DRAFT Report California State University Hayward

DEPARTMENT OF STATISTICS
BS DEGREE IN STATISTICS
Spring 2002

Recommendations

As set forth in the mission statement and the goals and objectives for the programs in Statistics, the objectives and specific competencies listed below constitute the learning objectives for the BS degree program in Statistics at CSUH. All courses are expected to address a specific and overlapping subset of these overall objectives and to develop specific competencies as required for entry-level statisticians working in the field of statistics.

Overall Objectives for the BS in Statistics

1. *To appreciate the basics of modeling and error analysis, particularly those enhanced by the use of cutting-edge computer technology and the most modern computer-intensive statistical methods.*
2. *To carry out accurate and careful statistical analyses of real world problems and to express these analyses as the meaningful oral and written communication of statistical ideas.*
3. *To experience working in teams and to receive cordial, meaningful, and persistent advising appropriate to the students' career goals.*
4. *To develop an appreciation of the ethical and legal issues involved in the collection and dissemination of statistical information and experimental design.*

Specific Competencies

1. Skill in the use of cutting-edge computer technology for collecting, cleaning, and managing data.
2. Skill in the use of cutting-edge computer technology for exploratory data analysis and the graphical display of data.
3. Skill in the use of cutting-edge computer technology for inference and simulation studies
4. Skill in the application of statistical methods and probability modeling to problems outside the classroom.
5. Skill in critically evaluating the uses of statistics and experimentation, including appropriate applications and deceptive or erroneous reasoning.
6. Skill in simulating probabilistic outcomes to assess theories appropriate to real-life examples or hypothetical situations.
7. Skill in the language of statistical methodology so that appropriate communication can take place.

Pedagogy/Best Practices

We undertake the following as useful pedagogical strategies to be encouraged in the achievement of the objectives listed above:

1. Faculty teaching statistics courses are committed to meeting these objectives for all students.
2. The statistics BS degree incorporates material and assignments with relevance to students' lives, experiences, and future employment, including concrete topics and applications.
3. Statistics course materials are presented in a variety of accessible ways that are understandable to students (e.g., lecture, small group problem-solving, everyday applications, projects).

4. The contributions, experiences, and perspectives of various demographic groups are incorporated into the statistics courses, when appropriate.
5. Learning in statistics is viewed as a developmental continuum, and students are given instruction and support appropriate to their developmental needs.
6. Department and/or institutional support and advising are provided for students at risk for not completing the BS requirements in statistics.

Assessment tools

Skills are introduced, practiced, and tested throughout the Statistics major. Reinforcement of these skills continues throughout the students' career at CSUH. Until we can develop a more specific capstone course, Statistics majors enroll, at or near the end of their program, in the course STAT 4601, Regression. This course was chosen because it comes at the end of a sequence of required courses and require the following:

1. A term project requiring passing work in skills 1-7 is required to pass this capstone course and graduate. Students prepare the project during the quarter with drafts and progressively more complex computer programs. These projects are evaluated for all skills noted above. These projects are suitable for demonstrating to a prospective employer the skills of the student concerning programming, writing, data handling, and project completion. The student may combine this project with other course work, such as a probability exam, to form a summative portfolio of work in statistics.
2. An objective final has been developed by the faculty to test the acquisition and retention of statistical language and knowledge of necessary techniques as outlined in the Goals and Objectives. Rather than test the hands-on nature of the education emphasized in 1 above, the objective test evaluates the understanding of the student in using the procedures and language studied during the program. This measure will permit tracking of deficiencies in the program by discipline area.

Departmental Support

We offer appropriate and necessary support for students to complete the Statistics BS requirements, including:

1. Advising to ensure that students select courses appropriate to their needs, interests, and preparation.
2. An adequate number and variety of approved courses each quarter with a variety of instructors over time.
3. Periodic review of courses for frequency of offering (at least once a year), number of sections, variety of perspectives represented, and achievement of departmental objectives (at least at the five-year-review).
4. Active intervention for students identified as at risk of not completing the requirements.

Student Preparation

Students who major in Statistics must take a foundation of coursework in mathematics through two quarters of calculus. Deficiencies in calculus are noted in early coursework, and students are encouraged to repeat calculus and/or the required background courses.

Attachments:

Mission Statement, Goals and Objectives, Instructions and Evaluation Forms for course projects, and Draft Objective Exam.

Department of Statistics Mission and Goals Statement

The mission of the CSUH Department of Statistics is to provide excellent education by professional statisticians in statistical theory and practice, and probability theory and modeling appropriate to the various constituencies it serves.

GOALS:

- *For students in all statistics and applied probability courses, as appropriate to the level:* To provide hands-on experience in the use of cutting-edge computer technology for collecting, cleaning, and managing data; exploratory data analysis; graphical display of data; inference; design of experiments; simulation studies; and the most modern computer-intensive statistical methods.
- *For lower-division students and students not majoring in the sciences:* To provide instruction in concepts such as randomness; basic concepts of research design (experiments and surveys); the interpretation of data from experiments and surveys; and summarization and communication of statistical results.
- *For undergraduate majors and minors in the mathematical, physical, life, social, and administrative sciences and engineering:* To provide instruction in statistical methods and probability modeling that will equip students to apply statistical methods in job markets where such education and training is required, or to pursue graduate work with optimal preparation for success.
- *For postbaccalaureate students:* To provide instruction in mathematical statistics and probability theory and in advanced statistical methodology and practice that will prepare them for leadership in the statistical profession,whether through employment in key positions or through further study and research in top-rated Ph.D. programs; to provide them with the depth of understanding necessary for lifelong learning in the field; and to encourage their participation in research under faculty guidance.
- *For all students studying to enter statistical practice:* To provide training in the meaningful oral and written communication of statistical ideas; experience working in teams.
- *For all graduate and undergraduate students in Statistics programs:* To provide cordial, meaningful, and persistent academic advice based on students' career objectives; regular events for networking with professional statisticians, including alumni/ae of our programs; assistance, from the point of view of professional statisticians on our faculty, in applying for employment or Ph.D. study; and encouragement to become participating members of professional statistical organizations.

GOALS AND OBJECTIVES:

- *For students in all statistics and applied probability courses, as appropriate to the level:* To provide hands-on experience in the use of cutting-edge computer technology for collecting, cleaning, and managing data; exploratory data analysis; graphical display of data; traditional and Bayesian inference; design of experiments; simulation studies; and the most modern computer-intensive statistical methods. Exercises in quantitative literacy and applying statistics to experimental data, demonstrating the scientific method.

UNDERGRADUATE STATISTICS COURSES–Skills

COURSE	Data cleaning and management	Exploratory data analysis	Graphical data display from:	Interval estimation	Inference	Hands-on computational experience	Written summary
1000 2010	Entry into data sheet	Graphical and probabilistic	Stem and leaf, dot, histograms, box, control and scatter	One-sample mean, large and small sample	One-sample means, slope, may include two-sample	Lab demonstrations and or assignments	Problem solution summaries
3010 3031	Entry into data sheet Transforming data	Graphical, experimental, and probabilistic	Normal, QQ, dot, histograms, box, scatter, means, and interaction	Two-sample means, slope, differences in means	Two-sample means, ANOVA (one- and two-way), nonparametric introduction	Lab demonstrations and assignments	Problem solution summaries and assignment summary and discussion
3601 3502 4612	Entry into data sheet indep, dep variables Transforming data	Graphical and probabilistic	Stem and leaf, dot, histograms, box and scatter	One- and two-sample means, differences in means, slope	One- and two-sample, regression, ANOVA	Lab demonstrations and assignments	Problem solution summaries and assignment discussion
3503 4000	Entry into data sheet Dummy variables	Graphical, experimental, and probabilistic	Stem and leaf, dot, histograms, box, scatter, mean and interaction	Two-sample means, slope, differences in means; variances	Two-way and higher ANOVA, dependence in models, covariates	Assignments exploring models and demonstrations of modeling changes	Problem solution summaries and assignment discussion
3900 4950	Entry into data sheet Cleaning and manipulating data	Graphical and investigative	Histograms, box, QQ, scatter, means and interaction	One- and two-sample means, differences in means	Applications of a variety of tests, parametric and nonparametric	Assignments in data manipulation and analysis of experimental and survey data	Problem solution summaries and assignment discussion
3910 4910	Creating pseudo random data	Graphical and investigative	Stem and leaf, dot, histograms, box and scatter	Moment confidence intervals	Applications specific to a stat content area	Advanced assignments in data manipulation and analysis of experimental and survey data	Computer code commenting and output summary
3510 3415	Specialized data entry	Graphical and probabilistic	Stem and leaf, dot, histograms, box and scatter	Means for differing sample types	Decisions based on probabilistic or sampling schemes	Assignments appropriate to course content	Computer code commenting and output summary

4515 4601 4610	Transforming and manipulating data, ranking data	Graphical, experimental, probabilistic, and investigative	Histograms, box, QQ, scatter, means, residuals, trends, curvature	Bonferroni, parameter estimates	Models and parameter estimates or rank structure	Advanced assignments prepared for portfolio development	Senior project development and written reports
48XX Exploratory Data Analysis	Transforming and manipulating data, ranking data; outlier detection	Graphical, experimental, and investigative	Histograms, box, QQ, scatter, means and interaction	Any as descriptive only	NA	Advanced assignments appropriate to exploration and cleaning of data	Senior project development and written reports
48XX Statistics for troubled data	Transforming and manipulating data, ranking data	Graphical, experimental, probabilistic, and investigative	Normal, QQ, dot, histograms, box, scatter, means, and interaction	Differences in means, odds ratios	Regression and ANOVA parametric structure	Advanced assignments appropriate to analysis of scientific and medical data	Senior project development and written reports

UNDERGRADUATE PROBABILITY AND APPLIED PROBABILITY COURSES Skills

COURSE	**Random variables**	**Pseudo random variables**	**Graphical displays**	**Variable transformation**	**Application to inference**	**Hands-on practice**	**Proofs and derivations**
2088	Binomial, uniform, multinomial, other discrete distributions	Demonstrated generation on computer	For verification	Discrete only	Decision making, likelihood	Applications and problems assigned	Few demonstrated
3401	Discrete and continuous distributions	Demonstrated generation on computer	For area of definition and change of variable	Linear transformations	Populations, parameter foundations, probabilistic alternatives	Problems regularly assigned	Means, variances of standard discrete variables demonstrated
3402	Continuous and jointly continuous	As time permits, demonstrated computer generation	For area of definition and change of variable	Joint transformations, nonlinear, moments	Parametric estimation and distribution of statistics, mle, conditioning	Problems regularly assigned	Means, variances derived for discrete and continuous variables, distributions for sums, transformed variables, and standard statistics
4401	Binomial, Poisson, exponential	Demonstrated generation on computer	For arrival times and random behavior demonstrations	Conditioning and marginality, moment generating	Stochastic data, arrival times, service times	Problems regularly assigned	Basic stochastic theorems proved

GRADUATE STATISTICS COURSES Skills

COURSE	Data cleaning and management	Exploratory data analysis	Graphical data display	Interval estimation	Inference	Hands-on computational experience	Written summary
6501 6502	Robust statistics to protect against errors	Robust methods	Q-Q plots moments	Derivations for various statistics	Derivations for various statistics	Advanced simulations to demonstrate	Problem summaries and proofs
6510	NA	Post-hoc comparisons	Means plots, interaction plots	Derivations for comparisons	Derivations for expected mean squares	Some examples applied	Proofs and discussions of fixed versus random effects
6515	Major focus of multivariate data entry and outlier identification	Cluster Analysis, outlier identification	Q-Q plots, multivariate specialties as star plots, faces	Bonferroni intervals, multivariate intervals	Wald and Wilkes statistics, largest eigenvalue	Extensive use of modern computational methods	Written reports suitable for portfolio and job search
6601	Extensive discussions with robust statistics as protection	Modern methods as spline models, smoothing	Extensive use of modern graphical techniques	Bootstrapping and jack knife as means to produce	Modern techniques using bootstrapping methodology, etc.	Extensive program development in S+ and other languages	Program documentation required and problem discussion
68XX Survival analysis	Utilization of cross tabulations to detect errors	Log linear and logistic models explored	Survival distribution, histogram for censored data	Odds ratio estimates	Model estimates and associated inference	Extensive modeling using SAS life test and logistic regression	Model summaries and descriptions of results
68XX Discrete data analysis	Utilization of cross tabulations to detect errors	Log linear and logistic models explored	Projections through time, other attribute	Confidence intervals for proportions	Inference for model parameters	Extensive computer modeling in SAS	Discussion of models developed; final project
68XX Time series	Attention to problematic, noisy data	Extensive use of smoothing techniques	Longitudinal, seasonal, and periodic plots	Parametric and future intervals	Model parameter inference	Extensive computer modeling in SAS and/or S+	Final project and discussion of models developed; computer code comments
68XX Neural networks	Extensive reliance on cleaning and imputing	Imputation of missing data, scaling, binning, transforming	Distributions, fits, residual analysis	NA	Optimization of parameters	Extensive computer modeling using SAS and S+	Discussion and comparison of models, oral presentation

GRADUATE PROBABILITY COURSES Skills

COURSE	Random variables	Pseudo random variables	Graphical displays	Variable transformation	Application to inference	Hands-on practice	Proofs and derivations
6401-02 Probability Theory	Development of important distributions	Use of uniform in deriving	Importance in transformations of variables	Most important ones in statistics derived	Theoretical development	Many homework problems and in-class problems	Demonstrated and required in homework and on examination
68XX Reliability	Development of important distributions	Weibull, gamma, exponential, and others	As demonstrations	Relationship between common distributions	Common applications to manufacturing	Many problems solved on and off the computer	Demonstrated and required in homework and examination
68XX Decision Theory	Priors and posteriors	Demonstrations of theory	Prior and posterior distributions	Conjugate and approximate priors, future values	Effect of loss functions on decision	Problems solved and applied to realistic situations	Demonstrated and required in assignments
68XX Bayes Theory	Priors and posteriors	Demonstrations of theory	Prior and posterior distributions	Conjugate and approximate priors, future values	Posterior inference using decision theory, loss functions	Programs developed and various priors tested by students, homework in S+	Derivations of posterior distributions

- *For lower-division students and students not majoring in the sciences:* To provide instruction in concepts such as randomness; basic concepts of research design (experiments and surveys); the interpretation of data from experiments and surveys; and summarization and communication of statistical results.

Lower division students and students not majoring in the sciences: Attitudes and exposure

COURSE	Randomness	Research design	Interpretation of data from experiments	Interpretation of data from surveys	Summarization and communication of statistical results
1000 2010	Basic probability Random number tables	Discussion, examples, and word problems	Populations, samples, scientific, ethical, and logical principles related to experimentation discussed	Populations, samples, generalization of results (where applicable)	Written summaries discussed and required in problem solution. Varieties of examples presented and discussed.

For undergraduate majors and minors in the mathematical, physical, life, social, and administrative sciences and engineering: To provide instruction in statistical methods and application and probability modeling that will equip students to apply statistical methods in job markets where such education and training is required, or to pursue graduate work with optimal preparation for success.

Undergraduate majors and minors in the mathematical, physical, life, social, and administrative sciences and engineering: Skills

COURSE	Statistical methods and application	Probability modeling	Preparation for work force application of statistics	Preparation for graduate work in their field where statistics is required
3010 3031	Entry into data sheet Transforming data	Graphical, experimental, and probabilistic	Normal, QQ, dot, histograms, box, scatter, means, and interaction	Two-sample means, slope, differences in means
3601 3502 4612	Entry into data sheet Indep, dep variables Transforming data	Graphical and probabilistic	Stem and leaf, dot, histograms, box and scatter	One- and two-sample means, differences in means, slope

- *For postbaccalaureate students (mainly M.S. and certificate students in Statistics and Applied Mathematics), as appropriate to specific student objectives:* To provide instruction in mathematical statistics and probability theory and in advanced statistical methodology and practice that will prepare them for leadership in the statistical profession, whether through employment in key positions or through further study and research in top-rated Ph.D. programs; to provide them with the depth of understanding necessary for lifelong learning in the field; and to encourage their participation in research (and, where appropriate, publication) under faculty guidance.

Post-baccalaureate students: Attitudes, life-long learning, and research

PROGRAM	Mathematical Statistics and Probability Theory	Advanced Statistical Methodology	Pratice
Applied Statistics Certificate	Regression theory Statistics 4601, prerequisite courses with probability	Statistical methods, and two from: multivariate, analysis of variance, nonparametrics, undergraduate seminar	Data analysis using statistical programming packages
Mathematical Statistics Certificate	Prob Theory I, Math Stat I&II	Prob Theory II, Analysis of Variance, Statistical Computation, or Graduate Seminar	Problem solution, stochastic modeling
Master's Degree in Statistics	Prob Theory I, Math Stat I & II	Three from: Prob Theory II, Analysis of Variance, Statistical Computation, or Graduate Seminar	Graduate Seminar or Coop Ed

- *For all students studying to enter statistical practice:* To provide training in the meaningful oral and written communication of statistical ideas; experience working in teams; an appreciation of the ethical and legal issues involved in the collection and dissemination of statistical information; and the opportunity for co-operative education.

Students studying to enter statistical practice: Attitudes, awareness of regulations, and communication of ideas

PROGRAM	Oral and written communication of statistical ideas	Mathematical statistics and probability theory	Team experience	
Minor in Statistics				
BS in Statistics				
Applied Statistics Certificate		Regression theory, Statistics 4601, prerequisite courses with probability	Statistical methods, and two from: multivariate, analysis of variance, nonparametrics, undergraduate seminar	Data analysis using statistical programming packages
Mathematical Statistics Certificate		Prob Theory I, Math Stat I&II	Prob Theory II, Analysis of Variance, Statistical Computation, or Graduate Seminar	Problem solution, stochastic modeling
Masters Degree in Statistics		Prob Theory I, Math Stat I & II	Three from: Prob Theory II, Analysis of Variance, Statistical Computation, or Graduate Seminar	Graduate Seminar or Coop Ed

- *For all graduate and undergraduate students in statistics programs:* To provide cordial, meaningful, and persistent academic advice based on students' career objectives; regular events for networking with professional statisticians, including alumni/ae of our programs; assistance, from the point of view of professional statisticians on our faculty, in applying for employment or Ph.D. study; and encouragement to become participating members of professional statistical organizations.

All graduate and undergraduate students in statistics programs

INDIVIDUAL CHECK SHEET FOR PAPERS IN STAT 4601

Title and Introduction

Title 2pts.	Intro-duction: problem 3pts.	Data intro-duction 3pts.	How data collected 1pt.	Adequacy of the sample 1pt.	Summary statistics 2pts.	Identi-fication of hold out sample 2pts.	Graphs y*x. 1pt.	Inter-actions/ trans-forms 3pts.

Methodology

Techniques used in analysis 5pts.	Best two or three models and graphs 5pts.	Summary table, comparison measures, models 5pts.	Including the hold-out data sets 4pts.	Outlier identi-fication/dis-cussion 2pts.	Outlier removal/ model assessment 2pts.	Assumption check for final model overall quality 2pts.	Normality check, residuals 2pts.

Assessment and Summary

Shape analysis residuals versus X vars. 3pts.	Shape analysis residuals versus Y 2pt.	Conclusion, overall quality 3pts.	Recom-mendation of model 3pts.	Drawbacks of the model 3pts.	Benefits of the model 3pts.	Future study 5pts.	Final paragraph 4pts.

Appendix and Extra Details

Organ-ization overall 5pts.	SAS® code, complex-ity 5pts.	Annotate code, care, clarity 4pts.	Any extra details 2pts.	Data included in Appendix 1pt.	Attention to form of paper 2pt.	Attention to present-tense witing 5pts.	Appear-ance of report 3pts.	Size of report, require-ments 2pts.

REGRESSION ANALYSIS
STATISTICS 4601/6509

SPRING 2002

INSTRUCTOR:	J. A. Norton	GRADUATE WRITING ASSISTANT	Beth Ochsner
CLASS MEETS:	MW 4-5:50 p.m.	OFFICE HOURS:	M 3:00-3:50 p.m.
		M 10:00-10:30 a.m., W 12:10-12:40 p.m., and by appointment	
OFFICE PHONE:	510-885-3431	MESSAGE:	510-885-3435
FAX:	510-885-4714		
EMAIL:	jnorton@csuhayward.edu	baochsner@yahoo.com	
TEXT:	*Applied Regression Analysis and Other Multivariable Methods*, Kleinbaum and Kupper		
	Computer manual for SAS® or SPSS®		
ADDITIONAL TEXTS:	*Principles of Regression*, SAS® Institute		
	Regression Analysis, Draper and Smith		

Regression analysis is the basis for many models in statistics. The goal of this course is to model complicated response variables and to write reports that describe the problem, the analysis, and the results in an accessible way. All computer work will be demonstrated in SAS®. Regression graphs and model assessment will be a strong component of the course work. College coursework requires two to four hours of work outside class for each hour inside class.

Grades will be based on two papers written for the course (70%), weekly in-class writing assignments (10%), and a final exam (30%). The first paper will be an application of a simple linear regression and correlation. The final paper requires regression computing techniques and a complete analysis of a large data set. Students are encouraged to seek out data for analysis that are particularly interesting. Some data sources will be available. Data must be approved in advance and must be applicable to multiple regression theory or the design of experiments.

Papers must be original. Sources for all data must be cited. Students must do their own work. Evidence of academic dishonesty will result in failure.

The final exam will consist of two portions. The first portion is a closed book multiple-choice exam that covers the basic topics, language, and reasoning of statistical regression and modeling. These questions come from the text and the lectures. The second portion is an open book discussion of a topic relevant to the course and level of the course.

TENTATIVE DATES FOR PAPERS:

Draft 1,	Paper 1	Wednesday, April 17
This assignment written in-class.		Read Chapters 2 and 3 to prepare.
Final version	Paper 1	Wednesday, May 1
Draft 1,	Final Paper	Wednesday, May 15
Final version	Paper 2	Monday, June 10
Final Exam		Monday, June 10, 4 p.m.

Papers will be submitted electronically through turnitin.com.

If you have a documented disability and wish to discuss academic accommodations, or it you would need assistance in the event of an emergency evacuation, please contact the instructor as soon as possible.

Statistics 4601
J. Norton

April 2, 2002

FIRST REGRESSION REPORT

Introduction: Assume that a firm (or another department in your company, in either case give the firm or department a name) consults you. The goal of the consultation is to develop a regression model to predict their weekly sales, service, or other reasonable product (which you should name) as a response to advertising costs in the sales market area. The final deliverable product from you is a report and a prediction for sales when advertising expenses are $147,000 dollars. It is this report that you will be writing for project one.

Methodology: Assume the data are collected using random sampling. Please demonstrate your understanding of random sampling by constructing a description of how the data were collected that fits your project. That is, tell the (made-up) story of how the data were collected. In your report, be sure to discuss the steps you took in arriving at a model.

The data file lies on the i: drive under the name advert.dat. The SAS® program called advert.sas is also available, with all of the pieces necessary for your report. You might wish to change the code, the titles, or the variable names in the code for the final paper to match your project, but this change is not necessary. Also remember to comment on the SAS code in the final report, include it in an appendix, and include a copy of the data in a second appendix.

Discussion: In later papers, you will want to develop several models at this stage. Make the discussion clear and straightforward so that someone who hasn't had a formal course in regression can get the idea and generally follow most of your discussion, especially the introduction, discussion, conclusion, and summary. Include at least one labeled graph. The methodology and discussion sections may contain some statistical jargon if you define any that you use. You need not develop the theory of regression. Discuss the model, the assumptions, and the fit.

I made an imperfect effort to use active voice and present tense while generating this assignment. Your goal should be to do so as much as possible in your report.

Conclusion: The goal for the draft is to incorporate as many of the details as you need assistance with in preparation for the final paper. The conclusion must contain the final model in a form that will assist the people for whom you are producing this report. You should briefly report why you believe this model is a good one. If it is not a good model or if you think it may be improved, suggest some additional ideas that the sponsors might take to gather other relevant data. You should also report your final prediction and the quality of that prediction (accuracy).

Summary: Specifically, you must include the following sections in your paper:

- Introduction describing the project.
- Methodology describing the data, its collection, the steps you took in arriving at a model, and assessments you made concerning the model or models.
- Discussion of the best model or models and the quality of the fit.
- Conclusion describing the generality and usefulness of the model, as well as ideas for improving the model.
- Summary should draw the previous important ideas together in one final paragraph.
- Appendices containing at least one page of the data, the commented program, whether you used SAS® or SPSS®, additional graphs, and other details in later papers that are too specific for the paper itself.

Good luck!