STAT 659 is intended for a mixed audience of graduate students majoring in statistics and graduate students in various fields who need to analyze categorical data in their own research. The level of covered material will depend somewhat on the backgrounds of the students in class and will emphasize the analysis of categorical data using SAS and interpreting the results. Lectures are scheduled from 1:50 to 2:40 MWF in Blocker 457 for the local students. Online question and answer sessions will be held on Mondays from 6 pm to 7 pm. These may be attended by local students in Blocker 411.

### Course Information

<table>
<thead>
<tr>
<th>Time and Place:</th>
<th>Class: Monday, Wednesday, Friday, 1:50–2:40 pm, Blocker 457</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question &amp; Answer:</td>
<td>Monday, 6–7 pm, Blocker 411</td>
</tr>
<tr>
<td>Instructor:</td>
<td>Thomas Wehrly. (<a href="http://stat.tamu.edu/~twehrly">http://stat.tamu.edu/~twehrly</a>)</td>
</tr>
<tr>
<td>E-mail:</td>
<td><a href="mailto:twehrly@stat.tamu.edu">twehrly@stat.tamu.edu</a></td>
</tr>
<tr>
<td>Office Hours:</td>
<td>MWF 2:50–3:50 pm, or by appointment.</td>
</tr>
<tr>
<td>Grader:</td>
<td>Hyuneui Lee ⟨<a href="mailto:hyuneui@stat.tamu.edu">hyuneui@stat.tamu.edu</a>⟩, BLOC 459F. Office hours: Tuesday, 2–4 pm.</td>
</tr>
<tr>
<td>Class Web Pages:</td>
<td><a href="http://ecampus.tamu.edu">http://ecampus.tamu.edu</a> (for lectures, notes, homework assignments and discussion board).</td>
</tr>
<tr>
<td>Prerequisite:</td>
<td>Successful completion of STAT 601, 641, or 652 or equivalent. Students who have taken only STAT 651 should note that we will use many regression-based techniques that are mainly covered in STAT 652. Please talk to me if it has been several years since your last statistics course or if you have questions about the prerequisites.</td>
</tr>
<tr>
<td>Computing:</td>
<td>SAS statistical analysis software. Instructions for obtaining SAS can be found under SAS Files on the course web site.</td>
</tr>
<tr>
<td>Homework:</td>
<td>Homework will be assigned and collected regularly. Homework is worth 10% of the total term score. <em>Please see the homework policy below.</em></td>
</tr>
<tr>
<td>Exams:</td>
<td>There will be two midterm exams worth 25% each and a final exam worth 40%. <em>Please see the exam policy below.</em></td>
</tr>
<tr>
<td>Exam Dates:</td>
<td>Test 1: Wednesday, February 26. Test 2: Wednesday, April 9. Final Exam: Tuesday, May 6 (Local students, 3:30 pm to 5:30 pm).</td>
</tr>
</tbody>
</table>
Course Information

Grading scale: 
A: 88%–100%.
B: 76%–87%.
C: 64%–75%.

Disabilities Help: 
The Americans with Disabilities Act ensures that students with disabilities have reasonable accommodation in their learning environment. If you have a disability requiring accommodation, please contact me and also the Office of Disability Services in B118 Cain Hall, 845-1637.

Plagiarism: 
As commonly defined, plagiarism consists of passing off as one’s own ideas, words, writing, etc., which belong to another. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turn it in as your own, even if you should have the permission of that person. If you have any questions regarding plagiarism, please consult the latest issue of the Texas A&M University Student Rules, under the section "Scholastic Dishonesty."

Academic Integrity: 
“An Aggie does not lie, cheat, or steal or tolerate those who do.” You are expected to maintain the highest integrity in your work for this class, consistent with the university rules on academic integrity. Please see the homework and exam policies below for specifics.

Copyright: 
All the resources I provide for this course are copyrighted and may not be copied or distributed without my express, written permission.

Course Policies

Homework Policy: 
Homework assignments will be available under Assignments on eCampus. Homework solutions must be in a single portable document format (PDF) file and uploaded to WebAssign. The initial page of each homework should contain your TYPED name, course, section number, and e-mail address.

Your homework solutions must be your own work, not from outside sources, consistent with the university rules on academic integrity. I expect you to follow this policy scrupulously.

You may use:
- Your textbook and notes from class.
- SAS software.
- Notes, homework, etc., from a related class.
- References listed on the syllabus.
- Discussion with the instructor or grader.
- Voluntary, mutual and cooperative discussion with other students currently taking the class. There will be an online discussion board available to facilitate this.

You may not use:
- Solutions manuals (printed or electronic).
- Solutions from previous classes or classes taught elsewhere.
- Copying from students in this class, including expecting them to reveal their solutions in “discussion”.

# Course Policies

**Exam Policy:** Local students will take the exams in the classroom under my supervision at the scheduled times. Distance students must be proctored according to the provided directions. You may obtain a proctor certification form under **General Information**.

Exams will be downloaded from WebAssign.

Exam solutions must be scanned into a single portable document format (PDF) file only and uploaded to WebAssign.

You should be identified on the initial page with your PRINTED name, course, section number, and e-mail address.

Your exam solutions must be entirely your own work, consistent with the university rules on **academic integrity**.

Each exam will be comprehensive, cumulative and closed book. Exams will emphasize the material covered most recently, but students are responsible for all the earlier material as well. I will not expect you to quote theorems and results explicitly but I do expect you to demonstrate that you can make use of them. Specifically, you will need to:

- Show all your work. This does not necessarily mean showing every individual algebraic step – but it must be clear what those steps are.
- Identify (by number, name or description) any theorems, results, examples or homework problems you use.
- Identify clearly the part of the SAS output that you are using in your solution.

You will be allowed a specified number of formula sheets and the use of a calculator.

Copies of old exams will be available for you to review.

**Makeup Policy:** This is based on university policy.

- If you must miss an exam due to illness or circumstances beyond your control, notify me or the Statistics Department (979-845-3141) (before, if feasible, otherwise within two working days after you return). See me as soon as possible to schedule a make-up exam.
- An Incomplete (I) will be given only in the event that circumstances beyond your control cause prolonged absence from class and the work cannot be made up. This grade is not to be given because you feel that you have too much other work or study or because you think that you will not earn an acceptable grade in the course.
Instructions for Using the eCampus Website

You may access streaming videos of the lectures and all other information for the course at the eCampus website. You may access eCampus either directly http://ecampus.tamu.edu or via http://howdy.tamu.edu.

After reaching the course web site, take a look at the resources. Information concerning the course will be posted under Announcements.

Please look/watch for further instructions concerning the use of eCampus, watching videos, attending the Q&A sessions, submission of homework and exams, and using SAS.

The recorded (streaming video) lectures will be found under Class Videos. You will find links to the printed lecture notes, SAS program files, homework assignments, and other course material. Under Links there is a link to an errata sheet for the textbook and many other useful links.

Course Outline

<table>
<thead>
<tr>
<th>Topic</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction to Categorical Data and Goodness of Fit</td>
<td>1</td>
</tr>
<tr>
<td>A. Binomial data</td>
<td></td>
</tr>
<tr>
<td>B. Multinomial data</td>
<td></td>
</tr>
<tr>
<td>C. Poisson data</td>
<td></td>
</tr>
<tr>
<td>D. Likelihood-based inference for these models</td>
<td></td>
</tr>
<tr>
<td>E. Testing goodness of fit for multinomial data</td>
<td></td>
</tr>
<tr>
<td>F. Models with for overdispersed data</td>
<td></td>
</tr>
<tr>
<td>2. Contingency Tables</td>
<td>2</td>
</tr>
<tr>
<td>A. Structure of tables</td>
<td></td>
</tr>
<tr>
<td>B. Comparing proportions</td>
<td></td>
</tr>
<tr>
<td>C. Relative risk and odds ratio</td>
<td></td>
</tr>
<tr>
<td>D. Types of studies</td>
<td></td>
</tr>
<tr>
<td>E. Chi-squared tests</td>
<td></td>
</tr>
<tr>
<td>F. Testing independence for ordinal data</td>
<td></td>
</tr>
<tr>
<td>G. Exact inference for small samples</td>
<td></td>
</tr>
<tr>
<td>H. Three-way contingency tables, marginal and partial association</td>
<td></td>
</tr>
<tr>
<td>3. Generalized Linear Models</td>
<td>3</td>
</tr>
<tr>
<td>A. Components of GLMs</td>
<td></td>
</tr>
<tr>
<td>B. GLMs for binary data</td>
<td></td>
</tr>
<tr>
<td>C. GLMs for count data</td>
<td></td>
</tr>
<tr>
<td>D. Poisson regression</td>
<td></td>
</tr>
<tr>
<td>E. Model inference based on the likelihood</td>
<td></td>
</tr>
<tr>
<td>F. Model checking and model selection</td>
<td></td>
</tr>
</tbody>
</table>

Test 1
### Course Outline

<table>
<thead>
<tr>
<th>Topic</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4. Logistic Regression with Dichotomous Responses</strong></td>
<td>4</td>
</tr>
<tr>
<td>A. Interpreting logistic regression</td>
<td></td>
</tr>
<tr>
<td>B. Inference for logistic regression</td>
<td></td>
</tr>
<tr>
<td>C. Logistic regression with qualitative predictors</td>
<td></td>
</tr>
<tr>
<td>D. Multiple logistic regression</td>
<td></td>
</tr>
<tr>
<td><strong>5. Building and Applying Logistic Regression Models</strong></td>
<td>5</td>
</tr>
<tr>
<td>A. Strategies in model selection</td>
<td></td>
</tr>
<tr>
<td>B. Classification tables and ROC curves</td>
<td></td>
</tr>
<tr>
<td>C. Model checking</td>
<td></td>
</tr>
<tr>
<td>D. Other topics in logistic regression</td>
<td></td>
</tr>
<tr>
<td><strong>6. Multicategory Logit Models</strong></td>
<td>6</td>
</tr>
<tr>
<td>A. Logit models for nominal responses</td>
<td></td>
</tr>
<tr>
<td>B. Cumulative logit models for ordinal responses</td>
<td></td>
</tr>
<tr>
<td><strong>Test 2</strong></td>
<td></td>
</tr>
<tr>
<td><strong>7. Loglinear Models for Contingency Tables</strong></td>
<td>7</td>
</tr>
<tr>
<td>A. Loglinear models for two-way and three-way tables</td>
<td></td>
</tr>
<tr>
<td>B. Inference for loglinear models</td>
<td></td>
</tr>
<tr>
<td>C. Loglinear-logistic connection</td>
<td></td>
</tr>
<tr>
<td>D. Association graphs and collapsibility</td>
<td></td>
</tr>
<tr>
<td>E. Modeling ordinal associations</td>
<td></td>
</tr>
<tr>
<td><strong>8. Models for Matched Pairs</strong></td>
<td>8</td>
</tr>
<tr>
<td>A. Comparing dependent proportions</td>
<td></td>
</tr>
<tr>
<td>B. Logistic regression for matched pairs</td>
<td></td>
</tr>
<tr>
<td>C. Comparing margins of square contingency tables</td>
<td></td>
</tr>
<tr>
<td>D. Symmetry and quasi-symmetry for square tables</td>
<td></td>
</tr>
<tr>
<td>E. Analyzing rater agreement</td>
<td></td>
</tr>
<tr>
<td>F. Bradley-Terry model for paired preferences</td>
<td></td>
</tr>
<tr>
<td><strong>9. Modeling Correlated, Clustered Responses</strong></td>
<td>9, 10</td>
</tr>
<tr>
<td>A. Generalized linear mixed models</td>
<td></td>
</tr>
<tr>
<td>B. Marginal models versus conditional models</td>
<td></td>
</tr>
<tr>
<td>C. Marginal model with generalized estimating equations</td>
<td></td>
</tr>
<tr>
<td>D. Random effects modeling of clustered categorical data</td>
<td></td>
</tr>
<tr>
<td><strong>Final Exam</strong></td>
<td></td>
</tr>
</tbody>
</table>