Math 341 / 641
Fall 2023 (4 credits)
Course Syllabus

Adam Kapelner, PhD.

Queens College, City University of New York

document last updated Wednesday 15^{th} November, 2023 11:28pm

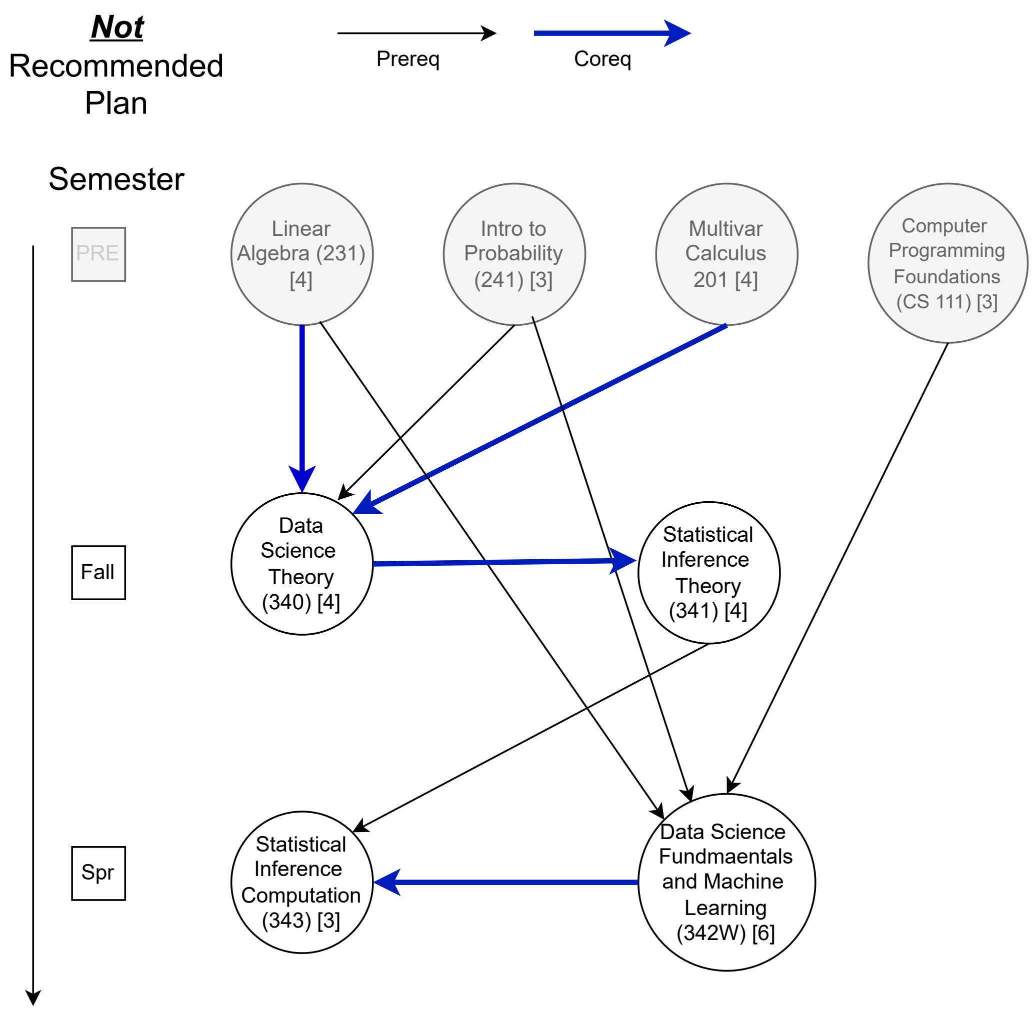
Course Homepage	`https://github.com/kapelner/QC_Math_341_Fall_2023`
Slack Homepage	`https://QCMath341Fall2023.slack.com/`
Contact	kapelner@qc.cuny.edu
Office	604 Kiely Hall
Lecture Time and Loc	Mon and Wed 6:30 – 8:20PM / KY 277
Instructor Office Hours and Loc	see course homepage
TA / Office Hours	Javendeen Naipul / see course homepage

Course Overview

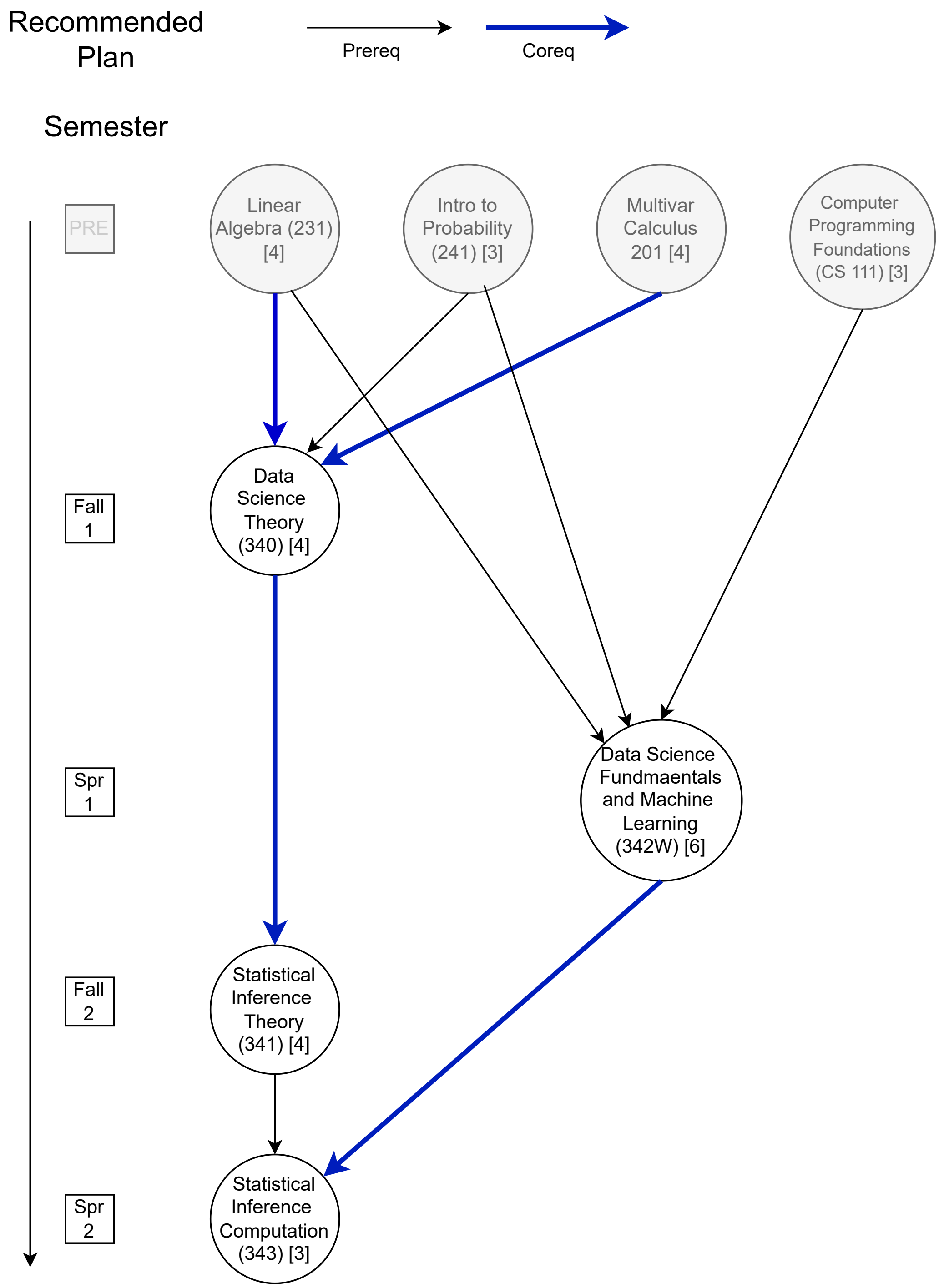
MATH 341 / 641. Statistical Theory for Data Science. 4 hr.; 4 cr. Coreq.: MATH 340 or 640. Sampling, point estimation via method of moments and maximum likelihood, loss and risk, bias-variance decomposition, frequentist confidence intervals, frequentist hypothesis testing, the p-value, type I/II errors, power, the testing problem, the model selection problem, confounding, randomized experimentation and causality, Bayesian inference, prior distribution specification, conjugacy, Bayesian point estimation, credible regions, posterior predictive distribution, Beta-Binomial and Normal-Inverse-Gamma model, consistency, Cramer-Rao lower bound, sufficiency, identifiability, delta method. Not open to students who are taking or who have received credit for MATH 341. Students cannot receive credit for both: MATH 341 and 641. Fall, Spring

The Four Data Science Core Classes

This course is one of the four data science core courses but it does not cover any typical "data science" topics. Instead, it is designed to provide theoretical skills for Math 341 and 343. Math 340 and 342W are designed to be standalone courses and the other two courses rely on topics covered therein. Thus there is an order the classes need to be taken. Below are two plans, the first is over two semesters and hence it is not *not recommended* as it will be a *very* heavy workload. The second is over four semesters and it is the recommended plan as I believe it will allow students to absorb the material more effectively:



Examining the above, we note that MATH 341 and 343 form a series of two statistics courses: the first, theoretical with traditional topics and the second, computational with modern topics. Both heavily rely on the theoretical topics taught in MATH 340.


Recommended Plan
Prereq
Coreq
Semester
PRE
Fall 1
Spr 1
Fall 2
Spr 2
Linear Algebra (231) [4]
Intro to Probability (241) [3]
Multivar Calculus 201 [4]
Computer Programming Foundations (CS 111) [3]
Data Science Theory (340) [4]
Data Science Fundmaentals and Machine Learning (342W) [6]
Statistical Inference Theory (341) [4]
Statistical Inference Computation (343) [3]

Tentative Day-by-Day Schedule

Lectures and their topics with rough time estimates per topic are below:

Day 1 [20min] Surveys, populations, sampling, sample size n, representativeness, SRS; [30min] definition of parameter, parameter space, inference, statistic, the three goals of statistical inference: point estimation, theory testing and confidence set construction; [20min] Sampling with/without replacement, iid assumption, equivalence of populations and data generating processes (DGPs) [55min] estimators, point estimation metrics: biasedness and unbiasedness, loss functions (absolute, squared error, others), risk function, mean squared error (MSE), bias-variance decomposition of MSE, maximum risk for iid Bernoulli model;

Day 2 [15min] Comparison of bad estimator with sample average [60min] Introduction to hypothesis testing, the intellectual honesty of assuming the null hypothesis H_0, the theory you wish to prove H_a, right, left and two-sided tests of single parameters, test statistics, null distributions, retainment and rejection regions, the two test outcomes, statistical significance [15min] the Binomial exact test of one proportion in the iid Bernoulli DGP [15min] size, level, Type I and II errors, 2×2 confusion table of testing errors, scientific convention of α

Day 3 [10min] Facts about normal distributions [30min] Motivation of exact Z tests for one sample with known variance [5min] equivalence of size and level [15min] Fisher's p-value and calculations under normal sampling distributions; [25min] Definition of power, assuming a true sampling ditribution to compute the probability of Type II errors and power [15min] Power function and its inputs [10min] Power convention in study design

Day 4 [20min] θ_0 in one-sided, one-sample tests and a discussion about what "retain" vs "accept" means in the context of scientific theories [15min] the exact two-sample z-test under different variances and shared variance [5min] review of moments, definition of sample moments; [15min] the system of equations yielding the method of moments estimators (MME); [10min] MME for the expectation, MME for the variance; [20min] MME for the two parameters in the iid Binomial DGP, nonsensical MME values

Done updating until here

Day 4 [15min] The power function for a one-sided z-test; [15min] the iid normal DGP; [15min] exact one-sample z-test with known variance, its power function; [30min] the naive estimator for σ^2, proof of its biasedness, definition of asymptotic unbiasedness, Bessel's correction and the unbiased sample variance estimator S^2; [20min] motivation of the t statistic, Student's t distribution, the one-sample t-test; [20min] the concept of two populations, two sample testing for mean differences, H_0 specification for left-sided, right-sided, two sided tests

Day 5 [15min] the exact two-sample z-test under different variances and shared variance; [20min] the exact two-sample t-test under shared variance, the pooled standard deviation estimator; [25min] the approximate t-test (Welch-Satterthwaite approximation) under different variances, the Behrens-Fisher distribution; [5min] review of moments, definition of sample moments; [15min] the system of equations yielding the method of moments estimators

(MME); [10min] MME for the expectation, MME for the variance; [20min] MME for the two parameters in the iid Binomial DGP, nonsensical MME values

Day 6 [10min] MME for the iid Uniform DGP with one unknown endpoint parameter; [10min] definition of likelihood, the equivalence of the likelihood function with the JDF/JMF; [10min] definition of argmax, equivalence of argmax under strictly increasing functions; [10min] definition of maximum likelihood estimators (MLE) and estimates, the log likelihood; [10min] MLE for the iid Bernoulli DGP; [10min] MLEs for the iid Normal DGP; [10min] MLE for iid Uniform DGP with one unknown endpoint parameter; [5min] variances of MME and MLE for that cases

Day 7 [5min] definition of relative efficiency and comparison of two estimators; [10min] the nonexistence of a minimum MSE estimator; [5min] definition of uniformly minimum variance unbiased estimators (UMVUEs); [10min] definition of the Cramer-Rao Lower Bound (CRLB), definition of Fisher Information

Day 8 [55min] proof of the CRLB, definition of the score function, expectation of the score function is zero; [10min] proof that the sample average is the UMVUE for the iid Bernoulli DGP; [15min] proof that the sample average is the UMVUE for the iid normal DGP; [10min] definition of asymptotically normal estimators; [15min] definition of consistent estimators, continuous mapping theorem, Slutsky's theorem

Day 9 Midterm I Review (Monday, Oct 2)

Day 10 Midterm I (Wednesday, Oct 4)

Day 11 [10min] asymptotic normality of asymptotically normal estimators when using the estimator for its standard error; [5min] statement of main theorem for MMEs and MLEs: consistency, asymptotic normality, asymptotic efficience of the MLE; [5min] review of Taylor series; [40min] proof that the MLE is asymptotically normal and asmyptotically efficient; [5min] definition of the Wald test, the one-proportion z-test as a Wald test; [5min] the one-sample t-test as an approximate one-sample z-test (Wald test);

Day 12 [15min] derivation of the MLE, Fisher Information and a Wald test for the iid Gumbel DGP with known scale parameter; [15min] introduction to confidence sets, interval estimators, coverage probability; [20min] definition of the confidence interval (CI), CI construction via hypothesis test inversion; [10min] comparison of hypothesis testing with CI construction; [15min] approximate CIs for the iid normal DGP under the four assumptions; [10min] CI for one proportion; [10min] CIs for MLEs; [10min] demonstration that MSE improvements improve all three statistical inference goals, illustration of all three goals

Day 13 [10min] meaninglessness of single inferences; [5min] odds-against reparameterization, odds-against point estimation; [15min] univariate delta method; [10min] CI for odds-against via delta method, CI for log-mean; [10min] risk ratio versus proportion difference; [20min] multivariate delta method; [5min] CI for the risk ratio; [25min] statistical significance vs. clinical / practical significance of the effect

Day 14 [30min] Problems and limitations with frequentist CIs and testing, valid interpretation of frequentist CIs, the frequentist p-value; [35min] review of definition of conditional probability, Bayes Rule, Bayes Theorem; [10min] marginal and conditional PMFs; [10min] Bayes rule for two rvs; [10min] anatomy of the Bayes identity: the likelihood, prior, prior predictive distribution and posterior distribution

Day 15 [40min] example posterior calculation with discrete parameter space and principle of indifference; [15min] Bayesian point estimation with the maximum a posteriori (MAP) estimate, conditions for equivalence with the MLE; [25min] Proof that Bayesian Inference is iterative in the data; [15min] uniform prior for the bernoulli iid model; [10min] Bayesian point estimation with the posterior median and posterior expectation

Day 16 [20min] derivation of general beta posterior for the bernoulli iid model, intro to beta distribution, beta function, gamma function; [5min] point estimation with beta posterior; [10min] all legal shapes of the beta distribution; [35min] the beta-binomial bayesian model; prior parameters (hyperparameters) and posterior parameters, point estimates; [10min] definition of conjugacy, beta-binomial conjugacy; [10min] pseudodata interpretation of the prior parameters; [10min] shrinkage estimators and the beta-binomial posterior expectation as a shrinkage estimator

Day 17 [15min] One-sided and two-sided credible regions (CRs); [5min] CR for beta-binomial model; [10min] high density regions; [10min] decisions in the Bayesian framework for one-sided hypothesis testing, Bayesian p-values; [15min] beta-binomial examples; [20min] two approaches for two-sided testing in the Bayesian framework; [30min] posterior predictive distribution formula, example for one future observation in the beta-binomial model

Day 18 [15min] mixture and compound distributions; [65min] the betabinomial distribution as an overdispersed binomial, example with birth data, proof of the general posterior predictive distribution for the beta-binomial model; [10min] Laplace and Haldane priors; [20min] Informative priors for the beta-binomial model, example with baseball batting averages, shrinkage in informative priors, empiral Bayes estimation

Day 19 Midterm II Review (Monday, Nov 6)

Day 20 Midterm II (Wednesday, Nov 8)

Day 21 [10min] definition of odds, reparameterization of the binomial with odds; [5min] PDF change of variables formula, proof that prior of indifference for binomial probability is not prior of indifference for odds; [15min] Jeffrey's prior specification concept; [10min] PDF/PMF decomposition into kernel and normalization constants; [10min] definition of Fisher information, computation of Fisher information for the binomial distribution; [30min] Definition of Jeffrey's prior, derivation of Jeffrey's prior for the beta-binomial model, verification that it robust to reparameterizations of the binomial model's parameter; [10min] proof of Jeffrey's prior satisfies Jeffrey's prior specification concept;

Day 22 [10min] derivation of Poisson model; [15min] derivation of the Poisson model's conjugate prior via kernel decomposition (the Gamma); [20min] Gamma shapes and properties; [15min] pseudodata interpretation of hyperparameters in the gamma-poisson model;

[20min] derivation of the shrinkage point estimator for the gamma-poisson model; [10min] CRs for the gamma-poisson model;

Day 23 [20min] uninformative priors for the gamma-poisson model [45min] derivation of the posterior predictive distribution being extended negative binomial in the gamma-poisson model; [20min] kernel decomposition of the normal PDF;

Day 24 [35min] derivation of the normal-normal conjugate model, pseudodata interpretation of the hyperparameters; [20min] Haldane prior for normal-normal model, point estimation in the normal-normal model, Jeffrey's prior derivation, shrinkage estimator; [40min] derivation of the normal posterior predictive distribution for the normal-normal model; [10min] derivation of the inversegamma distribution, properties of the inverse gamma distribution [35min] normal-inversegamma model, laplace prior, pseudodata interpretation of the hyperparameters, haldane prior

Day 25 [75min] The two-dimensional normal-inverse-gamma (NIG) distribution, its kernel, its use in bayesian inference for the conjugate NIG-NIG model; [15min] Marginal mean T distribution in the NIG posterior; [15min] Marginal variance inverse-gamma distribution in the NIG posterior

Day 26 [55min] derivation of the Student's T posterior predictive distribution in the NIG-NIG model; [30min] Sampling from the NIG distribution; [25min] the kernel of the semiconjugate NIG model

Day 27 [35min] concept of assuming DGP models, discussion of what models are, discussion of many model candidates; [15min] model selection via largest likelihood, asymptotic bias of the log-likelihood estimator with substituted MLEs; [15min] the AIC metric, AIC model selection algorithm, penalizing complexity; [10min] the AICC metric; [10min] the BIC metric; [10min] Akaike weights

Day 28 Final Review

This is more of a typical mathematics theory course than the rest of the data science series. But we will still attempt to keep our eye on developing ideas and concepts for helping to make decisions in the real world. Thus we may make limited use of computation using the `R` statistical language.

Corequisites

MATH 340 or an intermediate probability class that covers similar topics. Critical is coverage of the Central Limit Theorem, convergence in probability, Slutsky's theorem, Cochran's theorem.

Course Materials

Textbooks: For the first 2/3 of the class, we will be doing classical or "Frequentist Inference". Here, I will be referencing Larry Wasserman's *All of Statistics: A concise course in statistical inference* which can be purchased on Amazon and Casella and Berger's *Statistical Inference* which can be purchased on Amazon. There is no excuse not to have these books. They are *required*. However, I will not ususally be teaching "from the book" — most of the material in the class comes from the lecture notes. The textbooks are a way to get "another take" on the material and they will only cover about only half of the material done in class. For the other half, you will have to make use of other resources. I also recommend Rice's *Mathematical Statistics and Data Analysis*, 3rd edition which can be purchased on Amazon as well but I will not reference it during class. When the class transitions to the last third, we will switch to "Bayesian Inference". A good introductory book is *Introduction to Bayesian Statistics* by William M. Bolstad First Edition. It can be purchased used on Amazon. This is *recommended*. The Bayesian Bible is *Bayesian Data Analysis* by Andrew Gelman et al. which can be purchased on Amazon as well. This latter book is very dense and we will only scratch the surface of its topics in this class. Note that these books constitute a means to get "another take" on the material. However, most of the material in the class comes from the lecture notes.

Popular Books: For the first Frequentist section of the class, I recommend *The Lady Tasting Tea* by David Salsburg which can be purchased on Amazon. For the Bayesian section of the class, I recommend *The Theory that Would not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy* by Sharon B. McGrayne which can be purchased on Amazon.

Computer Software: During lectures, there will be demoes using `R` which is a free, open source statistical programming language and console. You can download it from: `http://cran.mirrors.hoobly.com/`. I do not expect you to do *any* programming. If `R` will be part of the homework, I will give you the code to run and expect you to interpret the results based on concepts explained during the course.

Calculator: You can use a TI-84, 85, 89 or any calculator which you wish. I strongly suggest you use Wolfram Alpha and its smartphone app.

The 641 Section

You are the students taking this course as part of a masters degree in mathematics. Thus, there will be extra homework problems for you and you will be graded on a separate curve.

The Use of Slack and Github as a Learning Management System

As the course homepage is updated (e.g. a new homework assignment is posted), you will hear about it in slack. Each assignment will have its own channel. You can feel free to discuss things with your fellow students there. If you are asking me a question, you must do so in the `#discussions` channel for a general questions or the assignment-specific channel (e.g. `#HW03`) so other students can see the question and benefit from the answer. If you pm me, I will not answer and just ask you to move it to the public channel. Do not be afraid to ask questions. There are many people who will have your same question!

Slack is a wildly successful company that recently got bought by Salesforce because businesses *really* use it. Pretend you are working at one of these businesses: no posting about random stuff; keep things professional!

We will not be using any features of github for learning management. Do not open "issues" on github!

Announcements

Course announcements will be made via slack in the `#general` channel (not on email). I am known to send a few slack messages per week on important issues. The Slack workspace `#general` channel is also connected to the course homepage via an integration. So every time I change the homepage (e.g. to release a new homework or upload notes or a video), you will get a notification.

I can't stress the following enough: if you are not on slack, you will miss all class announcements!!! Slack notifies you when there are messages. You may wish to mute all channels except for #general. See this article for how to do that.

Class Meetings

There are 28 scheduled meetings. Of these, 23 will be lectures, 2 will be midterm exams which are in class and 3 will be review periods during the meeting before the exams (see lecture schedule section above). The exam schedule is given on page 12. The last class of the semester will be rescheduled to be a review period that is conveniently before the final. We will discuss later in the semester.

Reschedule of Two Class Meetings

Due to the Jewish holidays, I cannot be here to teach Tuesday, September xx and Tuesday, September yy. Because my policy is to give an entire lecture period as review before exams, we will reschedule those periods to be night periods before the exams.

Homework

There will be 6–9 theory homework assignments. Homeworks will be assigned and placed on the course homepage and will usually be due a week later in class. Homework will be **graded** out of 100 with extra credit getting scores possibly > 100. I will be doing the grading and will grade an *arbitrary subset of the assignment* which is determined after the homework is handed in.

Homework must be handed in by emailing it to me as a PDF. You must do one of two things:

- **Print out the homework and handwrite your answers in the alotted space for each question. Then scan your homework as a PDF. There are a ton of good photo-PDF apps for both iPhone and droid.**

- **Open the PDF on your device and use a PDF-editing program to electronically handwrite your answers and save the PDF.**

I will NOT accept homework that is not atop the original rendered homework PDF file. Remember to write your name. There are no regrades during this pandemic semester. Homework must be at maximum 5MB.

You are encouraged to seek help from me if you have questions. After class and office hours are good times. You are highly recommended to work with each other and help each other. You must, however, submit your own solutions, *with your own write-up* and in *your own words*. There can be no collaboration on the actual *writing*. Failure to comply will result in severe penalties. The university honor code is something I take seriously and I send people to the Dean every semester for violations.

Philosophy of Homework

Homework is the *most* important part of this course.[1] Success in Statistics and Mathematics courses comes from experience in working with and thinking about the concepts. It's kind of like weightlifting; you have to lift weights to build muscles. My job as an instructor is to provide assistance through your zone of proximal development. With me, you can grow more than you can alone. To this effect, homework problems are color coded green for easy, yellow for harder, red for challenging and purple for extra credit. You need to know how to do all the greens by yourself. If you've been to class and took notes, they are a joke. Yellows and reds: feel free to work with others. Only do extra credits if you have already finished the assignment. The "[Optional]" problems are for extra practice — highly recommended for exam study.

Time Spent on Homework

This is a three credit course. Thus, the amount of work outside of the 4hr in-class time per week is 6-9 hours. I will aim for 7.5hr of homework per week on average. However, doing the homework well is your sole responsibility since I will make sure that by doing the homework you will study and understand the concepts in the lectures and you won't have all that much to do when the exams roll around.

Late Assignment Policy

Late homework will be penalized 10 points per business day (Monday–Friday save holidays) for a maximum of five days. *Do not ask for extensions*; just hand in the homework late. After five days, **you can hand it in whenever you want** until *the last scheduled class meeting according to the official academic calendar*. As far as I know, this is one of the most lenient and flexible homework policies in college. I realize things come up. Do not abuse this policy; you will fall far, far behind.

[1] In one student's observation, I give a "mind-blowing homework" every week.

LaTeX Homework Bonus Points

Part of good mathematics is its beautiful presentation. Thus, there will be a 1–5 point bonus added to your theory homework grade for typing up your homework using the LaTeX typesetting system based on the elegance of your presentation. The bonus is arbitrarily determined by me.

I recommend using overleaf to write up your homeworks (make sure you upload both the hw#.tex and the preamble.tex file). This has the advantage of (a) not having to install anything on your computer and thus not having to maintain your LaTeX installation (b) allowing easy collaboration with others (c) alway having a backup of your work since it's always on the cloud. If you insist to have LaTeX running on your computer, you can download it for Windows here and for MAC here. For editing and producing PDF's, I recommend TeXworks which can be downloaded here. Please use the LaTeX code provided on the course homepage for each homework assignment.

If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

Since this is extra credit, do not ask me for help in setting up your computer with LaTeX in class or in office hours. Also, **never share your LaTeX code with other students** — it is cheating and if you are found I will take it seriously.

Homework Extra Credit

There will be many extra credit questions sprinkled throughout the homeworks. They will be worth a variable number of points arbitrarily assigned based on my perceived difficulty of the exercise. Homework scores in the 140's are not unheard of. They are a good boost to your grade; I once had a student go from a B to and A- based on these bonuses.

Examinations

Examinations are solely based on homeworks (which are rooted in the lectures)! If you can do all the green and yellow problems on the homeworks, the exams should not present any challenge. I will *never* give you exam problems on concepts which you have not seen at home on one of the weekly homework assignments. There will be three exams and the schedule is below.

Exam Schedule

- Midterm examination I will be on [see course homepage] with the first review session on the class meeting prior
- Midterm examination II will be [see course homepage] with a review on the class meeting prior.
- The final examination will be [see course homepage] with a review TBA.

Exam Policies and Materials

I allow you to bring any calculator you wish but it cannot be your phone. The only other items allowed are pencil and eraser. I do not recommend using pen but it is allowed. Food is not allowed during exams but beverages are allowed.

I also allow "cheat sheets" on examinations. For midterm I, you are allowed to bring *one* 8.5" $\times$ 11" sheet of paper (front and back). Two sheets single sided are not allowed. Midterm II, you are allowed to bring *two* cheat sheets. On these sheets of paper you can write anything you would like which you believe will help you on the exam. For the final, you are allowed to bring *three* 8.5" $\times$ 11" sheet of paper (front and back). Six sheets single sided are not allowed. I will be handing back the cheat sheets so you can reuse your midterm cheat sheets for the final if you wish.

Cheating on Exams

If I catch you cheating, you can either take a zero on the exam, or you can roll the dice before the University Honor Council who may choose to suspend you.

Missing Exams

There are no make-up exams. If you miss the exam, you get a zero. If you are sick, I need documentation of your visit to a hospital or doctor. Expect me to call the doctor or hospital to verify the legitimacy of your note.

Accommodations for Students with Disabilities

Candidates with disabilities needing academic accommodation should: 1) register with and provide documentation to the Special Services Office, Frese Hall, Room 111; 2) bring a letter indicating the need for accommodation and what type. This should be done during the first week of class. For more information about services available to Queens College candidates, contact: Special Service Office; Director, Miriam Detres-Hickey, Frese Hall, Room 111; 718-997-5870 (Monday – Thursday 8:00 a.m. to 5:00 p.m. & Friday 8:00 a.m. to 4 p.m.).

Class Participation

This portion of your grade is assessed based on your level of interaction during the course lectures e.g. asking / answering questions. Participation on slack counts towards this total.

Grading and Grading Policy

Your course grade will be calculated based on the percentages as follows:

Homework	10%
Midterm Examination I	23%
Midterm Examination II	23%
Final Examination	39%
Class participation	5%

The semester is split into three periods :

(a) From the beginning until midterm I. Midterm I covers material during this time.

(b) From midterm I to midterm II. Midterm II covers material in this period only.

(c) From midterm II until the final. The final is cumulative and covers all course material.

Each of the periods is assessed evenly. Thus, each period must count the same towards your grade. Since there is 75% of the grade allotted to exams, there is 25% allotted to each period. Thus, the final is upweighted towards the material covered in the third period. In summary, the final will have $5/35$ points $\approx 14\%$ for the first period's material, $5/35$ points $\approx 14\%$ for the second period's material and $25/35$ points $\approx 71\%$ for the last period's material. A good strategy for the final is to just study the material after Midterm II and minimal studying for the previous material.

The Grade Distribution

As this is a small and advanced class, the class curve will be quite generous. Last year, it was approximately 40% A's and 40% B's. If you do your homework and demonstrate understanding on the exams, you should expect to be rewarded with an A or a B. C's are for those who "dropped out" somewhere mid-semester or who cannot demonstrate basic understanding. I have never given an F in this class. Don't give me a reason to change this tradition!

Checking your grade and class standing

You can always check your grades in real-time using `http://qc.gradesly.com`. You will enter in your QC ID number (or CUNYfirst email address). I will provide you with your password by email after the first assignment is completed.

Auditing

Auditors are welcome. They are encouraged to do all homework assignments. I will even grade them. Note that the university does not allow auditors to take examinations.