

# **SPPH 400/521 WB1 (3 credits): Statistics for Health Research Course Syllabus 2024**

**SCHEDULED DAYS & TIMES:** Every other Friday, 9:30am-12:30pm.

\* See “weekly schedule” document for list of class meetings and topics covered each class

**INSTRUCTORS:** Mike Marin

**EMAIL:** [mike.marin@ubc.ca](mailto:mike.marin@ubc.ca)

**OFFICE HOURS:** My office hours will be held by appointment. The teaching assistant will also have weekly office hours open for all. Day and time for the office hours and Zoom links will be posted on the course website.

**DISCUSSION BOARD:** The course website has a discussion board for posting questions, etc. Please post all (non-personal) questions related to the course here. Others will benefit from seeing your question, and the answer to it. The discussion board will be monitored by the teaching team, although students are encouraged to answer each other’s questions as well.

**TEACHING ASSISTANTS:** The teaching assistants will be introduced at the beginning of the course. They will hold office hours, monitor discussion boards, assist in the grading of assignments and examinations, etc.

## **COURSE PHILOSOPHY AND OBJECTIVES:**

This course will introduce basic statistical methodology used in health research. Greater emphasis will be placed on conceptual understanding, with relatively less emphasis on computation.

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 their appropriate use.
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.

\* See “learning outcomes” document for a full list of course learning outcomes

**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 you are familiar with the material covered in the module 1 notes.

**COURSE READINGS:** We provide a set of detailed notes developed for the course. These notes are more of a textbook, than a set of lecture slides. The course notes will be posted on the course website for you. You can print or save a copy for yourself. If you would like to pass these notes on to someone else outside of SPPH, please discuss this with me first.

**EQUIPMENT REQUIREMENTS:** A calculator will be required for examinations. A simple calculator will be sufficient as long as it is able to take logs, exponentiate, etc. It is recommended that you download/install a copy of R & RStudio for your personal computer and bring your computer to Zoom class.

**STATISTICAL COMPUTING:** 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.

### **COURSE EVALUATION:**

5 Homework Assignments – 25%

“Table One” Assignment – 5%

One Midterm Test – 30%

Final Examination\* – 40%

\* Must pass the final exam to pass the course

\* If you believe an assignment grade is incorrect, please provide in writing an explanation of what you believe was marked incorrect and submit to the instructor within 1 week to request a regrading of your work. Your entire assignment will be reviewed for accuracy of grading.

### **LEARNING APPROACH:**

We will be using a “blended learning” approach. The one-sentence explanation of this is that you are expected to do some *pre-lecture work to review material and begin work on a learning activity* outside of class time, and then the in-class time will be spent *actively engaging with the course material*, mainly through the continuation of *learning activities* worked on before coming to the class. The learning activities will have many questions in them, and I will let you know ahead of time which questions we will focus on in class discussion.

Rather than spending the entire class time delivering course content and then having you go away and work on problems on your own, the plan is to leave out some of the content-delivery, and replace that time with some active engagement with the material. This will require you to read material and watch videos prior to lectures, and come to class prepared. This approach has the added benefit of having more contact with the instructors as you work through the more challenging part of the material, rather than having the instructors there for the most basic of the material. We are replacing a more dated approach where students *come to class to get the notes*, with a more modern approach where students *come to class to deepen their understanding of the material*.

There are 4 types of class materials that you should study each week: readings, video tutorials, R-tutorials, and video lectures. By design, readings, video lectures and video tutorials complement each other. They are linked and cross-referenced to some extent. Readings give you a somewhat full exposition of the topic. Tutorials focus on methods. Lectures review ideas. Readings, and video lectures should be studied in preparation for weekly classes.

R tutorials are different. R tutorials walk you over examples mentioned in video-lectures and video-tutorials. They give you a freedom of converting ideas into calculations. Without memorizing formulas, or even writing a script. These video tutorials start installing the software, and assume no familiarity with R or with programming in general. You can find the videos by going to the URL <http://www.youtube.com/marinstatlectures>

### **ASSIGNMENTS:**

The assignments are designed to help students master the concepts presented in class and gain experience in applied data analysis and interpretation, and are formative in nature. Students are encouraged to discuss the assignment and share their ideas, and assignments may be submitted individually, or in groups of up to 3 people.

- We will be working with an online homework system, which will be introduced in class. Many answers will be directly submitted to the system online. Other portions of your assignment will be typed up, saved in a PDF document, and submitted online through the course website.
- Through the online homework system, you will get your own personal set of data that will be worked on over the duration of the course. Your variables will be the same as others, but your actual data and observed values will differ from your classmates.
- Assignments should be typed (or neatly written). This is not a thesis, but it should still look like something you are proud to have your name on. Some marks will be allocated to clarity of presentation.
- Make sure your approach to a problem is clearly outlined. A clear explanation of what you are doing and why is more important than any numerical answer provided.
- 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. Do NOT include every piece of information from software output...you must select what is relevant to include.
- Due dates for assignments are posted on the course website and in the weekly schedule. Slight adjustments may be made to the dates, when necessary. Students may submit up to 2 assignments up to 1-week late without needing to ask for an extension. If you will need additional time to complete more than 2 assignments due to extenuating circumstances, just discuss this with me at least 1-day before the due date. Late submissions (after the 2 'free' extensions) without having discussed with me will be deducted 25% per day.
- Please review the document outlining "*How Assignments Are Evaluated*", to get a better understanding of the scales we will use in assigning a label/category to your work

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

The midterm test will take place during regular school term. The final exam will be scheduled for shortly after lectures end. 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. More info regarding exams will be provided when exams are nearing.

#### **COURSE NOTES:**

A set of course notes will be posted on the web for you. You can print or save a copy for yourself. You will be reading these prior to lectures, and may want to bring a copy with you to class. These notes are detailed, and more of a textbook, rather than a set of lecture slides.

#### **COPYRIGHT:**

All materials of this course (course handouts, lecture slides, assessments, course readings, etc.) are the intellectual property of the Course Instructor or licensed to be used in this course by the copyright owner. Redistribution of these materials by any means without permission of the copyright holder(s) constitutes a breach of copyright and may lead to academic discipline.

If you wish to record the class in any way, please discuss this with me first.

## **USE OF GENERATIVE AI:**

The use of Chat GPT or other generative AI tools is permitted in this course.

If you use generative AI to get ideas and/or partial answers for an assignment and/or to generate any text for a draft or final version of any part of an assignment, you must declare that you have used it. Please provide this declaration in the first line/paragraph of submission of your work. You must also add a couple sentences describing the extent to which it was used, and you must save any generated text from this tool (in a separate document) in case it is requested. A TA or the instructor may ask you to provide the generated text in order to help with grading decisions.

## **RESPECTFUL ENVIRONMENTS:**

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions.

Details of the policies and how to access support are available on [the UBC Senate website](#).

UBC and all Members of the UBC Community share responsibility for ensuring and maintaining an environment that is free from Discrimination. UBC regards Discrimination as a serious offence that is subject to a wide range of remedial or disciplinary measures, including dismissal or expulsion from UBC.

(<https://universitycounsel.ubc.ca/policies/discrimination-policy/>)

SPPH is committed to providing a positive education experience free from discrimination. If you have had an experience in this course where you feel unsafe, have been mistreated or have witnessed mistreatment, please let us know. If you want to raise this beyond the course instructor the School recommends the following. You may contact your academic supervisor, the [education manager for your program](#) or the [Associate Director-Education](#). You may also report your concerns to the Faculty of Medicine Office of Respectful Environments, Equity, Diversity & Inclusion (REDI) at <https://mistreatmenthelp.med.ubc.ca/>. Both SPPH and the REDI Office have.

It is my intent that students from diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as an asset. It is my intent to present course material and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. It is important to note that the goal of statistical analysis is to use data to make generalizations about individuals. In doing so, we often categorize people into group "A" or "B" in order to make a generalization about those in group "A" relative to group "B". This is not to suggest that everyone naturally fits into either "A" or "B", but grouping people into a category based on certain characteristics is necessary for attempting to make useful generalizations. Your suggestions on ways to improve the effectiveness of the course for you personally or for other students or student groups, presented in a constructive and helpful way, are encouraged and appreciated. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make appropriate arrangements.

**LAND ACKNOWLEDGEMENT:** *UBC's Point Grey Campus is located on the traditional, ancestral, and unceded territory of the xwməθkwəy̓əm (Musqueam) people. The land it is situated on has always been a place of learning for the Musqueam people, who for millennia have passed on their culture, history, and traditions from one generation to the next on this site.*

## Course Topics:

The course consists of 8 modules, each containing ideas that fit together, 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.

In SPPH 521, we will begin with the Normal distribution (page 39 of module 2 notes).

### Module 1 – Samples:

- Introduction, course outline and course objectives. Definitions of statistics. Observational and experimental studies.
- 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:

- Definitions of probability, odds, and terminology. Axioms of probability. Methods of assigning probabilities. Independence and conditional probabilities.
- 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.
- Binomial and Poisson probability distributions for discrete random variables.
- 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. One and two sided confidence intervals, the underlying logic and estimation
- 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:

- Independent populations: The two-sample t-test (equal and non-equal variances), analysis of variance. Multiple comparisons procedures. Checking assumptions of parametric tests. Non-parametric tests: Wilcoxon rank-sum test (aka Mann-Whitney U test), Kruskal-Wallis analysis of variance for ranks
- 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)