

SURV699Q/SURVMETH 895

Prediction Theory

Days: Thurs, September 8-December 15, 2011
13 sessions (no class November 24)

Time: 3:30pm- 6:10pm
(10 minute break at approx. 4:45pm)

Place: JPSM	MPSM
1218 Lefrak Hall	300 Perry
University of Maryland	University of Michigan
College Park MD	Ann Arbor MI

Text: *Finite Population Sampling and Inference: A Prediction Approach* (2000) by R. Valliant, A.H. Dorfman, and R.M. Royall

Instructor: Richard Valliant
rvalliant@survey.umd.edu
Telephone: 301-405-0932 Fax: 301-314-7912

Webpage

<https://ctools.umich.edu/portal>

Class notes and datasets will be posted here.

Print the notes each week before coming to class.

See the last page of this syllabus for instructions on how to establish a “Friend” account at UMich that will allow you to use Ctools.

Office Hours by appointment

Your questions will usually be most quickly answered by email.

Teaching assistant: none

Homework

Answers can be emailed, faxed to JPSM (301-314-7912), or paper copies put in my JPSM mailbox.

Assessment:

Homework Assignments (70%)

Homework is due (typically) 1 week after it is assigned.

Simulation project (20%), due on Dec. 15

Class participation (10%)

Simulation project

The simulation project will be designed to illustrate the properties of some estimators covered in the class. You will have some discretion in what you include but the final content will be approved by the instructor. The simulations will be written in the R language, which you will learn during the course of the semester. Your project and results will be described in a brief (15-20 pages) technical report. The R code should be included in an appendix (that does not count as part of the 15-20 pages).

Due dates:

Description of your proposed project, Nov. 17

Feedback from instructor, Nov. 23

Final report, Dec. 15

Course Topics: Text Book

Most but probably not all of these topics will be covered. There will likely be some substitutions during the semester based on current interests of the instructor.

1. Introduction to Prediction Theory

Sampling Theory and the Rest of Statistics

Prediction Theory

Which Approach to Use—Prediction or Probability Sampling?

Why Use Random Sampling?

2. Prediction Theory under the General Linear Model

General Prediction Theorem

The *BLU* Predictor under Some Simple Models

Asymptotic Normality of the *BLU* Predictor

Ignorable and Non-ignorable Sample Selection Methods

Comparisons with Design-based Regression Estimation

3. Bias-Robustness

Design and Bias

Simple Polynomial Framework and Balanced Samples

Expansion and Ratio Estimators and Balanced Samples

Bias-Robust Strategies

Weighted Balance

BLU Estimators Unbiased under Weighted Balance

Methods of Selecting Balanced Samples

Simulation Study of Weighted Balance

Robustness and Design Based Inference

4. Robustness and Efficiency

General Linear Model

BLU Predictor under the General Linear Model with Diagonal Variance Matrix

Results for Probability Proportional to x Sampling and x -balance

Results for Probability Proportional to $x^{1/2}$ Sampling and $x^{1/2}$ -balance
Results for Equal Probability Systematic Sampling and Simple Balance
Sample Size Determination

5. Variance Estimation

Homoscedastic Through the Origin Model
Variance Estimators for the Ratio Estimator
Variance Estimation of a Linear Function of the Parameter
The “Sandwich” Variance Estimator
Adjustments to the Sandwich Estimator
Jackknife Variance Estimator
Variance Estimation for Totals
Misspecification of the Regression Component
Comparisons with Design-based Variance Estimation

6. Stratified Populations

Stratification with Homogeneous Subpopulations
Stratified Linear Model and Weighted Balanced Samples
Optimal Allocation for Stratified Balanced Sampling
Allocation to Strata in More Complicated Cases
Contrasts between Strata
More than One Target Variable
Some Empirical Results on Strata Formation
Variance Estimation in Stratified Populations
Stratification in Design-based Theory

7. Models with Qualitative Auxiliaries

Factors, Levels, and Effects
Generalized Inverses
Estimating Linear Combinations of the Y 's
One-way Classification
Two-way Nested Classification
Two-way Classification without Interaction
Two-way Classification with Interaction

8. Two-stage Sampling for Clustered Populations

An Intracluster Correlation Model for a Clustered Population
A Variance Component Model
Estimation of Model Parameters
ANOVA Estimators
Maximum Likelihood Estimators
General Class of Estimators When There is a Common Mean
Two Alternative Estimators of the Total
Optimal Allocation of Second-stage units given a Fixed Set of First-stage Sample Units
Optimal First and Second Stage Allocation Considering Costs
Estimation under a More General Regression Model

Estimation under a Post-stratification Model

9. Robust Variance Estimation in Two-Stage Sampling

The Working Model and a General Class of Estimators

Variance Estimation for an Estimated Total—Known Cluster Sizes

A Non-Robust ANOVA Estimator

Robust Variance Estimator for the General Class of Estimators

Variance Estimation for an Estimated Total—Unknown Cluster Sizes

Variance Estimation for a Ratio of Estimated Totals

Variance Estimation for the Post-stratified Estimator

10. Alternative Variance Estimation and Confidence Interval Construction Strategies

Jackknife

unstratified and stratified cases

Miller residual formula

consistency

application to ratio and regression estimators

multistage sampling

11. Special Topics

Distribution Function And Quantile Estimation

Estimation of $F_N(\cdot)$

Estimation in the Presence of Outliers

Nonparametric Estimation of Totals

Course Topics: R programming

Programming in R will be interspersed throughout the course. There will be several class meetings devoted almost entirely to use of R. Topics will include:

- Workspaces
- Objects
 - Vectors, matrices, lists, data frames, attributes
- Importing/exporting data
- Scripts
 - Editing scripts with Tinn-R
- Commands
- Writing functions
 - Debugging using browser
- Writing simulation programs
- Packages
 - foreign, ggplot2, MASS, rpart, survey, among others

Datasets

The course will use (at least) six datasets:

- cancer
- co60
- co70
- grade3
- hospital
- labor

These are described in Appendix B of the text. They can be downloaded from the class website.

How to Set Up a Friend Account for Guest Access to U of M Computing Resources

How to Set Up a Friend Account

You will need to have an e-mail address that does not end in “umich.edu” in order to set up a Friend account.

1. Use your web browser to go to this URL: <https://weblogin.umich.edu/friend/>
2. Enter the full e-mail address of the person who will use the Friend account in the box, and click the **Request** button.
3. You will see confirmation that your Friend request has been submitted.
4. A message with further instructions will be sent to the e-mail address that you entered. Click the **Create Friend Account** link in the message to go to the web page where you can continue creating your Friend account. (Depending on which e-mail program you use, the message may be formatted differently.)
5. Select a password and enter it in the **Password** box. Retype the password in the **Confirm Password** box to confirm your selection. Then click the **Set Password** button.
6. You will see a confirmation screen letting you know that your Friend account has been created.
7. You will receive an e-mail message confirming the creation of your Friend account and providing you with information about it. We recommend that you save this message for reference.

Using Your Friend Account

You can use your Friend account login name and password to log in to U of M web-based computing resources at this URL: <https://weblogin.umich.edu/>

Your login name is your full e-mail address, the one you used when creating your Friend account.

After you log in, you will see a list of services. Click any service link to open that service.

Note that you can only use services that you have been authorized to use. If you want to use a service that you have not yet been authorized to use, contact the service provider and ask for authorization.

*For more information about setting up a Friends account:

<http://www.itd.umich.edu/itcsdocs/all.php> (click on S4316)

To Access C tools

Access Ctools at the following link: <https://ctools.umich.edu/portal>, use the account you established following the instructions above to log in to Ctools