University of Guelph
College of Engineering and Physical Sciences
Department of Mathematics and Statistics

STAT*2040 Statistics I

Winter 2024
Credit Weight: 0.5

CALENDAR DESCRIPTION: This course focuses on the practical methods of Statistics and the topics include: descriptive statistics; univariate models such as binomial, Poisson, uniform and normal; the central limit theorem; expected value; the t, F and chi-square models; point and interval estimation; hypothesis testing methods up to two-sample data; simple regression and correlation; introduction to analysis of variance. Assignments will deal with real data from the natural sciences and involve the use of statistical software for computing and visualization.

PREREQUISITES: 1 of 4U Calculus and Vectors, Advanced Functions and Calculus, OAC Calculus, MATH*1080
RESTRICTIONS: STAT*2060, STAT*2080, STAT*2120, STAT*2230.

INSTRUCTOR: Jeremy Balka x54481 OFFICE: 550 MacN EMAIL: jbalka@uoguelph.ca
LECTURES: Section 1: MWF 1:30–2:20 in WMEM 103
Section 2: MWF 8:30–9:20 in MACN 105

OFFICE HOURS: 9:30-11:00 MWF. (I have class until 9:20, so if I'm talking with students after class I may be a little late getting back to my office.)

ONLINE OFFICE HOURS: I will have at least one online office hour per week. Details will be posted to Courselink.

LEARNING OUTCOMES: After successful completion of the course, students should be able to:

- Explain the fundamental concepts of samples and populations, statistics and parameters, observational studies and experiments, and sampling bias.
- Create and properly interpret numerical and graphical data summaries.
- Discuss fundamental probability concepts and carry out basic probability calculations.
- Carry out probability calculations for various discrete and continuous probability distributions, and choose the appropriate probability distribution in different scenarios.
- Explain statistical inference concepts and methods, including concepts related to sampling distributions, confidence intervals, and hypothesis tests.
- Choose an appropriate statistical inference procedure in a variety of situations, carry out the procedure, and effectively communicate a proper interpretation of the results.
- Use statistical software to create plots and carry out the calculations for various statistical inference procedures.
- Read published statistical studies and describe the results of their statistical inference procedures.

COURSE TEXT: There is no publisher text and no extra costs in this course. I'll be releasing draft chapters of the 1.11 version of my pdf text (Introductory Statistics Explained) as we progress through the semester. I've been using the 1.10 version in this course for a number of years, and you can get away with using that if you prefer. But I'm updating to 1.11 and the chapters I release through the semester will be the official course text. Please let me know if you find any typos or other errors. Suggestions for improvements (areas that require greater clarification, topics for short videos etc.) are welcome.

There is *lots* of video support, in different formats. I have over 100 shorter videos on specific topics and many video solutions to exercises. Links to the supporting videos are found in the text and exercises.

LECTURE CONTENT AND FORMAT:
To a great extent, lectures will be a summary of the text. If my lectures sound similar to the text, it is because my text is based on my lectures (not the other way around).

On Courselink I provide partially completed notes ("lecture outlines") for each topic in pdf format. I will complete these notes during lectures, and I assume that you have these notes with you. Bringing these notes to lectures will greatly reduce your writing burden. Some students prefer to come to class and take minimal notes, relying instead on the online text. That is a reasonable approach. I will leave it up to you to decide what works best for you.

N.B. Lectures are but a part of the course. From the Undergraduate Calendar:

A credit weight of [0.50] indicates 10-12 student effort hours, including class time, on academic tasks associated with the course.

Ten to twelve hours. Each week. Let's round up and call lectures 3 hours per week. As *part* of those other hours, it is expected that students work through the exercises each week.

GRADING SCHEME:

- 20% Data analysis assignments. There will be 2 assignments that require the use of R software, and each of them will be worth 10% of your final grade. The assignments will be submitted electronically in pdf format. (Full details will be posted on Courselink.) You may complete the data analysis assignments on your own, or in groups of 2 or 3. N.B. Students in this version of the course (the face-to-face offering) can only submit group work with other students in this offering (and not the DE offering). Deadline dates: Wednesday January 24 and Wednesday March 27.
- There will be two term tests. Your best test will count for 25% of your final grade, your worst for 15%. Term test dates:
  - Test 1: Saturday February 10. 3:30 – 5:00 pm. Location TBA.
  - Test 2: Saturday March 9. 3:30 – 5:00 pm. Location TBA.
- 40% Final exam. (An old-school, sit-down, in-person two-hour exam.) Saturday April 20. 8:30-10:30 am. Location TBA

ASSIGNMENT AND EXAM POLICIES:

- **Any assignment not submitted by the deadline will not be marked and will receive a grade of 0.** Contact me as soon as possible if you run into an issue that prevents you submitting the assignment by the deadline.
- You may submit each assignment individually or as a group of 2 or 3. You may discuss approaches to assignment questions with others, but your submitted assignment must be your own work. Copying any part of another student's work is considered academic misconduct. (Please read the section on academic misconduct in the Undergraduate Calendar.)
- The term tests and final exam are to be completed individually.
- If you miss a term test for a valid reason (that you let me know of as soon as possible), your final exam will be reweighted to make up for the missed term test.

STATISTICAL SOFTWARE: We will use R (or R Studio) for our statistical analyses in this course. Use of R will be required on the assignments and tests. R is available in the computer pools on campus, and can be downloaded (for free) from https://www.r-project.org/.

EXERCISES: There will be exercises and solutions (including some video solutions) available on Courselink for all topics in the course. Although these do not count for marks directly, they are *required* exercises. It is best to work through these questions after each lecture, in order to consolidate the information we talked about in lecture. It is assumed that students are working through these exercises as we proceed through the course.

OPTIONS FOR GETTING HELP IN THE COURSE:

- Talk to me after class or during my office hours.
- Post on the discussion board on Courselink.
- Go to the Statistics Learning Centre (in the Science Commons on the 3rd floor of the library). The hours of operation are:
  - Monday and Wednesday: 9:30am - 3:30pm
  - Tuesday and Thursday: 10am - 4pm
  - Friday: 9:30am - 2:30pm
- Attend the R drop-in help (9:30-11:30 each weekday in SSC 1305). It is staffed by graduate student TAs who can help you with R.
- Attend the Supported Learning Group (SLG) sessions. For more information, go to https://www.lib.uoguelph.ca/writing-studying/studying-resources-workshops/slgs
- Haphazardly search for help online, often getting terrible advice from people who don't know what they are talking about. (There is lots of good help online, but there are also lots of *wrong and terribly misleading* statistics resources, and when you are starting out it can be hard to know the difference.)

GENERAL STATEMENTS THAT APPLY TO ALL COURSES:

A list of standard University of Guelph statements that apply to all undergraduate courses can be accessed here: https://mathstat.uoguelph.ca/node/534.

TOPIC OUTLINE: For the most part I cover the course material in lecture. But there will be times when I refer you to the text. Some students prefer to base their studies on the text. While we will cover this list quite closely, it is tentative and we may skip the occasional topic here or there.

- Chapter 1: *Introduction to Statistics.*
  Descriptive statistics, inferential statistics.
- Chapter 2: *Gathering Data.*
  Population, variables, samples, quantitative and qualitative data, bias, simple random samples, experiments, observational studies, lurking variables, confounding.
- Chapter 3: Descriptive Statistics.
  Bar charts, pie charts, histograms, stem-and-leaf displays, summation notation, numerical measures of central tendency, numerical measures of variability, numerical measures of relative standing, boxplots, methods for detecting outliers, linear transformations.
- Chapter 4: Probability.
  Simple events, events, Venn diagrams, combinations, unions, intersections, complements, mutually exclusive events, conditional probability, addition rule, multiplication rule, independent events, Bayes' theorem.
- Chapter 5: *Discrete Random Variables and Discrete Probability Distributions.*
  Random variables, discrete and continuous random variables, probability distributions, expected value and variance of a discrete random variable, the binomial distribution, the Poisson distribution, the hypergeometric distribution.
- Chapter 6: *Continuous Random Variables and Continuous Probability Distributions.*
  Properties of continuous random variables and continuous probability distributions, the uniform distribution, the normal distribution, descriptive methods for assessing normality (normal quantile-quantile plots).
- Chapter 7: *Sampling Distributions.*
  The concept of the sampling distribution, the sampling distribution of the sample mean, the central limit theorem.
- Chapter 8: *Confidence Intervals.*
  Basic concepts of confidence intervals, confidence intervals for a population mean when sigma is known, confidence intervals for a population mean when sigma is unknown, sample size determination.
- Chapter 9: *Hypothesis Testing.*
  Motivation for hypothesis testing, null and alternative hypotheses, significance levels, p-values, Type I and Type II errors, power, tests of hypothesis on a single mean.
- Chapter 10: *Inference for Two Means.*
  Confidence intervals and hypothesis tests for the difference between population means, independent sampling, paired difference experiments.
- Chapter 11: *Inference for Proportions.*
  Confidence intervals and hypothesis tests for a single proportion, sample size determination, confidence intervals and hypothesis tests for the difference between two population proportions.
- Chapter 12: *Inference for Variances.*
  Confidence intervals and hypothesis tests for a single variance, confidence intervals and hypothesis tests for two population variances.
- Chapter 13: *Chi-square tests for count data.*
  Chi-square goodness of fit tests. Tests for one-way tables, chi-square tests of independence.
- Chapter 14: *One-Way Analysis of Variance.*
  Using one-way ANOVA to test for a difference in population means. (One-way ANOVA extends the two-sample pooled-variance t test to more than two groups.)
- Chapter 15: *Simple Linear Regression and Correlation.*
  Least squares regression, model assumptions, inference on the slope, the correlation coefficient, the coefficient of determination, estimation and prediction using the fitted line.