Mathematical Statistics With R —

Stat 5810, Section 001 & Stat 6810, Section 001

Spring 2025 (2 Credits)

Instructor: Dr. Jürgen Symanzik
Office: AnSc 313
Phone: 435–797–0696
e–mail: juergen.symanzik@usu.edu
Web: `https://www.usu.edu/math/symanzik/`
`https://www.usu.edu/math/symanzik/teaching/2025_stat5810/stat5810.html`

Office Hours: Monday (M) & Wednesday (W) 3:30pm – 4:30pm in AnSc 313,
Thursday (H) 9:00am – 10:00am (via Zoom), and by appointment.

Classes & Rooms:
MWF 2:30pm – 3:20pm, W 1/22 – F 4/4, 2025 (tentatively): AnSc 320 (face–to–face).

Please visit the course Web page listed above for emergency announcements, e.g., when Canvas is unavailable. Otherwise, visit Canvas frequently for lecture notes, data sets, R code, etc. — in particular if you miss our face–to–face lectures for any reason. All (additional and updated) materials, announcements, discussions, recordings, etc. from Canvas are part of the course materials. Not seeing one of these in time does not serve as an excuse for not getting point deductions for the course. Deadlines may change or unexpected new regulations and requirements may occur. It is your responsibility to make sure to receive all announcements in time.

Detailed Class Schedule:
For a 2–credit course, we need 28 lectures/lecture days (in contrast to 41 lectures/lecture days for a 3–credit course). Those days are marked as "Lecture 01" to "Lecture 28" in the overview below:

Week	Monday	Wednesday	Friday
1	1/6: No class	1/8: No class	1/10: No class
2	1/13: No class	1/15: No class	1/17: No class
3	1/20: No class	1/22: Lecture 01	1/24: Lecture 02
4	1/27: Lecture 03	1/29: Lecture 04	1/31: Lecture 05
5	2/3: Lecture 06	2/5: Lecture 07	2/7: Lecture 08
6	2/10: Lecture 09	2/12: Lecture 10	2/14: Lecture 11
7	2/17: No class	2/19: Lecture 12	2/21: Lecture 13
8	2/24: Lecture 14	2/26: Lecture 15	2/28: Lecture 16
9	3/3: Lecture 17	3/5: Lecture 18	3/7: Lecture 19
10	3/10: No class	3/12: No class	3/14: No class
11	3/17: Lecture 20	3/19: Lecture 21	3/21: Lecture 22
12	3/24: Lecture 23	3/26: Lecture 24	3/28: Lecture 25
13	3/31: Lecture 26	4/2: Lecture 27	4/4: Lecture 28
14	4/7: Backup	4/9: Backup	4/11: Backup
15	4/14: Backup	4/16: Backup	4/18: Backup
16	4/21: Backup		

Note: "No class" means guaranteed no class that day. I have marked the dates in the three last weeks of the semester as "Backup", e.g., in case we miss lectures because of weather conditions, sickness, or any other reason. If nothing goes wrong, our tentative last lecture date will be on F 4/4/2025. Backup lectures (if any) will be held via Zoom (and not face–to–face). There is no need to stay in Logan after F 4/4/2025 for a backup lecture in this course.

Course Objectives:

This course will look into selected concepts from MATH 5710 (and 5720) that are typically explored theoretically and on paper in those two courses — but here we will use R and its numerous packages to find computer–based solutions. Simulation will be a major component to bring more meaning to some of the theorems and results from MATH 5710 (and 5720). This course does not intend to replace MATH 5710/5720 or STAT 6710/6720, but rather has its place in between those two course sequences and is ideal for everyone who likes doing things in R, but sometimes has problems to fully understand what some of the theorems really mean.

Prerequisites:

STAT 5050 ("Introduction to R") with a C− or better. MATH 5710 ("Intro to Probability") with a C− or better.

Recommended (or to be taken concurrently) is MATH 5720 ("Intro to Mathematical Statistics"). Alternatively, STAT 6710/6720 can serve as higher–level replacements for MATH 5710/5720. Knowledge from STAT 5550 ("Statistical Visualization I") and STAT 5080 ("Data Technologies") would be beneficial, but is not required.

Moreover, you should be familiar with a tool such as R Markdown, knitr, or Sweave that allows you to combine text, R code, graphics, and numerical results in high–quality documents. LaTeX is a plus, but is not formally required at the 5000 level, but it will be required at the 6000 level of this course.

IDEA Center Learning Objectives:

Objective 1) Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories).

Objective 4) Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.

Objective 13) Learning appropriate methods for collecting, analyzing, and interpreting numerical information.

Topics: (additional topics will be added as time permits)

This course will revisit (and extend) selected topics from MATH 5710/5720 with a focus on R:

1. Using R to Simulate Events.
2. Basic Probabilities, Distributions, and Random Variables via R.
3. Transformations of Random Variables via R.
4. The Inverse CDF Sampling Technique.
5. Establish Convergence and Limit Theorems via R.
6. Permutation Tests.
7. The Jackknife Estimator.
8. Bootstrap Methods.
9. The Expectation–Maximization (EM) Algorithm.
10. Others (as time permits).

For MS and PhD students majoring in Statistics, it is important to learn LaTeX — from basic document preparation, over the inclusion of R graphics into your LaTeX documents to advanced topics such as Sweave (`https://stat.ethz.ch/R-manual/R-devel/library/utils/doc/Sweave.pdf`), knitr (`https://yihui.org/knitr/`), and the LaTeX bibliography BibTeX (`http://www.bibtex.org/`). LaTeX is essential for graduate work (at the MS and PhD level) and will be used for many theses, dissertations, and scientific publications. **Therefore, LaTeX will have to be used for all homeworks, projects, presentations, etc. at the 6000 level of this course.**

Course Format and Lecture Activity Points:

The course will be offered in a face–to–face format, implying that you should attend most of the lectures in person. There will also be lecture recordings via Kaltura that will be posted in the "Media Gallery" in Canvas in the evenings of a lecture day. I will also try to open a Zoom link through Canvas at the time of planned in–class lecture activities so that you can (partially) participate in those activities and earn some credits for your remote participation. However, there will be limitations to what can be captured via the Kaltura recordings and what can be done via the Zoom link — so try to attend the lectures in person whenever possible.

Under this setup, it will be possible to conduct activities during the lectures and award credit for participating in those activities, ideally in person, but alternatively via Zoom and/or the Canvas "Discussions". You may be asked to answer some short questions during a lecture, do some programming and then submit your code into the Canvas "Discussions", or fill out a worksheet on paper. Some of these activities are stand–alone while others build

on each other during the course of a lecture (and thus the need to be able to participate in real time via Zoom).

Points earned through such activities count as Lecture Activity Points (LAPs). Short stand–alone activities will result in one LAP and typically will not be announced in advance. Extended sequences of activities may result in up to 10 LAPs in a lecture and will usually be announced in the preceding lecture and/or a Canvas announcement.

LAPs will be awarded on a done (answered) / not done (not answered) basis and not on correctness. This should encourage everyone to think about a question or try some coding without the pressure to do things immediately right when trying.

I cannot say in advance at this time how many lectures there will be with such activities, but I would estimate somewhere around 10 ± 5 lectures. Some of the stand–alone activities may be answered prior to a lecture when posted in a Canvas "Discussion". Extended sequences of activities typically cannot be answered prior to a lecture as they may require some of the lecture materials and answers to some of the earlier activities in such a sequence of activities.

Points from LAPs will contribute to 10% of your course grade. You will obtain 10 points for 90% or more of all possible LAPs, 9 points for 80% up to 90% of all possible LAPs, 8 points for 70% up to 80% of all possible LAPs, 7 points for 60% up to 70% of all possible LAPs, 6 points for 50% up to 60% of all possible LAPs, and 0 points for less than 50% of all possible LAPs.

So, you can miss (for private reasons) some lectures where LAPs are provided and still get the full number of course points for LAPs. However, you should not miss too many of the lectures where up to 10 LAPs will be provided. Thus, make sure that you do not miss an announcement in class and/or in Canvas before a lecture with a large number of LAPs is coming up — or simply plan on attending all lectures in person.

In case of an excused absence, e.g., for medical reasons, family emergencies or funerals, court appointments, university–approved travel, etc., please provide some supporting information and your LAP score will be adjusted according to the number of lectures with LAPs you could attend. Private reasons such as travel, most family events (such as weddings), etc. do not count as an excused absence.

Homework Assignments:

There will tentatively be 3 HW assignments for this course, roughly one every three weeks. Each HW assignment will include a value (typically 20–100 points) that it will be scored out of. **HW assignments will contribute to 90% (for Stat 5560), respectively 60% (for Stat 6560), of your course grade.** The value of each HW assignment will be roughly proportional to its importance and the amount of work involved.

You will be allowed to discuss general approaches to questions on the HW assignments with other students, but each student must write and submit their own R code and comments. Any students caught sharing R code or other parts of their homework submissions will fail the class.

Unless otherwise stated on the HW assignment sheet, all homework assignments have to be submitted electronically via Canvas. HW assignments will be posted tentatively in weeks 6, 9, and 13 of the semester and will be tentatively due in weeks 9, 13, and 16 of the semester (see the the *Detailed Class Schedule* above how these weeks approximately

translate to dates). **You will have at least 2 weeks after the last lecture to finalize and submit the final HW assignment.**

There will be no (in–class or take–home) quizzes, midterm exams, or final exams. Nevertheless, this will be a very challenging course that requires a lot of individual time to work on the assignments. Just attending classes will not be enough to pass this course! **In addition, you will have to do a lot of individual reading of textbooks, online documentation, and help pages, and search for available information on the web.**

No Excuse Needed Late Homework Submission Policy:

Each student has **3 tokens** for late homework submissions. These tokens can be used in multiple ways, e.g., one token at a time for each of the 3 HWs or all 3 tokens can be used for a single late HW submission. You will need 1 token if your HW is 1 min – 24 hours late; you will need 2 tokens if your HW is $>$ 24 hours – 48 hours late; and you will need all 3 tokens if your HW is $>$ 48 hours – 72 hours late. Once you have used up all 3 tokens, late HWs will count as 0 points, even if such a HW is only one additional minute late. As an example, if you used 2 tokens early on, you will have only 1 token left and can submit your next HW only up to 24 hours late. If you submit that HW more then 24 hours late, it will count as 0 points. All times are based on the Canvas time stamp that is created when a HW gets submitted — and not on your local computer clock.

It is your responsibility to keep track of the number of tokens you have used. Record that information on your phone or check how you can extract the submission dates and times in Canvas on your side. I will not be able to tell you shortly before the next submission deadline how many tokens you have left. In fact, I only plan to use that information when I determine final course grades. You may therefore get assigned a score $>$ 0 points for a certain HW submission initially, but that score may get adjusted to 0 points if it turns out that this was for a late submission and you had already used up all your tokens.

You will not get any credit for any unused tokens, so you could use the remaining ones for proofreading, checking that all figures and R code are included, etc., in particular for the later HWs — and if you still have tokens left at that time!

As the name indicates, there is no excuse needed when you use a token, be it for a minor sickness, personal travel due to family events, or just needing more time as you couldn't finish that HW by the deadline. If you are officially absent from USU when a HW is due, e.g., due to travel with a USU sports team or band, attending a conference, doing fieldwork, etc., I need an official statement from a coach or supervisor — and no token will be used. However, we then need to determine on a case–by–case basis when that particular HW will be due for you — and when you would have to use additional token(s) to further extend your extended deadline.

You can resubmit your HW as often as you want in Canvas prior to the deadline. Apparently, if you resubmit it after the deadline, one (or more) tokens will be used. In that case, also let me know via e–mail at the time of the deadline that you are still working on your HW beyond the original deadline and plan to resubmit a new version later on. Once a HW has been graded on my side, you will no longer be allowed to resubmit a revision of that HW (or in case you did, you still would get the points based on the last submission prior to the original deadline).

Project (Stat 6560 only):

There will be one major project towards the end of the semester. This will include the preparation of a final project report and a short presentation of your work for the other students in this course. The project will be done in a small group of students. **The project will account for 30% of your course grade.**

Grading Scheme:

I do not specify any definite grade cutoffs in advance, but I typically assign the same grade to all students that fall into the same narrow cluster of points that extends below one of the commonly used grade cutoffs (i.e., 93% for A, 90% for A−, 87% for B+, 83% for B, 80% for B−, 77% for C+, 73% for C, 70% for C−, 67% for D+, 60% for D, and <60% for F). So, for example, if one student earned 92.9% of points and the next higher student earned 93.1% of points, both students almost certainly will get an A grade. However, if one student earned 92.9% of points and the next higher student earned 95.0% of points, this will almost certainly translate into A− and A grades, respectively.

Textbooks:

Chihara, Laura M., and Hesterberg, Tim C. (2022) *Mathematical Statistics with Resampling and R, 3rd Edition*, Wiley, Hoboken, NJ,
`https://www.wiley.com/en-jp/Mathematical+Statistics+with+Resampling+and+R%2C+3rd+Edition-p-9781119874034`,
`https://github.com/lchihara/MathStatsResamplingR`, and
`https://CRAN.R-project.org/package=resampledata3`.

Meyer, Mary C. (2019) *Probability and Mathematical Statistics: Theory, Applications, and Practice in R*, SIAM, Philadelphia, PA,
`https://doi.org/10.1137/1.9781611975789`.

Ramachandran, Kandethody M., and Tsokos, Chris P. (2021) *Mathematical Statistics with Applications in R, 3rd Edition*, Academic Press / Elsevier, London / San Diego, CA / Cambridge, MA / Oxford,
`https://doi.org/10.1016/C2018-0-02285-9`,
`https://shop.elsevier.com/books/mathematical-statistics-with-applications-in-r/ramachandran/978-0-12-817815-7`, and
`https://educate.elsevier.com/book/details/9780128178157`.

These textbooks are resources for extended reading, but it is not necessary that you buy any of these books. Unfortunately, the USU library only provides limited online access to these books. If you plan to work in the area of Mathematical Statistics with R for your MS or PhD degree, you should consider to purchase some of these books for an ongoing use beyond this course.

Software:

We will primarily be using R (`http://cran.r-project.org/`), a free software environment for statistical computing and graphics. Please install a recent version of R, i.e., 4.4.2 or later, on your own computer so we can exchange code. Working with version 4.4.0 or 4.4.1 should still be OK. I would discourage the use of version 4.3.3 or older. Also install the most recent version (2024.12.0 Build 467, released on 2024-12-11) of RStudio (`https://posit.co/`) as a front end to R. **Note that some versions of RStudio released earlier in 2024 contained bugs that prevented them from opening certain Rnw files, similar to the ones used in class. If you get error messages when you try to open**

some of the files from class, check whether you have the most recent version of RStudio installed — and if not, upgrade to the most recent version. For Windows, also install Rtools 4.4 (`https://cran.r-project.org/bin/windows/Rtools/`).

For most work in this course, using the *tinytex* R package is all that is needed, but if you plan to work on more complicated LaTeX documents later on (such as your MS thesis or your PhD dissertation), you should also install a full version of LaTeX that can be used in connection with RStudio. My personal recommendation is TeX Live (`https://www.tug.org/texlive/`).

Courtesy:

Please turn off cell phones and similar devices before class, and please keep conversations to a minimum during lectures. Please do not read/reply to your e–mails or browse other web pages than the ones discussed during class.

I will not keep track if you come to class or not. However, I would highly recommend to attend all lectures.

Americans with Disabilities Act:

If a student has a disability that will likely require some accommodation by the instructor, the student must contact the instructor and document the disability through the Disability Resource Center (DRC – `https://www.usu.edu/drc/`), preferably during the first week of the course. Any requests for special considerations relating to attendance, pedagogy, taking of examination, etc. must be discussed with and approved by the instructor. In cooperation with the Disability Resource Center, course materials can be provided in alternative formats — large print, audio, or Braille.

Note:

The above schedule and procedures in this course are subject to change in the event of extenuating circumstances.