Today's Agenda

Wk 1, Tuesday Hr 1

- Welcome to Stat 302

- Policy
- Course Schedule

Hr 2 (Review 1)

- Samples and Populations
- Descriptive and Inferential Statistics
- Normal and T distributions

What is this class about?

My assumption is that most of you are senior undergraduates in health sciences, and that you are taking this course mostly because it is a graduation requirement.

My intention is to bring maximal value to that assumed demographic.

By the end of this course, successful students will be able to:

1. ***Interpret*** the results and graphs of a range of popular statistical methods.
2. ***Determine*** which of these methods, if any, is appropriate for a given data-based problem.
3. ***Evaluate*** the validity of key assumptions for each of these methods.
4. ***Communicate*** their needs when problems are beyond their statistical expertise.

These goals could describe many statistical courses. The differences between each course are the methods in question.

For Stat 302, these methods are
Analysis of Variance (AnOVa),
Multiple Regression,
General Linear Models, and
Logistic Regression.

Compared to previous offerings of Stat 302, this course will be more focused on statistical practice (sampling, experimental design, validity, interpretation) aspects than methodology (mathematics, proofs).

Material on partial correlation and Analysis of Covariance have been shortened to make room for greater coverage of logistic regression at the end of the course.

Today's Agenda

Hr 1

- Welcome to Stat 302

- Policy

- Course Schedule

Grading policy

Minimum scores for letter grades

A+ : 90%	B-: 68%
A: 85%	C+: 64%
A-: 80%	C: 60%
B+: 76%	C-: 55%
B: 72%	D: 50%
	F: 0%

Grades are based on

2 Midterms x 20% = 40%

4 Assignments x 5% = 20%

1 Final x 40% = 40%

Grades will be available on Coursys (I'm told).

Handing in assignments, late policy

- There are some drop boxes labeled ‘Stat 302’ outside the stats lab, on the main floor of Shrum Science Centre K. All assignments are to be handed in there by 4:30pm of the due date.

- Assignments not in the drop-box when they are picked up in the drop box will not be graded.

-Assignments are graded by TAs, and solutions will be e-mailed out after the due date.

Contact:

E-mail: jackd@sfu.ca

Course website: TBA, but will archive notes.

Office Hours: Tentatively 1-2pm Tuesday and 4-5pm Thursday at K10564

The **stats lab** is available for free tutoring in the through K9510, 9:30 – 4:30 M-F, and sometimes later hours. This room also has computers with R installed for you to use.

About the textbook:

The textbook, Applied Regression Analysis and Other Multivariable Methods 5th ed. , by Kleinbaum et. al is OPTIONAL.

Optional means *You can succeed in this course without the textbook*.

* No assignment or exam question will refer to information that is only in the textbook. All the information you need will be in the lecture notes, the assignments, or the assigned readings.

* Textbook chapters/sections for each week are meant to be for supplementary reading. They will have a lot of the same information as the notes.

* There are required readings as part of the assignments, but these are freely available to SFU students.

Regarding software:

For assignments, you will need to interpret output from the statistical software R.

I will also be providing the data sets and R code in order to produce this output.

This way if you want to explore the data to a greater depth or get better acquainted with R, you can, but for the assignments, you need only copy and paste R code.

You may also use SAS, JMP, SPSS, or Excel to get these results if you are already proficient in one of these programs.

If you are using the textbook, the example output is SAS output.

Differences in the terms used between SAS and R will be given to you in your assignments.

R is open source and freely available for Windows, Mac, and Linux at https://cran.r-project.org/

It is also available on the Play Store for Android phones.

Today's Agenda

Hr 1

- Welcome to Stat 302
- Policy
- **Course Schedule**

Course plan:

Wk 1: Lectures Jan 5 and 7.

Course introduction, review topics.

Optional Reading - Textbook Ch.3.

Optional Reading – Online Stat Book, available at http://onlinestatbook.com/2/index.html

Review sections from Online Stat Book:

Descriptives,
random variables,
normal/z distribution,
t-distribution,
chi-squared distribution,
F-distribution,
point estimation,
hypothesis testing,
Type I and II error,
power, and sample size.

Wk 2: Lectures Jan 12 and 14

Correlation vs association,
Pearson's R, Spearman rho,
non-linearity,
the 'window' effect in regression,
regression and the bivariate normal.

Required reading: Installing and using R.

Optional reading: Textbook Ch. 4, Ch. 6

Wk 3: Lectures Jan 19 and 21

Simple regression,
inferences about slope and intercept,
hypothesis testing about the slope,
estimating means and individual points,
correlation vs causation.

Required Reading: Rubin on Causality.
Optional reading: Textbook Ch. 5.1, 5.2, 5.4, 5.8, 5.9, 5.10, and Textbook Tables 5.2, 5.3.

Wk 4: Lectures Jan 26 and 28

Assignment 1 due Jan 26

AnOVa review.

Tukey's HSD (Honestly Significant Difference).

Two-way AnOVa.

Optional reading: Textbook Ch.7, 17.1, 17.2, 17.3.

Wk 5: Lecture Feb 2,

MIDTERM 1 Feb 4

(Class withdraw deadline Feb 5*)

AnOVa and multiple regression – the R-squared connection.

Midterm review

No reading.

Wk 6: READING WEEK.

Required reading: The case against Null Hypothesis testing.

Optional reading: Textbook Ch. 8, 9.1, 9.2, 9.5, 9.6, 9.7.

Wk 7: Lectures Feb 16 and 18
(Assignment 2 due Feb 18)

Multiple regression:

- co-linearity,
- perturbations,
- correlation matrix,
- uniqueness,
- non-linearity revisited.

Optional reading: Textbook Ch. 10.1, 10.2, 10.3, 10.5, 10.8. and Ch. 14

Wk 8: Lectures Feb 23 and 25

Multiple regression:

Graphing and interpretation.
Model selection,
AIC and Mallow's c,
stepwise regression,
all subsets,
principles of hierarchy and parsimony.

Optional reading: Ch. 11.

Wk 9: Lectures Mar 1 and 3

Transformations.
Dummy variables,
    Interactions,
    Interpretation.
AnCOva. (Analysis of Covariance, if time permits)

Optional reading: Ch. 12.1, 12.2, 12.3, 12.4, 12.5, 12.7.

Wk 10: Lectures Mar 8 and 10
(Assignment 3 due Mar 10)

Diagnostic plots,
Influence and outliers,
cross-validation.
Missing data,
    missing at random / not at random.

Required reading: TBA Paper on Meta Studies or on Funnel Plots for Publication Bias.

Wk 11: Lecture Mar 15,

MIDTERM 2 Mar 17

Imputation.
Midterm review.

No reading.

Wk 12: Lectures Mar 22 and 24

Logistic regression:

odds,

odds ratio,

confidence intervals.

Required reading: Multiple Imputation by van Buuren.

Wk 13: Lectures Mar 29 and 31
(Assignment 4 due Mar 31)

Logistic regression and dummy variables, multiple logistic regression.

Optional reading: Textbook Ch 22.1, 22.2, 22.3, 22.4

Wk 14: Lectures Apr 5 and 7

Partial correlation (if time permits).
Final exam review.

(End of classes Monday, April 11)

*Note about the drop deadline.

Officially the deadline is Monday, Feb 8, but that's a statutory holiday and the office will be closed.

Also see:
http://www.sfu.ca/students/deadlines/spring2016.html

Regarding notes:

All the course notes will be in a fill-in-the-blank system. Before each lecture, I will e-mail out notes like I did with these ones, but with blanks to be filled in during class. However, some calculation and diagrams are drawn in class, and there will be entirely blank slides appear in the notes.

I recommend printing these notes so that 4-6 slides appear on a single page, because there will be 1200-1400 such slides over the course of the semester.

There are single-slide breaks between every 10-15 minutes of material. On these break slides, I like to include pictures of cute/funny animals with stupid stats puns. If there are any animals that you feel uncomfortable seeing (mice, reptiles, fish, birds, whatever), please e-mail me a request not to include those.

Regarding collaboration, honesty, and plagiarism:

None of the assignments or exams for this course are recycled from previous sources. Anyone claiming to have a test bank for this offering of this course is lying.

Please include the names of your collaborators on your assignments. This way, the markers will understand when some solutions look very similar that there wasn't blind copying.

You are encouraged to work together to do the computational and analytical portions of the assignments. However, all written work is expected to be solely yours.

Copying the writing of another student, *or using services to write assignments on your behalf* will be considered academically dishonest and will be dealt with as appropriate in SFU's academic dishonesty policy.

The use of proofreading and essay skills services, such as those in the Student Learning Commons, is perfectly fine.