Concordia University

Department of Economics

ECONOMICS 221: Statistical Methods I

Fall Term 1999

Section: Class Hours:
Name: Office Hours:
Office Address: Telephone:

General Description

The general purpose of the course is to introduce students of economics to the theory of probability and statistical methods and how these are applied to economic models. The principal aim is to study statistical inference based upon information collected from a suitably chosen sample from a (larger) population: given information in the sample, what can be inferred about the population? Probability and statistics are particularly useful in applied economics where conclusions must be reached on the basis of incomplete information.

It is necessary to understand the theoretical basis of a statistical method in order to understand its limitations, when it may be applied and how it may be interpreted. Nonetheless, the central direction of the course is toward the application of theory to practical situations arising in economics.

The material of the course is required for Econ 222 and the subsequent study of econometrics.

Textbook

The text for the course is D.L. Harnett and J.L. Murphy, *Statistical Analysis for Business and Economics, First Canadian Edition*, 1993, Don Mills, Ontario: Addison-Wesley, Chapter 1-7, 9 and 10.

Assignments and Required Reading

One aspect of a statistical training in economics is that students must learn how to solve problems and how to present solutions to problems. Practice is the only effective means to achieve these ends. For this reason, various problems will be assigned each week and students must hand in their solutions for correction and marking, before a deadline set by the instructor.

Another aspect of a statistical training involves writing: writing clearly, in well-constructed sentences, is a necessary part of any worthwhile statistical report. Writing about statistics is also a

means to broaden a student's understanding of probabilistic reasoning and may even brighten the otherwise technical labour of 221! All sections of 221 are therefore required to read the book by Peter L. Bernstein, *Against the Gods*, New York: Wiley, 1998 and to write a 3-page report on a selected part of it. 15% of the final grade is based on this report: due 99-11-30.

Grading

Students will be awarded a final grade on the basis of the following distribution:

Assignments	20%
Report on Required Reading	15%
Common Mid-term Test (Week #7)	20%
Common Final examination	45%

The mid-term test will take place on during Week #7 and will cover all class material up to the end of Week #5. The common final examination will cover the whole course, but relatively more weight will be given to topics covered after the mid-term test.

Weekly Calendar

Week	*From*	*To*
1	Tuesday 07 September	Monday 13 September
2	Tuesday 14 September	Monday 20 September
3*	Tuesday 21 September	Monday 27 September
4	Tuesday 28 September	Monday 04 October
5	Tuesday 05 October	Monday 11 October
6**	Tuesday 12 October	Monday 18 October
7**	Tuesday 19 October	Monday 25 October
8	Tuesday 26 October	Monday 01 November
9	Tuesday 02 November	Monday 08 November
10	Tuesday 09 November	Monday 15 November
11	Tuesday 16 November	Monday 22 November
12	Tuesday 23 November	Monday 29 November
13**	Tuesday 30 November	Monday 06 December

Week # 7 is Mid-term Test Week: classes for revision and review only.
*Assignments begin during week # 3.
**No assignments in weeks #'s 6, 7 and 13. There will be 8 assignments in total, including the Book Review. Thus 7 assignments will take form of statistical exercises. The Book Review must be submitted on or before November 30th, 1999. There will also be two computer tasks: to establish a data bank and the computation of means and variances.

Detailed Weekly Schedule

Week 1:

a) Introduction and class organization (see *ECONOMICS 221: Statistical Methods I*).

b) A general description of Helliwell (1996), noting the variables, errors and direction of causality, as an example of a linear economic model (overhead transparency).

c) Introduction to the data of an economic model: data are arranged in columns, the constant has an artificial variable which is either zero (US state) or one (Canadian province). The column arrangement is important for computation.

d) Mathematical review: $e'e = N$, $e'x = \acute{O}x_i$, $e'y = \acute{O}y_i$, μ_x, μ_y; $e'(x - \mu_x) = e'(y - \mu_y) = 0$: sum of deviations about the mean is zero, always. TASK: set up computer account and transfer data to it. READING: text chapter 1.

Week 2:

a) Econometric model building: Helliwell (1996) used to amplify the discussion of week 1 b) and c) using week 1 d). Model simplification to $y_i = \mu + \mathring{a}_i$ (i = 1,2,...,N) or $y = e\mu + \mathring{a}$ using $e'(x - \mu_x) = 0$ (handout: *Mathematical Review for Econ 221*).

b) Further discussion of *Mathematical Review*.

c) Discussion of data for economic models: frequency distributions, relative frequency distributions. Discrete and continuous data. Short discussion of sources of data.

d) Further discussion of models: populations, parameters, samples and estimates. Example of mean and variance. Statistical inference. Other characteristics of distributions. TASK: from a given population X, find μ_x and $\acute{o}_x^2$. Then draw k random samples and calculated $\overline{x}$ and s_x^2 for each sample, and hence find the distribution of $\overline{x}$ and s_x^2.

Week 3: Probability: errors in economic data and models (see: *Models of Errors*). Probability and relative frequency. A priori probability. Example: tossing two dice. Sample space (S) and events (A,B,C), complements of events (A^c, B^c, C^c). Sets and events, disjointness. Axioms, $P(A|S)$, $P(A \cup B|S)$, $P(A \cap B|S) = P(B|S)\ P(A|B \cap S) = P(A|B \cap S) = P(A|S)P(B|A \cap S)$, the reversal rule and Bayes' theorem, independence. ASSIGNMENT 1: exercises in probability READING: text chapter 3.

Week 4:

a) Probability continued: counting rules, discussion of mutual exclusivity and independence and of Bayes' theorem with examples. Bivariate probability, marginal and conditional probability.

b) Random variables: discrete random variables and expectations. Rules for linear transformations. Bivariate probability and covariance. Covariance and

independence. ASSIGNMENT 2: further exercises in probability and introduction to expectations. READING: text chapter 4.

Week 5: Continuation of Week 4 b) and introduction to discrete probability distributions: binomial distribution, mean and variance mean and variance of proportions. Hypergeometric and Poisson distributions via sampling with and without replacement and $n \to \infty$, $ð \to 0$ such that $nð = m$, a constant. ASSIGNMENT 3: further exercises in expectations and exercises on the binomial distribution and the use of binomial tables. READING: text chapter 5, except 5.4 and 5.9.

Week 6: Continuous random variables: the normal distribution, the standardized normal distribution, the binomial distribution as n gets larger, normal approximation and continuity correction for the binomial distribution. READING: text chapter 6 except 6.7 , 6.8 and 6.9.

Week 7: MID-TERM TEST WEEK

Week 8: Sampling distributions of a normal population: $\bar{x}$ and s_x^2 and their sampling distributions. The $÷^{-2}$, F- and t- distributions (handout: *Sampling Distributions*). ASSIGNMENT 4. READING: text chapters 7.1 - 7.5, 7.8 and 7.9.

Week 9: a) Sampling distributions from an unknown population distribution: $X \sim [\mu, ó^2] \Rightarrow \bar{X}_n \sim [\mu, ó^2/n] \Rightarrow$ (via CLT) $\{\sqrt{n}\,(\bar{X} - \mu)/ó\}$ trends in distribution to N(0,1). Finite population correction factor. Compute demonstration of CLT. READING: text chapters 7.6 and 7.6.

b) Introduction to estimation: properties of good estimators and condition on linear estimators for unbiasedness; BLUE (minimum variance unbiased linear estimators). ASSIGNMENT 5. READING: text chapters 9.1 - 9.5.

Week 10: Interval estimation: the interpretation of confidence intervals, confidence intervals for μ_x from $N(\mu_x, ó_x^2)$ when $ó_x^2$ is known and when $ó_x^2$ is unknown. Determining the size of a sample, level of confidence. ASSIGNMENT 6. READING: text chapters 9.6 - 9.9.

Week 11: Interval estimation: confidence intervals: ð in B(ð,n), $ó^2$ in $N(\mu, ó^2)$. ASSIGNMENT 7. READING: text chapter 10.

Week 12: Hypothesis testing: Tests on μ, ð and $ó^2$.

Week 13: Review. BOOK REVIEW: due 17:00 hrs, Monday 99-11-30.

SIMPLIFIED ECONOMIC MODELS:

ECON 221 and ECON 222

- The model:

$$(1) \quad y_i = \alpha_o + \alpha_1 x_i + \alpha_2 w_i + \alpha_3 z_i + \alpha_4 m_i + \varepsilon_i$$

$$i = 1,2,.....n$$

is the most advanced model of ECON 222. In (1) the average value of the ε_i in the population is assumed to be zero.

- A simplified form of the model (1) is

$$(2) \quad y_i = \alpha + \beta x_i + \varepsilon_i$$

$$= \alpha 1 + \beta x_i + \varepsilon_i , i = 1,2,...,n.$$

- Again the average value of the ε_i in the population is zero Model (2) may be simplified further:

$$y_i = \alpha + \beta\mu_x + \beta(x_i - \mu_x) + \varepsilon_i$$

$$(3) \quad y_i = \mu_y + \eta_i ,$$

$$= \mu_y 1 + \eta_i , i = 1,2,...,n$$

where $\mu_y = \alpha + \beta\mu_x$ and $\eta_i = \beta(x_i - \mu_x) + \varepsilon_i$ which has a population average of zero. Model (3) is the central model of Econ 221.

AN ECONOMIC MODEL

J.F. Helliwell (1996), "Do National Borders Matter for Quebec's Trade?"
Canadian Journal of Economics, XXIV, 507-522

$$\log S_{ij} = á_o + á_1 \log GDP_i + á_2 \log GDP_j + á_3 \log (\text{distance i to j})$$

$$+ á_4 (\text{Interprovincial}) + å_{ij}$$

S_{ij} = value of shipments from exporting region i to importing region j in millions real $CDN

GDP = real GDP

Distances are measured in miles; logs are natural logarithms; a circumflex denotes estimate of what lies beneath it.

For Quebec 1990 (number of observations = 78)

$\hat{a}_o$: -6.70 (±1.5952)

$\hat{a}_1$: 1.15 (±0.0827)

$\hat{a}_2$: 1.06 (±0.0828)

$\hat{a}_3$: -1.08 (±0.0947)

$\hat{a}_4$: 3.31 (±0.2043)

84.7% of the total variation in $\log S_{ij}$ is explained by the corresponding variation in:

$\hat{a} + \hat{a}_1 \log GDP_i + \hat{a}_2 \log GDP_j + \hat{a}_3 \log$ (distance i to j) + $\hat{a}_4$ (Interprovincial), leaving 15.3% of the total variation and in $\log S_{ij}$ left unexplained and attributable to $\hat{e}_{ij}$.

The border effect = 27.4 = antilog 3.31.

Using annual data for real personal consumption and real disposable income for Canada from the period 1971 to 1996, and considering the linear regression model of the form

(3) $C_i = á + â Y_i + å_i,$

the following results are obtained:

(4) $C_i = 162{,}270 + 0.3865 Y_i + å_i.$

(4,022.62) (0.0122)

In (4) Y_i explains 97.65% of the total variation in C_i.

Standard errors are in parenthesis beneath the estimates $\hat{a}$ and $\hat{b}$.