School of Mathematical Sciences

INTRODUCTION TO MATHEMATICAL STATISTICS
Semester 1, 2004

OUTLINE FOR WEEKS 1 AND 2

PREPARATION:

Lecture Notes: You will be given a copy of summary lecture notes in the form of two overhead slides to a page. These notes can also be downloaded from MyUni or from my homepage `http://www.maths.adelaide.edu.au/people/psolomon/teaching.html`
I thoroughly recommend that you read the relevant sections of the textbook Wackerley, Mendenhall & Scheaffer (hereafter WMS) before the lecture, and come prepared to take additional notes. In the lectures, we will discuss the notes and work through a number of problems and examples.

Chapter 1 of WMS gives some basic ideas about statistics, and should be read carefully by students who have not done Statistical Practice I or an equivalent course. You should also come prepared to ask questions (if you have any!) about this material in the first Tutorial in Week 2.

Chapter 2 of WMS will be covered in the first two weeks. We will not cover WMS Section 2.6 on 'counting methods' in detail, but we will use similar ideas to develop probabilities directly when needed. Section 2.12 on Random Sampling will be covered later in Section 3 of our notes. We will also cover the idea of 'chance odds' in the lectures, which are not covered in WMS.

Lecture	Section	Overheads	Title	WMS	Pages
1	1	1-1:1-8	Introduction	1	
	2.1	2-1:2-5	Notation and Axioms	2.1-2.3	19-24
	2.2	2-6	Equally likely outcomes	2.4	25-37
2		2-7:2-9	Odds		not covered
	2.3	2-10:2-12	Interpretations of prob.		
	2.4	2-13:2-23	Conditional probability and independence	2.7-8	50-57
3	2.5	2-24:2-27	Named distributions		
	2.6	2-28:2-33	Sequences of events	2.9	59-65
4	2.7	2-34:2-43	Bayes' Rule	2.10	67-69

TUTORIAL 1/ASSIGNMENT 1:

You should attempt these questions prior to the Tutorial in Week 2, then use the Tutorial time to obtain clarification of any difficulties. Try to bring your lecture notes to the Tutorial.
The following questions provide a check of whether you understand the material presented so far and whether you can apply the ideas to new situations. Note

that the Tutorial sheets will often contain additional practice problems that you are expected to complete in your own time. If you would like further practice, attempt some of the odd-numbered exercises in WMS for which solutions are provided at the back of the book. Note too that there is a copy of the Solutions Manual to WMS held on Reserve in the Barr-Smith Library.

The multiple choice Questions 1 to 6 below are to be done in pairs or groups in the Tutorial. **The two questions, 9 and 10, marked with an asterisk * are part of Assignment 1 which is due by 10am on Friday 2 April, Week 5.** (Two additional assignment questions will be drawn from Tutorial 2 sheet which we will work on in Week 4.) You may, of course, seek advice about the assignment questions in the Tutorial. Note that some questions of the Tutorial questions introduce new ideas which we will develop more fully later in the course.

1. In a study of the effects of acid rain, a random sample of 100 trees from a particular forest are examined. Forty percent of these show some signs of damage. Which of the following statements are correct?

   (a) 40% is a parameter.

   (b) 40% is a statistic.

   (c) 40% of all trees in this forest show signs of damage.

   (d) More than 40% of the trees in this forest show some signs of damage.

   (e) Less than 40% of the trees in this forest show some signs of damage.

2. Refer to the previous question. Which of the following statements are correct?

   (a) The sampling distribution of the proportion of damaged trees is approximately normal.

   (b) If we took another random sample of trees, we would find that 40% of these would show some signs of damage.

   (c) If a sample of 1000 trees was examined, the variability of the sample proportion would be larger than for a sample of 100 trees.

   (d) This is a comparative experiment.

   (e) None of the above.

3. A randomly selected student is asked to respond 'yes', 'no' or 'maybe' to the question: 'Do you intend to vote in the next election for president of the student union?' The sample space is $\{\text{yes}, \text{no}, \text{maybe}\}$. Which of the following represent a legitimate assignment of probabilities to this sample space?

   (a) .4, .4, .2

   (b) .4, .6, .4

   (c) .3, .3, .3

   (d) .5, .3, -.2

   (e) none of the above.

4. In a population of students, the number of calculators owned, X, has the following distribution: $P(X = 0) = 0.2$, $P(X = 1) = 0.6$ and $P(X = 2) = 0.2$. The mean of this distribution is

   (a) 0
   (b) 2
   (c) 1
   (d) 0.5
   (e) the answer cannot be computed from the information given.

5. The Central Limit Theorem states that

   (a) The sample mean is unbiased.
   (b) The sample mean is approximately normal.
   (c) The binomial distribution is skewed.
   (d) The sample standard deviation is approximately normal.
   (e) None of the above.

6. You have measured the systolic blood pressure of a random sample of 25 employees of a company located near you. A 95% confidence interval for the mean systolic blood pressure for the employees of this company is $(122, 138)$. Which of the following statements gives a valid interpretation of this interval?

   (a) 95% of the sample of employees have a systolic blood pressure between 122 and 138.
   (b) 95% of the population of employees have a systolic blood pressure between 122 and 138.
   (c) If the procedure were repeated many times, 95% of the resulting confidence intervals would contain the population mean systolic blood pressure.
   (d) The probability that the population mean blood pressure is between 122 and 138 is 0.95.
   (e) If the procedure were repeated many times, 95% of the sample means would be between 122 and 138.

7. **Probability questions** (You may like to do this in pairs or groups in the Tutorial.)

   (a) A fair coin is tossed four times. Write down all possible sequences of heads and tails. What is the probability of each sequence? Find the probability that two heads occur in a row. Use the basic rules of probability to justify your answers.
   (b) Now determine how many outcomes lead to each possible number of heads, and hence obtain the distribution of Y, the number of heads. Can you name the probability distribution of Y? What is the 'expected number' of heads?
   (c) Get each person in your group to toss a coin four times and count the number of heads. Collate the numbers for the class on the board and see how close these come to the theoretical probabilities.

(d) Two people compare birthdays. What is the probability that they have a common birthday (assume 365 days in the year)?

(e) If there are three people, what is the probability that they have a birthday in common? (Note it is easier to work out the probability that they are all different.)

8. A person is given a random breath test. If their blood alcohol level is over the legal limit, there is a 95% chance that the test will in fact show positive. If the person's blood level is under the legal limit, there is only a 2% chance that the breath test will show positive.

   (a) In 2000, 4% of people tested were over the legal limit. What is the probability that a person is over the limit and provides a positive test?

   (b) What then is the probability that a person who shows positive is actually over the legal limit in 2000?

   (c) What is the probability that the test will give a 'correct' result?

   (d) Campaigns to reduce drunk driving have been so successful that nowadays we may assume that only 1% of people tested are over the legal limit. Find the probability that a person who shows positive is over the legal limit under this assumption.

   (e) (Optional: harder.) During a particular long weekend, 1000 people are tested. If all of these 1000 are under the legal limit, what is the distribution of the number of them who show positive? Provide an estimate for the number who will show positive, and find the probability that at most 10 will give a positive test.

   Justify any approximations you use in determining this.

9. * Suppose that the probability of exposure to 'flu during an epidemic is 0.6. Experience has shown that a serum is 80% successful in preventing an inoculated person from acquiring the 'flu if exposed to it. A person not inoculated faces a probability of 0.90 of acquiring the 'flu if exposed to it.

   Two people, one inoculated and one not, perform a highly specialised task in a business. Assume that they are not in the same location, are not in contact with the same people, and cannot expose each other to the 'flu. What is the probability that at least one of them will get the 'flu?

10. * Assume the random variable X has a geometric distribution with probability p and probability function $P(X = k) = (1 - p)^{k-1}p$, for $k = 1, 2, \ldots$.

    (a) Show that $P(X \leq k) = 1 - (1 - p)^k$.

    (b) Let k_1 and k_2 be two positive integers and suppose that X takes a value greater than k_1. Show that

    $$P(X > k_1 + k_2 | X > k_1) = P(X > k_2).$$

    This is called the **memoryless** property of the geometric distribution.

    [Hint: the intersection of the events $X > k_1$ and $X > k_1 + k_2$ is the event that $X > k_1 + k_2$.]

Associate Professor Patty Solomon