ILRST 515: STATISTICAL RESEARCH METHODS

INSTRUCTOR: WILLIAM M BRIGGS

1. Class Organization

Instructor	William M Briggs, brigg1wm@cmich.edu.
Text (free)	*Breaking The Law of Averages.* Download a free PDF at `wmbriggs.com/book/`.
Project	Each student will complete a project and project report, to include gathering data that can be manipulated and analyzed by the methods taught in class.

2. Class Goal

Statistics has undergone in the last 20 years a number of profound changes, both in its fundamental philosophy, and in the practical way in which computations are carried out. The good news is that these changes have made it possible to solve problems that were impossible to solve before. The bad news is that these changes have not yet made their way into every field. The concentrations of physics, astronomy, various engineering disciplines, meteorology, and so on, have all seen the new methods of statistics. The various fields of business, sociology, psychology, medicine, and so on have not seen as much. Which means that students have to learn both the old ways and the new ways. This means hard work on your part, but work which will pay off, because learning the old and and the new will help you really appreciate—and assimilate—the ideas.

The *most important* goal of this class is for you to be able to *understand* the meaning and interpretation of probability, probability models, and the results from statistical analyses. We will not concentrate on hand calculations as is more traditional. We will instead commit to understanding output from standard computer programs and how statistics is—and should be—really used.

It is crucial for us to understand the definitions and interpretations of probability, randomness, and statistics; why they are different, and how they are related. The definition and the modern (Bayesian) and classical interpretation of probability models will be given (statistics is the creation and use of probability models!).

Statistics is all about making sense of evidence to argue for or against some claim. For real life claims, there is *never* a set of fixed procedures that you can "just follow" to get an answer. You will always have to think hard about every problem.

This is not a hard-core math class. You will not be proving any theorems or working through a ton of abstract formulae. There are very few equations that you will be asked to remember. Most of the math you will see will be presented fully born; and your job will be to understand the philosophy behind it, usually explained through pictures. You will have to generalize the situations learned in

class and in homework and apply them to novel scenarios. This kind of activity takes a lot of work, so be prepared to think.

If you want to know how you are doing in the class—a frequent question—look at the assigned material. If you can understand and work the problems, you are doing well; if not, you are doing poorly.

The outline below will be followed, though not necessarily exactly. We'll slow down or speed up as warranted.

3. Class Outline

Lesson 1	The logic of probability introduced and defined: the first two rules, addition and multiplication with discrete probabilities. The mystical term *randomness* is demystified. Statistics itself defined.
Lesson 2	Probability: the second two rules (total probability and other rules), breaking statements and events apart and Bayes's theorem.
Lesson 3	How to count; and how counting leads to a simple probability distribution, the binomial.
Lesson 4	Writing distributions using the short-hand notation of variables. How to picture probability graphically. The biggest distribution of them all, the normal, is shown in all its glory.
Lesson 5	Introduction to the R open source and freely available software package. We go through the simple (and sometimes complicated) syntax using basic examples. A guide to downloading and installing the software (for Linux, Mac, and even Windows) will be given. R is infinitely expandable and has a growing number of packages for specific types of data analysis.
Lesson 6	Some generalizations of the normal, how to easily compute it using software.
Lesson 7	The top 3 ways of screwing up your data; how to really store and manipulate your data, and how to make simple summaries.
Lesson 8	Probability up to now has been written like we knew all about it. We almost never do, however, so we have to find a way to estimate the probabilities we use. The two main philosophies of how this is done are given for the normal distribution.
Lesson 9	Same thing as last week, except for the binomial distribution, where we define the "probability of a probability."
Lesson 10	How can you tell if one thing, like an ad campaign, gives better results than an another thing, such as a different ad campaign. This, understandably, is called testing. The times when testing makes sense, and when it doesn't, are detailed. The "p-value" and magic number are introduced. All this is for the normal distribution.
Lesson 11	The same as last week, except for the binomial.
Lesson 12	How to cheat at statistics. Also, how other people can cheat you in surveys and polls.
Lesson 13	A straight line is shown; then a probabilistic straight line is contrasted. The later is called a regression line.
Lesson 14	More on regression; different kinds and types of variables. How testing and estimation are often simple cases of regression. Lots and lots of regression examples, paying strict attention to the interpretation of the model parameters.
Lesson 15	More on regression; same as before except for logistic regression..
Presentation	Each student will present his project, including graphs, tables, explanations, a regression model (logistic or linear), and must demonstrate an understanding of the software output, including the frequentist and Bayesian interpretations.