Geog 2000: Introduction to Geographic Statistics

Geog 2000 Spring 2011

Course Syllabus

Instructor: Dr. Paul C. Sutton
Office Hours: By appt or Tue and Thur 9-9:50 or stop by my office and take your ***chances*** ☺
Office: #129 Boettcher West Phone: (303) 871-2399 E-mail: psutton@du.edu

Lecture: T Th 2:00 – 2:50 (Boettcher West 126)
Lab: T Th 3:00 – 3:50 (Boettcher West 126)
Required Texts: 1) *The Cartoon Guide to Statistics* by Larry Gonick
2) *How to Lie with Statistics* by Darrell Huff
3) *An introduction to Statistical Problem Solving in Geography*
By Chapman McGrew Jr. and Charles B. Monroe

Course Description

This course introduces the basic concepts of probability and statistics with an emphasis on applications and an ongoing focus regarding the nature and problems associated with spatial or geographic data. Topics covered include: 1) Characterizing univariate and bivariate data, 2) basic ideas of probability and probability distribution functions with particular emphasis on the normal distributions and important spatially relevant non-normal distributions, 3) Sampling distributions and Hypothesis testing , 4) Chi-Square tests with non-parametric data; and, 5) Correlations and Ordinary Least Squares linear Regression. Some hand calculations will be performed to emphasize conceptual understanding; however, mastering the use of a statistical software package (JMP) is an important part of this course that is mostly covered in the lab sections where we work on the problems sets. The lab sessions will center on: 1) Helping students complete the problem sets, 2) Using the JMP software, and 3) Administering the three exams.

A Note on Universal Design (UD) Principles: Research has shown over and over again that people learn in different ways. Pedagogic principles of UD are about teaching in a way that serves as many of the learning styles of the different students in the course as is reasonably possible. Some of us are visual learners, some of us are auditory learners, some of us are kinesthetic learners, and various combinations of strengths and weakness also exist. I will be teaching this class in a way that incorporates as many of the best practices of UD as I can. I encourage you to come visit me in my office and tell me a little bit about yourself so I can try to teach in a way that will work for you. My methods may seem unusual (I will bring toys to class), and I won't necessarily treat all students in the same way during class. If you have any questions don't hesitate to ask me. Some principles of UD are the idea of teaching in different modalities of instruction (e.g. visual, auditory, kinesthetic), and; variation in the nature of instruments used to measure mastery of the material. Some principles I have incorporated in this course that you may notice are the following:

1) Three very different texts covering much of the same material (A light broad concept book – *How to lie with statistics*; a deceptively dense, example focused practical book – *The Cartoon Guide to Statistics*; and a traditional Introductory Statistics Book – *The McGrew and Monroe book*.).
2) Individual and collaborative assignments (The three exams require individual efforts; however, they will look a lot like the problem sets which I will encourage you to complete in a collaborative manner with other students).

3) A web site with many, many links to many cool things. I teach another version of this class on-line and have harvested many URLs to videos of other instructors teaching the same statistical principles and formulas as I do (e.g. The Khan Academy). I also have cool little Java applets which graphically demonstrate many of these ideas. Some worked out problems sets will also be posted to this web site. This web site will also have powerpoint summaries of the book chapters prepared by former students. Check out these web resources for this class at:

   http://urizen.geography.nsm.du.edu/~psutton/Sutton_Courses/Geog_2000_Intro_Geog_Stats/Geog2000_Sutton.html

4) I will also incorporate some pseudo-kinesthetic approaches to some classic probability problems in the course. One will involve having everyone flip coins. Another will be little 'skits' in a 'Let's Make a Deal' game show format to attempt to clarify a fascinating and wonderful problem known as the "Monty Hall Problem" that is associated with contingent probability. I did not actually understand this problem until I actually physically simulated it with playing cards. Our minds work in mysterious ways.

Method of Grading

Exam #1 `	**(Thursday - Week #4) (covers Chs 1-4)**	**20%**
Exam #2	**(Thursday - Week #8) (covers Chs 5-9)**	**20%**
Exam #3	**(Last day of class) (Cumulative w/ emph Ch 9-12)**	**20%**
Four Problem Sets	**(Assigned weeks 1, 3, 5, & 7: DUE weeks 3, 5, 7, & 9)**	**40%**

Instructor's Note: The three exams will draw heavily from the four problem sets. Studying statistics is a funny thing. Some people just 'get it' but **MOST** people **don't**. DO THE PROBLEM SETS with eternal and constant vigilance. By taking care of that 40% of your grade the other 60% will be easy. If you don't do the problem sets you will be ***doomed*** (unless of course you are one of those *rare* people who somehow 'get it'). Based on my empirical statistical estimates only about 3% of you can get away with 'getting it'. Just something for you to consider ☺. Also, READ THE BOOKS ***several*** times. Two are short, seemingly 'light' reads. They are not really 'light' in any sense of the word. They are DENSE. There is deep information presented in perhaps an almost too 'easy to read' format. The book alternatives are horrifyingly boring though. If you are wise you will come to love and cherish these books. Graduate students beg me for them but I've stopped giving my books to them because they never give them back - and those that do have scribbled copious notes in them. These are great books. Read them a lot. Buy two copies and put one in your bathroom (These books are less than half the price of most of the other books I have considered). Learning statistics works better if you do it in a lot of short sessions rather than cramming it in all at once. It's actually fun that way too. These books let you take statistics in small doses and will make you think and go back after you have thought. But remember and live by this: Do the Problem Sets first – the text is only a supplement to your comprehension of this material. **Work** the problem sets to your own mind's understanding to ACE this class.

Tentative Schedule of Lecture Topics

Week 1 : The basic Idea of statistics and Basics of Data description

Topics covered this week include basic data description and fundamental ideas and concepts of statistical inference as a process. "*How weird is that*?" A folksy wording of what is an essential statistical question. Measures of central tendency and spread: means, medians, modes, ranges, standard deviations. Graphical devices such as the histogram and the stem and leaf diagram. Reading: Chapter 1 & 2: from the Cartoon Guide. Read all of Darrell Huff's How to Lie with statistics**. *Problem Set #1 assigned today and due in two weeks (as all other problem sets are due two weeks from their assignment).***

Week 2 : Probability

Flipping coins and rolling dice. Approaches to probability: Classical, Relative Frequency, Subjective, and Monte Carlo. Ideas of Independence and conditional probability will also be covered. Reading: Chapter 3 in Gonick's cartoon guide.

Week 3 : Random Variables and Probability Distribution Functions

Discrete vs. continuous random variables. Nominal, ordinal, interval, and ratio measurement scales and some of their impacts on how we do statistics. Probability Distribution functions. Reading: Chapter 4 of the cartoon guide. ***Problem Set #1 is collected . Problem Set #2 is assigned.***

Week 4 The Binomial and the Normal Distribution

(NOTE: 1st Mid-Term Exam on Thursday April 14th in Lab) This week I will be at the AAG (Geography's annual conference). A TA will review for exam I and administer it on Thursday. Reading: Chapter 5 of the Cartoon Guide.

Week 5 : Sampling and Confidence Intervals

We will explore the discrete binomial distribution and the continuous normal distribution. Understanding the similarities and differences of these two distributions and the concept of normalization and z-scores. The sampling distribution of the mean (yow – this takes a lot of explaining – no kidding- But it's REALLY important). The central limit theorem and its relation to the aforementioned. The student's 't' distribution (really just the Normal with less degrees of freedom – note: the tails are longer – watch out for long tails). Types of Sampling Design: Simple Random, Cluster, Stratified, and others. Confidence Intervals and Standard Errors. Reading: Chapter 6 & 7 Gonick. ***Problem set #2 is collected. Problem Set # 3 is assigned***.

Week 6 : Hypothesis Testing and Comparing Two Populations

Type I and Type II Error. The Null Hypothesis and Hypothesis Formulation from a statistical perspective. Comparing Success Rates. Independence of Means tests. Paired comparisons (usually a type of 't' test). Reading: Chapter 8 & 9: Gonick's Cartoon Guide.

Week 7 : Experimental Design and Spatial Data Problems and Prospects

Learn how to design experiments ***first***, then gather data, and then answer interesting questions. The special nature of spatial data will also be covered: pitfalls, problems, prospects, and potential. Reading: Chapter 10 from the cartoon Guide. ***Problem set #3 is collected. Problem Set # 4 is assigned***.

Week 8 : Linear Regression (aka Ordinary Least Squares)

(NOTE: 2nd Mid-Term Exam on Thursday May 12th in Lab) Conceptual introduction to Ordinary Least Squares Simple Linear Regression. Scatterplots, slopes, intercepts, and High School Algebra class revisited. Some review on Monday for the second exam. Reading: Chapter 11& 12 from the cartoon Guide.

Week 9 : Linear Regression

Scatterplots, Correlation and Causality, Parameter estimation, Y = m*x + b up the wazooo. Confidence intervals and analysis of residuals. ***Problem Set # 4 is collected***.

Week 10: The Chi-Square Test and other Statistical Methods

The Chi-Square test (and the Chi-Square Distribution) and its application to several spatial and non-spatial experimental designs involving nominal data. A broad overview of statistical methods beyond the scope of this course (multi-variate regression, ANOVA, logistic regression, Cluster analysis, spatial interpolation via Kriging, , etc). This material is NOT in the reading but WILL be on the 3rd Exam.

Note on Final Class Grade: A single score out of 100 will be created from your grades on the problem sets, and the three exams. This single composite score is the ***only*** *score that will be assigned a letter grade. No lab or exam will be assigned a letter grade. Consequently a 40% on an exam will be much more useful than a zero. Keep that in mind when you are taking exams. I predict that I will be grading on a curve. However, if everyone gets 95 out of 100 in the class you'll all get A's. I anticipate that the standard 90 and up is an A, 80-89 is a B, 70-79 is a C, 60-69 is a D, less than 60 is an F grading scale will be a conservative means of evaluating your interim performance in the class. However, if and when a curve is produced an 85 could be an A. Feel free to contact me with any questions regarding this grading procedure.*