The University of Washington
Center for Statistics and the Social Sciences

C&SSS 594

"Distributional Methods with Application to the Measurement of Inequality"

Winter 2003

Outline

Meeting: Tuesday and Thursday, 1:30 pm to 3:00 pm

Location: 255 Mary Gates Hall

Professor:

Mark S. Handcock, C014B Padelford Hall, 221-6930

Office Hours:

Mondays: 2:00pm - 3:00pm
Wednesday: 2:00pm - 3:00pm

Other times by arrangement. Clearly composed questions sent to the address "handcock@u" will receive written replies

Motivation and Synopsis

This course provides an introduction to modern statistical methods for comparing distributions. Social science research relies on these methods any time comparisons are made between groups. When the attribute of interest is continuous, for example racial differences in life expectancy, or earnings differences between men and women, the traditional methods make comparisons in terms of means, medians and standard deviations. Traditional methods, however, provide a weak and unnecessarily restrictive framework for comparison. Consider the earnings distribution in the United States. Over the past 30 years, median real earnings have declined by about 10and the variance in earnings has risen dramatically. Hidden behind these summary statistics are a range of important questions. Have the upper and lower tails of the earnings distribution grown at the same rate? Can we determine the role played by the decade-long freeze in the minimum wage? Is there anything more to the narrowing of the gender wage gap than the convergence in median earnings between the two groups? The information we need to answer these questions is there in the data, but inaccessible using traditional statistical methods such as regression and Gini index summaries.

With the emergence of Exploratory Data Analysis (EDA, Chambers, *et. al.* 1983; Tukey 1977) and the development of high speed computing and graphical user interfaces, there has been a movement towards more *nonparametric* and *distribution-oriented* analytic methods. A prominent feature of these methods is the use of *graphical displays.* Graphics exploit the power of our visual senses to convey information in a direct way.

Objectives of the Course

In this course we will start from scratch and introduce practical *nonparametric, distribution-oriented* and *graphical* analytic methodological tools to aid social science research.

We will follow the topics of traditional methods courses: univariate and multivariate summaries; simple and multivariate regression. These will be supplemented by quantile regression, methods for categorical data and an overall emphasis on *distributional comparisons*.

These methods aim to bridge the gap between exploratory tools and parametric restrictions. The goal is to present the concepts, theory and practical aspects of the methods in a coherent fashion, with a minimum of statistical prerequisites.

The course will have an applied focus on the development of tools for research in the social sciences. The course will involve the practical application of the ideas and their implementation through statistical software to make them accessible to social scientists.

This course is part of the curriculum of the Center for Statistics and the Social Sciences (CSSS), with funding from the University Initiatives Fund. The CSSS is includes faculty members from the Department of Statistics and a broad-range of social science disciplines including Anthropology, Economics, Geography, Political Science, and Sociology. This curriculum is been developed to complement and strengthen the quantitative methods course offerings for social science students at both the undergraduate and graduate levels.

Syllabus of the Course

The syllabus of the course will develop on the following weekly schedule. The some later topics may not be reached and we will make choice among them toward the end of the quarter.

1. Introduction and Motivation

1.1 Motivation
1.2 Principles of comparison
1.3 Description and summarization
1.4 Graphical displays
1.5 Numerical summary measures
1.6 Limitations
1.7 Organization of rest of the course

2. The Relative Distribution

2.1 Basic distributional concepts
2.2 The relative distribution
2.3 Using a known reference distribution
2.4 History and literature

3. Location, Scale and Shape Decomposition

3.1 Decomposing the relative distribution
3.2 Further decomposition of shape
Exercises

4. Application: White Men's Earnings 1967–1997

4.1 Background
4.2 Data
4.3 Findings
4.4 Discussion

5. Summary Measures

5.1 Motivation
5.2 Measuring distributional divergence
5.3 Two measures of distributional divergence
5.4 Effect summary statistics
5.5 Measures motivated by hypothesis testing
5.6 Measuring distributional polarization

6. Application: Earnings by Race and Sex: 1967–1997

6.1 Background
6.2 Data
6.3 Findings
6.4 Discussion

7. Adjustment for Covariates

7.1 Compositional adjustment
7.2 Comparison of composition-adjusted distributions
7.3 Further decomposition by location/shape
7.4 Adjusting for multiple covariates
7.5 Categorical contrasts

8. Application: Comparing Wage Mobility in Two Eras

8.1 Background
8.2 Data
8.3 Findings

9. Inference for the Relative Distribution

9.1 Estimation when the reference distribution is known
9.2 Estimation when both distributions are unknown
9.3 Estimation for a pooled reference group
9.4 Estimation when the data are censored
9.5 Estimation when the data are weighted
9.6 Confidence intervals and confidence bands

10. Inference for Summary Measures

10.1 Inference for two measures of distributional difference
10.2 Measures motivated by hypothesis testing
10.3 Inference for the median relative polarization
10.4 Computing standard errors
10.5 Statistical properties of estimates of the upper and lower indices
10.6 Tests of significance and multiple comparisons
10.7 Bootstrap confidence intervals and achieved significance level

11. The Relative Distribution for Discrete Data

11.1 The discrete relative distribution
11.2 Application: men's and women's hours worked
11.3 Inference when the reference distribution is known
11.4 Inference for the discrete relative distribution
11.5 Grouped data
11.6 Inference for the relative polarization indices

12. Application: Changes in the Distribution of Hours Worked

12.1 Background
12.2 Data
12.3 Findings
12.4 Discussion

13. Quantile Regression

13.1 Estimation of quantiles
13.2 Motivation for quantile regression
13.3 Linear quantile regression
13.4 Nonparametric quantile regression

Structure of the Course

There will be two lectures and one "laboratory" session per week.

The objective of the laboratory is to help students develop the skills necessary to analyze real data using these methods and hence learn the underlying statistical concepts.

This course is part of the curriculum of the Center for Statistics and the Social Sciences (CSSS) (`http://www.csss.washington`), with funding from the University Initiatives Fund. The CSSS is includes faculty members from the Department of Statistics and a broad-range of social science disciplines including Anthropology, Economics, Geography, Political Science, and Sociology. This curriculum is been developed to complement and strengthen the quantitative methods course offerings for social science students at both the undergraduate and graduate levels.

Reading list

[**RD**] Handcock, Mark S. and Morris, Martina (1999)
Relative Distribution Methods in the Social Sciences, Springer–Verlag: NY
Required.

[**SM**] Simonoff, Jeffrey S. (1996)
Smoothing Methods in Statistics, Springer–Verlag: NY
Recommended.

[**VR**] Venables, William N., and Ripley, Brian D. (1999)
Modern Applied Statistics with S-Plus. 3rd. Ed. Springer–Verlag: New York.
Recommended.

Intranet Resources

The course will make extensive use of the web as a source of statistical information and for the course materials. Much of the conduct of the class will take place via the web.

The home page for this class is: `http://www.stat.washington.edu/handcock/594`

It will contain lecture notes, homework, solutions and course information. Please consult it regularly.

Computer Usage and Software

The computer is the scientific laboratory of the applied researcher in quantitative fields. As such this course requires the student to develop a degree of comfort and competence "in the lab".

Questions and problems with computing in the course should be directed to me.