SPPH 400 (Distributed Learning) – Statistics for Health Research
Course Syllabus 2014

DAYS & TIMES: All face-to-face meetings will be held in room B104 of The School of Population and Public Health building, located at 2206 East Mall, on UBC Campus. Please see the course schedule below for a list of on-line and face-to-face meetings.

LOCATION: Face-to-face meetings in Room B104, SPPH Building

INSTRUCTOR: Mike Marin

EMAIL: marin.mike@gmail.com, mmarin@stat.ubc.ca

OFFICE: School of Population and Public Health
University of British Columbia
SPPH Room 263

OFFICE HOURS: Since much of the course is completed from a distance, most office hours will be held on-line. The TA will hold weekly on-line tutorials. Further, a discussion forum on the course website will help with asking and answering questions.

TEACHING ASSISTANTS: The teaching assistant (TA) for this course will be introduced at the beginning of the course. The TA will offer on-line tutorial sessions reviewing course material. They will also offer support for the statistical software R. They will assist in the grading of assignments and examinations. Scheduling of on-line tutorials will be done within the first few weeks of the course, to be worked out between the students and the TA.

COURSE SCHEDULE: The schedule below is meant to help you pace your work in this course. We may deviate slightly from the schedule below, and please note that the assignment dates are there as a guide but are not hard deadlines. Given the nature of a DL format, we are a bit more flexible with deadlines. Also, on the course website I will give regular reminders of which material you should be reading over, between our face-to-face meetings.

<table>
<thead>
<tr>
<th>Week Starting:</th>
<th>Mode</th>
<th>Topic</th>
<th>Assessment (Approximate Dates)</th>
</tr>
</thead>
</table>
| Sept 1         | On-line | * Read Module 1  
Read Module 2 |                                    |
| Sept 8         | In Person | Discuss Modules 2/3 |                                    |
| Sept 15        | On-line | Re-read Module 2/3  
Read Module 3 | Assignment #1 due |
| Sept 22        | On-line | Read Modules 4/6 | Assignment #2 due |
| Sept 29        | On-line | Read Modules 4/6  
Read Module 6 and possibly Module 5 |                                    |
| Oct 6          | On-line | Read Module 6 and possibly Module 5 |                                    |
| Oct 13         | In Person | Discuss Modules 4/6 and Possibly module 5 | Midterm #1 |
| Oct 20         | On-line | Re-read Module 6  
Read Module 5 | Assignment #3 due |
| Oct 27         | On-line | Read Module 5  
Read Module 7 | Assignment #4 due |
| Nov 3          | On-line | Read Module 7 | Assignment #5 due |
| Nov 10         | In Person | Discuss Modules 5/7 |                                    |
| Nov 17         | On-line | Re-read Module 5 |                                    |
| Nov 24         | On-line | Re-read Module 5  
Read Module 7 | Assignment #5 due |
| Dec 1          | On-line | Read Module 7 | Final Exam, date TBA |

* Our face-to-face meetings take place on Sept 12&13, Oct 18&19, Nov 14&15, each day at 9:00am-12:30pm
* Only on September 12, we will meet from 1:30pm-5:00pm
* Note that module 5 and 6 are reversed in order, and module 6 will be presented BEFORE module 5
* Note that we will share face-to-face days with SPPH 502 (us in the morning, 502 in the afternoon)
COURSE PHILOSOPHY AND OBJECTIVES:

This course will introduce students to basic statistical methodology used in health research. By the end of this course, students will be able to:

1. Choose and create effective graphical, tabular, and numerical summaries of univariate and bivariate data.

2. Distinguish between basic methods for selecting samples and understand the impact of the sampling method on the choice of statistical analysis and generalizability of results.

3. Identify commonly used basic statistical methods and the circumstances under which their use is appropriate.

4. Understand the notion of sampling variability and sampling distributions.

5. Calculate and interpret confidence intervals and p-values and understand their limitations.

6. Select and carry out an appropriate method of analysis to compare the means or proportions of two or more populations, and provide an interpretation of the results of such an analysis.

7. Conduct simple linear regression analyses, assess their validity, and interpret the results. Understand the extension of these ideas to multiple linear regression.

8. Recognize situations where the opinion of an experienced statistician is required.

Greater emphasis will be placed on conceptual understanding with relatively less emphasis on computational ability. For example, recognizing that a "paired t-test" is the appropriate method of analysis in a given situation would be worth more than the calculations associated with conducting the test. Correspondingly, applying an "independent sample t-test" when it is inappropriate will receive little or no credit, even if the computations are done correctly.

PREREQUISITE:

No previous courses in statistics or previous computing experience are required. However, students will be expected to be familiar with algebra (intermediate level) and simple graphing techniques. Further, it will be assumed that students are familiar with the material presented in the “Stats Prep” course/course notes. No class time will be spent on the topics presented in the stats prep notes, up to page 52; students should review these topics on their own time if required.

RECOMMENDED COURSE TEXTBOOK:


This text provides fairly good explanations for most of the main ideas of the course. Some of the topics will be covered in greater depth during the lectures, tutorials and in the provided set of course notes. Purchase of the textbook is optional. It will provide most useful for practice problems.
SOME ADDITIONAL REFERENCE TEXTS:


5. The Cambridge Dictionary of Statistics in the Medical Sciences B.S. Everitt

6. Modern Applied Statistics with S by Venables and Ripley

EQUIPMENT REQUIREMENTS:

A calculator (preferably with some statistical functions) will be required for examinations. It is recommended that you download a free copy of R for your personal computer.

STATISTICAL COMPUTING:

In the course, you are welcome to use any statistical software you like or are familiar with, although R will be the main software used in the course. Lectures will present R code/output, and for exams you will be expected to be familiar with interpreting statistical output from R.

I have created a set of video tutorials to teach the use of R. These videos start right form the beginning with installing the software, and assume no familiarity with R or with programming in general. While the video tutorials are general in nature, they are intended to serve this course. You an find the videos by searching “MarinStatsLectures” on YouTube, or by going to the URL listed below. Make sure to subscribe to my YouTube channel to have access to all videos in a neat and organized fashion.

http://www.youtube.com/user/marinstatlectures

You will also be directed to an online SAS workshop, which you can work through if you like. Information regarding this workshop will be sent out at a later date.

COURSE EVALUATION:

5 Assignments 30%
1 Midterm Test 30%
Final Examination: 40%

ASSIGMENTS:

The assignments are designed to help students master the concepts presented in class and gain experience in applied data analysis and interpretation. Students are encouraged to work together
in teams and discuss their ideas but you will likely find that you will learn the material most thoroughly by first attempting to work out the solutions yourself. Each student is expected to submit his or her own written solutions (prepared by him/herself). Academic dishonesty will not be tolerated.

- Assignments should be neatly written (or typed). This is not a thesis, but it should still look like something you are proud to have your name on.

- Make sure your approach to a problem is clearly outlined and explained. A clear explanation of what you are doing and why is more important than any numerical answer you provide.

- When preparing solutions related to data analysis, include only those parts of the computer output that are relevant to your answer and highlight or underline the specific items of interest. Alternatively, transcribe those items to another page if you prefer. Choosing the relevant pieces of information is more important than simply providing a long list of computer output.

**MID-TERM TESTS AND FINAL EXAMINATION:**

The midterm and final can be written at a distance, with an arranged invigilator. This person should be someone in a position of authority and whose identity can be verified. If in doubt, just ask me if the individual you are considering is suitable. It is a good idea to start considering possible invigilators early in the term. Also, for those in or around Vancouver, I will arrange a date to come in and write the exams if that is preferable for those individuals.

Books may not be used during tests or exams. Students may bring a formula sheet with any relevant formulas or properties written on it. Statistical tables, when necessary, will be provided with exams.

**COURSE NOTES:**

A set of course notes I have written will be posted on the website for you. You can print or save a copy for yourself. It is a good idea to read these prior to lectures, and to bring a copy with you to class. These notes are detailed, and more of a textbook, rather than a set of lecture slides.

**NOTE:** Please do not use the printers in the computing lab for printing course notes.

**Course Topics:**

The course is broken down into a set of 8 modules. Each module contains ideas that fit together, at least to some extent, and most modules build on the ideas presented in previous modules. **Modules 1 and the first half of 2 will NOT be covered in lectures.** The notes for modules 1 and part of 2 are provided so that you may refresh on this material, if necessary. Material in modules 1 and 2 will we reviewed in the Stats Prep Course. In SPPH 400, we will begin with the Normal distribution (page 39 of module 2 notes).
Module 1 – Samples:

- Summary of univariate data using numerical and graphical methods. Measures of location and dispersion. The standard deviation used as a unit of measurement.
- Summary of bivariate data using graphical methods.
- Methods of sampling and types of bias.
- Data collection and management.

Module 2 – Probability and Probability Distributions:

- Probability trees, Bayes’ Theorem. Diagnostic tests: sensitivity, specificity, positive predictive value, negative predictive value, false positive, false negative and prevalence.
- Random variables. Mean and standard deviation of a random variable.
- Common probability distributions for discrete random variables: Binomial and Poisson.
- Linear transformations of random variables, and the properties of the mean and variance
- Introduction to continuous probability distributions. The normal distribution, standardizing, and properties of the normal distribution. Chebychev’s inequality.
- The Central Limit Theorem, and the sampling distribution of a mean/proportion.
- Normal approximation to the Binomial and the Poisson.

Module 3 – Confidence Intervals and Hypothesis Tests:

- The role of a sampling distribution in statistical inference
- Student’s t-distribution. Estimation, one and two sided confidence intervals and the underlying logic behind a confidence interval
- One and two-sided hypothesis testing, definition and limitations of p-values.
- Brief mention of Bayesian methods
- Statistical vs. scientific significance
- Types of errors, power and sample size calculations

Module 4 – Types of Variables and Hypothesis Tests:

- Definition of bivariate data, outcome and explanatory variables
- Parametric vs. non-parametric tests
- Appropriate statistical methods for the type of outcome variable you want to analyze

Module 5 – Statistical Inference for a Continuous Outcome and Qualitative Explanatory:

• Dependent populations: The paired t-test, repeated measures ANOVA. Non-parametric tests: Wilcoxon signed-rank test, Friedman’s test.
• Brief discussion of two-way ANOVA, and randomized block designs

Module 6 – Statistical Inference for two qualitative variables:

• Analysis of 2x2 tables: The Chi-square test of independence, and Fisher’s exact test.
• McNemar’s test for paired data.
• Measures of association for 2x2 tables: Risk difference and the number needed to treat, relative risk (risk ratio), odds ratios, confidence intervals for odds ratios.
• Brief discussion of the two proportions hypothesis testing

Module 7 – Statistical Inference for two quantitative variables:

• Pearson’s and Spearman’s correlation
• Simple linear regression. Interpretations and tests for model parameters.
• Model assumptions and regression diagnostics

Module 8 – Multiple Linear Regression and Extensions:

• Discussion of multiple linear regression
• The idea of ‘adjusting’ for other variables in a regression model
• Extensions of the linear model for different types of outcome variables (logistic, Poisson and Cox regression)