

# ISTA 116: Statistical Foundations for the Information Age

Mihai Surdeanu

Last Revised January 9, 2013

## 1 Course Information

### 1.1 Course Description

Data is everywhere in the modern world. Each of us has access to data on virtually any topic imaginable via the internet, and public and private institutions are making use of data in decision-making more than ever before. This trend will only accelerate. However, access to data is not the same thing as access to information. The purpose of this course is to develop some of the foundational skills needed to consume data and create information. The main theme in the course is understanding the variability inherent in data, and the inherent uncertainty associated with conclusions drawn from data.

The first part of the course is dedicated to descriptive statistics: taking a data set, visualizing and summarizing its key features. How often does each value occur, what are the “typical” values, how much variation does it contain, and what are the relationships among the variables? In the second part, we develop the foundations of inferential statistics, namely the basic principles of probability and the behavior of random samples. Finally, we discuss how to evaluate how strong the evidence is for simple claims about the data, such as “homeopathic headache medicine reduces pain more effectively than sugar”, or “children’s income is correlated with parent’s income”.

Throughout the semester, the lab component of the class will introduce computational techniques (using the statistical language R) to make our machines do the

tedious work of counting, arithmetic, and generating random numbers. That is, until they rebel and enslave us.

## 1.2 Prerequisites

MATH 109C or MATH 112 or placement beyond the level of College Algebra.

## 1.3 Locations and Times

Lectures:

Monday/Wednesday 4:30pm - 5:45pm in Saguaro Hall, Room 202

Labs:

Tues 12:00 – 1:50 in ECE 229

Tues 2:00 – 3:50 in ECE 229

Tues 4:00 – 5:50 in Shantz 338

## 1.4 Texts

The main text for the lectures will be:

Griffiths, Dawn. (2008). *Head First Statistics*. Sebastopol, CA: O'Reilly. <http://universityofarizona.worldcat.org.ezproxy1.library.arizona.edu/title/head-first-statistics/oclc/297556511> (available for free electronically through UA library)

There is no text for the lab component, but the lab instructors will provide pdf notes each week. In addition to the lab notes, the following is a good reference for R:

Owen, W. J. (2010). *The R Guide, ver. 2.5*. Department of Mathematics and Computer Science, University of Richmond, Richmond, VA. <http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf> (available for free electronically)

For those who would like to go a bit deeper (in either statistics or R, but especially R), the following book has good coverage:

Cohen, Y. and Cohen, J. (2008). *Data and Statistics with R*. Wiley & Sons. (available free through the UA library).

## 1.5 Instructor Information

Mihai Surdeanu  
Email: [msurdeanu@email.arizona.edu](mailto:msurdeanu@email.arizona.edu)  
Office: Gould-Simpson 811  
Office Hours: M/W 9:15 – 10:30

## 1.6 Lab Instructors

Nathan Dykhuis  
Email: [ndykhuis@cs.arizona.edu](mailto:ndykhuis@cs.arizona.edu)  
Office: Gould-Simpson 918  
Office Hours: M/W 3:30 – 4:30

Richard Reilly  
Email: [richardr@email.arizona.edu](mailto:richardr@email.arizona.edu)  
Office: Gould-Simpson 228  
Office Hours: Tue 2:30 – 4 and F 12:30 – 2

## 2 Course Policies and Grading

### 2.1 Grading

Grades are based on homework, quizzes, one midterm, and a comprehensive final exam. The grading scheme is as follows:

Component	Weight
Homework	300 pts
In-Class Quizzes	150 pts
Web Assignments	50 pts
Term Paper	100 pts
Midterm Exam	200 pts
Final Exam	200 pts
Total	1000 pts

Grade	Point Range
A	900-1000
B	800-899
C	700-799
D	650-699
E	0-649

### Grade Disputes

Disputes about grades on a particular assignment or exam will be entertained for two weeks from the day the assignment is returned, or 1 day before grades are due, whichever is sooner. These will be resolved by re-grading the entire assignment or exam. Note that this can result in a lower grade in the event that new mistakes are discovered. The final exam will be graded and made available for review by students within 24 hours of its completion, to allow time for any requested regrades.

**No negotiations about individual students' letter grades will be entertained once final grades are assigned, except as permitted by the policy stated above.**

### 2.2 Homework Assignments

There will be 7 homework assignments, consisting of a mixture of paper-and-pencil and computer-based problems. Individual assignments are worth 50 points. Assignments are graded for correctness; however, genuine effort is nearly always worth partial credit.

## Collaboration Policy

Students are encouraged to work together, both in class / lab / office hours and otherwise, to understand problems and general approaches for solutions. However, **final write-ups of solutions must be done individually**. Any collaboration that takes place outside section or office hours must be identified, along with the nature of the collaboration (e.g., “we worked together”, “she helped me”, “I helped him”). **Copying another person’s answers, work or code is not permitted, regardless of collaboration status, and will be treated as a case of academic dishonesty.**

## Late Policy

Assignments are due electronically via d2l by the stated deadline. Permission for an extension must be granted by the lab instructor *in advance* of the deadline in order to receive full credit for a late assignment. The first request by a given student is likely to be granted; the probability decreases with each subsequent request. No homework will be accepted once solutions are posted online.

## Calculation of Final Homework Grade

Each student’s final homework grade will be based on the six highest homework scores, provided all but at most one assignment is completed. “Completion” is judged by the lab instructors, and is based on an honest attempt to answer nearly all of the questions. If more than one assignment is incomplete, the homework grade will be the total number of points earned, divided by the total points possible, times 300.

## 2.3 Quizzes

There will be a short (approximately 5 min) quiz during most classes starting in week 2, based on the reading for that class. Each quiz will consist of 1-2 straightforward (multiple choice, fill-in, etc.) questions, typically involving simple definitions. Quizzes are worth 10 pts each, some of which are awarded just for showing up. The 15 highest quiz grades count toward the final grade.

**Students may be administratively dropped from the class if they miss more than 2 quizzes in the first 3 weeks, or more than 1/3 of all quizzes by the administrative drop deadline, unless prior arrangements have been made with the instructor.**

## 2.4 Web Assignments

Web assignments will be due in d2l (generally) during weeks when there is no written homework due. These are intended to provide extra practice with statistical concepts, divorced from computational concerns. These assignments can be redone up to 3 times; the highest score will count. Students are encouraged to make at least their first attempt without any reference materials.

## 2.5 Exams

One in-class midterm exam will be given during the lecture on March 6th. A comprehensive final exam will be given on May 3<sup>rd</sup> from 3:30 to 5:30 PM. The format of the exams will be mixed, and may include multiple choice, short verbal answer, and “math problems”. Each exam is graded out of 200 pts. The higher of the two scores will be upweighted at the end of the semester to be out of 225 points, and the lower downweighted to be out of 175.

### Permitted Materials

Students will be allowed to use one double-sided  $8\frac{1}{2} \times 11$ ” page of notes for the midterm, and two pages for the final. A simple, self-contained hand calculator may also be used. **No other electronic devices of any kind, including cell phones, iPods, laptops, Borg implants, etc., will be permitted during the exams.**

### Make-up Policy

**Exams may only be made up in case of a serious, unanticipated emergency.** The student or an appointed proxy must contact the instructor in advance of the exam. Clearly worded documentation, from a doctor, dean, etc., demonstrating

physical inability to take the exam at the appointed time, will be required to grant a make-up appointment. No make-ups will be granted for personal reasons, including travel or personal hardship. Any make-up exams that are granted must be scheduled for after the original time.

## 2.6 Term Paper

There will be one short (roughly 4-6 page) paper due on April 26<sup>th</sup>. The purpose of the paper is to find a data set on your own and say something about it. This could involve a critique of the reporting of statistics in the media, or it could involve the use of a data set to address a new question that interests you.

## 3 University Policies

### Classroom Behavior

Students are expected to behave respectfully toward each other and to the instructor and TAs. Disrespectful behavior includes the use of cell phones or other electronic devices in the classroom during class hours.

The Arizona Board of Regents Student Code of Conduct is here: <http://dos.web.arizona.edu/uapolicies/scc5308abcd.html#sccphilosophy>

ABOR Policy 5-308, prohibits threats of physical harm to any member of the University community, including to oneself. See: <http://policy.web.arizona.edu/~policy/threaten.shtml>.

### Special Needs and Accommodations

Students who need special accommodation or services should contact the

Disability Resources Center  
1224 East Lowell Street, Tucson, AZ 85721  
(520) 621-3268  
FAX (520) 621-9423  
email: [uadrc@email.arizona.edu](mailto:uadrc@email.arizona.edu)  
web: <http://drc.arizona.edu/>.

You must register and request that the Center or DRC send official notification of your accommodations needs as soon as possible. Please plan to meet with the instructor by appointment or during office hours to discuss accommodations and how the course requirements and activities may impact your ability to fully participate. The need for accommodations must be documented by the appropriate office.

### Student Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected

to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://dos.web.arizona.edu/uapolicies/>.

### **Confidentiality of Student Records**

See <http://www.registrar.arizona.edu/ferpa/default.htm>

### **Subject to Change Statement**

Information contained in this syllabus, other than the grade and absence policy, may be subject to change by the instructor, with advance notice.

## 4 Schedule

### 4.1 Tentative Schedule of Lecture Topics

Topic	Approx. Dates	Readings	Add'l Lab Readings
Introduction	Jan 9	Course Syllabus Ch. 1: pp. 2-4	Owen Ch. 1, 2.1-2
Categorical Data	Jan 14, 16, 23	Ch. 1: pp. 5-16, Lecture Notes	Owen 3.3, 4.1
Frequency Distributions	Jan 28	Ch. 1 pp. 17-43	
Central Tendency	Jan 30, Feb 4	Ch. 2	Ch. 4.2
Variability	Feb 6, 11	Ch. 3	
Sampling	Feb 13	Ch. 10: pp. 415-431	
The Logic of (Classical) Hypothesis Testing	Feb 18, 20	Lecture Notes	
Probability Basics	Feb 25, 27	Ch. 4: pp. 127-155	
Catch-up Day	Mar 4	—	
<b>MIDTERM EXAM</b>	<b>Mar 6</b>	—	
Conditional Probability	Mar 18, 20	Ch. 4: pp. 156-195	
Discrete Random Variables	Mar 25, 27	Ch. 5	
The Geometric and Binomial Distributions	Apr 1	Ch. 7: pp. 269-305	Owen 6.1, 6.3.1
Hypothesis Testing: Binary Data	Apr 3, 8	Ch. 13: pp. 521-544	
Continuous Random Variables: the Normal Distribution	Apr 10	Ch. 8	Owen 6.3.2
Distributions of Sums & Means; Bootstrapping	Apr 15, 17, 22	Ch. 9: pp. 361-380, Lecture Notes	
Confidence Intervals	Apr 24	Ch. 12	
Correlation & Regression	Apr 29, May 1	Ch. 15	
<b>FINAL EXAM</b>	<b>May 3</b>	—	

## 4.2 Tentative Schedule of Lab Topics

Date	Topics
Jan 15	Intro and R Fundamentals
Jan 22	R Fundamentals Cont'd, Tables, Plotting Bar Charts, Pie Charts
Jan 29	Contingency Tables
Feb 5	Histograms & Density Curves
Feb 12	Summary Statistics, Visualization, Boxplots
Feb 19	Sampling
Feb 26	Scripts and Functions
Mar 5	<i>Midterm Review</i>
Mar 12	Simulating Random Experiments
Mar 19	Simulating Conditional Probability
Mar 26	PDF, CDF, and random functions
Apr 2	Sampling Distributions: Proportions
Apr 9	Normal Distribution
Apr 16	Sampling Distributions: Sums & Means
Apr 23	Bootstrapping; Confidence Intervals
Apr 30	<i>Catch Up &amp; Review</i>

### 4.3 Homework and Web Assignments

All assignments are due in the d2l dropbox by 5:00 P.M. on the indicated day.

Assignment	Topics	Due Date
HW 0	Installing/Starting R	Jan 18
WA 1	Course Policies	Jan 18
HW 1	R Fundamentals	Jan 25
WA 2	Categorical Data	Feb 1
HW 2	Contingency Tables	Feb 8
WA 3	Central Tendency & Variability	Feb 15
HW 3	Distributions & Summary Statistics	Feb 22
WA 4	Probability Basics	Mar 1
	<i>Midterm week</i>	
HW 4	Probability Continued	Mar 15
WA 5	Discrete RVs	Mar 22
HW 5	Simulation	Mar 29
WA 6	Normal Probabilities	Apr 5
HW 6	Hypothesis Testing	Apr 12
WA 7	Sampling Distributions & Confidence Intervals	Apr 19
HW 7	Sampling Distributions	Apr 26
	<i>Review week</i>	