than 80 miles per gallon on the highway?

    10% 5% 2% 1% 0.5%

28. *(Adapted from a problem written by L. Heyer)* A manufacturer has developed a new fishing line which he claims has a mean breaking strength of 15 kg or more. The government suspects fraud and conducts a test to see if the evidence points to the contrary. They take a random sample of 25 lines and measure the breaking strength of each of them.

    (a) Formulate the null and alternative hypotheses.

    (b) It is found that the sample has a mean breaking strength of 14.8 kg and a standard deviation of 0.5 kg. Estimate the p-value for this data assuming that the breaking strengths are normally distributed.

    (c) At which of the following significance levels can the manufacturer's claim be rejected?

    10% 8% 5% 2% 1%

29. A neurologist is investigating the effect of a drug on the response time of rats. The response time for rats not injected with the drug is normally distributed with a mean of 1.2 seconds. She injects 150 rats.

    (a) Formulate the null and alternative hypotheses.

    (b) The response time of the drug injected rats has a mean of 1.11 seconds and a standard deviation of .5 seconds. Estimate the p-value.

    (c) At which of the following significance levels would the neurologist conclude that the drug affects the response time of rats?

    10% 5% 2% 1% 0.5%

30. *(Adapted from a problem written by L. Heyer)* The owner of a gas station that performs headlight inspections for the state is trying to decide whether or not to continue the service. She figures that to make the operation profitable, the station must average in excess of 15 inspections per week. Unless there is evidence that this is the case, the service will be discontinued. Data for the past year (52 weeks) yield a sample average of 16.7 inspections per week with a sample standard deviation of 4.5. Is this strong enough evidence to cause the owner to retain the inspection service? Test the relevant hypotheses using the p-value method. Indicate which of the following significance levels would result in rejection of the null hypothesis.

    10% 8% 5% 2% 1%

31. A researcher conducts a test of $H_0 : \mu \leq 23.2$ versus $H_a : \mu > 23.2$. The variable in the population is normally distributed and its standard deviation is not known. The test statistic used is $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$. Find the rejection region in each case where n is the sample size used and α is the significance level at which the test is carried out.

   (a) $n = 10$; $\alpha = 5\%$

   (b) $n = 30$; $\alpha = 1\%$

   (c) $n = 18$; $\alpha = 10\%$

32. A researcher conducts a test of $H_0 : \mu = 350$ versus $H_a : \mu \neq 350$. The variable in the population is normally distributed and its standard deviation is not known. The test statistic used is $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$. Find the rejection region in each case where n is the sample size used and α is the significance level at which the test is carried out.

   (a) $n = 12$; $\alpha = 5\%$

   (b) $n = 61$; $\alpha = 1\%$

   (c) $n = 20$; $\alpha = 10\%$

Tests on the value of p

33. Albert and Bertha independently conduct statistical tests of $H_0 : p \geq .45$ versus $H_a : p < .45$. Albert uses a random sample 100 items and will reject H_0 if $\hat{p} < .4$. Bertha uses a random sample of 1000 items and will reject H_0 if $\hat{p} < .4$. Who has the lower significance level?

34. *(Adapted from a problem written by L. Heyer)* Prior to the elections, a random poll of 147 people showed that 46 favored George Bush, 50 favored Clinton, 22 favored Ross Perot, and 29 were undecided. We want to test the hypothesis that Perot's support was less than or equal to 10%. Find the p-value for the data in the test. Should the hypothesis be rejected at a 5% significance level?

35. *(Adapted from a problem written by L. Heyer)* A civic researcher is investigating the distribution of household sizes in her community. She suspects that at most 45% of all households have two or fewer members and unless she has evidence to the contrary she will work with that belief. A survey of 25 households shows that 14 of them have two or fewer members.

   (a) What is the p-value of this test?

   (b) Is the researcher's null hypothesis rejected at level 0.05?

36. A test is conducted of $H_0 : p \leq .6$ versus $H_a : p > .6$. Find the rejection region in each case.

   (a) Sample size is 100; significance level is 5%; test statistic used is $\hat{p}$.

   (b) Sample size is 1500; significance level is 1%; test statistic used is $\hat{p}$.

   (c) Significance level is 8%; test statistic used is $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$.

   (d) Significance level is 10%; test statistic used is $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$.

37. A test is conducted of $H_0 : p = .9$ versus $H_a : p \neq .9$. Find the rejection region in each case.

   (a) Sample size is 120; significance level is 15%; test statistic used is $\hat{p}$.

   (b) Sample size is 1600; significance level is 1%; test statistic used is $\hat{p}$.

   (c) Significance level is 5%; test statistic used is $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$.

   (d) Significance level is 10%; test statistic used is $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$.

Focus on p-values

38. Elizabeth and Fred each independently conduct a test of $H_0 : \mu = 30$ versus $H_a : \mu \neq 30$. Elizabeth gets a p-value of 0.093 and Fred gets a p-value of 0.032. Who has more compelling evidence for H_a?

39. Aaron and Bill each conduct a statistical test of $H_0 : \mu \leq 23$ versus $H_a : \mu > 23$ in a population that has a normally distributed variable with standard deviation equal to 2.

   (a) Aaron uses a sample of 25 individuals and gets $\bar{X} = 24$. What is Aaron's p-value?

   (b) Bill uses a sample of 36 individuals and also gets $\bar{X} = 24$. What is Bill's p-value?

   (c) Who has more compelling evidence for the alternative hypothesis?

40. *(Adapted from a problem written by K. Alexander)* Researcher Smith performs an hypothesis test of H_0: $\mu \leq 20$ versus H_a: $\mu > 20$. The population variable is normally distributed with a standard deviation of 10. She uses a sample of size 40 and gets $\bar{X} = 22.5$.

(a) Calculate the p-value.

(b) If Smith finds a mistake in her data and actually $\bar{X} = 22.7$, would this make the p-value larger or smaller than what you calculated in (a)?

(c) If the sample size were actually 60 instead of 40, but otherwise the results were the same, would the p-value be larger or smaller than what you calculated in (a)?

(d) Suppose the null hypothesis is, in fact, true. If another researcher, Jones, redoes the test with 40 new observations, what is the probability he gets a p-value smaller than the one you calculated in (a)?

(e) Jones gets a p-value of 0.015. Who has stronger evidence for the alternative hypothesis, Smith or Jones?

41. *(Written by L. Heyer)* A hypothesis test is performed of H_0: $\mu \geq 10$ versus H_a: $\mu < 10$ in a population where the variable is normally distributed and σ is not known. A sample size of 25 is used and the value of $T = \frac{\bar{X}-\mu}{(s/\sqrt{n})}$ is calculated from the data to be -2.1. If the null hypothesis is in fact true, and another sample this size were taken, what is the probability that the value of T in this sample would be smaller than -2.1?

42. *(Written by K. Alexander)* When the machine at SugarBuzz Children's Cereals Corp. is properly adjusted, it puts a mean of 18.2 ounces of cereal in each box. When it gets out of adjustment, the mean may be higher or lower. Assume the actual quantity of cereal in a box is normally distributed. Once a day, the hypothesis that the machine is properly adjusted is tested, using a sample of 15 boxes of cereal. On Monday the sample mean is 18.04 ounces, and on Tuesday it is 18.27 ounces. The sample variance is the same on both days. On which day is the p-value larger?

43. *(Written by K. Alexander)* Judy at Acme Product Testing has the job of testing the hypothesis that the mean calcium content of Max-Vita vitamin tablets is at least 500 mg (that is, the null hypothesis is H_0: $\mu \geq 500$). She does the test on Monday using a sample of 30 tablets, and finds a sample mean of 479.3. She calculates that the result has a p-value of .043.

(a) On Tuesday, Judy repeats the test, this time with a sample of 60 tablets. Remarkably, she finds the same sample mean and sample variance that she found on Monday. Is Tuesday's p-value

(i) .043 again, (ii) more than .043, or (iii) less than .043?

(b) On Wednesday, Judy repeats Monday's test, using 30 tablets. She gets a sample mean that is larger than Monday's 479.3, but has the same sample standard deviation. Is Wednesday's p-value

(i) .043 again, (ii) more than .043, or (iii) less than .043?

(c) On Thursday, Judy repeats Monday's test on another brand, Slapdash Vitamin Tablets. She uses 30 tablets and gets the same sample mean of 479.3, but a much larger sample variance (probably due to the weak quality control at Slapdash). Is Thursday's p-value

(i) .043 again, (ii) more than .043, or (iii) less than .043?

44. *(Adapted from a problem written by K. Alexander)* A survey several years ago showed that at least 60% of USC students are Republicans. Bill is not so sure that this is still true, so he sets up a hypothesis test using a random sample of USC students.

(a) Formulate the null and alternative hypotheses.

(b) Bill picks a sample size for the test and finds that 48% of the students he surveys are Republicans. He calculates the p-value to be .0036. Later, Carla repeats Bill's test with another sample of the same size. Assuming the null hypothesis is true and 60% of USC students are Republicans, what is the probability that Carla's sample has 48% or fewer Republicans?

45. *(Written by K. Alexander)* Dick and Jane each have a nickel, and each tests the hypothesis that his/her coin is fair (that is, that each time it is tossed there is a probability of 1/2 that it will come up heads). Dick tosses his coin 100 times and gets 62% heads. Jane tosses her coin more than 1000 times and gets 62% heads also.

(a) Who has the smaller p-value, or are the p-values the same?

(b) Dick is able to reject the null hypothesis (that $p = 1/2$) at the 5% level. Can you tell from this whether or not Jane will be able to reject the null hypothesis at the same 5% level?

Type I errors

46. Emmanuel has a coin that he suspects might be weighted so that its probability of landing heads is not 0.5. He decides to test the coin by flipping it 100 times and recording the number of heads.

    (a) Formulate the null and alternative hypotheses.

    (b) He conducts the test at a 10% significance level. If the coin, in fact, is not weighted what is the probability that his test will lead him to incorrectly conclude that it is weighted?

47. A company has developed a new drug to treat headaches and it is seeking approval from the FDA to market it. There is concern that the drug may cause seizures and death in a significant proportion of people that use it and before approving the drug the FDA requires that a test be done to demonstrate that this is not the case. The null and alternative hypotheses are:

    H_0: The drug causes seizures and death.

    H_a: The drug does not cause seizures and death.

    Weighing the benefits of curing a headache against the seriousness of approving the drug when it may in fact cause seizures and death should the FDA conduct the test at a 5% or a 1% significance level?

48. A certain disease occurs in 2% of the population. It is suspected that higher levels of chromium in the water may increase the rate of occurrence of this disease. One hundred different locations are identified that have higher levels of chromium in the water. At each of these locations a test is done of $H_0 : p = 0.02$ versus $H_a : p > 0.02$ where p represents the proportion of the population in that location that acquire the disease. The tests are performed at the 5% level. Suppose that, in actual fact, the higher levels of chromium in the water have no effect on the rate of occurrence of this disease and that $p = 0.02$ in each of these locations. In how many of the locations would we expect the tests to reveal incorrectly that the alternative hypothesis, $p > 0.02$, is true? Explain.

49. A drug company is making its case to the FDA to approve a new drug it has developed to reduce the tendency for heart attacks among people with high blood pressure. It cites 5 studies that were conducted by independent researchers in each of which the evidence was significant at the 5% level that the drug helps to reduce heart attacks.

    (a) Does the fact that there are 5 studies that are all significant at the 5% level make the evidence more compelling than if there was only one such study?

    (b) What the company isn't mentioning is that another 95 similar studies were done in which the evidence was not significant. Does knowledge of this weaken the drug company's case or do we still have evidence at the 5% level that the drug reduces heart attacks?

Type II errors and power

50. *(Adapted from a problem written by L. Heyer)* A certain toothpaste manufacturer claims in a television commercial that at least 4 out of 5 dentists recommend its product. In order to test this claim, a consumer protection group plans to randomly sample 400 dentists and determine whether or not they recommend that toothpaste.

    (a) If the test is to be performed at a 99% significance level, find the rejection region.

    (b) Suppose that in actual fact only 75% of all dentists recommend the manufacturer's toothpaste. What is the probability that the consumer protection agency will correctly identify that the manufacturer's claim is incorrect?

    (c) Now suppose that in actual fact only 60% of all dentists recommend the manufacturer's toothpaste. What is the probability that the consumer protection agency will correctly identify that the manufacturer's claim is incorrect?

51. Suppose you are interested in conducting the statistical test of H_0: $\mu \leq 200$ versus H_a: $\mu > 200$ in a population whose standard deviation is 80. You will reject H_0 if the sample mean of a random sample of 100 items is more than 215.

    (a) Suppose the true value of μ is 207. What is the probability that you will reject H_0 when you conduct the test?

    (b) Suppose now the true value of μ is 220. What is the probability that you will reject H_0 when you conduct the test?

52. Suppose you are interested in conducting the statistical test of H_0: $\mu \geq 20$ versus H_a: $\mu < 20$ and you know that the population variable is normally distributed with standard deviation equal to 3. You will conduct the test at the 5% level. Unknown to you (of course) the alternative hypothesis is, in fact, true. Indeed, the actual value of μ is 19.

    (a) If you use a sample of size 10, what is the probability that the data you collect will lead you to reject the null hypothesis? (Hint: First choose a test statistic and find the rejection region. This is done assuming H_0 is true. Then, using the fact that the true value of μ is 19, find the true distribution of $\bar{X}$ and use that to find the probability that the test statistic falls in the rejection region.)

    (b) If you use a sample of size 40, what is the probability that the data you collect will lead you to reject the null hypothesis?

53. Under normal operations a firm that makes batteries has a 5% or less defective rate. Periodically, quality control samples a number of batteries to test that the defective rate is still less than or equal to 5%. They perform the test at a 90% significance level. Suppose the true defective rate has increased to 6%.

    (a) If they sample 200 batteries what is the probability that they will correctly conclude that the defective rate has increased to above 5%? (Hint: First choose a test statistic and find the rejection region. This is done assuming H_0 is true. Then, using the fact that the true value of p is 0.06, find the true distribution of $\hat{p}$ and use that to find the probability that the test statistic falls in the rejection region.)

    (b) If they sample 400 batteries what is the probability that they will correctly conclude that the defective rate has increased to above 5%?

54. USC uses thousands of fluorescent lightbulbs every year. Suppose the brand of bulb it currently uses has a mean lifetime of 900 hours. A manufacturer claims that its new bulbs, that cost the same as the brand that USC is currently using, have a mean lifetime of more than 900 hours. USC decides to test the new brand. If the test reveals evidence that the mean lifetime is greater than 900 hours then they will purchase the new brand. Whatever sample size USC decides to use in their test, the company offers to provide the funds to double that sample size. Why would the company want to ensure that USC uses a large sample size in their test, or is their offer purely selfless?

Solutions to Selected Problems

Please note: These solutions have not been double-checked for accuracy.

Sheet I - Solutions

1. (a) The student is male.
   (b) The student is either a junior or a senior.
   (c) The student is neither a junior nor a senior.
   (d) The student is both a junior and a senior (presumably the empty set).
   (e) The student is a female business student.
   (f) The student is a senior and hasn't taken a math course.

2. (a) $B \cap D$
   (b) $B^c \cap D^c$
   (c) $(B \cap D^c) \cup (B^c capD)$
   (d) $A \cup B$
   (e) $A \cup C$
   (f) $A^c \cap B^c$
   (g) AB and CD

3. (b) {F1, F2, F3, M1, M2, M3, V1, V2, V3, V4} 10 elements total.
   (d) If you want to take into account the order in which the marbles are chosen you can choose your sample space to be:

	2nd			
1st	B	G1	G2	G3
B	x			
G1		x		
G2			x	
G3				x

This has 12 elements total.

Or, if you don't want to take into account the order in which the marbles are chosen then you can choose your sample space to be:

	B	G1	G2	G3
B	x			
G1	x	x		
G2	x	x	x	
G3	x	x	x	x

This has 6 elements total.

   (e) {AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB}

4. (a) No.
   (b) Yes.
   (c) No.
   (d) No.
   (e) Yes.

5. (a) No.
   (b) Yes.
   (c) Yes.
   (d) Yes.

7. (a) .08
   (b) .72
   (c) .2
   (d) .8
   (e) Yes.
   (f) No.
   (g) Yes.

8. $.62 + .53 - .3 = .85$

9. $.4 + .6 - .8 = .2$

10. (a) $.19 + .44 = .63$
    (b) $1 - .19 = .81$

11. (a) $.35 + .39 = .74$
    (b) $1 - .35 = .65$
    (c) $1 - 10^{-10}$

12. (a) 1/13
    (b) 1/4
    (c) 0

13. (a) $\frac{13}{52} = \frac{1}{4}$
    (b) $\frac{12}{51}$
    (c) $\frac{13}{52} = \frac{1}{4}$

14. (a) $\frac{1}{13}$
    (b) $\frac{1}{13}$
    (c) $\frac{1}{13}$

15. (a) 0.5
    (b) 0.5
    (c) 0.5
    (d) 1
    (e) 0.5

16. (b) $\frac{22}{30} = \frac{11}{15}$
    (c) $\frac{18}{24} = \frac{3}{4}$
    (d) $\frac{18}{22} = \frac{9}{11}$

(e) $\frac{10}{10} = 1$. If at least one of the marbles is green then they must be different colors since there is only one green marble and therefore you can't get two green marbles.

(f) $\frac{10}{22} = \frac{5}{11}$

18. (a) $\frac{.2}{.4} = \frac{1}{2}$

(b) $\frac{.2}{.5} = \frac{2}{5}$

(c) 1

(d) 0

(e) $P(A|B) = P(A)$ so A and B are independent. The other pairs are dependent.

(f) A and C are mutually exclusive.

19. (a) $\frac{190}{280} = \frac{19}{28}$

(b) $\frac{110}{800} = \frac{11}{80}$

(c) No, since $(\frac{600}{1000})(\frac{110}{1000}) \neq \frac{80}{1000}$.

20. (a) Reading across the rows we have: 1, 5, 0, 6; 1, 6, 1, 8; 3, 19, 4, 26; 5, 30, 5, 40.

(b) $\frac{5}{40} = \frac{1}{8}$

(c) $\frac{4}{26} = \frac{2}{13}$

(d) No, since $(\frac{5}{40})(\frac{26}{40}) \neq \frac{3}{40}$

21. $(.53)(.83) = .4399$

23. (a) Independent. Not mutually exclusive.

(b) Not independent. Are mutually exclusive.

(c) $A = \{HHH, HHT, HTH, HTT\}$; $C = \{HTH, HTT\}$. $C \subset A$

24. (a) $P(A) = \frac{8}{20} = .4$; $P(B) = \frac{8}{20} = .4$; $P(B|A) = \frac{7}{19}$.

(b) No.

(c) No.

(d) $(\frac{8}{20})(\frac{7}{19})$

(e) C and D are mutually exclusive and not independent.

(f) $P(C) = P(D) = \frac{8\times 12}{20\times 19} = .2526$

(g) $E = C \cup D$

(h) $P(E) = P(C) + P(D) = .5053$

25. (a) $(\frac{4}{52})(\frac{3}{51})$

(b) $(\frac{48}{52})(\frac{47}{51})$

(c) $1 - (\frac{48}{52})(\frac{47}{51})$

26. (a) $(\frac{13}{52})(\frac{12}{51})(\frac{11}{50})$

(b) $(\frac{13}{52})(\frac{12}{51})(\frac{13}{50})$

(c) $(\frac{39}{52})(\frac{38}{51})(\frac{37}{50})$

(d) $1 - (\frac{39}{52})(\frac{38}{51})(\frac{37}{50})$

(e) $3\frac{(13)(39)(38)}{(52)(51)(50)}$

(f) $\frac{(39)(38)(37)}{(52)(51)(50)} + 3\frac{(13)(39)(38)}{(52)(51)(50)}$

28. (a) $\frac{(11)(10)(9)}{(30)(29)(28)}$

(b) $\frac{(19)(18)(17)}{(30)(29)(28)}$

(c) 1 - (b)

29. (a) 0.001

(b) 0.001

(c) $(.9)^3 = .729$

(d) 1 - (c)

31. (a) $(\frac{8{,}999{,}999{,}999}{9{,}000{,}000{,}000})^{1{,}000{,}000{,}000} = .89$

(b) 1 - (a) = .11

32. Assuming that one going out is independent of the others going out the probability is: $1 - (.1)^3 = .999$

33. (a) $(.35)^4 = .015$

(b) $1 - (.35)^4 = .985$

(c) $1 - (.4)^4 = .9744$

Sheet II - Solutions

1. (a) $\binom{20}{5} = 15504$

   (b) $\binom{18}{3} = 816$

2. (a) $\binom{7}{2} = 21$

   (b) $\binom{2}{1}\binom{5}{1} = 10$

3. $\binom{8}{2}\binom{12}{3} = 6160$

4. $\binom{10}{3}\binom{15}{4} = 163800$

5. $\binom{100}{3} = 161700$

6. $\binom{100}{3} = 161700$

7. $\binom{6}{2} = 15$

8. $\binom{6}{2} = 15$

9. $\binom{8}{3}$ or $\binom{8}{5} = 56$

10. $\binom{8}{2}$ or $\binom{8}{6} = 28$

11. (a) $52 \times 52 = 2704$

    (b) $4 \times 4 = 16$

    (c) $\frac{16}{2704} = \frac{1}{169}$

12. (a) $\binom{12}{5} = 792$

    (b) $\frac{1}{\binom{12}{5}} = .0013$

    (c) $\binom{7}{5} = 21$, $\frac{\binom{7}{5}}{\binom{12}{5}} = .0265$

13. (a) $\binom{12}{6} = 924$

    (b) $\binom{7}{6} = 7$

    (c) $\frac{\binom{7}{6}}{\binom{12}{6}} = .0076$

14. (a) $\binom{42}{6} = 5245786$

    (b) $\frac{1}{\binom{42}{6}} = 1.9 \times 10^{-7}$

    (c) 1.9×10^{-6}

15. (a) $\binom{8}{4} = 924$

    (b) $\binom{6}{3}\binom{2}{1} = 40$

    (c) $\frac{\binom{6}{3}\binom{2}{1}}{\binom{8}{4}} = .57$

16. (a) $\binom{30}{3} = 4060$

    (b) $\binom{20}{2}\binom{10}{1} = 1900$

    (c) $\frac{\binom{20}{2}\binom{10}{1}}{\binom{30}{3}} = .468$

17. (a) $\frac{\binom{2}{1}\binom{5}{1}}{\binom{7}{2}} = .48$

    (b) $\frac{\binom{5}{2}}{\binom{7}{2}} = .48$

18. $\frac{\binom{15}{2}}{\binom{50}{2}} = .086$

19. (a) $(.1)^5 = 10^{-5}$

    (b) $(.1)^2(.9)^3 = .00729$

    (c) $(.9)(.1)(.9)^2(.1) = .00729$

    (d) $\binom{5}{2} = 10$

    (e) $10(.1)^2(.9)^3 = .0729$

20. (a) $(.05)^3(.95)^{17} = 5.23 \times 10^{-5}$

(b) $(.95)^4(.05)(.95)^{10}(.05)^2(.95)^3 = 5.23 \times 10^{-5}$

(c) $\left(\begin{array}{c} 20 \\ 3 \end{array} \right) = 1140$

(e) $1140(.0.5)^3(.9)^{17} = .060$

21. (a) $(\frac{1}{3})^6(\frac{2}{3})^4 = 2.71 \times 10^{-4}$

(b) same as (a)

(c) $\left(\begin{array}{c} 10 \\ 6 \end{array} \right) = 210$

(d) $210(\frac{1}{3})^6(\frac{2}{3})^4 = .057$

22. $\left(\begin{array}{c} 8 \\ 2 \end{array} \right)(.2)^2(.8)^6 = .294$

Sheet III - Solutions

1. (a) qualitative
   (b) quantitative and discrete
   (c) quantitative and continuous
   (d) qualitative
   (e) quantitative and discrete but could be treated as continuous
   (f) quantitative and continuous
   (g) quantitative and continuous but could be treated as discrete

2. (a) .04
   (b) .93
   (c) .55
   (d) .51

4. (a) $(.9)^{12}$
   (b) $\binom{12}{4}(.25)^4(.75)^8 = .19$

5. (a) $\frac{17}{150}$
   (b) $\frac{17}{150}$
   (c) $\frac{3}{17}$
   (d) $\frac{17}{150}\frac{133}{149} + \frac{133}{150}\frac{17}{149} = .202$

6. (a) 4%
   (b) 50.9%
   (c) 0
   (d) .15
   (e) 1
   (f) 1.1
   (g) 1.36
   (h) $\mu - 2\sigma = -1.62$; $\mu + 2\sigma = 3.82$; 92%

7. (a) Could be; the fractions are all positive and they sum to 1.
   (b) Could not be; one of the fractions is negative.
   (c) Could not be; the fractions sum to more than 1.

9. (a) By symmetry $\mu_x = 5$; $\mu_y = 1$; $\mu_z = 5$
   (b) $\sigma_x < \sigma_y = \sigma_z$. The distributions of y and z are just shifts of one another so they have the same spread or sd. A greater proportion of the x-values are close to or equalt to the mean than is the case with the y and z values so $\sigma_x < \sigma_y$.

10. (a) $\frac{1}{4}$
    (b) $\sqrt{2}$
    (c) (iii)
    (d) $\frac{4}{3}$
    (e) $\sqrt{\frac{2}{3}}$
    (f) $\frac{3}{4}$
    (g) $\frac{9}{16}$

11. (a) $\frac{1}{2}$
    (b) $y = \frac{1}{2}x + \frac{1}{2}$; $y = -\frac{1}{6}x + \frac{1}{2}$
    (c) 1.10
    (d) $\frac{3}{4}$

12. (a) 0.8
    (b) -20
    (c) -7.8
    (d) 0.8

13. (a) 0.393
    (b) 0.607
    (c) 0.368
    (d) 0.477

14. (a) $\frac{3}{4}$
    (b) $\frac{11}{32}$
    (c) 0
    (d) $\sqrt{\frac{1}{5}}$

15. (a) $\frac{9}{16}$
    (b) $\frac{8}{15}$
    (c) $\frac{\sqrt{11}}{15}$

Sheet IV - Solutions

1. (a) $\frac{1}{1024}$
   (b) 600 gallons
   (c) 167 gallons

2. (a) $\frac{1}{4}$
   (b) $6 - \sqrt{18} = 1.76$
   (c) 2
   (d) 2; $\sqrt{2}$

3. If we consider it to be distinct outcomes when the same marbles are drawn but in a different order then this problem can be done as follows:

   (a)

	B1	B2	B3	R1	R2	W
B1		2	2	1	1	1
B2	2		2	1	1	1
B3	2	2		1	1	1
R1	1	1	1		0	0
R2	1	1	1	0		0
W	1	1	1	0	0	

   (b) {(B1, B2), (B1, B3), (B2, B1), (B2, B3), (B3, B1), (B3, B2)}
   (c) $\frac{6}{30} = \frac{1}{5}$
   (d) Everything that's not listed in (b).
   (e) $\frac{4}{5}$
   (f)

r	$P[X = r]$
0	1/5
1	3/5
2	1/5

   (g) $E[X] = 1$; half the marbles in the bag are blue, therefore if one did this experiment many times one might expect that on average half of the marbles (i.e. 1) selected would be blue.

   If, on the other hand, we consider it to be the same outcome when the same marbles are drawn in a different order then this problem can be done as follows. Notice that all the probabilities we calculate have the same values.

   (a)

	B1	B2	B3	R1	R2	W
B1		2	2	1	1	1
B2			2	1	1	1
B3				1	1	1
R1					0	0
R2						0

   (b) {(B1, B2), (B1, B3), (B2, B3) }
   (c) $\frac{3}{15} = \frac{1}{5}$
   (d) Everything that's not listed in (b).
   (e) $\frac{4}{5}$
   (f)

r	$P[X = r]$
0	1/5
1	3/5
2	1/5

   (g) $E[X] = 1$; half the marbles in the bag are blue, therefore if one did this experiment many times one might expect that on average half of the marbles (i.e. 1) selected would be blue.

4. (b)

r	$P[X = r]$
-1	5/9
1	4/9

   (c) $E[X] = -\frac{1}{9}$; This is not a fair game. If you played many many times then on average you would lose $\frac{1}{9}$th of a dollar per game.

5. (a) 5
   (b) There are two outcomes: DE and DF (One might also count this as four outcomes if the same children chosen in a different order are considered to be distinct outcomes: DE and ED and DF and FD)
   (c) Let's take the approach that the same children chosen in a different order is the same outcome. In this case we get:

	A	B	C	D	E	F
A		5	5	6.5	7.5	7.5
B			5	6.5	7.5	7.5
C				6.5	7.5	7.5
D					9	9
E						10

   (d)

r	$P[X = r]$
5	1/5
6.5	1/5
7.5	2/5
9	2/15
10	1/15

   (e) $E[X] = 7.17$

6. (a) E.g. (1, 4, 5), (3, 4, 8). Let's choose our outcomes so that order is unimportant. In other words, so that two selections of the same numbers made in different orders are considered to be the same outcomes. In this case the number of outcomes is ${}^{10}C_3 = 120$.
   (b) $\{(1, 2, 3)\}$
   (c) $\frac{1}{120}$
   (d) $\{(1, 2, 4), (1, 3, 4), (2, 3, 4)\}$
   (e) $\frac{3}{120} = \frac{1}{40}$
   (f) $\frac{1}{8}$
   (g) $\frac{1}{12}$
   (h)

r	$P[X = r]$
3	1/120
4	3/120
5	6/120
6	10/120
7	15/120
8	21/120
9	28/120
10	36/120

   (i) $E[X] = 8.25$

7.

r	$P[X = r]$
P	$(.9772)^5 = 0.891$
F	$1 - (.9772)^5 = 0.109$

8. (a) 13; 0.1
   (b) 13; 0.1
   (c) 0.2
   (d) Other ways are: first bag contains 7 and the second bag contains 5 or both bags contain 6
   (e) 0.37
   (f) 10, 11, 12, 13, 14
   (g)

r	$P[X = r]$
10	0.09
11	0.3
12	0.37
13	0.2
14	0.04

   (h) 11.8

Sheet V - Solutions

1. (a) Hypergeometric distribution with $N = 52$; $A = 13$; $N - A = 39$; $n = 5$
   (b) $\dfrac{\binom{13}{3}\binom{49}{2}}{\binom{52}{5}} = \frac{11}{85}$
   (c) $P[X = 3] + P[X = 4] + P[X = 5]) = .143$
   (d) $(\frac{13}{52})(5) = 1.25$

3. (a) $1 - \left(\dfrac{\binom{7}{4}\binom{8}{1}}{\binom{15}{5}} + \dfrac{\binom{7}{5}}{\binom{15}{5}}\right) = 0.90$
   (b) $\dfrac{\binom{3}{2}\binom{12}{3}}{\binom{15}{5}} + \dfrac{\binom{3}{3}\binom{12}{2}}{\binom{15}{5}} = 0.24$

5. (a) $\dfrac{\binom{7}{5}}{\binom{20}{5}} = 0.0014$
   (b) $\dfrac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}} + \dfrac{\binom{12}{4}\binom{8}{1}}{\binom{20}{5}} + \dfrac{\binom{12}{5}}{\binom{20}{5}} = 0.704$

6. (a) Binomial distribution with $n = 8$ and $p = 0.05$.
   (b) $1 - ((.95)^8 + 8(.05)(.95)^7) = .06$
   (c) 0.4
   (d) 0.62

7. (a) .558
   (b) .068

8. (a) .5576
   (b) .36
   (c) 1

9. $1-((.607)^8+8(.393)(.607)^7+28(.393)^2(.607)^6) = 0.670$

10. (a) X has a hypergeometric distribution with $N = 60$, $n = 5$, $A = 20$, $N - A = 40$.
    (b) Y has a binomial distribution with $n = 5$ and $p = \frac{1}{3}$.
    (c) If you wrote the distributions out explicitly you will see that they are almost identical.

11. (a) X has a hypergeometric distribution with $N = 52$, $n = 3$, $A = 13$, $N - A = 39$.
    (b) Y has a binomial distribution with $n = 3$ and $p = \frac{1}{4}$.
    (c) The two distributions are almost identical.

12. (a) $4(.85)^3(.15) = 0.368$
    (b) We are selecting 4 calls at random from among all Los Angeles 911 calls. The set of all these calls is finite so the distribution of the number of calls in the sample that are not emergencies is hypergeometric. However, we don't know the value of N so in our answer to (a) we used the binomial distribution instead. This is an approximation but should be a very good one since we expect that N is large compared with $n = 4$.

14. (a) 17.5 minutes
    (b) 4.33 minutes
    (c) one third of the time
    (d) $\frac{1}{3}$
    (e) $\frac{1}{2}$
    (f) $\frac{8}{9}$

15. (a) $\frac{1}{3}$
    (b) $1 - ((\frac{2}{3})^{10} + 10(\frac{1}{3})(\frac{2}{3})^9 + 45(\frac{1}{3})^2(\frac{2}{3})^8 + 120(\frac{1}{3})^3(\frac{2}{3})^7) = .441$
    (c) 3.33

16. (a) .4906
    (b) .4991
    (c) .0162
    (d) .9484
    (e) .0233

17. (a) .0142
    (b) .9398
    (c) .00001
    (d) .9997
    (e) 0

18. (a) .84
    (b) .39
    (c) .67

19. (a) .67 (or -.67)
    (b) -.13
    (c) -.52
    (d) 1.28 (or -1.28)
20. (a) .4772
    (b) .9987
    (c) .1069
    (d) .6709
21. (a) .3446
    (b) .8185
    (c) .1151
    (d) .7257
22. (a) .0013
    (b) .0228
    (c) .6902
23. Let w denote the warranty. We require that $\frac{w-81{,}200}{4{,}300} = -2.33$. Thus $w = 71,181$ miles.
25. (a) $(2)(.4772) = .9544$
    (b) We need to find the area under a standard normal curve from $\frac{5.28-5.321}{.017} = -2.41$ to $\frac{5.32-5.321}{.017} = -.06$. This area is: $.4920 - .0239 = .4681$
    (c) We require $2.575 = \frac{5.32-5.3}{\sigma}$. Thus $\sigma = .0078$ inches.
26. (a) .0159
    (b) .0159
    (c) .9841
    (d) $3(.7486)(.2514^2) = .1419$
27. (a) $\frac{.4772}{.9772} = .49$
    (b) Let X be the number selected that weigh less than 4 oz. X has a binomial distribution with $n = 8$ and $p = .5 - .4772 = .0228$. We need to find $P(X \geq 2) = 1 - (.9772)^8 - 8(.0228)(.9772)^7 = .013$.

Sheet VI - Solutions

1. (a) .28
   (b) $.19/.2 = .95$
   (c) No, because the answers to (b) and (c) are not the same.
   (d) No.

3. (a) $\frac{308}{453} = 70\%$
   (b) $\frac{246}{330} = 75\%$
   (c) No, since (b) and (c) show us that {Income is less than 40K} and {Sex is Female} are dependent.

4. Reading across the rows the contingency table reads:

   1st row: 72, 9, 81

   2nd row: 24, 3, 27

   3rd row: 96, 12, 108

5. (a) X takes on the values -1, 0 and 1 with probabilities .2, .1 and .7 respectively.
   (b) Y takes on the values 0 and 1 with probabilities .6 and .4 respectively.
   (c) Yes.
   (d) Yes, since all the entries in the body of the table are products of the row and column totals.

6. (a) Same as 5(a)
   (b) Same as 5(b)
   (c) Yes.
   (d) No, since, e.g., $\{X = 0\}$ and $\{Y = 0\}$ are not independent.

8. (a)

	Y			
X	0	1	2	
0	.36	.18	.06	.6
1	.18	.09	.03	.3
2	.06	.03	.01	.1
	.6	.3	.1	

   (b) There are many possibilities. Here is an example.

	Y			
X	0	1	2	
0	.3	.2	.1	.6
1	.25	.05	0	.3
2	.05	.05	0	.1
	.6	.3	.1	

10. (b)

r	Heads	Tails
$P[X = r]$	.5	.5

   (c)

r	1	2	3	4	5	6
$P[Y = r]$	.17	.17	.17	.17	.17	.17

   (d)

	Y						
X	1	2	3	4	5	6	
H	.08	.08	.08	.08	.08	.08	.5
T	.08	.08	.08	.08	.08	.08	.5
	.17	.17	.17	.17	.17	.17	

   (e) Yes.

13. (a) Same distribution. Independent.
    (b) Same distribution. Dependent.
    (c) Same distribution. Dependent.
    (d) Different distributions. Independent.
    (e) Different distributions (since they take on different values). Dependent (the value of one completely determines the value of the other).

Sheet VII - Solutions

3. (a) $\{Y = 4\} = \{X = 2\}$ so $P[Y = 4] = P[X = 2] = .1$
   (b)

s	$P[Y = s]$
3	.5
3.5	.3
4	.1
4.5	.1

4. (a) $X = 2$
   (b) $\{1 < X < 2\}$
   (c) $P[2 < Y < 4] = P[1 < X < 2] = \frac{1}{2}$
   (d) $P[4 < Y < 6] = P[2 < X < 3] = \frac{1}{3}$

5. (a) .0625
   (b) $1 - \frac{1}{2}(20)(0.025) = 0.75$
   (c) The people that gain more than 10 kg are the ones that gain more than 22 lb. The proportion of people that do this is 0.45.
   (d) $y = \frac{1}{2.2}x$ so $\mu_y = \frac{1}{2.2}\mu_x = 9.09$ kg and $\sigma_y^2 = (\frac{1}{2.2})^2\sigma_x^2 = 25$ kg^2.

8. Let X denote the number of bars in the sample that weigh less than 4 oz. X is binomially distributed with $n = 20$, and $p = 0.5 - 0.4772 = 0.0228$. Let $\hat{p}$ denote the proportion of bars in the sample that weigh less than 4 oz. Notice that $\hat{p} = \frac{1}{20}X$.
   (a) We are being asked to find $P[\hat{p} = .1]$. This is equal to $P[X = (20)(.1) = 2] = 190(.0228)^2(.9772)^{18} = .065$
   (b) We need to find $P[\hat{p} < .2]$. This is equal to $P[X < 4] = (.9772)^{20} + 20(.0228)(.9772)^{19} + 190(.0228)^2(.9772)^{18} + 1140(.0228)^3(.9772)^{17} = 0.9990$
   (c) $E[X] = 20(.0228)$ so $E[\hat{p}] = \frac{1}{20}20(.0228) = 0.0228$.
   (d) $\text{SD}[X] = \sqrt{20(.0228)(.9772)}$ so $\text{SD}[\hat{\text{p}}] = \frac{1}{20}\sqrt{20(.0228)(.9772)} = \sqrt{\frac{(.0228)(.9772)}{20}} = .033$

9. (a) $\{23 < Y < 26\} = \{6 < X < 7\} = \frac{1}{5} = .2$
   (b) Y is uniformly distributed on the interval $[20, 35]$ so $P[23 < Y < 26] = \frac{3}{15} = .2$

10. It's a uniform distribution on $[5, 25]$.

12. Y has a standard normal distribution.

13. Let $Z = \frac{X-3000}{100}$. Every outcome of the experiment has a Z-value associated with it. The distribution of Z is standard normal.
   (a) $P[3035 < X < 3230] = P[\frac{3035-3000}{100} < Z < \frac{3230-3000}{100}] = P[.35 < Z < 2.3] = .4893 - .1368 = .3525$
   (b) $P[2940 < X < 3010] = P[-0.6 < Z < 0.1] = .2257 + .0398 = .2655$
   (c) $P[X < 2850] = P[Z < -1.5] = .5 - .4332 = .0668$
   (d) $P[X < 3325] = P[Z < 3.25] = .5 + .4994 = .9994$

14. Let x denote the height of adult women in inches and let $z = \frac{x-63.5}{2.5}$. The variable z has a standard normal distribution.
   (a) The women that are taller than 6 feet = 72 inches have z-values that are greater than $\frac{72-63.5}{2.5} = 3.4$. From the normal tables we see that the proportion is .5 - .4997 = .0003.
   (b) This is the proportion that have z-values between -1.4 and 3.4. From the tables we get: .4192 + .4997 = .9189.
   (c) From normal tables we see that the tallest 10% of women have z-values that are bigger than 1.28. The height, x that corresponds to $z = 1.28$ is given by $\frac{x-63.5}{2.5} = 1.28$ i.e. $x = 66.7$ inches, i.e. 5 feet 6.7 inches tall.

15. (a) .4
   (b) .5
   (c)

value of w	proportion of population with that value
0	.2
1	.3 + 0 = .3
2	.2 + .2 = .4
3	.1

16. (a) $P[W = 5] = P[X = 0 \text{ and } Y = 5] + P[X = 10 \text{ and } Y = -5] = (.1)(.6) + (.2)(.4) = .14$
   (b) $P[W < 10] = 1 - P[W \geq 10] = 1 - ((.7)(.6) + (.2)(.6)) = .54$
   (c)

r	-5	0	5	10	15
$P[W = r]$	.04	.28	.14	.42	.12

20. (a)

r	0	1	2	3	4
$P[W = r]$	.04	.2	.37	.3	.09

(b) $E[W] = 2.2$; $\text{Var}[W] = .98$

(c) Yes, it is true and yes, it should be true since it is always the case that $E[X+Y] = E[X] + E[Y]$.

(d) Yes, it is true and yes, it should be true since X and Y are independent means that $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

21. (a)

r	0	1	2	3	4
$P[W = r]$	.1	.2	.3	.2	.2

(b) $E[W] = 2.2$; $\text{Var}[W] = 1.56$

(c) Yes, it is true and yes, it should be true since it is always the case that $E[X+Y] = E[X] + E[Y]$.

(d) No, it is not true and no, it should not necessarily be true. We only know that $\text{Var}[X+Y] = \text{Var}[X]+\text{Var}[Y]$ when X and Y are independent. This is not the case here.

23. (a) $E[W] = 2(75) - 30 = 120$; $\text{SD}[W] = 2(8) = 16$; $\text{Var}[W] = 4(64) = 256$

(b) $E[V] = 2(75) + 6(42) = 402$; $\text{Var}[V] = \text{Var}[2X] + \text{Var}[6Y] = 4(64) + 36(100) = 3856$; $\text{SD}[V] = \sqrt{3856} = 62.1$

24. (a) $E[W] = 15 + 25 - 10 = 30$; $\text{Var}[W] = \text{Var}[X+Y] = 9 + 9 = 18$; $\text{SD}[W] = \sqrt{18} = 4.24$

(b) $E[V] = 5(15) - 2(25) + 1 = 26$; $\text{Var}[V] = \text{Var}[5X - 2Y] = 25(9) + (-2)^2(9) = 261$; $\text{SD}[V] = \sqrt{261} = 16.16$

25. Let L denote the number of lemonade sales and let B denote the number of brownie sales. We are told that $\mu_L = 25$; $\sigma_L = 5$; $\mu_B = 18$; $\sigma_B = 6$. Let R denote his total revenue.

$R = (0.75)L + (1.25)B$

$\mu_R = (.75)(25) + (1.25)(18) = \$41.25$

$\sigma_R^2 = (.75^2)(25) + (1.25^2)(36) = 70.3125$

$\sigma_R = \sqrt{70.3125} = \$8.39.$

No, it probably is not reasonable to assume that L and B are independent. One might expect that if one is large the other would also tend to be large because that would mean that the entrepreneur had had a lot of customers that day.

26. Let B denote her sales from bakery goods, M denote her sales from material goods, and I denote her net income after tax and paying her expenses.

$I = B + (.9)M - 2000.$

$\mu_I = 5500 + (.9)(500) - 2000 = \$3,950;$

$\sigma_I^2 = 1000^2 + (.91)^2(100)^2 = 1.008,281;$

$\sigma_I = \$1,004.13$

28. (a) This is the area under the standard normal to the left of $\frac{5-5.5}{.3} = -1.67$ and to the right of $\frac{6-5.5}{.3} = 1.67$. This comes to: $1 - (2)(.4545) = .095$

(b) Let X_1 denote the amount in the first can and let X_2 denote the amount in the second can. X_1 and X_2 are independent and normally distributed with a mean of 5.5 ml and an sd of 0.3 ml. After mixing the two cans and pouring the contents back again into the cans, the amount in the cans will then be $\bar{X} = \frac{1}{2}(X_1 + X_2)$. $\bar{X}$ is an average of normal variables so it is also normally distributed. Its mean is $\frac{1}{2}(5.5) + \frac{1}{2}(5.5) = 5.5$. Its variance is $(\frac{1}{2})^2(.3)^2 + (\frac{1}{2})^2(.3)^2 = .045$ so its sd is $\sqrt{.045} = .21$. So the probability that they are unacceptable is the area under a standard normal to the left of $\frac{5-5.5}{.21} = -2.38$ and to the right of $\frac{6-5.5}{.21} = 2.38$. That comes to $1 - (2)(.4913) = .0174$. Notice that the answer to (b) is less than the answer to (a). If you mix the paints before using them they are more likely to be closer to the correct color - the average of two cans is more likely to be close to the mean than just one can alone.

Sheet VIII - Solutions

2. The weight of a box, X, is the sum of the weights of the 2000 nails that it contains, i.e. $X = X_1 + X_2 + \ldots X_{2000}$, where X_i is the weight of the i'th nail in the box. So $E[X] = 2000E[X_i] = 2000(3.22) = 6,440$ grams and $\text{SD}[X] = \sqrt{2000}\text{SD}[X_i] = \sqrt{2000}(0.14) = 6.3$ grams.

3. (a) The probability of winning 50 cents or more on a single play is $.05 + .05 = .1$. If you play 25 times, the probability of winning 50 cents or more at least once is 1 minus the probability of always winning less than 50 cents which is $1 - (.9)^{25} = .93$.

   (b) $\bar{X}$ denotes the average amount you win per play when you play 25 times. To find its expected value, variance and sd we will need to know the expected value and sd of W.

   $E[W] = (0)(.6) + (25)(.3) + (50)(.05) + (100)(.05) = 15$

   $\text{Var}[W] = (15)^2(.6)+(10)^2(.3)+(35)^2(.05)+(85)^2(.05) = 587.5$

   $\text{SD}[W] = \sqrt{587.5} = 24.2$

   $E[\bar{X}] = E[W] = 15$ cents

   $\text{SD}[\bar{X}] = \frac{\text{SD}[W]}{\sqrt{25}} = 4.84$ cents

   $\text{Var}[\bar{X}] = \frac{\text{Var}[W]}{25} = 23.5 \text{ cents}^2$

   (c) The expected value of $\bar{X}$ is 15 and its sd is 4.84 so 50 cents is more than 7 standard deviations away from its expected value. Yes, that is highly improbable and therefore it would be very surprising when it happened.

6. The total amount of time, X, that you will be able to study at night at your desk is the sum of the lifetimes of the dozen lightbulbs that you purchased. I.e. $X = X_1+X_2+\ldots X_{12}$ where the X_i are iid normal random variables with mean 600 and sd 100.

   (a) $E[X] = 12(600) = 7,200$ hours.

   (b) X has a normal distribution with an expected value of 7,200 hours and an sd equal to $\sqrt{12}(100) = 346.4$ hours. Thus, $P[X > 7000] = P[Z > \frac{7000-7200}{346.4}] = P[Z > -.58] = .5 + .2190 = .7190$

7. Let n denote the number of cookies that should be put into each package. Let X denote the total weight of these packages. X is a sum of n iid variables that are each normally distributed with a mean of 14.22 grams and an sd of 0.29 grams. It follows that X is also normally distributed with a mean of $14.22n$ and an sd of $\sqrt{n}(0.29)$. We need to find n so that this normal distribution is positioned so that 95% of the area lies to the right of 340 grams. Using the standard normal tables we see that the z-score of 340 must be somewhere between -1.64 and -1.65. We'll use -1.64. Thus: $\frac{340-14.22n}{\sqrt{n}(0.29)} = -1.64$. It follows that $340 - 14.22n = -.4756\sqrt{n}$ and $14.22n - .4756\sqrt{n} - 340 = 0$. This is a quadratic equation in $\sqrt{n}$. Using the quadratic formula we get: $\sqrt{n} = \frac{.4756\pm\sqrt{.2262+4(14.22)(340)}}{28.44} = \pm 4.9$ so $n = 4.9^2 = 24.01$. This is very close to 24 so we might want to put 24 cookies in a package. However, if we really want to be certain that at least 95% of the packages weigh at least 340 grams then we should round up and put 25 cookies in each package.

8. Let X denote a player's score when they roll the die 30 times. X is the sum of 30 iid variables whose distributions have probability $\frac{1}{6}$ on each of the integers 1 through 6. The expected value of such a variable is 3.5 and its sd is 1.71. So $E[X] = 30(3.5) = 105$ and $\text{SD}[X] = \sqrt{30}(1.71) = 9.37$. Moreover, by the Central Limit Theorem X is approximately normally distributed. Thus, $P[X \geq 110] \approx P[Z \geq \frac{110-105}{9.37}] = P[Z \geq .53] = .2981$.

   Note: In this problem X is a discrete variable. It takes on only integer values between 30 and 180 inclusive. We are approximating its distribution by a normal distribution that is continuous. Thus, to get a better estimate we should use the continuity correction. Let $\tilde{X}$ be a normally distributed variable with the same mean and sd as X. $P[X \geq 110] \approx P[\tilde{X} \geq 109.5] = P[Z \geq \frac{109.5-105}{9.37}] = P[Z \geq .48] = .3156$.

9. Let X denote the sum of the spans of all 1200 people. X is approximately normally distributed with a mean of $1200 \times 64 = 76,800$ inches and a standard deviation of $\sqrt{1200} \times 4 = 138.6$ inches. Thus $P[X \geq 77,000] = P[Z \geq \frac{77,000-76,800}{138.6}] = P[Z \geq 1.44] = .5 - .4251 = .0749$

10. (a) X is the sum of the number of candies in the first bag, X_1, and the number of candies in the second bag, X_2. Since X is the sum of only two iid variables we can't use the central limit theorem, i.e. X is *not* approximately normally distributed. Instead, we need to work out the desired probabil-

ity from first principles. Our answer will be exact since no approximation is being used. $P[X > 12] = P[X_1 = 6 \text{ and } X_2 = 7] + P[X_1 = 7 \text{ and } X_2 = 6] + P[X_1 = 7 \text{ and } X_2 = 7] = (.4)(.3) + (.3)(.4) = (.3)(.3) = .33$.

(b) In this case, because X is the sum of two hundred iid variables, we *can* use the central limit theorem and conclude that X is approximately normally distributed. Our answer will be approximate because the distribution of X is only approximately normal. To find the mean and sd of X we first need to find the mean and sd of the individual X_i's:
$\mu = 5(.3) + 6(.4) + 7(.3) = 6$
$\sigma = \sqrt{1^2(.3) + 0^2(.4) + 1^2(.3)} = \sqrt{.6}$
It follows that X has a mean of $200(6) = 1200$ and an sd of $\sqrt{200}\sqrt{.6} = 10.95$. So $P[X > 1200] \approx P[Z > \frac{1200-1200}{10.95}] = .5$
Note: X is a discrete variable. It takes on only integer values between 1000 and 1400 inclusive. We are approximating its distribution by a normal distribution that is continuous. So, to get a better estimate we should use the continuity correction. Let $\tilde{X}$ be a normal variable that has the same mean and sd as X. $P[X > 1200] \approx P[\tilde{X} > 1200.5] = P[Z > \frac{1200.5-1200}{10.95}] = P[Z > 0.05] = .5 - .0199 = .4801$

11. (a) Since $\bar{X}_2$ is the average of only two iid variables we can't use the central limit theorem. Instead we work out explicitly all the different ways that we could get $\bar{X}_2 < 0.5$. Our answer will be exact since we are not approximating the distribution of $\bar{X}_2$ by anything.
$P\bar{X}_2 < 0.5] = P[X_1 = 0 \text{ and } X_2 = 0] = (.6)(.6) = .36$

(b) Since $\bar{X}_{100}$ is the average of 100 iid variables we conclude by the central limit theorem that it is approximately normally distributed. To find its mean and sd we first need to find the mean and sd of the indivdual X_i's:
$\mu = 0(.6) + 1(.3) + 2(.1) = .5$
$\sigma = \sqrt{(.5)^2(.6) + (.5)^2(.3) + (1.5)^2(.1)} = \sqrt{.45}$
Thus, $\bar{X}_{100}$ has a mean of .5 and an sd of $\frac{\sqrt{.45}}{\sqrt{100}} = .067$. It follows that $P[\bar{X}_{100} < 0.5] \approx P[Z < \frac{0.5-0.5}{.067}] = 0.5$
Note: $\bar{X}_{100}$ is a discrete variable. It takes on the values $0, 0.01, 0.02, 0.03, \ldots 2$. We are approximating its distribution by a normal distribution that is continuous. Therefore, to get a better estimate we should use the continuity correction. Let $\tilde{X}$ be a normally distributed variable with the same mean and sd as $\bar{X}_{100}$. $P[\bar{X}_{100} < 0.5] \approx P[\tilde{X} < 0.495] = P[Z < \frac{0.495-0.5}{.067}] = P[Z < -0.07] = .5 - .0279 = .4721$

12. (a) Let X denote the number of heads obtained when the coin is tossed 100 times and let $\hat{p} = \frac{1}{100}X$ denote the percentage of heads obtained. X has a binomial distribution with $n = 100$ and $p = 0.5$. $np(1-p) = 100(0.5)(.5) = 25 > 5$ so by the central limit theorem $\hat{p}$ has an approximately normal distribution with mean $p = .5$ and sd $\sqrt{\frac{p(1-p)}{n}} = .05$. We will use the continuity correction so let $\tilde{p}$ denote a normally distributed variable with the same mean and sd as p. The values taken on by $\hat{p}$ are $0, 0.01, 0.02, \ldots 1$. Thus: $P[\hat{p} > 60\%] \approx P[\tilde{p} > .605] = P[Z > \frac{.605-.5}{.05}] = P[Z > 2.1] = .5 - .4821 = .0179$

(b) Let $\hat{p}$ here denote the percentage of heads obtained when the coin is tossed 1000 times. $np(1-p) = 1000(0.5)(0.5) = 250 > 5$ so by the central limit theorem $\hat{p}$ has an approximately normal distribution with mean $p = .5$ and sd $\sqrt{\frac{p(1-p)}{n}} = .016$. Let $\tilde{p}$ denote a normally distributed variable with the same mean and sd as $\hat{p}$. The values taken on by $\hat{p}$ in this case are $0, 0.001, 0.002, \ldots 1$. Thus: $P[\hat{p} > 60\%] \approx P[\tilde{p} > .6005] = P[Z > \frac{.6005-.5}{.016}] = P[Z > 6.28] = .5 - .5 = 0$.

13. Let $\hat{p}$ denote the percentage in a package that are defective. $np(1-p) = 300(.1)(.9) = 27 > 5$ so by the central limit theorem $\hat{p}$ has an approximately normal distribution with mean .1 and sd $\sqrt{\frac{(.1)(.9)}{300}} = .017$ The values taken on by $\hat{p}$ are $0, \frac{1}{300}, \frac{2}{300}, \ldots, 1$. Let $\tilde{p}$ denote a normally distributed variable with the same mean and sd as $\hat{p}$. $P[\hat{p} > .15] \approx P[\tilde{p} > .15 + \frac{1}{600}] = P[Z > \frac{.1517-.1}{.017}] = P[Z > 3.04] = .5 - .4988 = .0012$

Sheet IX - Solutions

1. Let X denote the number of men that are chosen when 8 employees are selected at random. X has a hypergeometric distribution with $N = 50$, $A = 15$ and $n = 8$. We want to find $P[X \geq 2] = 1 - P[X = 0] - P[X = 1] = 1 - .0438 - .1879 = .7683$

2. Let X denote the number of passengers that show up when 267 people are sold reservations. X has a binomial distribution with $n = 267$ and $p = .94$. We want to find $P[X > 255]$. Use the normal approximation. This is legitimate since $np(1-p) = 267(.94)(.06) > 5$. Let Y be a normally distributed random variable with mean equal to $np = 250.98$ and standard deviation equal to $\sqrt{np(1-p)} = 3.88$. Then $P[X > 255] \approx P[Y > 255.5] = P[Z > 1.16] = .5 - .3770 - .123$.

3. (a) $.5 - .2734 = .2266 = 22.66\%$

   (b) Let the time be t. We require $-1.28 = \frac{t-450}{40}$. It follows that $t = 398.8$. Any boy that runs a mile in less than 398.8 seconds is deemed to be "excellent."

   (c) $\bar{X}$ has a normal distribution with mean 450 seconds and standard deviation $\frac{40}{\sqrt{25}} = 8$ seconds. Thus $P[\bar{X} < 428] = P[Z < -2.75] = .5 - .4970 = .003$.

   (d) The answer to (c) is rather small. Rather than believe that such an unlikely event occurred by chance, it seems more reasonable to believe the alternative explanation that it occurred because of King Henry's excellent athletics program.

4. (a) Let ΣX denote the combined weight of the 20 hopper cars. ΣX is normally distributed with mean equal to 20μ and standard deviation equal to $\sqrt{20}(1.5) = 6.71$. We require that $-2.33 = \frac{1500-20\mu}{6.71}$. It follows that $\mu = 75.78$ tons.

   (b) The probability that any one randomly selected hopper car is dangerously overfull is equal to $.5 - .4772 - .0228$. Let X denote the number of hopper cars that are dangerously overfull when 20 cars are filled. X has a binomial distribution with $n = 20$ and $p = .0228$. Thus $P[X \geq 1] = 1 - P[X = 0] = 1 - (.9772)^2 0 = .37$.

5. (a) $\mu = (0)(.6) + (10)(.3) + (20)(.1) = 5$; $\sigma^2 = (25)(.6) + (25)(.3) + (225)(.1) = 45$; $\sigma = \sqrt{45} = 6.71$

   (b)

total winnings	probability
0	$(.6)(.6) = .36$
10	$2(.6)(.3) = .36$
20	$2(.1)(.6) + (.3)(.3) = .21$
30	$2(.3)(.1) = .06$
40	$(.1)(.1) = .01$

   (c) Let X denote the total amount won each day. X is the sum of the winnings of 500 plays. Since it is a sum of more than 30 (500) iid variables X has an approximately normal distribution with expected value equal to $n\mu = (500)(5) = 2500$ and sd equal to $\sqrt{n}\sigma = \sqrt{500}(6.71) = 150$. We want to find the amount, w, for which $P[X \leq w] = .99$. From the normal tables (look up .49 in the body of the table) we get that the z-score of w is 2.33. It follows that $\frac{w-2500}{150} = 2.33$. Solving yields $w = 2500 + (150)(2.33) = 2,849.50$.

6. (a) We find the area under the distribution function between 10 yrs and 20 yrs. This area is $1/4 = 0.25$.

   (b) Let X be the number that last longer than 10 yrs when 20 are selected at random. X has a binomial distribution with $n = 20$ and $p = 0.25$. Thus $P[X \geq 4] = 1 - (0.75)^{20} - 20(0.75)^{19}(0.25) - 190(0.75)^{18}(0.25)^2 - 1140(0.75)^{17}(0.25)^3 = .775$

   (c) $\bar{X}$ is approximately distributed with mean equal to μ (the mean of the distribution) and sd equal to $\frac{\sigma}{\sqrt{200}}$ (where σ is the sd of the distribution). To find μ and σ we note that the equation of the line is $y = -.005x + .1$. Thus
   $\mu = \int_0^{20}(-.005x^2 + .1x)\,dx = \frac{20}{3}$
   $\sigma^2 = \int_0^{20}(-.005x^3 + .1x^2)\,dx - (\frac{20}{3})^2 = \frac{200}{9}$
   $\sigma = \frac{10\sqrt{2}}{3}$
   Thus $P[\bar{X} \geq 7] \approx P[Z \geq 1] = .5 - .3413 = .1587$

7. (a) $10(\frac{6}{50}) = 1.2$

   (b) $$1 - \frac{\binom{44}{10}}{\binom{50}{10}} - \frac{\binom{44}{9}\binom{6}{1}}{\binom{50}{10}} = .344$$

8. (a) Let X denote the number of seeds in a package that germinate. X has a binomial distribution with $n = 100$, $p = .8$. We want to find $P[X \geq 70]$. We use the normal approximation. Let Y be a normally distributed r.v. with mean equal to 80 and sd equal to 4. $P[x \geq 70] \approx P[Y \geq 69.5] = P[Z \geq -2.63 = .5 + .4957 = .9957$.

   (b) Let X denote the number of seeds among the ten selected that germinate. The ten seeds are selected at random from the package that contins 100 seeds total, 85 of which will germinate. So X has a hypergeometric distribution with $N = 100$, $A = 85$, $N - A = 15$, and $n = 10$. We want to find $P[X \geq 7] = P[X = 7] + P[X = 8] + P[X = 9] + P[X = 10] = 0.959$.

9. (a) Let ΣX be the total number of gallons in 100 randomly selected barrels. ΣX is normally distributed with mean equal to 5010 gallons and sd equal to $(10)(0.5) = 5$ gallons. Thue $P[\Sigma X \geq 5000] = P[Z \geq -2] = .5 + .4772 = .9772$.

   (b) Let ΣX now denote the total number of gallons in 20 randomly selected barrels. ΣX is normally distributed with mean equal to 20μ and sd equal to $\sqrt{20}(0.5) = 2.24$ gallons. We require that $\frac{1000-20\mu}{2.24} = -2.33$. Thus $\mu = 50.261$ gallons.

10. (a) The experiment here is buy a bag of candies, i.e. select 50 candies from an infinite pool of candies of assorted colors. Let X denote the number that are red. X has a binomial distribution with $n = 50$ and $p = .3$. We want to find $P[12 \leq X \leq 15] = P[X = 12] + P[X = 13] + P[X = 14] + P[X = 15] = .43$

    (b) The experiment here is buy two bags of candies, i.e. select 100 candies at random from an infinite pool of assorted colors. Let X denote the number of candies that are orange. X has a binomial distribution with $n = 100$ and $p = .05$. We want to find $P[X = 0] = (.95)^{100} = .0059$.

    (c) The experiment now is to select 5 candies from a finite pool consisting of 18 red, 3 orange, 11 yellow, 12 green, and 6 blue candies. Let X denote the number selected that are red. X has a hypergeometric distribution with $N = 50$, $A = 18$, $N - A = 32$, $n = 5$. We want to find $P[X \geq 3] = P[X = 3] + P[X = 4] + P[X = 5] = .24$.

11. (a) A shipment is rejected if more than one telephone out of five randomly selected telephones is rejected. It looks like our experiment here is to sample 5 telephones from a finite shipment consisting of 20 telephones. Let X be the number of telephones from the 5 sampled that are defective. At first glance it looks like X has a hypergeometric distribution. However, the catch is that different shipments have different numbers of defective telephones so we are unable to identify A (it's different for different shipments). The way to think of this problem is that each of the 5 telephones selected has a 5% chance of being defective. Therefore X has a binomial distribution with $n = 5$ and $p = 0.05$. We want to find $P[X > 1] = 1 - P[X = 0] - P[X = 1] = .023$.

    (b) Now the experiment is to select 5 telephones from a particular shipment consisting of 18 good phones and 2 defective phones. Let X denote the number of phones selected that are defective. This time X has a hypergeometric distribution with $N = 20$, $A = 2$, $N - A = 18$, and $n = 5$. We want to find $P[X > 1] = P[X = 2] = .053$

    (c) We want to find: P[shipment contains no defective phones *given* a random selection of 5 of its phones contains no defective phones]. This is the same as: P[shipment contains no defective phones *and* a random selection of 5 of them contains no defective phones] *divided by* P[a random selection of 5 phones from any shipment contains no defective phones]. If a shipment contains no defective phones then certainly any random selection of 5 phones from it will contain no defective phones. So the numerator is simply the proportion of all shipments that contain no defective phones, namely $(0.95)^{20}$. The denominator can be calculated as in (a) and comes to $(0.95)^5$. So the answer here is $(0.95)^{15} = 0.46$.

Sheet X - Solutions

1. (a) $\bar{X}$ is normally distributed with $E[\bar{X}] = \mu$ and $\mathrm{SD}[\bar{X}] = \frac{0.01}{\sqrt{5}} = 0.00447$. Thus $P[\bar{X} > \mu + 0.005 \text{ or } \bar{X} < \mu - 0.005] = 2P[\bar{X} > \mu + 0.005] = 2P[Z > \frac{\mu+0.005-\mu}{0.00447}] = 2P[Z > 1.12] = 2(0.5 - 0.3686) = 0.2628$

   (b) Let X denote the number of pistons sampled that have diameters greater than $\mu + 0.005$. X has a binomial distribution with $n = 5$ and p equal to the proportion of pistons whose diameters are greater than $\mu + 0.005$ which is equal to $P[Z > \frac{\mu+0.005-\mu}{0.01}] = P[Z > 0.5] = 0.3085$. Thus $P[X \geq 3] = 10(.3085)^3(.6915)^2 + 5(.3085)^4(.6915) + (.3085)^5 = .1745$ and $P[M > \mu + 0.005 \text{ or } M < \mu - 0.005] = 2(.1745) = .349$.

   (c) If our goal is to estimate μ to within ± 0.005 then we are more likely to do that with $\bar{X}$ than with M so $\bar{X}$ would be the better estimator.

2. (a) Yes, this is an unbiased estimator of T. The variable x has mean $\frac{T}{2}$ and sd equal to $T/\sqrt{12}$, so $\bar{X}$ is approximately normally distributed with mean equal to $\frac{T}{2}$ and sd equal to $\frac{T/\sqrt{12}}{\sqrt{100}} = \frac{T}{20\sqrt{3}}$. Thus $P[T-(.1)T < 2\bar{X} < T + (.1)T] = P[.45T < \bar{X} < .55T] \approx P[\frac{.45T-.5T}{T/20\sqrt{3}} < Z < \frac{.55T-.5T}{T/20\sqrt{3}}] = P[-1.73 < Z < 1.73] = .9164$

   (b) No, $X_{\max}$ will always be less than the true value of T. It is a biased estimator of T. $P[T - (.1)T < X_{\max} < T + (.1)T] = P[X_{\max} > (.9)T] = 1 - (.9)^{100} = .9999734$

   (c) If our goal is to estimate the value of T to within 10% of its true value then the answers to (a) and (b) indicate that we are much more likely to do that if we use $X_{\max}$. However, $X_{\max}$ is biased. It will always provide an underestimate of the true value of T. Perhaps we could multiply it by some number greater than 1 and make it unbiased that way.

3. (a) Let X denote the number in the sample whose value is greater than m. X has a binomial distribution with $n = 9$ and $p = 0.5$. Thus $P[M > m] = P[X \geq 5] = 0.5$.

   (b) $P[M < m] = P[X \leq 4] = 0.5$.

   (c) No. It follows from (a) and (b) that the median of M is m but we cannot conclude from this that $E[M] = m$.

Sheet XI - Solutions

1. (a) $\bar{X}$ is normally distributed with mean equal to 450 seconds and sd equal to $\frac{40}{\sqrt{25}} = 8$ seconds. So $P[\bar{X} \le 428] = P[Z \le -2.75] = .5 - .4970 = 0.003$.
   (b) The probability of it happening by chance is so small (0.3%) that the alternative explanation that it happened because the students at King Henry have a tendency to do the one-mile run faster than students nationwide seems more plausible.

2. (a) Let X denote the number of patients selected that become paralysed, X has a binomial distribution with $n = 100$ and $p = 0.2$. We will use the normal approximation since $np(1-p) = 16 > 5$. So $P[X \le 10] \approx P[Z \le \frac{10.5-20}{\sqrt{16}}] = P[Z \le -2.375] = .5 - .4913 = 0.0087$.
   (b) The probability of this occurring by chance is so small (only 0.87%) that the alternative explanation that the patients that take vitamin supplements have a lesser tendency towards paralysis seems more plausible.

3. (a) $H_0 : \mu \le 80$; $H_a : \mu > 80$
   (b) $H_0 : \mu \ge 80$; $H_a : \mu < 80$

4. (a) $H_0 : \mu \le 100$; $H_a : \mu > 100$
   (b) $H_0 : \mu \ge 100$; $H_a : \mu < 100$

5. (a) $H_0 : \mu = 450$ versus $H_a : \mu < 450$ where μ is the average one-mile run time of King Henry's students.
   (b) 0.003

6. (a) $H_0 : p = .2$ versus $H_a : p < .2$ where p is the proportion of patients that are given vitamin supplements that become paralysed within 12 months.
   (b) 0.0087

7. (a) .0808; (b) .003; (c) .7157

8. (a) .1646; (b) .4594; (c) 0

9. (a) p-value = .0475. This is significant at the 5% level and H_0 would be rejected.
   (b) p-value = 2(.0475) = .095. This is not significant at the 5% level and H_0 would not be rejected.

10. .001, .021, .047

11. .001

12. 10%, 5%

13. none of them

14. .05 = 5%

15. (a) $\bar{X}$
   (b) $\bar{X} > 215$
   (c) The significance level of the test is the probability of landing in the rejection region assuming H_0 is true. Assuming H_0 is true then $\bar{X}$ is normally distributed with expected value equal to 200 and standard deviation equal to $40/\sqrt{10} = 12.65$. Therefore $\alpha = P[\bar{X} > 215] = .5 - .3830 = 11.7\%$.

16. (a) $\hat{p}$
   (b) $\hat{p} > .35$ or $\hat{p} < .25$
   (c) Assuming H_0 is true $\hat{p}$ is approximately normally distributed with expected value equal to 0.3 and sd equal to .032. Thus $\alpha = P[\hat{p} > .35$ or $\hat{p} < .25] = .1188 = 11.88\%$.

17. (a) $H_0 : \mu = 10.4$ versus $H_a : \mu > 10.4$ where μ is the mean number of miles driven by people in LA.
   (b) Any value of $\bar{X}$ that is greater than x where $\frac{x-10.4}{3.1/\sqrt{1000}} = 1.645$ i.e. $x = 10.56$.

18. (a) $H_0 : p = \frac{1}{5}$ versus $H_a : p > \frac{1}{5}$ where p is the probability that Deborah "senses" how many fingers Elizabeth is holding up.
   (b) Let X denote the number of times out of 100 trials that Deborah "senses" correctly. If H_0 is true then X has a binomial distribution with $n = 100$ and $p = \frac{1}{5}$. This is approximately normal (since $np(1-p) = 16 > 5$) with mean equal to $100(\frac{1}{5}) = 20$ and sd equal to $\sqrt{16} = 4$. Thus, for convincing evidence we need $X > x$ where $\frac{x-20}{4} = 1.282$ i.e. $X > 25.128$ i.e. $X \ge 26$.

19. (a) $H_0 : \mu = 5.3$ versus $H_a : \mu \ne 5.3$ where μ is the average diameter of the pistons being produced by the machine that day.
   (b) Type I error.
   (c) Type II error.

20. (a) $H_0 : \mu \ge 48$; $H_a : \mu < 48$.
   (b) .0367

(c) 10%, 8%, 5%

21. (a) $\bar{X} < 496.3$
    (b) $\bar{X} < 498.09$
    (c) $Z < -2.326$
    (d) $Z < -1.645$

22. (a) $\bar{X} < 30.45$ or $\bar{X} > 33.55$
    (b) $\bar{X} < 30.94$ or $\bar{X} > 33.06$
    (c) $Z < -1.75$ or $Z > 1.75$
    (d) $Z < -2.576$ or $Z > 2.576$

23. (a) $H_0 : \mu = 4.1$; $H_a : \mu \neq 4.1$
    (b) $Z = \frac{\bar{X}-4.1}{(.05/\sqrt{10})}$; rejection region is: $Z < -1.96$ or $Z > 1.96$ OR use $\bar{X}$; rejection region is: $\bar{X} > 4.131$ or $\bar{X} < 4.069$
    (c) If using Z: $Z = -1.9$ does not lie in rejection region. If using $\bar{X}$: $\bar{X} = 4.07$ does not lie in rejection region. Either way conclusion is that the machine is properly adjusted.

24. (a) $T = 2$
    (b) $T = 2.5$
    (c) Nathan has more compelling evidence (larger value of T). Notice that this is the case even though Nathan's $\bar{X}$-value is closer to 15.

25. (a) between 0.05 and 0.025
    (b) less than 0.005
    (c) between 0.025 and 0.01
    (d) .5 - .4066 = .0934 (Use normal tables since when $\nu = 149$ the t-distribution is essentially a standard normal distribution.

26. (a) between 0.05 and 0.02
    (b) less than 0.01;
    (c) between 0.02 and 0.01
    (d) 2(.5 - .4066) = .1868

27. (a) $H_0 : \mu \leq 80$; $H_a : \mu > 80$.
    (b) $T = 1.45$ so the p-value is between 0.1 and 0.05.
    (c) 10% only.

28. (a) $H_0 : \mu \geq 15$; $H_a : \mu < 15$
    (b) $T = -2$ so the p-value is between 0.05 and 0.025.
    (c) 10%, 8%, 5%

29. (a) $H_0 : \mu = 1.2$; $H_a : \mu \neq 1.2$
    (b) $T = -2.2$ so the p-value = 2(.5 - .4861) = .0278 (use normal tables since $\nu = 149 > 120$).
    (c) 10%, 5%

30. $H_0 : \mu \leq 15$; $H_a : \mu > 15$; $T = 2.72$; p-value is less than 0.005 ($\nu = 51$ is not in table so be conservative and use $\nu = 40$ to give you an upper estimate of the p-value). Reject H_0 at all the significance levels.

31. (a) $T > 1.833$
    (b) $T > 2.462$
    (c) $T > 1.333$

32. (a) $T > 2.201$ or $T < -2.201$
    (b) $T > 2.660$ or $T < -2.660$
    (c) $T > 1.729$ or $T < -1.729$

33. Bertha.

34. $H_0 : p \leq .1$; $H_a : p > .1$; $\hat{p} = .15$; $Z = 2.02$; p-value = 0.0217. Yes, reject H_0 at the 5% signficance level.

35. (a) $H_0 : p \leq .45$; $H_a : p > .45$; $\hat{p} = .56$; $Z = 1.11$; p-value = 0.1335
    (b) No.

36. (a) $\hat{p} > 0.681$
    (b) $\hat{p} > 0.629$
    (c) $Z > 1.41$
    (d) $Z > 1.282$

37. (a) $\hat{p} > .939$ or $\hat{p} < .861$
    (b) $\hat{p} > .919$ or $\hat{p} < .881$
    (c) $Z > 1.96$ or $Z < -1.96$
    (d) $Z > 1.645$ or $Z < -1.645$

38. Fred because he has the smaller p-value. The smaller the p-value the more compelling the evidence is for H_a.

39. (a) .0062; (b) .0013; (c) Bill has more compelling evidence.

40. (a) .0571; (b) smaller; (c) smaller; (d) .0571; (e) Jones.

41. The answer here is the p-value which is between 0.025 and 0.01.

42. On Tuesday $\bar{X}$ is closer to 18.2. Since the variances are the same on both days this means that Tuesday has less compelling evidence for H_a therefore the p-value is larger.

43. (a) (iii) p-value is less than 0.043
    (b) (ii) p-value is larger than 0.043
    (c) (ii) p-value is larger than 0.043

44. (a) $H_0 : p \geq .6$; $H_a : p < .6$
    (b) We are being asked for the p-value and that is 0.0036.

45. (a) The distribution of $\hat{p}$ when $n = 1000$ is much more focussed about the mean of 50% than when $n = 100$. Therefore the probability of getting a value that is as extreme or more extreme than 62% is much smaller when $n = 1000$ than when $n = 100$. Therefore Jane has the smaller p-value.
    (b) Yes, Jane can reject the null as well. If Dick's p-value is less than 5% then Jane's must also be.

46. (a) H_0 : coin is not weighted, i.e. $p = 0.5$; H_a : coin is weighted, i.e. $p \neq 0.5$.
    (b) 10%

47. If H_0 is true and our test leads us to incorrectly conclude that H_a is true then we would be exposing people to seizures and death. That is very serious, so we need to have a small probability of making such a mistake. Therefore we should choose $\alpha = 1\%$.

48. We would expect to get significant results in 5% of them i.e. in 5 of them.

49. (a) Yes. If H_0 were true then an event that had less than 5% chance of occurring actually occurred 5 times. If the studies were independent then the probability of this is less than $(0.05)^5 = 3.13 \times 10^{-7}$ - pretty convincing evidence for H_a!
    (b) Yes, this considerably weakens the drug company's case. If H_0 were true and 100 studies were done at the 5% level then we'd expect 5 of them to produce significant results in support of H_a simply by chance. Ignoring the other 95 studies is like taking a sample and throwing away any data that does not support H_a. If you did that then of course you would end up with a sample that produces significant results in support of H_a!

50. (a) Let's use $\hat{p}$ as the test statistic. Assuming H_0 is true, $\hat{p}$ has an approximately normal distribution with mean 0.8 and sd 0.02 and the rejection region is: $\hat{p} < 0.8 - 2.326(0.02) = 0.75$.
    (b) Now, in fact, the true value of p is 0.75 so the true distribution of $\hat{p}$ is approximately normal with mean 0.75 and sd 0.022. So $P[\text{reject null}] = P[\hat{p} < 0.75] = P[Z < 0] = .5$.
    (c) Now the true distribution of $\hat{p}$ is approximately normal with mean 0.6 and sd 0.024. So $P[\text{reject null}] = P[\hat{p} < 0.75] = P[Z < 6.3] \approx 1$

51. (a) If the true value of μ is 207 then the true distribution of $\bar{X}$ is normal with expected value equal to 207 and sd equal to $80/\sqrt{100} = 8$. Therefore $P[\bar{X} > 215] = P[Z > 1] = .1587$.
    (b) The distribution of $\bar{X}$ is now normal with expected value equal to 220 and sd equal to 8 as in (a). Thus $P[\bar{X} > 215] = P[Z > -.625] = .7357$.

52. (a) Let's use $\bar{X}$ as the test statistic. If H_0 is true then $\bar{X}$ has a normal distribution with mean 20 and sd $3/\sqrt{10} = 0.95$, so the rejection region is: $\bar{X} < -1.645(0.95) + 20 = 18.44$. Now, we are told that in fact the true value of μ is 19. Therefore the true distribution of $\bar{X}$ is normal with mean 19 and sd 0.95. Thus $P[\text{reject null}] = P[\bar{X} < 18.44] = .5 - .2224 = .2776$.
    (b) We'll use $\bar{X}$ as the test statistic again. This time, if H_0 is true then $\bar{X}$ has a normal distribution with mean 20 and sd $3/\sqrt{40} = 0.47$, so the rejection region is: $\bar{X} < 19.23$. Now, the true distribution of $\bar{X}$ is normal with mean 19 and sd 0.47, so $P[\text{reject null}] = P[\bar{X} < 19.23] = .5 + .1879 = .6879$.

53. (a) Let's use $\hat{p}$ as the test statistic. Assuming H_0 is true, the distribution of $\hat{p}$ is approximately normal with mean = 0.05 and sd = $\sqrt{\frac{(0.05)(0.95)}{200}} = 0.015$ so the rejection region is: $\hat{p} > 1.282(0.015) + 0.05 = 0.069$. Now, we know that the true value of p is 0.06, so the true distribution of $\hat{p}$ is approximately normal with mean 0.06 and sd = $\sqrt{\frac{(0.06)(0.94)}{200}} = 0.017$ so $P[\text{reject null}] = P[\hat{p} > 0.069] = .5 - .2019 = 0.2981$.

(b) Let's use $\hat{p}$ as the test statistic again. This time, assuming H_0 is true, the distribution of $\hat{p}$ is approximately normal with mean 0.05 and sd 0.011 so the rejection region is: $\hat{p} > 1.282(0.011) + 0.05 = 0.064$. Now, we know that the true value of p is 0.06, so the true distribution of $\hat{p}$ is approximately normal with mean 0.06 and sd 0.012, so $P[\text{reject null}] = P[\hat{p} < 0.064] = .5 - .1293 = 0.3707$.

54. Their offer is not selfless. If the manufacturer's brand does indeed last longer than 900 hours, then a test using a larger sample size is more likely to reveal this than a test using a smaller sample size.