STAT 270: Introduction to Probability and Statistics

Introduction to **Surrey Version** Spring 2007

Instructor: Larry Weldon, weldon@sfu.ca

Surrey Campus contact info:
Office Hours at Surrey Campus M 1100-1200, F 1030-1130 Rm 4366
Burnaby Campus contact info: 604-291-3667
Office Hours at Burnaby Campus T 1130-1220, R 1330-1420. Rm K 10565

Course Schedule:
Lectures MWF 1230-1320 in Rm 3350
Open Tutorial Workshop MWF
(MW 1130-1330 Rm 3130 and F 930-1130 Rm 3100)

Course Text: Devore, Jay L. (2004) Probability and Statistics: for engineering and the sciences. Sixth Edition. Brooks/Cole, Belmont, CA.

Evaluation:	Final Exam	40% (or 60% without midterm, if to student's advantage)
	Midterm	20%
	Term Work	40%

Midterm, and Quizzes: The midterm will be at the end of week 8, Friday, March 2. It will be based on all material covered by then, including Ch 1-5 of the text. On Friday of other weeks, there will usually be a quiz, an assignment, or an oral Q&A in class. These will tend to involve the material just covered. Note that questions on mid-term tests and the final examination will seem more challenging since they involve a wide range of material rather than the material recently discussed in class. I will provide more guidance about this in the course.

Course Content:

The course is intended to provide a basis for future study of statistics as well as stand-alone introduction to the subject. This dual purpose requires some compromise – the theoretical basics must be covered as prerequisite to future stats courses, and this limits the breadth of the subject than can be demonstrated in this course. Nevertheless, the text does cover the most basic theory, and the examples and exercises in the text and assignments will suggest considerable breadth.

The text is used for both STAT 270 and STAT 285. In STAT 270 we will cover Chapters 1-8 and most of Ch 9. We will follow the text quite closely. Note that some of the theory is introduced via problems and applications, both in the text and in lectures; this is part of the core material you need to study.
What follows is an overview guide to these 9 chapters:

Overview Guide to the Chapters Covered in STAT 270

Ch 1: Overview and Descriptive Statistics: In order for data to provide useful information, we need to consider how the data was collected, how it should be summarized, and what the summary reveals. In this chapter, we focus mainly on the data summary aspect, which we call "descriptive statistics". The other parts will be covered in later chapters.

Ch 2: *Probability*: Data collection often involves some random selection of objects to measure, as well as random errors of measurement. To relate this data to the population that was its source, we need to make allowance for the randomness of the selection and the randomness of the measurement itself. It turns out that the best we can do is to describe the information inferred from data using probabilities. This chapter introduces probability itself as well as the most useful models for relating probabilities to data.

Ch 3: *Discrete* Random Variables and Probability Distributions: Some phenomena (like the heads count of coin tosses) are described by discrete variables, and some (like the time it takes a sprinter to cover 100m) are described by continuous variables. The way the probabilities are described looks quite different for these two types of variables, and to avoid confusion, as well as to emphasize this theoretical difference, we separate the introduction of models of the two types into two chapters. In this chapter we discuss the models for probabilities associated with discrete measurements. We will see that there are many connections among the models for discrete random variables, and also between the discrete and continuous random variables. To understand these connections deepens your understanding of the models, and prepares you for wise application of them. Probability models are useful both for summarizing data and for predicting precision.

Ch 4: *Continuous* Random Variables and Probability Distributions: As will be explained in this chapter, the description of continuous random variables requires calculus (derivatives and integrals), whereas for discrete random variables differences and sums were enough. The famous "Normal" random variable is continuous. Its slightly complex description is compensated by its widespread applicability, since it tends to pop up in any description of sums or averages. It turns out that even discrete data analysis tends to involve the Normal distribution, even though the Normal random variable is continuous.

Ch 5: *Joint* Probability Distributions and Random Samples: We first discuss the connection between random sampling and measurement with error. Once this is explained, we can begin to consider more realistic inference problems. When items are measured, it is seldom sufficient to take only one measurement per item. Often the interest is in the relationship among the different measurements, and so we need to construct descriptions of multiple measurements jointly. A useful tool in this construction is the idea of *conditioning*: what happens to one variable in situations where we constrain the other variable to a certain value. When conditioning does not produce anything interesting, we describe the variables as *independent* (more explanation to come!) It turns out that "independence" is actually very interesting ...

Ch 6: Point Estimation: The most common approach to statistical inference (from sample to population) in sampling contexts is to decide on an appropriate class of parametric models for the variability in the data, then estimate the parameters in the model, and finally to express what these estimates imply about the underlying population. An important part of this process is this: for a sample (X_1, X_2, ..., X_3), and a model for the frequency distribution $F(x|\theta)$, how do we use the X's to estimate θ? This is the focus of this chapter.

Ch 7: Interval Estimation: Because statisticians are experts in coping with unexplained variability (or "random" variability), for every point estimate there is an even more informative interval estimate (known as a "confidence interval"), an interval likely to include the true parameter value. There are some exact methods for this and some very useful approximations.

Ch 8: Hypothesis Testing: Often when data is collected, our purpose is to see if the result is what was expected, or is it surprising? In the latter case, we may need to modify our understanding since it apparently led us to a false expectation. Of course, the difficulty in this process is to be able to distinguish unexplained variation with variation caused by some identifiable factor.

Ch 9: Inference Based on Two Samples: The techniques of chapters 7 and 8 are extended to enable comparison of two samples.

Addendum: In addition to these topics, I will introduce you to the following key concepts and tools of statistics, even though we will not be able to cover them in depth: simulation, time series, nonparametric smoothing, analysis of variance. regression analysis, and quality control. If you pace yourself throughout the course, you will be able to stay on top of the material. As the material is to some extent cumulative, it is a good idea to keep up. If you plan on reading the theory of one chapter each week, that will allow a little time for the supplementary applications material, and studying for tests and the final exam.

Statistical Software: An introduction to statistics without statistical software is like an introduction to cycling without a chance ride a bicycle. You can use any software you like in this course, but I will use MINITAB and R to demonstrate various concepts and tools. MINITAB is available on the SFU servers (both Surrey and Burnaby). R is free to download and is very powerful, but a bit harder to learn. My intention is to have you use statistical software without needing to learn to write programs. The small amount of time needed to learn to use statistical software will more than be offset by the time saving of the calculations it facilitates. Some assignments will require use of statistical software. However, the tests and exams will not use statistical software, but will assume a familiarity with what it can do.

(KLW 07/01/06)