# ECONOMICS 321
# INTRODUCTION TO STATISTICS

SPRING SEMESTER 2006
BILGER 150
TR 12:00-1:15

JAMES MONCUR
SAUNDERS HALL RM 530
TEL: 956-7143 OR 956-7847
EMAIL: jmoncur@hawaii.edu
OFFICE HOURS: TR 11:00 - 12:00
AND BY APPOINTMENT

Text: **David S. Moore,** The Basic Practice of Statistics **(4th ed.)** New York: W.H. Freeman & Company, 2007.

The great 19th century Prime Minister of Great Britain, Benjamin Disraeli, once claimed that "There are three kinds of lies: lies, damned lies and statistics." This course aims to inform you somewhat on the latter, without reference to the former.

You may or may not come away from the course with a love of statistics and may or may not wish to make statistics a central part of your professional life. But you can hardly avoid some contact with the subject. Managers and consultants encounter statistical procedures daily, and recent emphases on quality control have even brought statistics to the production line. As consumers, you may have occasion to look up a Consumer's Report evaluation of a major purchase--and to wonder about the validity of their ratings. And as citizens, we face a daily barrage of polls, politicians, bureaucrats and other nonsense deserving healthy skepticism.

Most of you will come to agree that statistics is not the easiest subject to learn. To help you gain an intuitive understanding of the subject, a computer is almost indispensable. Although many assignments can be done with a moderately sophisticated calculator, some will require you to use spreadsheets such as Excel or Quattro Pro or Lotus 1-2-3. On these occasions, I expect the computer output to be submitted as an appendix to your written description of the problem and display of the computer output. Computer output without comment will be discounted heavily.

## GRADING

Your grade for this course will depend on homework exercises, two mid-term exams (100 points each) and a comprehensive final (200 points). Problem sets will be assigned each day, and will be due at the beginning of class on the due date. We will collect and grade six of these homework assignments, chosen arbitrarily and not announced beforehand. At a maximum of 20 points each, the best five of the six graded papers will constitute 100 points in your semester grade. In order to be fair to all class members, **I will not accept papers submitted after class time on the due date**.

Your final grade will thus be based on a potential 500 points.

## AN OUTLINE OF THE SEMESTER'S WORK

There are two branches of the study of statistics: descriptive and inferential. We begin with descriptive statistics, concerned with summarizing the information contained in a body of data. Data can be described with graphs, or in numerical form–for example, the mean $\bar{X} = \Sigma X_i / n$ measures the central tendency of the individual $X_i$ values. Some types of data are best described by correlation or regression concepts. We also consider measures of variability and other properties of the data. This material, covered in chapters 1 - 7 of our text, can be messy but is conceptually straight forward. Next we consider the production of data and concepts of probability (chapters 8 - 13) as represented by the normal,

binomial and several other specific distributions, and segueing into sampling distributions, which form a bridge between descriptive and inferential statistics. Whereas the initial study of probability distributions refers to individual variables, sampling distributions refer to the probability distribution of statistics such as the sample mean $\bar{X}$ .

Inferential statistics makes up most of the remainder of the course, 14 - 28. With this material fully absorbed, we move on to inferential statistics, the point of which is to infer something about, say, the population mean $\mu$ (which we rarely if ever know) from information contained in the sample mean $\bar{X}$ . The point here is to be able to make a statement about the likelihood that a particular sample mean truly represents the mean of the population from which we think the sample is drawn. The basic concept is applied to means, variances, differences between means and variances, and regression parameters.

For economists, the highlight of the course is the material on regression. For example, standard economics implies that the quantity demanded of, say, aloha shirts depends on the price p of such shirts, prices of other kinds of shirts, of related clothing, of Hawai'i vacations, and other factors $p_1, p_2, \ldots, p_m$ , and on the consumer's income $y$. One can express this general relationship in an equation like $q = f(p, p_1, p_2, \ldots, p_m, y)$ . Assuming the relationship to be linear, we can write $q = \beta_0 + \beta_p p + \beta_1 p_1 + \cdots + \beta_m p_m + \beta_y y$ . According to the theory of demand, the higher the pric lower will be the quantity bought. Hence estimates of this relationship should find $\beta_p < 0$ so that the demand curve slopes downward, other things the same.

Our first problem here is to estimate the $\beta's$ . One collects sample data $q, p, p_1, \ldots, p_m, y$ $\beta$ . The estimates are purely descriptive. Once the $\beta's$ are estimated, we must then try to infer something about the significance of the estimates: how accurate are they? Without complete data on the entire population of aloha shirt consumers, we will never know for sure whether $\beta_p < 0$ , as demand theory implies it will be, or $\beta_p \geq 0$ . If the latter, the price of aloha shirts has no bearing on the quantity purchased, or is positively related to the quantity purchased, and demand theory is refuted for aloha shirts. So we approach this inference by asking, what is the probability that $\beta_p$ differs from zero?

One can pose similar hypotheses about all the $\beta's$ , of course. Using these concepts, we arrive at inferences concerning the truth or falsity of propositions derived from economic theory or from past history or conventional wisdom or wherever. These concepts are the source of all the fun from statistics and economics.

## Schedule of Topics

Be prepared for scheduling adjustments, to be announced in class. Read each day's assignment before class and be prepared with questions and discussion.

| Week | Assignment | Subject |
|---|---|---|
| 1 | Course introduction; Chapters 1 and 2 | Exploring and describing data |
| 2 | Chapter 3 | Normal distribution(s) |
| 3 | Chapters 4 and 5 | Correlation and regression |
| 4 | Chapters 5 | Regression |
| 5 | Chapters 8 and 9 | Sampling & experiments |
| 6 | Review of Chapters 1 to 9, EXAM I | |
| 7 | Chapters 10 and 11 | Probability concepts & sampling distributions |
| 8 | Chapter 12 | Rules of probability |
| 9 | Chapters 14 and 15 | Interval estimates (confidence itnervals) |
| 10 | Chapters 15 and 16 | Tests of significance: derivation and use |
| 11 | Review of Chapters 10 to 16, EXAM II | |
| 12 | Chapters 18 and 19 | Inference: population mean and standard deviation |
| 13 | Chapters 20 and 21 | Inference: population proportions |
| 14 | Chapter 23 | Chi-square tests; inference for regression |
| 15 | Chapter 28 | Multiple regression |
| | Comprehensive Final Exam | (Tuesday May 8, 12:00-2:00) |