OVERVIEW OF PROPOSED CHANGES IN STATISTICS COURSES

1. Add ST-351, Principles of Statistics II.

This would be a 3 credit hour sequel to ST-251, Principles of Statistics. Pre-requisites are ST-251, 271, or 301. Format would be 3 hours of lecture and 1 hour of recitation/lab per week.

Justification---This would be a general elective for any interested student and a required course for undergraduates minoring in statistics. *Note* that this would essentially replace ST- 401 for undergraduates and we will need Mathematics input on this. Initial enrollment would likely be no more than 30, but it could expand to 100. This course would be (initially) offered in the Spring Semester only, would replace the current spring semester offering of ST-401, and thus would not increase the Division's teaching load.

The rationale for the change is primarily to maintain a "second" undergraduate statistics course, necessary once ST-401 is dropped. A secondary reason is to increase the number of undergraduates taking a minor in Statistics.

2. Drop ST-401, Statistical Analysis.

3. Add ST-511, Introductory Statistics for Researchers.

This would be a 3 credit hour course for graduate students with little or no statistics background, or those who have not taken statistics in several years. The only pre-requisite is MA-143 or its equivalent, and the format would be 3 hours of lecture per week.

Justification---This would likely become a required course for graduate students who would be required to take ST-515 but are ill prepared. The probable enrollment initially would be around 30 to 40, but would be likely to grow considerably. This course would be (initially) offered in the Fall Semester only and would replace the current Fall Semester offering of ST-401, and thus it would not increase the Division's teaching load.

The rationale for the change is primarily to prepare graduate students for ST-515, thus allowing ST-515 to be a class with minimal review of basic statistics.

4. Add ST-515, Statistics Methods for Researchers.

This would be a 3 credit hour course for graduate students having a solid understanding of basic statistical concepts and elementary methods. The pre-requisite is ST-251, 271, 301, or 511. The format would be 3 hours of lecture per week.

Justification---This would become a required course for many graduate students. It will essentially replace ST-401 for graduate students and will differ from the current ST-401 in that the amount of time spent reviewing basic concepts and elementary methods will be minimal, less coverage of design, and more coverage of modeling different types of response variables. The probable enrollment is likely to be around 100 students.

This course will (likely) increase the department teaching load. We will need additional resources. One solution may be to hire additional instructors to teach ST-150 to free a professor's time for teaching ST-515.

ST-351 PRINCIPLES OF STATISTICS II

- **Background** A sequel to ST-251 (Principles of Statistics), is proposed. A tentative proposal is to drop ST-401 (Statistical Analysis), which is usually made up of 85-90% graduate students, and replace it with ST-515 (Statistics Methods for Researchers), a class more clearly directed at graduate students.

  However, ST-401 is a required course for undergraduates in Mathematics taking either the Actuarial Science option or the Statistics option. It is also a required course for undergraduates in Geography and Cartography. Additionally, it is a requirement for undergraduates seeking a minor in Statistics. A replacement for ST-401 appears necessary.

  Another reason for a sequel is the perception by some members of the Statistics faculty that the current ST-251 syllabus is too ambitious, in containing too many topics for a 3-credit course. This perception was voiced by several faculty in an ad hoc group meeting to discuss ST-150 and 251.

  The syllabus for ST-251 should be slightly changed. One-way ANOVA should be dropped in ST-251 and χ^2 tests should be made optional. ST-351 would include both topics. The balance between concepts and methods would differ between ST-251 and ST-351 as well. The ratio of concepts to methods would be 75:25 for ST-251 and 30:70 for ST-351.

  A third reason for a sequel is to attempt to increase the number of undergraduate students taking statistics and those seeking a minor in Statistics. With the exception of ST-301, there is no other 300 level statistics course. Making the jump from ST-251 to ST-422 is perhaps perceived to be greater than it is, but the jump to ST-401 is in fact sizeable, especially once the student realizes that 85-90% of the class are graduate students. ST-351 would be perceived as a gentler next step.

- **Pre-requisites** ST-251, 271, or 301.

- **Format** 3 one-hour lectures and a one-hour recitation/lab.

- **Text** *Introduction to the Practice of Statistics*, 3rd Ed. by Moore and McCabe.

- **Topics**

  1. (1.5 weeks) Overview of ST-251 with re-emphasis on basic concepts
  2. (1.5 weeks) More probability, particularly conditioning
  3. (1 week) Chi-square, goodness of fit, tests of independence and homogeneity
  4. (2 weeks) One-way ANOVA, including contrasts, multiple comparisons
  5. (2 weeks) Two-way ANOVA, including Randomized Block Design
  6. (4 weeks) Multiple regression, including ANOVA and ANCOVA as special cases, diagnostics
  7. (1 week) Nonparametrics: sign test, rank-sum test, signed rank test, Kruskal-Wallis test
  8. (1 week) Logistic regression

ST-511 INTRODUCTORY STATISTICS FOR RESEARCHERS

- **Objectives:** Graduate students who successfully complete this course (with an A or a B grade) will:
  --- have a good understanding of the foundations of statistics and how it is applied in research;
  --- be able to apply basic statistical methods in their own research;
  --- be well-prepared to take more advanced courses in applied statistical methods.

- **Audience:** Non-major graduate students

- **Pre-requisite:** MA-143 (Pre-calculus Algebra and Analytic Geometry) or its equivalent.

- **Format and "Philosophy"**: 3 hours lecture. Case Studies will be used to demonstrate most of the major topics with the intent being to emphasize statistical methods as logically sound procedures for learning from data.

- **Comparison and contrast with other courses**

  1. Like Statistics 251 in discussing different procedures for generating data, covering basic estimation and testing terminology, and presenting one and two-sample *t* confidence intervals and tests for means, one and two-sample *z* confidence intervals and tests for proportions, χ^2 tests, linear regression, and 1-way ANOVA.

  2. Differing from ST-251 in the considerably greater coverage of probability, emphasis on statistics for research, and an introduction to more sophisticated statistical models, including logistic regression, nonlinear regression, and Poisson regression.

  3. Like Statistics 301 in presentation of basic terminology, coverage of probability, and use of basic methods.

  4. Differing from ST-301 in that calculus will not be required and the introduction statistical models will contain more sophisticated models.

Text: Lecture notes will be prepared and available online for students.

ST-511 TOPICS

1. (1.5- 2 weeks) Data

(a) *Sources of Data*: quantitative and categorical data, measurement, experiments, observational studies, quasi-experiments, probability samples versus census.

(b) *Interpretation of Data*: randomization, causation versus association, confounding/lurking variables, controls, single and double blinds.

2. (5 weeks) Probability and Random Variables

(a) *Probability*: random phenomena, outcomes and events, sample space, mutually exclusive events, independent events, conditional probability.

(b) *Random Variables*: discrete versus continuous random variables, probability mass and density functions (e.g., binomial, Poisson, normal), cumulative distribution functions, expectation, conditional distributions.

(c) *Sampling & Statistics*: random samples, statistics, sampling distributions, Central Limit Theorem

3. (3.5 - 4 weeks) Statistical Inference

(a) *Scientific Method*: deductive versus inductive reasoning, inference.

(b) *Point Estimation*: estimators, bias, standard error.

(c) *Interval Estimation*: confidence intervals for one sample mean and one sample proportion

(d) *Testing*: hypotheses, test statistics, P-value, Type I and Type II errors, power, caveats.

(e) *Model Fit and Evaluation*: assumption checking, robustness, diagnostics

4. (5 weeks) Statistical Modeling

(a) *Binary response variables*: 2-sample case with inference, logistic regression.

(b) *Categorical response variables*: contingency tables, χ^2 tests of homogeneity and of independence

(c) *Continuous response variables*: 2-sample case with inference, *k*-sample case, simple linear regression, nonlinear regression

(d) *Count responses*: 2-sample case with inference, Poisson regression.

ST-515 STATISTICAL METHODS FOR RESEARCHERS

- **Objectives**: Graduate students who successfully complete this course (with an A or a B grade) will know how to:

  --- effectively size up data analysis problems---to identify objectives, experimental/sampling units, separate response from explanatory variables when appropriate, to always distinguish association from causation.

  --- analyze "most" kinds of data (excluding temporal, spatial, multivariate, and ordinal); in particular the student will know common techniques for modeling continuous, count, or binary response variables as a function of quantitative and categorical predictor variables.

  --- effectively communicate results.

- **Audience**: non-major graduate students

- **Pre-requisite**: ST-251, 271, 301, or 511.

- **Format and "Philosophy"**: 3 hours lecture with Case Studies used to demonstrate most of the major modeling techniques. The emphasis will be on learning from data: finding patterns, modeling relationships, interpreting models and succinctly communicating results.

- **Overview of Models**

Response y

Predictor x	"Continuous"' (Normal)	Binary (Binomial/Multinomial)	Count (Poisson)
Quantitative	Linear Regression	Logistic Regression	Poisson Regression
Categorical	"ANOVA" models	Logistic Regression (and χ^2 Analyses)	Poisson Regression ("Loglinear"' models)
Mixture	"ANCOVA" models and GLM	Logistic Regression	Poisson Regression

- **Contrast with Current ST-401**

1. Almost no review of ST-251 level material
2. Extension from "Normal only" responses to Binomial and Poisson
3. Less experimental design (e.g., no Latin Square, no Split Plot)
4. Less random and mixed effects coverage [no E(MS) calculations]

- **Text:** *Applied Linear Statistical Models, 4th Ed*. (1996) by Neter, Wasserman, Nachtsheim, and Wasserman.

ST-515 TOPICS

1. (10 weeks) Models for continuous response: Standard Linear Models
   (a) Brief review of normal distribution and idea of conditional probability
   (b) Multiple linear regression (with matrices)
      i. EDA: variable screening, scatter-plot matrices, transformations
      ii. Estimation of β's; $\mu_{y|x_1,x_2,\ldots x_k}$, (least squares); Prediction
      iii. Testing: *t*-tests, partial F-tests
      iv. Diagnostics: normality, constant variance, outliers, influential points, collinearity and some remedies (transformations, WLS)
      v. Model selection: criterion, algorithms
   (c) ANOVA models
      i. One-way ANOVA (with 2-sample *t*-test as special case)
      ii. Testing for treatment effects (fixed and random cases)
      iii. Estimating effects (linear contrasts with multiple comparison techniques)
      iv. Randomized Block Designs (with paired *t*-test as a special case)
      v. Two-way ANOVA (balanced, mostly, introduction to random and mixed effects)
      vi. Factorial treatment combinations
      vii. Nonparametrics: Kruskal-Wallis/Wilcoxon rank-sum
   (d) ANCOVA models with emphasis on use to control for continuous confounders
   (e) General linear models: showing how all the above are special cases of GLM

2. (3 weeks) Models for binary/polytomous response (primarily Logistic Regression)
   (a) Brief review of binomial distribution; some discussion of multinomial distribution
   (b) Exploratory Data Analysis (EDA)
   (c) Estimation of β's via maximum likelihood (briefly, w/o calculus)
   (d) Testing
   (e) Diagnostics (including ROC)
   (f) Model Selection (AIC/BIC/Change in Deviance)
   (g) Two-way tables: degree of association measures; χ^2 test

3. (1 week) Models for count response or rare events (generalized linear model with Poisson and log link)
   (a) Brief discussion of Poisson distribution (including its use as an approximation to the binomial)
   (b) EDA
   (c) Estimation of β's via maximum likelihood (briefly, w/o calculus)
   (d) Testing
   (e) Diagnostics
   (f) Model Selection (AIC/BIC/Change in Deviance)

4. (1 or 2 weeks) Summary
   (a) A general data analysis project with a written report.
   (b) Emphasizing EDA (on the front end) and Communication (on the back end)
   (c) What hasn't been discussed: stochastic processes (time + space), multivariate, more complicated designs.

SCHEMATIC OF PROPOSED COURSE CHANGES IN THE DIVISION OF STATISTICS

	Introductory (Concepts)	Second Semester (Methods)	Specializations
G			**Design** (ST-351 or 515)
R			**Nonparametrics**
A			(ST-351 or 515)
D	**ST-511**	**ST-515**	**Multivariate** (ST-351 or 515)
U	(MA-143)	(ST-251, 271,	[**Qualitative Data** – WSU
A		301, or 511)	(ST-351 or 515)]
T			[**Time Series** – WSU
E			(ST-351 or 515)]
U	**ST-301**	**ST-351**	**Sampling** (ST-251,271, or 301)
N	(Calculus)	(ST-251, 271,	
D		or 301)	[**Geostatistics** (ST-301)]
E	**ST-271**		
R	(Survey		[**Econometrics** (ST-251
G	Calculus)		and Calculus)]
R			
A	**ST-251**		
D	(MA-143)		

( ): Prequisite
[ ]: Cross Listed in Statistics