

STAT 1000
Basic Statistical Analysis I
Winter 2009

Calendar Description

(Formerly 005.100) An introduction to the basic principles of statistics and procedures used for data analysis. Topics to be covered include: gathering data, displaying and summarizing data, examining relationships between variables, sampling distributions, estimation and significance tests, inference for means. *Not to be held with STAT 1001, STAT 2220 (or the former 005.222).* Prerequisite: Any grade 12 or 40S Mathematics, or equivalent.

Teaching Philosophy and Goals

It is the desire of the Department of Statistics to present this course in a manner that emphasizes and illustrates the statistical analysis arising from “real-world” applications. Upon completion of this course students can proceed in many directions: to further intensive study of statistics, to one or more additional courses in statistics, to the use of statistical methods in other fields of study, to being a consumer of statistical information in daily life. It is our objective to serve all of these diverse directions.

The course is designed to include basic topics deemed crucial for problem formulation and understanding of the foundations of statistical thinking and reasoning. The concepts of statistical analysis will be stressed. The course will place an emphasis on the development of critical thinking skills.

To aid in the analysis of data, extensive and intelligent use will be made of the computer — with virtually every assignment involving the computer in some fashion. The computer package that has been selected for this course, JMPse (student edition), is easy to use and is available for use with Macintosh or Windows systems. The package also has many advanced statistical features that you will find useful in subsequent courses.

Whenever possible, we will attempt to bring real-life examples and data into the classroom. This will be done using, as appropriate, videotape clips showing statisticians at work, newspaper articles, in class demonstrations and experiments, and the like.

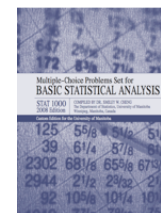
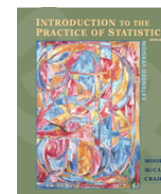
We are interested in feedback from you. If you can think of ways in which this course could be improved, please let us know.

Text & Supplementary Material (Required)

Introduction to the Practice of Statistics (IPS), David S. Moore, George P. McCabe and Bruce A. Craig, 6th edition, W.H. Freeman, New York, 2009. There are two options for purchasing the required material:

Option 1: (ISBN 1-4292-2628-5) This is the full package of a new textbook with all the assorted supplementary material. This includes the following components: (i) the book and a CD to accompany the book (containing statistical applets, tables, data sets, supplementary material and companion chapters); (ii) the StatsPortal access card (which gives you 1-year access to the electronic version of the book, with associated tools such as StatTutor, the Study Guide and the JMPse Manual), (iii) the JMPse CD.

Option 2: (ISBN 1-4292-2926-3) This option includes all of the items in Option 1, except for the textbook (although the electronic version of the textbook is included). This may be a good option if you happen to have a copy of the book, or do not wish to have a hard copy, but would like access to the electronic supplements and the JMPse CD. This option must be asked for at the check-out counters in the bookstore.



Optional Material

Multiple-Choice Problems Set for Basic Statistical Analysis I, STAT 1000, Compiled by Dr. Smiley W. Cheng. (2008 Edition) The Department of Statistics, University of Manitoba; Winnipeg, Manitoba, Canada (ISBN: 0-536-78492-2). You should find this “problems set” to be useful when studying for the term test and final exam. It contains term tests and final exams for this course for recent years. It also contains a collection of multiple-choice questions from previous introductory courses in statistics.

Computer Package

JMPse (version 6). As noted above, this statistical software is bundled with the text. It can be installed on either Macintosh or Windows machines. The system requirements are on the CD.

In addition, there are many computers on campus that can be used for running JMPse. In particular, the Department of Statistics has a number of Macintosh computers and PC's in the Statistics Lab (Room 311 Machray Hall) that you may use and the software is also available on the PC's in the “open area” ACN computer labs. More details will be given in class.



Mark Breakdown

Assignments	15%
Term Test	35%
Final Examination	50%

There are no firm cut-offs for grades. However, *subject to the caveat in the paragraph below*, the following are “guarantees” to you: A+ (≥ 90), A (≥ 80), B+ (≥ 75), B (≥ 70), C+ (≥ 65), C (≥ 60), D (≥ 50). This means, for example, that if you obtain 80% or more, you will receive no worse than an A.

However, there is an additional requirement for obtaining a C or a D in the course. **To obtain a grade of C or better, you must obtain at least 45% on the final examination; to obtain a D you must obtain at least 35% on the final examination.**

Assignments

The assignments in this course will be done in **WebAssign**, an on-line assignment system. There will be 12 weekly assignments. Assignments will be **due at 11:59 p.m.** on the due dates.

It is very important that you complete all 12 assignments. However, only the best 10 of the 12 assignments will be used to determine your Assignment Mark. **Consequently, assignment extensions will not be given.**

It is important that you do lots of problems on a regular basis. For extra practice you should do questions from the text; the answers to most of the odd-numbered questions are given in the back of the book and many detailed solutions are given in the *Study Guide*.

Test and Examination

The term test will be 1.5 hours in duration and will be held on Saturday, February 28, commencing at 9:30 a.m., as noted in the **Aurora Student Online Class Schedule**. It will cover the material in Modules I and II. The final examination will be 2 hours in duration and will be scheduled by the Student Records Office. It will cover Modules I–V, with emphasis on Modules III, IV & V.

The term test will be all multiple-choice. The final examination will contain both multiple-choice questions and a written component, in an approximate 70:30 ratio. For the test and examination: (i) non-programmable hand-held calculators are permitted (graphing calculators are not permitted), (ii) electronic devices, such as cell phones or headphones, are prohibited, (iii) statistical tables will be provided, if required, (iv) NO formulas will be provided.

Videotape Series

There are two series of videotapes (*Statistics: Decisions Through Data* and *Against all Odds: Inside Statistics*) that are quite closely related to the text. Your instructor may show some of these during the classes. However, two copies of the complete set of the second title are also available for viewing in the Science Library; tapes and headphones may be obtained at the Reserve Desk. Each tape is approximately 26 minutes in length.

Demonstrator Hours

In the Statistics Lab in Room 311 Machray Hall (which contains a number of computers), graduate students and senior undergraduate students in statistics are available to help you at the following times:

Mondays & Thursdays	9:00 a.m.–4:00 p.m.
Tuesdays & Wednesdays	9:00 a.m.–7:00 p.m.
Fridays	9:00 a.m.–12:00 noon

Voluntary Withdrawal

Note that the voluntary withdrawal date is March 19, 2009 (by which time you will have received your marks for the test and several assignments).

Academic Dishonesty

It is important that you understand what constitutes academic dishonesty and that you are familiar with the very serious consequences. Links to resources that describe academic dishonesty (including plagiarism, cheating, inappropriate collaboration and examination impersonation) can be found at:

<http://www.umanitoba.ca/faculties/science/student/webdisciplinedocuments.html>, or through the Faculty of Science home page at:

<http://www.umanitoba.ca/faculties/science>

Typical penalties imposed within the Faculty of Science for academic dishonesty are also described.

Course Content

The following is a non-exhaustive list of topics. Most of these are covered in the text. The course covers material in Chapters 1–8. However, some material is omitted: decomposing time series (pp. 19–21), density estimation (p. 71), scatterplot smoothers (pp. 92–93), transforming relationships (pp. 119–121), residual plots (pp. 128–129), data mining (pp. 136–137), §2.5, capture-recapture sampling (pp. 220–221), §4.4, §4.5, the continuity correction (pp. 326–327), the Weibull distributions (pp. 344–346), the bootstrap (p. 368), §6.4, the power of the t-test (pp. 433–435), inference for non-normal populations (pp. 435–440), §7.2–§7.3, §8.2, plus-four confidence interval for a single proportion (pp. 491–493).

Module I Looking at Data (3.5 weeks)

Chapters 1 and 2 (omit §2.5)

Unit 1 Examining Distributions

Types of Variables

- quantitative, categorical, continuous, nominal, ordinal

Displaying Distributions with Graphs

- graphs for categorical variables, stemplots, histograms, examining distributions, dealing with outliers, time plots

Describing Distributions with Numbers

- mean, median, quartiles, percentiles, interquartile range, range, variance and standard deviation
- five-number summary and boxplots
- the 1.5 x IQR rule for suspected outliers, modified boxplots
- resistant measures
- changing the units of measurement

Introduction to **JMPse** — a Statistical Computer Package

Unit 2 Density Curves and Normal Distributions

- mean and median of a density curve
- normal distributions
- 68-95-99.7 rule
- standardizing observations (z -scores)
- normal distribution calculations
- use of normal quantile plots

Unit 3 Examining Relationships

Scatterplots, Correlation, Regression

- association, response variable, explanatory variable
- examining scatterplots
- adding categorical variables to scatterplots
- categorical explanatory variables
- correlation, properties of correlation
- least-squares criterion and least-squares regression line
- r^2
- residuals, outliers, influential observations
- cautions about correlation and regression
- association vs. causation, lurking variables
- extrapolation, causation, confounding, establishing causation

Module II Producing Data (1.5 weeks)

Chapter 3 (§3.1–§3.2)

Unit 4 Design of Experiments

- observations vs. experiment
- experimental units, subjects, treatments
- factors, levels
- placebo effect, control group, bias
- randomization, principles of experimental design
- statistical significance
- randomized comparative experiments
- matched pairs design, block design

Unit 5 Sampling Design

- populations and samples
- voluntary response sample
- simple random sample
- probability sample, stratified random sample, multistage sample
- undercoverage, nonresponse, response bias

Term Test covers material in Modules I and II.
The test is on Saturday, **February 28, 2009**
from 9:30 a.m. to 11:00 a.m.

Module III A Bridge to Inference (2.5 weeks)

§3.3, §3.4, Ch. 4 (§4.1–§4.3), Ch. 5

Unit 6 Toward Statistical Inference

- sampling distributions
- bias and variability
- sampling from large populations
- why randomize?
- ethics

Unit 7 Randomness, Probability, Sampling Distributions

- randomness, the language of probability
- probability models, sample space, events
- some probability rules, independence
- discrete and continuous random variables
- normal distributions as probability distributions
- sampling distributions for counts and proportions
- binomial formula
- sampling distribution of a sample mean
- Central Limit Theorem

Module IV Introduction to Statistical Inference (2.5 weeks) Chapter 6 (§6.1–§6.3)

Unit 8 Confidence Intervals for a Population Mean (σ known)

- margin of error
- effect of sample size, confidence level, standard deviation (σ)
- effect of population size
- assumptions
- choosing the sample size

Unit 9 Tests of Significance

- hypotheses, test statistic, P -value, statistical significance
- tests for a population mean (σ known)
- two-sided tests and confidence intervals
- use and abuse of tests

Module V Additional Topics in Inference (2.5 weeks) §7.1, §8.1

Unit 10 Inference for the Mean of a Population (σ unknown)

- one-sample t -procedures — confidence intervals and tests
- matched pairs t -procedures
- robustness of the t -procedures

Unit 11 Inference for a Population Proportion

- sampling distribution of the sample proportion \hat{p}
- confidence intervals and tests
- choosing the sample size

Final Examination covers material in Modules I–V, with emphasis on Modules III, IV & V.

Module VI Transition (0.5 weeks)

Unit 12 Wrap-up & Looking Ahead

- warnings (non-random samples and inference, descriptive vs. inferential statistics, practical vs. statistical significance, etc.)
- re-iteration of some of the important ideas
- further courses in statistics, programs in statistics, careers in statistics
- a recently-introduced course in probability (STAT 2400); see below

After STAT 1000

After you have completed STAT 1000, you may want to take further courses in statistics, or possibly become a statistician! So, what courses should you take? Here are the options for your next course:

STAT 2000 (Basic Statistical Analysis II) is a continuation of STAT 1000.

This course, which is taken by students in many disciplines, covers more advanced statistical methods and leads to further courses in applied statistics. The only requirement to take this course is a grade of C or better in STAT 1000.

STAT 2400 (Introduction to Probability) introduces the basic concepts of probability from a rigorous mathematical perspective and provides a solid foundation for further courses in mathematical statistics. The prerequisites for STAT 2400 are a grade of C or better in STAT 1000 and in one of MATH 1700 or MATH 1690.

STAT 2000 and STAT 2400 are required courses in any honours or major program in statistics and are normally taken during the second year.

2008–2009 REGISTRATION ADVISORY

IMPORTANT NOTE TO STUDENTS FROM THE FACULTY OF SCIENCE:

It is your responsibility to ensure that you are entitled to be registered in this course. This means that you:

- have the appropriate prerequisites, as noted in the calendar description, or have permission from the instructor to waive these prerequisites;
- have not previously taken, or are concurrently registered in, this course and another that has been identified as “not to be held with” in the course description. For example, BIOL 1000 cannot be held for credit with BIOL 1020.

The registration system may have allowed you to register in this course, but it is your responsibility to check. If you are not entitled to be in this course, you will be withdrawn, or the course may not be used in your degree program. There will be no fee adjustment. This is not appealable. Please be sure to read the course description for this and every course in which you are registered.