

# STAT 100: Chance and Data Analysis

**Text:** *Statistics: A Guide to the Unknown, 4<sup>th</sup> ed.*, by R.Peck, G.Casella, G.Cobb, R.Hoerl, D.Nolan, R.Starback, H.Stern, publisher: Duxbury, 2006

## Course Outline: Keyed to the Text edited by Peck.

1. **Introduction.** Basics of Data Analysis. Distributions. Basics of Chance. Sports Leagues. Simulation. Moore xvii: Introduction. Cobb & Gehlbach 3-18: Statistics in the Courtroom.
2. **Graphics.** Gasoline Consumption. Zipf's Law. Time Series. Lambert & Pinheiro 293-306: To Catch a Thief: Detecting Cell Phone Fraud. Doganaksoy, Hahn & Meeker 339-358: Assuring Product Reliability and Safety. Kahn & Roseman 373-390: Advertising as an Engineering Science.
3. **Randomness.** Illusion of predictability. Stock Market Application. Cleary & Sharpe 359-372: Randomness in the Stock Market. Wilks 171-182: Statistical Weather Forecasting.
4. **Risk:** Insurance and Investment. Diversification. Variability. Cleary & Sharpe 359-372: Randomness in the Stock Market. Ashenfelter 407-423.
5. **Study Design.** Eddy & Smykla 211-226: The Last Frontier: Understanding the Human Mind. Weir 255-270: DNA Fingerprinting. Kahn & Roseman 373-390: Advertising as an Engineering Science.
6. **Models.** Zipfs Law. Gerow, Miquelle & Aramilev 105-118: Monitoring Tiger Prey Abundance in the Russian Far East. Matis & Kiffe 119-134: Predicting the African Bee Invasion. Madigan 135-148: Statistics and the War on Spam. Brookmeyer 197-210: Modeling an Outbreak of Anthrax. De Veaux & Edelstein 307-322: Reducing Junk Mail Using Data Mining Techniques.
7. **Sampling.** Randomized Response Technique. Hill 69-88: Evaluating School Choice Programs. Heilig, Hill & Karon 227-242: Leveraging Chance in HIV Research.
8. **Rare Events.** Lotteries. Average and actual returns. Assessment of Coincidences.
9. **Survival.** Aging of cars, cells, and people. Traffic Accidents.
10. **Spatial distributions.** Clusters Real and Imagined. Travelling Salesman. Epidemics.
11. **Quality Control.** Snee 323-338: Improving the Accuracy of a Newspaper: A Six Sigma Case Study of Business Process Improvement. Doganaksoy, Hahn & Meeker 339-358: Assuring Product Reliability and Safety.
12. **Data Mining.** Change point detection. Lambert & Pinheiro 293-306: To Catch a Thief: Detecting Cell Phone Fraud.
13. **Significance Testing.** Cobb & Gehlbach 3-18: Statistics in the Courtroom. (and previous selections).

## Lecture Schedule

T 1130-1320 AQ 3181

R 1130-1220 AQ 3182

MT I - Tuesday, Feb 2, 2010. MT II – Tuesday March 2, 2010

Final Exam: April 22, 2010 1200-1500. Location: TBA

Office Hours: T after class as needed and R 1230-1320. Office: K 10564

Email: weldon@sfu.ca

## Grading

Assignments – 20%

2 Midterms – 15% each

Final Exam – 50%

# STAT 100: Chance and Data Analysis

## More Info

Instructor Contact: Larry Weldon Office: K 10564  
Email: [weldon@sfu.ca](mailto:weldon@sfu.ca)  
Phone: There is no phone in the shared office K10564. Use email.  
Office Hours T 1320 as required, and R 1230-1330

Department Office: Department of Statistics and Actuarial Science  
K 10545, 778-782-3801  
Undergraduate Secretary: Charlene Bradbury

Statistics Workshop (SW):  
K 9516 (inside K 9510) - 778-782-4530  
Drop-in assistance from TAs  
Assignment submission and pickup  
TAs mark assignments  
The SW will be open most week days 930-1630.  
SW Manager: Robin Insley. K 10552. 778-782-3805

Midterms and Final Exam: Open Book, notes, etc.  
Dates of Midterms: MT1 – Feb 2, 2010: MT2 – Mar 2, 2010  
Final Exam – April 22, 2010

Computing:  
Although all statistical work is done these days using statistical software, I will attempt to introduce statistical concepts and techniques without requiring students to actually use statistical software. However, for students who are keen to learn this aspect of statistics, I will provide opportunities for students to do this.

Assignment Deadlines:  
Timely marking of 100+ assignments requires some careful organization. Assignment deadlines must be enforced for administrative reasons. Be forewarned - late assignments will not be marked.

Feedback:  
I welcome your feedback during the course – you can provide this either directly to me (in person or email) or via the Statistics Workshop.

Objective:  
My objective in this course is to convey the utility and charm of the discipline of statistics, while assuming a minimum of prerequisite knowledge. You will learn the basic concepts and techniques while extracting information from certain data sets.

**Introduction by David Moore:** Text pp xvii-xxiv

“Statistics is the Science of Learning From Data.” or,

Statistics is the Science of Extracting Information From Data.

Data need not be numerical. Opinions are data, phenomena are data, ...

**Where the data come from matters.**

The results of a study cannot be judged without knowledge of how the study was conducted.

Q1: What is the difference between an observational study and an experiment?

Q2: What advantage does an experiment have over an observational study?

Q3: Why does non-response in a survey sample cause problems for inferring valid information.

Q4: What sort of investigators can be expected to produce reliable information?

**Always Plot your Data.**

Q5: Why does the graph on page xxi suggest that many voters in Palm Beach County intended to vote for Al Gore rather than Pat Buchanan?

**Fighting the Curse of Variation**

(Statistics' principal job as a discipline is to provide tools to allow for variation in data when extracting information from that data.)

Q6: What beneficial effect does averaging have in data summary?

Q7: Why are large samples “better” than small samples?

Q8: What can be done in assessing a group difference if there are multiple characteristics varying among members of both groups (and we are interested in one particular characteristic)?

Q9: What is the role of *Probability* in describing the outcome of statistical studies?

**Statistics: A Guide to the Unknown**

Q10: Why do so many statistical studies seem to produce contradictory results?

**\*\*\*Assignment 1:** (Due in Boxes 4pm Tuesday January 12, 2010)

Give **brief** answers to the ten questions above. The answers may be found (or implied) in the introductory article by Moore (text pp xvii-xxiv).