# STAT 2000 Section A03

# Basic Statistical Analysis 2

# Winter 2016


**Instructor** Andrew Morris
333 Machray Hall
Telephone: 204-480-1073
Email: Andrew.Morris@umanitoba.ca

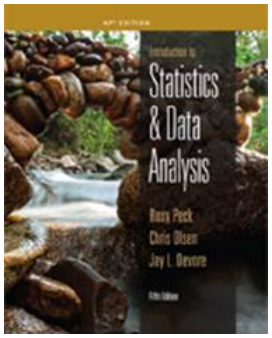

**Time** T/Th 10:00 a.m. – 11:15 a.m.
**Location** 204 Armes
**CRN** 20089

**Office Hours**
Tuesday 11:30 a.m. – 12:30 p.m.
Wednesday 11:30 a.m. – 12:30 p.m.
Thursday 11:30 a.m. – 12:30 p.m.
(or by appointment)

**Web Pages**
UM Learn: http://umanitoba.ca/umlearn
Statistics: http://umanitoba.ca/statistics

## Textbook & Resources

Peck, Roxy, Chris Olsen, and Jay Devore, *Introduction to Statistics and Data Analysis*, Fifth Edition. Cengage Learning, 2014 (ISBN: 978-1-3056-4983-5). This book has a bundled JMP access code for students who wish to use JMP on their own computers.
This course uses i►clickers for participation marks. Additional resources may be posted on UM Learn.
All course notes, assignments, tests, exams, practice exams and solutions are the intellectual property of your instructor. Reproduction or distribution of these materials is strictly forbidden without the consent of the Department of Statistics.




## Exam Information

The term test will be held **Wednesday, March 2, 2016 from 5:30 p.m. – 7:30 p.m.**. The final exam will be 3 hours in duration and will be scheduled by the Student Records Office. The final exam is cumulative, with emphasis on the material covered after the midterm exam.

***The term test will consist of only multiple choice questions, however, the final examination will contain both multiple choice questions and a written component, in an approximate 70:30 ratio.***
For the tests 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, and (iv) a formula sheet will be provided

## Evaluation

| | |
|---|---|
| Assignments | 10% |
| i▶clicker / Participation | 5% |
| Midterm Examination | 35% |
| Final Examination | 50% |

*Subject to the caveat in the paragraph below*, the following are the minimum percentage grades required to receive each of the various letter grades: A+ (90%), A (80%), B+ (75%), B (70%), C+ (65%), C (60%), D (50%).

There is an **additional requirement** for obtaining a C in the course: **to obtain a grade of C or better, you must obtain at least 50% on the final examination.**

For every i▶clicker response that you give, you will be awarded 1 point. For questions with a correct answer, an additional point will be awarded for selecting the correct response. Full marks (5/5) will be given if you receive at least 75% of the total possible i▶clicker points. Partial marks (3/5) will be given if you receive between 50% and 75%. No marks (0/5) will be given if you receive less than 50%. You are responsible for bringing your i▶clicker to class and ensuring that it has functional batteries. The assignments in this course will be done in UM Learn (formerly Desire2Learn). There will be several assignments, but the lowest assignment mark will not count towards your final grade. **Assignments will be due at 11:59 p.m.** on the respective due dates. **Note that assignment extensions will not be given to individual students.** Note that the voluntary withdrawal date is **March 18, 2016** (by which time you will have received your marks for the first two term tests and several assignments).

## Statistics Help Centre

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 (from January 11 until April 8):

| | |
|---|---|
| Monday | 9:30 a.m. – 4:30 p.m. |
| Tuesday | 9:30 p.m. – 7:00 p.m. |
| Wednesday | 9:30 a.m. – 7:00 p.m. |
| Thursday | 9:30 a.m. – 4:30 p.m. |
| Friday | 9:30 a.m. – 12:00 p.m. |

**Note:** The lab will be closed on holidays and during reading week (February 15 to February 19).

## 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; as well typical penalties) can be found at:
`http://umanitoba.ca/science/undergrad/resources/webdisciplinedocuments.html`.

See also:
`http://umanitoba.ca/student/resource/student_advocacy/cheating_plagiarism_fraud.html`.

# STAT 2000
# Basic Statistical Analysis II
# Winter 2016

## Course Content

$^*$ **indicates that a topic is not covered (or covered differently) in the textbook. Where there are any discrepancies between the way topics are covered in the course notes and in the textbook, please refer to the notes.**

**Module I: Inference for the Mean of a Single Population when $\sigma$ is Known or the Sample Size is Large; Inference for the Mean of a Single Population when $\sigma$ is Not Known** Ch. 9 (omit §9.2) and Ch. 10 (omit §10.3)

- Review of principles of statistical inference: testing and estimation, confidence intervals
- Statistical decisions: Type I and Type II errors and their probabilities, power of a test
- Review of $t$-distribution (comparison with normal distribution), tests and confidence intervals, robustness of $t$-procedures

**Module II: Inference for the Means of Two Populations** Ch. 11 (§11.1 & §11.2)

- Matched pairs $t$ procedures
- Inference for the equality of means in two populations when population variances are equal$^*$ and when population variances are unequal, assumptions of normality and independence

**Module III: Inference for the Means of Two or More Populations** Chapter 15 (§15.1)

- Graphical comparison of distributions
- Inference for the equality of means in two or more populations: introduction to ANOVA
- The $F$-distribution
- Equivalence of pooled 2-sample $t$-test and $F$-test$^*$

**Module IV: Probability and Discrete Distributions** Ch. 6 (omit §6.7) and Ch. 7 (omit §7.7 & §7.8)

- Review of probability concepts and rules
- Conditional probability
- Random variables, probability distributions, mean and variance of a random variable
- Mean and variance of the sum and difference of two independent random variables$^*$
- Distribution of the sum and difference of two independent normal random variables$^*$
- Review of binomial distribution
- Poisson distribution$^*$

**Module V: Analysis of Categorical Data and Goodness-of-Fit Tests**

Ch. 9 (§9.2), Ch. 10 (§10.3), Ch. 11 (§11.3) & Ch. 12

- Inference for a population proportion
- Power calculations*
- Inference for comparing two population proportions
- Inference for $(r \times c)$ two-way tables: tests of independence and homogeneity of proportions, chi-square test, expected values, degrees of freedom
- Equivalence of $Z$-test and chi-square test*
- Goodness-of-fit tests
- Binomial goodness-of-fit test*

**Module VI: Regression and Correlation** Ch. 13 & Ch. 14

- Inference in simple linear regression (slope, confidence intervals*, prediction intervals*)
- Analysis of residuals and use of diagnostic tools
- Correlation: inference, correlation vs. regression
- Equivalence of testing for zero correlation and testing for zero slope
- Multiple regression

**Module VII: Nonparametric Statistics**

- Sign tests for median and paired data*

* * * * * * * * * * * * * *

The following is a list of topics covered in the textbook that are **not** included in the course (and will not be tested):

Law of Total Probability (pp. 328 – 330), Bayes' Rule (pp. 330 – 332), estimating probabilities empirically using simulation (§6.7), geometric distribution (pp. 388 – 390), checking for normality and normalizing transformations (§7.7), using the normal distribution to approximate a discrete distribution (§7.8), an alernative to the large-sample $z$ interval (pp. 473 – 474), $\beta$ and power for the $t$ test (pp. 544 – 547), standardized residuals (p. 686), polynomial regression (pp. 704 – 706), interaction between variables (pp. 706 – 710), qualitative predictor variables (pp. 710 – 712).

**Some Additional Notes**

**Where there are any discrepancies between the way topics are covered in the course notes and in the textbook, please refer to the notes.**

**Unit 1**

- The order of the material for this unit is presented differently in the course notes and the textbook. In the notes, we cover confidence intervals for $\mu$ when $\sigma$ is known, hypothesis tests for $\mu$ when $\sigma$ is known, confidence intervals and hypothesis tests for $\mu$ when $\sigma$ is unknown, and then confidence intervals and hypothesis tests for $p$ (in Unit 5). In the textbook, all confidence intervals are covered, followed by all hypothesis tests (first for $p$, then for $\mu$).
- There margin of error $m$ is referred to in the textbook as $B$, the bound on the error of estimation.
- There is very little coverage in the textbook of hypothesis tests for $\mu$ when $\sigma$ is known. Please refer to the course notes for this material.
- The critical value method and confidence interval method for hypothesis tests are not covered in the textbook. Please refer to the course notes for this material.

**Unit 2**

- Only the conservative (unpooled) inference methods for independent samples are covered in the textbook. In practice, we will divide the higher sample standard deviation by the lower. If this quantity is greater than 2, we will use the conservative method. However, if this ratio is less than or equal to two, we will use the pooled method (please see the course notes).

**Unit 3**

- The textbook refers to the "treatment sum of squares" ($SSTr$) and the "mean squares for treatments" ($MSTr$). We refer to them as the "sum of squares for groups" ($SSG$) and the "mean square for groups" ($MSG$), respectively.
- The textbook denotes the total sum of squares as $SSTo$, whereas we simply use $SST$.
- The textbook calculates confidence intervals in the ANOVA setting (both for a single population mean and the difference between two population means) differently than we do. Please refer to the course notes.

**Unit 5**

- The textbook uses the rule that the normal distribution can be used to calculate approximate probabilities for $\hat{p}$ when no more than 10% of the population is included in the sample. We will assume it is safe to use the normal distribution approximation as long as $np \geq 10$, $n(1-p) \geq 10$, and the population is "large" compared to the sample.
- In calculating the required sample size for estimating the population proportion $p$ to within a specified margin of error, we use an educated estimate, denoted by $p^*$, as opposed to the textbook, which simply uses $p$.
- For chi-square tests of homogeneity and independence, the textboook uses the rule that all expected cell counts are at least 5. We will assume the chi-square approximation is appropriate as long as no more than 20% of cells have an expected count less than 5, and there are no expected cell counts less than 1.

**Unit 6**

- Much of the notation and many of the formulas differ from the textbook to the notes. Please refer to the course notes.

**Unit 7**

- The material in this unit is not covered in the textbook. Please refer to the course notes.