|  |  |  |  |
| --- | --- | --- | --- |
|  | **Department of Statistics**  Website : [www.stat.ubc.ca](http://www.stat.ubc.ca/) | 3182 Earth Sciences Bldg  2207 Main Mall  Vancouver, BC V6T 1Z4 | Telephone: (604)822-0570  Facsimile: (604) 822-6960 |

**STAT 404, Design and Analysis of Experiments**

**2023-2024, Term 1**

**Instructor: Professor Jiahua Chen**

### Lecture days, times and location:

* Image day Sept 5, 2023. No class, yet welcome to meet your professor
* First day of class: Sept 7, 2023, Last day of class: Dec 7, 2023.
* Tuesday and Thursday 3:30pm-5:00pm (except for holidays).
* Lectures will be held on Tuesday and Thursday at DMP110 (<https://learningspaces.ubc.ca/classrooms/dmp-310>)

### Labs, office hours, and other means of discussions:

* Four in-person Labs will be held throughout the term

<https://courses.students.ubc.ca/cs/courseschedule?pname=subjarea&tname=subj-course&dept=STAT&course=404>

* We do not take attendance of labs. Students may attend any one of the labs, particularly when we cannot hold all labs in some weeks due to holidays.
* Office hours, locations of TA and instructor will be posted on Canvas and Piazza.

**Course description:**

Statistics is a discipline that concerns the collection, organization, analysis, interpretation, and presentation of data. Its ultimate goal is to answer scientific questions based on data. Stat 404 introduces some standard experimental designs for data collection that are broadly used in the science community. Together, we discuss how the design principles are employed to increase the information content in the data, limit the influence of lurking factors, and separate the effects of uninterested factors. We will cover the statistical data analysis methods, including linear regression, hypothesis tests, multiple comparisons, confidence intervals, analysis of variance, analysis of covariance. We include standard experimental designs, including one-way layout, two-way layout, blocked, nested, factorial, and split plot designs. We discuss fixed and random effects in appropriate models.

**Textbook/course materials:** no required textbooks.

Lectures will be based on materials in

Wu and Hamada: *Experiments: Planning, Analysis and Parameter Designs optimization.*

Ajit C. Tamhane: *Statistical Analysis of Designed Experiments: Theory and Applications*.

Will Welch: Design and Analysis of Experiments: course notes.

Summary notes by instructor will be posted.

Slides in Jupytor together with its pdf version will also be posted.

**Delivery:**

This course involves in-person activities, which are the most effective way to participate in the discussions that constitute a critical part of your learning experience. Our sessions will not be recorded due to the nature of the conversations held during the class. However, if you happen to miss a class for any reason, the course material is provided as Jupyter slides, followed by clarification questions during the scheduled Zoom office hours.

While catching up on missed classes through Jupyter slides can be effective, consistently skipping classes can lead to other problems. This habit could potentially result in falling behind on the course's pace, feeling disconnected, and missing out on opportunities to engage with fellow students.

**Some Topics:**

1. Introduction. Three Principles of Design: Randomization, Replications, Blocking. Observational and Experimental data. Historical Figures in DoE. Types of Experiments. Terminology: Response, Treatment, Experimental Error, Experimental Units, Factors. Review of Basic Probability Theory: Axioms, Conditional Probability, Bayes' Theorem, Random Variable, Binomial, Normal, and Normal related distributions. Sample Mean and Variance.
2. Linear regression. The Name of Regression, Response and Covariates, Assumptions We Postulate: Covariates Are Not Random, Experiment Errors Are Independent and Normally Distributed with the Same Variance. Least Squares Estimation of the Regression Coefficient, the Predicted/Fitted Values, Residuals, Error Variance Estimation. Mean and Variance of LSE $\hat{\beta}$. Sum of Squares of Regression, Residual/Error, Total, and Their Relationship. F-Test for the Effectiveness of the Regression Model, R-Square Measures the Proportion of Variations Explained by the Regression, t-Test for Individual Regression Coefficient. Distributional Claims. General Form of Confidence Intervals. QQ-Normal Plot for Residuals.
3. Two-Sample Problem: t-Statistic and t-test under equal variance assumption and normal model, with reference distribution being t with $n\_1+n\_2-2$ degrees of freedom. F-test for equal variance assumption. Equivalent of an F-test (and its accompany analysis of variance) and the usual t-test for the equal mean hypothesis. For normal data but likely unequal variance in two sample problem, we introduce Welch’s t-test. In class demonstration of these tests, when normality assumption, equal variance assumption are violated. The emphasis on strict correct reference distribution and the useful test procedures. Permutation or randomization test that is justifiable without parametric model assumptions.
4. One-Way Layout and Multiple Comparison, Analysis of Variance, F-Test, Bonferroni and Tukey’s Methods, Simultaneous Confidence Intervals. Quadratic Forms. Random effects. Sample size calculation. Additional Information on Standard Distributions: Chi-Square, t- and F-Distributions. Non-Central Versions. Sample Size Calculation.
5. Experiment with blocking factors. Difference between the blocking factor and other factors. Rationales with design without replicates. Analysis of variance table. Multiple comparison issues and Bonferroni and Tukey’s Methods.
6. Factorial design with two factors and extensions. The structure of the design, analysis of variance, factorial model and the linear regression model. Random effects. Some algebraic and distributional results.
7. Factorial design at two-levels. analysis of variance. main effect, two-factor interaction, interaction plot, within cell replications, factorial model and the linear regression model.
8. Fractional factorial design. Defining contrast subgroup, alias patterns, resolution, normal plot, half normal plot, Analysis of variance table.
9. Analysis of covariance. A model with both qualitative factor and quantitative factor.

**Computing:**

Programming languageR will be used for all computations in this course.

Download and install R: <https://cran.r-project.org/>

R-Studio: <https://rstudio.com/products/rstudio/download/>

Add-on Packages: R is an extensible system and many people share useful codes they have developed as packages via CRAN and github. To install a package from CRAN, for example the [plyr](http://plyr.had.co.nz/) package for data aggregation, here is one way to do it in the R console (there are others).

install.packages("plyr", dependencies = TRUE)

**Official manual for R**: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

Do not be intimidated by the volume of this manual. Students can learn from examples given in classes and from each other.

**Jupyter**: This tool enables me to create lectures with embedded R codes and other content. I encourage students to install Jupyter for easy access to the slides and the ability to experiment with the codes. Instructions for installing Jupyter can be found in the link provided below. Remember, your fellow students can be your best teachers. Make sure to attend labs and interact with the students around you, or reach out to the TA for assistance..

The codes in my Jupyter slides can be used via copy-and-paste without a jupyter server.

<https://jupyter.org/install>

Avoid running programs on the university jupytor server during lectures. The system may crush.

It is also possible to read Jupytor files online. Try it yourself and teach each other. I am not the best resource.

i**Clicker cloud:**

We utilize iClicker Cloud to promote attendance, engage you during lectures, and assess your comprehension of the course content. Participating is key, and the emphasis is on engagement rather than having all the answers correct. There is ample leeway for missing lectures or providing incorrect responses to in-class questions before it significantly impacts your final grade. If you happen to miss a few lectures due to illness or a family emergency, there's no need to inform me individually, as keeping track of a large number of students can be quite challenging.

https://lthub.ubc.ca/guides/iclicker-cloud-student-guide/

**Assignments, mini-quizzes, midterm and final:**

There will be 10+ mini-quizzes. You will be asked to answer a few generally effortless questions regarding the contents of the latest lectures.

There will be 5 assignments. You are asked to write the solutions on papers, scan them into pdf and upload them to GradeScope.

In-person **midterm** will be held Oct 19 from 3:30-5:00pm in rooms to be announced.

**Final exam** date and time will be arranged at the university level.

**Final grade:**

A = {10% mini-quiz + 25% Midterm + 20% Assignment + 45% final exam}.

One must achieve 40%+ in the final exam to pass this course. Otherwise, the maximum final grade is 45% and the next rule will not apply.

Iclicker scores (60, 65, 70, 75, 80+) are first converted to B = (6, 7, 8, 9, 10). The final grade will be (A + B)/(A + 10) if it helps. A student with A = 60 and B = 10 will have their final grade (60+10)/110 = 64.

**Marking of Assignments**

II will instruct Teaching Assistants to apply a more rigorous approach when grading assignments. In simpler terms, the likelihood of receiving partial marks will decrease if the overall structure of the solution is flawed. The objective is to provide students with clearer feedback, marking incorrect answers as wrong rather than partially correct for partial credit.

Following this, we will adjust the assignment scores using a formula: 40 + [(A - 40) \* 60]^(0.5) if A > 40%, and A if A < 40%. This adjustment aims to alleviate stress. This process will be carried out for each of the five assignments independently. The ultimate goal is to foster a comprehensive understanding of the material.

**Teaching philosophy**

The instructor places a strong emphasis on explaining the underlying motivations behind various designs covered in this course. This involves justifying the recommended approaches for data analysis and effectively interpreting the results of statistical inference. While mathematical justifications will be presented, their testing will be moderate to allow some students to distinguish themselves. Simulation experiments will be employed to critically assess whether the stated statistical methods indeed possess the claimed statistical properties, particularly when certain required conditions regarding data distributions are violated.

To earn credit for this course, students are expected to possess a solid grasp of the fundamental concepts in experimental design. This requirement is reinforced by including questions about replication, blocking, and randomization in both the midterm and final exams.

For success in this course, students should be well-acquainted with the introduced designs, capable of conducting the relevant data analysis using basic computational tools rather than relying solely on an all-encompassing R function. Additionally, they should be able to interpret the outcomes of their data analyses.

To excel in this course, students need to comprehend the principles that underlie the various data analysis methods. They should be able to adeptly apply these principles to address questions in diverse yet analogous situations.

**Policies:**

We expect all students to attend all lectures. Late submission results in a 50% mark deduction. Students who miss midterm or an assignment due to illness need to provide a self-declaration with the Instructor as soon as possible. Failing to present a declaration results in a grade of zero. When a concession is granted, the final mark will be evened out on other parts of the course work. There will be no make-ups. **We calculate alternative marks following a general principle and wish not to have all scenarios deliberated and explicitly published.**

**UBC policies for Academic concessions:**

**<https://www.calendar.ubc.ca/vancouver/index.cfm?tree=3,329,0,0>**

**UBC policies related to exam issues**

<https://science.ubc.ca/students/advising/exams>

**We use: student declaration of academic concession.**

**UBC policy with regards to Academic Misconduct:**

<https://www.calendar.ubc.ca/vancouver/index.cfm?tree=3,54,111,0>

Acknowledging that UBC Vancouver is located on the traditional, ancestral, and unceded territory of the Musqueam people is an important way to remind learners that UBC and the people who study, work, live, and play within the institution have responsibilities that emerge from past and ongoing relationships with Indigenous host nations.