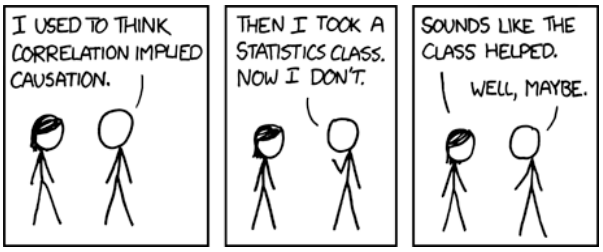
# Course Description

## Ve401 Probabilistic Methods in Engineering

## Spring 2022

### Hello and welcome to ECE4010J!

This is a first course on probability and statistics; no previous knowledge about these subjects is assumed. But by the end of this term, you should have learned quite a few things; enough to read and understand basic statistical methods used in research articles.




R. Munroe, *Correlation*, https://xkcd.com/552/

On a heuristic level, you will also have gained a clearer understanding of many popular misconceptions regarding probability and statistics as well as a more critical mindset regarding data and claims people make about data. On a more concrete level, the course introduces the basic concepts necessary for further studies in data science, data analytics, machine learning, optimization, etc.

Formally, this is a required course for students whose major is ECE, while it may be used as a flexible technical elective or to fill the "Advanced Mathematics" requirement for students in ME and MSE.

### Teaching team

Horst Hohberger has designed and given this course more than 25 times since 2008; it was never taught in exactly the same way twice! This is due to continuous student feedback, which is always very much welcomed. You can reach him at **horst@sjtu.edu.cn** or find him in his office, Room 441c of the Longbin Building.

Gao Jingnan is a junior student majoring in ECE. He is interested in computer vision and he is currently working on adversarial attack now. He can be reached at **gjn0310@sjtu.edu.cn**.

Lin Siyuan is a junior student majoring in ECE. He is interested in edge computing and machine learning. He can be reached at **Jeremy.l@sjtu.edu.cn**.

Wang Yangyang is a junior student majoring in ECE. He is interested in data science, especially in relevant computational methods. He can be reached at **wangyangyang@sjtu.edu.cn**.

Zhang Fan is a junior student majoring in ECE. This is the sixth time for him to work as a TA. He is interested in machine learning and finance. You can contact him at **zf0903@sjtu.edu.cn** to seek for help.

## What you need (prerequisites)

The most essential prerequisite to understanding is to be able to admit when you don't understand something.

Richard Saul Wurman

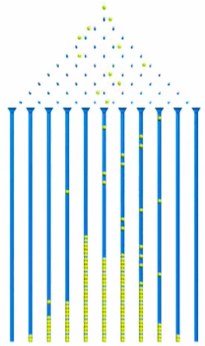
This presentation of the material assumes a solid grasp of multivariable calculus as well as some linear algebra (in particular: matrices, determinants, eigenvalue problems and diagonalization). So you should have taken and passed one of the following sequences: Vv115/116/215/216 or Vv156/255/256 or Vv186/285/286. If necessary, please review some core concepts, especially those involving integration in $\mathbb{R}^n$, which plays a very important role in the course.

If you have not passed one of the above courses or if you feel uncertain about any of the key concepts mentioned here, please do not hesitate to let me and/or the Teaching Assistants know! We will try to help as much as possible.

## What you will gain from this course

The course content is divided into approximately 1/3 probability theory and 2/3 statistics.

We start of a with a discussion of probability and its historical roots. From the principle of counting outcomes to derive probabilities of events (Cardano) we quickly progress to a more formal notions of probability functions on sigma-fields (Kolmogorov) leading to independence and conditional probabilities. This formal notion allows us to discuss various "strange" effects, such as the *Monty Hall Problem*, the *Two Children Paradox*, the *Two Envelope Paradox*, etc.



Random Influences on Measurements

We then bring the power of calculus to bear by introducing discrete and continuous random variables and random vectors. These supply the basis for our subsequent discussion of probability within statistics.

Probability theory is basically the discussion of randomness assuming that all information (in the form of probabilities) is known. For example, you may ask, "how likely is it that a fair coin comes up 'heads' at least 60 times in 100 tosses?" and probability theory will provide an answer.

On the other hand, statistics is more like physics: given an experiment (for example, a coin is tossed 100 times) and an outcome (for example, "heads" is observed 61 times) one asks about the underlying probabilities that lead to that outcome ("How likely is it that the coin is fair?"). Clearly, probability theory is necessary to provide answers to such a question. At the same time, a little thought shows that a general question such as the one about whether a coin is fair, cannot be answered without a lot more information. Statistics turns out to be more challenging than probability theory in this regard, since it involves actual, real-life experiments with often incomplete data and/or unknown parameters.

Furthermore, a broader reading of statistics is to provide a real-life answer to such questions. Can a statement such as "There is a 84% probability that the probability of obtaining heads for this coin differs from 50% by more than 5%" be translated to "experiments suggest that the coin is most likely not fair" or even, the simple statement of fact "the coin is not fair"?

It may seem that this would be an extreme simplification and possibly a non-justified statement (also known as "a lie") but consider this: in testing vaccines for the COVID-19 virus, large studies are conducted, not unlike coin tosses, and the final conclusion has to be distilled to "this vaccine is safe" or "this vaccine is unsafe."

While the interpretation of a statistical result is always subjective, the examples and discussions in our course will help you interpret statistical statements and apply them to real-life decisions.

On a more technical level, our course will give a modern and up-to-date perspective on topics including data exploration, parameter estimation, confidence intervals, hypothesis testing in many different contexts, linear regression and (if time permits) analysis of variance.

Along the way, you will learn to use *Mathematica*® software for statistical analysis and probability theory, you will read technical articles on statistics and you will work in teams to complete your own projects and reports.

## Syllabus

We will meet for 30 lectures of 90 minutes each over the 12 weeks of summer term.

| Lecture | Lecture Subject |
|---|---|
| 1 | Elementary Probability |
| 2 | Conditional Probability |
| 3 | Discrete Random Variables, Binomial and Geometric Distributions |
| 4 | Expectation, Variance and Moments of Random Variables |
| 5 | The Pascal, Negative Binomial and Poisson Distributions |
| 6 | Continuous Random Variables, Exponential and Gamma distributions |
| 7 | The Normal Distribution |
| 8 | Multivariate Random Variables |
| 9 | The Hypergeometric Distribution / Transformation of Random Variables |
| 10 | Reliability |
| 11 | Samples and Data |
| 12 | Parameter Estimation |
| 13 | Interval Estimation I |
| 14 | Interval Estimation II |
| 15 | **Midterm Exam*** |
| 16 | Fisher's Significance Test |
| 17 | Neyman-Pearson Decision Theory |
| 18 | Tests for the Mean and Variance |
| 19 | Non-parametric Tests |
| 20 | Inferences on Proportions and Variances |
| 21 | Comparison of Two Means |
| 22 | Paired vs. Pooled Tests; Non-Parametric Methods |
| 23 | Categorical Data |
| 24 | Simple Linear Regression |
| 25 | Simple Linear Regression |
| 26 | Multiple Linear Regression |
| 27 | Multiple Linear Regression |
| 28 | Analysis of Variance |
| 29 | Analysis of Variance |
| 30 | **Final Exam** |

* The schedule of the Midterm Exam is approximate and subject to change; please see the syllabus tab on Canvas for the actual date.

## Course outcomes

The *course outcomes* define a set of minimal skills or items of knowledge that you should feel confident in applying after completing Ve401. Naturally, the course will cover much more than just the outcomes listed below. However, they represent a cross-section of basic skills and are a useful guide to how well you have mastered the contents of the course in general.

A quantitative measurement of these outcomes (except the last one) will be provided by *Course Outcome Quizzes*, which will also contribute to the course grade. You will be also asked to report your subjective impression on attaining the outcomes in the Course Evaluation survey at the end of the term.

After completing Ve401, students should be able to:

i) Apply and interpret Bayes's theorem correctly in a given context.

ii) Calculate the Pearson correlation coefficient of two random variables.

iii) Transform and perform basic calculations on random variables.

iv) Create visualizations of data and make basic inferences on the nature of the data.

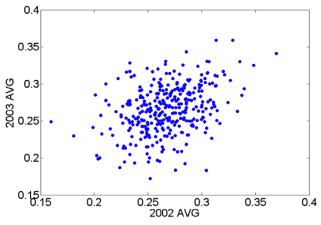
v) Estimate parameters of distributions and find confidence intervals.

vi) Perform Fisher and Neyman-Pearson tests.

vii) Perform and analyze a linear regression in an elementary context.

viii) Working as a team, read and understand a publication on a topic in probability or statistics and create a technical report based on a specified topic or task.

## Why there is no textbook




Luck vs. Skill in Baseball

The field of probability theory has existed for several hundred years and most elementary notions and paradoxes have by now been explored exhaustively. We will only discuss some selected foundational topics, which usually can be found in the beginning chapters of a wide variety of textbooks.

In contrast, the field of statistics is only slightly more than 100 years old and continues to develop. Many conventions and procedures are based on research and rules set down by the pioneers of the field (e.g., Fisher, Pearson, etc.) in the first half or middle of the 20$^{\text{th}}$ century. Certain methods (such as those due to Tukey) even stem from the 1970s or later.

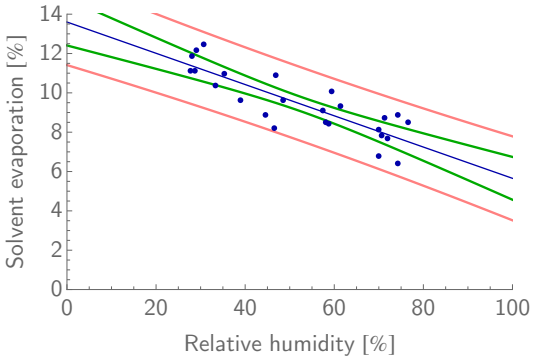
It is natural that these approaches are still evolving while being constantly re-examined by the scientific community. A typical example is the controversy surrounding hypothesis testing and P-values in scientific research that has been playing out over the past 20 years. Therefore, many existing textbooks in applied statistics (or statistics for engineers, statistics for the sciences etc.) contain material that is not up to date with current practice. Therefore, we will cite a variety of current sources and not rely on any single textbook.

You are encouraged to explore the range of textbooks available online and in the university library and compare them critically with the contents of our lecture slides. Discussion about similarities and differences is very much desired!

## Mathematica

In modern statistics, the use of computer software to perform complex and/or tedious calculations is indispensable. Our course will continually refer to the Mathematica software package, published by Wolfram Research.

This program has been chosen because (i) it is capable of symbolic manipulation of mathematical expressions and thereby complements other software such as MatLab; (ii) it is easy to learn; (iii) it has very good graphics capabilities; (iv) it is easy to mix programming features (loops, conditionals, etc.) with mathematical evaluations; (v) it is the "swiss army knife" of mathematical software and you should at some point in your studies become familiar with it.




Confidence and prediction bands for a simple linear regression.

In further courses in statistics and data science, you are likely to use other software, such as R or Python.

Mathematical software, ideally Mathematica, *will be required for some of the assignments and for the final exam.* If you wish, you may of course use a different software package, but it is up to you to verify that the necessary statistical functions can be performed. Hence, while you are strongly encouraged to use Mathematica (and the lecture slides provide support), it is not compulsory.

### Obtaining Mathematica

i) Visit

https://user.wolfram.com/portal/registration.html

and create a Wolfram ID. You must use an @sjtu.edu.cn email address and give your first and last names in pinyin (example: Xu Baishen enters last name: Xu and first name: Baishen).

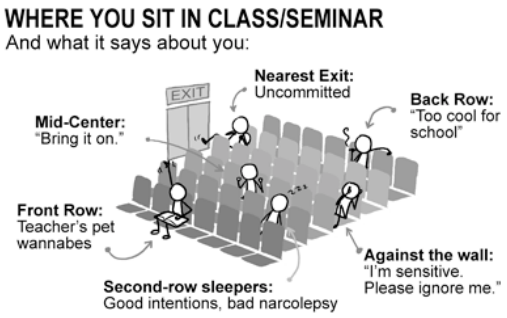
ii) Next, visit

https://user.wolfram.com/portal/requestAK/c51e79e5334a3600a4f740a2b3720961216dbc17

and request an Activation Number. Make a note of the activation number. You will be directed to a page where you can download the installation binaries for the most current version of Mathematica. The software binaries are available for Windows, Linux or macOS.

iii) After downloading, you can install the software. You will be asked to enter the Activation Number you noted above and you will need internet access. Mathematica will then run on a temporary two-week license. Your name will be checked against a list, and if successful, the license will automatically be extended for one year. *Therefore, it is very important that you enter your name properly when you request the Wolfram ID.*

## Lecture attendance and participation




J. Cham, *Where do you sit?*, phdcomics.com/comics/archive.php?comicid=1017

Our course will attempt to blend online and in-class teaching. The classroom is D415 in Dongshangyuan 东上院)) and online access will be implemented via Zhumu (see Canvas for details). You may participate either by coming to the classroom or by logging into the live video link.

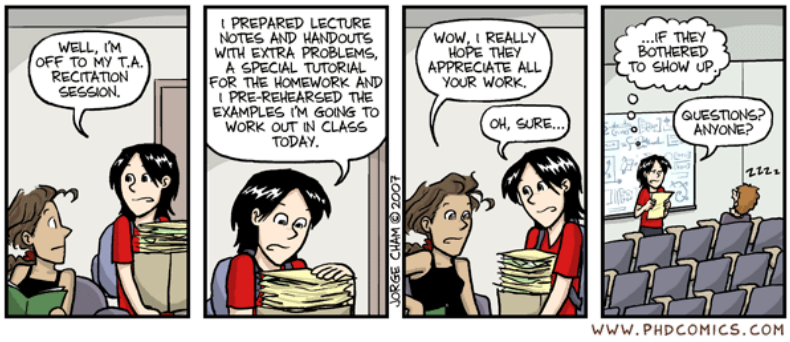
If you participate by joining online, please observe the rules of online etiquette: you should log in using your real name (in pinyin transliteration as appropriate), you should have your camera turned on and your microphone turned off, but be ready to answer questions, etc. Details will be given in class.

An effort will be made to provide recordings of the classes, but there is no guarantee regarding the sound and video quality or even that the recordings will be available for every session.

While attendance is *not mandatory in principle*, there will be a small percentage of the course grade reserved for class participation, which may be evaluated by whether you have participated in online class polls, whether you have followed online etiquette, etc.

## Recitation classes, office hours and Piazza

There will be recitation classes and office hours offered by the teaching assistants, as well as office hours offered by the instructor. None of these are mandatory but you may find them very helpful. The detailed times will be published on Canvas.




J. Cham, *Appreciation*, http://phdcomics.com/comics/archive.php?comicid=922

In addition, our class will use *Piazza* for you to pose questions online. Piazza provides a much better venue than traditional office hours, since questions and answers are visible to everyone and any student can participate in the discussion on a given topic. You can sign up for Piazza here:

https://piazza.com/sjtu.org/summer2021/ve401

## Course evaluation

Furthermore, at the end of the term there will be a course evaluation survey administered by the Undergraduate Education Office (UEO). The survey is completely confidential and only the collated comments and score summaries will be communicated to the instructor. Filling out the survey makes a small contribution to the course grade, so the whether or not you filled out the survey will also be communicated. Your honest and thoughtful opinions are very much appreciated!

## Coursework

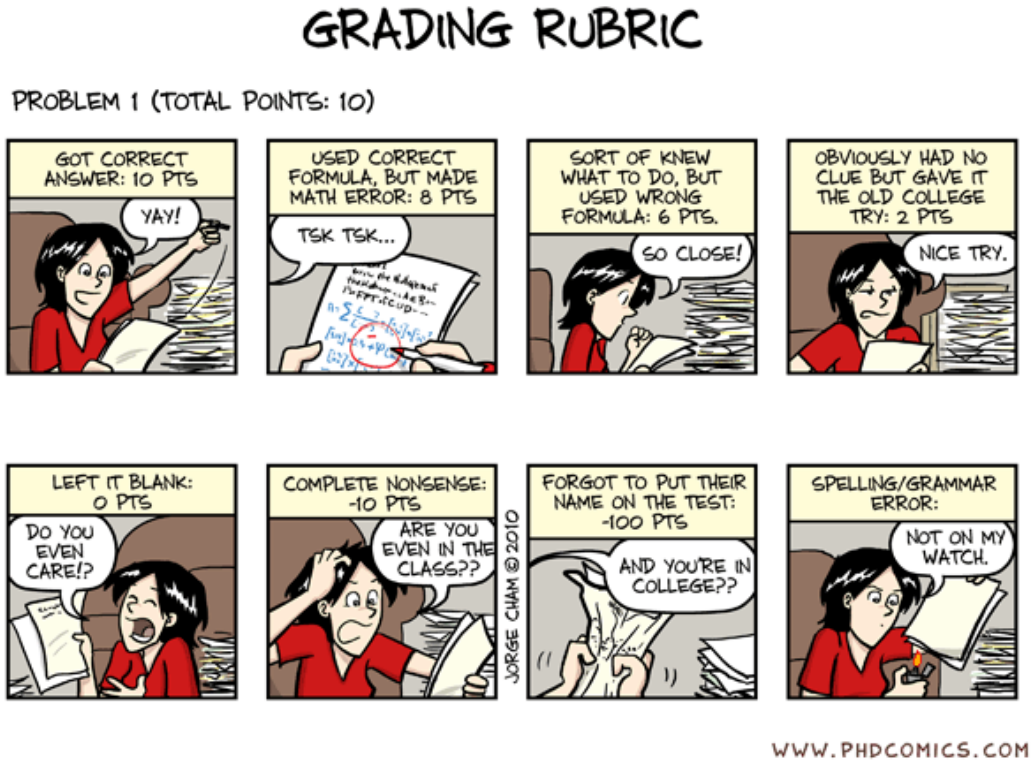
There will be weekly assignments. The goal of these assignments is for you to practice the lecture material and investigate some concepts by yourselves. the goal is *not* to evaluate you or to put undue pressure on you. Therefore, the coursework will not quantitatively contribute to the course grade.

**However,** you do need to attain at least 60% of the points of the coursework over the whole term. If you do not, you will **automatically fail** the course. On the other hand, it does not matter at all what percentage above 60 you attain. As mentioned, the course grade is not directly affected.

You are asked to hand in coursework on time. The due time is usually tied to a lecture period. If you hand it in later the same day, you may do so at the convenience of the teaching assistants (who may well refer you to the instructor if they are not able to accommodate you). After that, you must hand it in to the instructor personally and be prepared to explain the reasons for the delay.

You will be randomly assigned into groups of three students each; each group will jointly work on each assignment and hand in a single submission. To ensure that all members of each group contribute to the assignments, there will be a peer evaluation process. Details will be announced on Canvas. As for the course evaluation, participating in the peer evaluation process will make a small contribution to the total course grade.

In general, every member of an assignment group will receive credit for the assignments; however, if the peer evaluation process indicates that any members are not contributing to the assignments, the corresponding points for the affected assignments may be voided for them.




J. Cham, *Grading Rubric*, http://phdcomics.com/comics/archive.php?comicid=1319

## Term project

The term project forms important part of the course. You will be randomly assigned into groups of 4-5 members each to work on these over the course of 3-5 weeks. Details will be provided in the project description.

Just as for the coursework groups, there will be a peer evaluation process. The results will have the potential to positively or negatively affect the individual grades for the term project. Participation in the peer evaluation process will also affect the term project grade.

## Online modules

This course contains a certain number of online modules which need to be completed during the term. Typically, these are small case studies of specific problems in probability and statistics and involve some sort of activity such as watching a video or performing simulations with Mathematica. Collectively, these modules should make up about 20 credit hours (15 actual hours) of work, roughly equivalent to the amount of work required for a term project. (Earlier iterations of this course had two term projects.) The online modules will be completed collectively by the same 3-person groups assigned for the coursework.

## Examinations

There will be two exams, a midterm and a final. These will be given in-class and **you must personally attend on-campus** if it is possible for you to do so. If it is not, perhaps because you are situated outside of Shanghai or there is a medical reason preventing you from attending, please contact the instructor beforehand. The arrangements described in the framed box below will then need to be made.

For all exams you may use an english, monolingual dictionary in book form. Cell phones and electronic devices other than those specifically mentioned below are not permitted.

Paper for your exam answers and for notes will be provided for on-campus exams. You may not bring your own paper. You must, however, bring your own pens, rulers and other necessary writing materials. A pen and a ruler will be sufficient, there is no need for other stationery, though you can bring more if you like.

It is allowed to bring drinks and snacks to the exams as long as no disturbance of others is created.

For both exams, you may use a calculator of type "Casio fx-991CN X" or "Casio fx-82". If you are unable to obtain one of these two models, please contact the TAs or the instructor for approval of a different model.

For both exams you may bring one (and only one) laptop. The exams are **open-book, no-internet** exams. The laptop should have Mathematica or a similar statistical/mathematical software installed (e.g., MatLab, R, SPSS, SAS, etc.).

It is expected that you will have a calculator for the Midterm Exam and a Laptop for the Final Exam, although either can be brought for either exam.

You may use such software to perform calculations or obtain graphical representations. In the case of calculation results, please put **(Mathematica)** in brackets behind the result, e.g., you can state "$R^2 = 0.967$ (Mathematica)." Boxplots, histograms or similar graphics must be copied out by pencil into this exam booklet and labelled appropriately.

Please **retain your Mathematica .nb notebook file.** You will be asked to submit it to Canvas after the exam.

You may refer to the lecture slides or to any other course-related documents (including homework, recitation class notes, sample exams and their solutions, etc.) during this exam. You may also bring hand-written or printed-out notes to the exam.

You must **turn off your wifi, bluetooth and any other networking functionality**. You may not use a web browser, chat client (WeChat, QQ, etc.) or email client of any kind. **Please ensure that you have a dedicated PDF reader software.** Using a web browser to view PDF files is not permitted.

It is recommended that you restart your laptop and ensure that any any such programs are not running before the exam. If you are seen to be running a web browser, chat client, email client or similar software, or if your wifi/bluetooth/etc. is connected, you will be considered to be violating the Honor Code.

If you think it necessary, you may bring a power adapter and multi-socket extension cord to connect you laptop to a power outlet during the exam.

No item not mentioned above (esp. cell phones, tablet computers, backpacks) may be at your desk unless you are explicitly granted permission by the supervisors. In particular, **iPads and other tablets are not permitted**.

### Taking Examinations Remotely

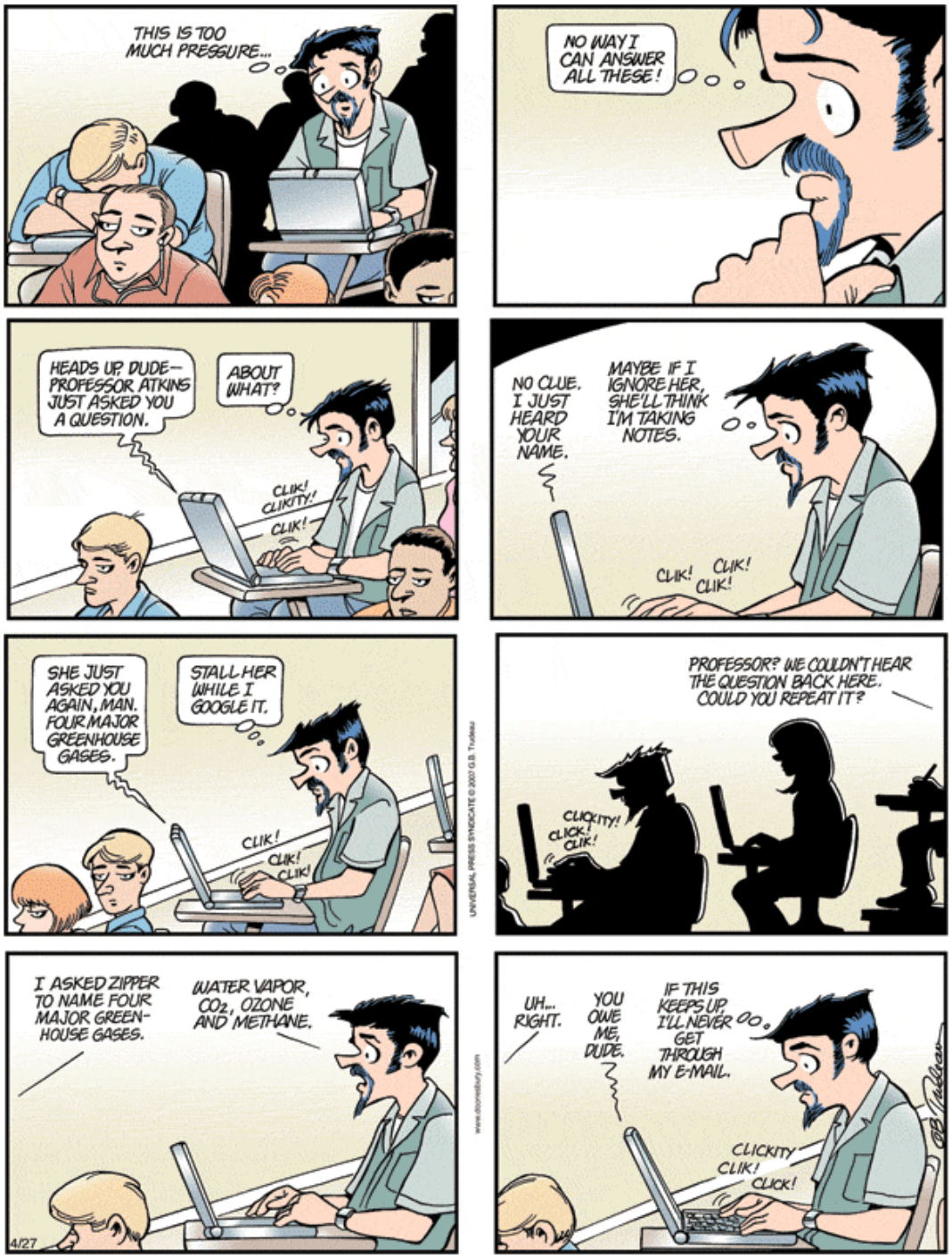
If you are unable to attend an exam in-person, but able to take the exam remotely via video link, you must ensure the following:

- Your exam location must be an enclosed space that is not entered by anyone else during the exam. The space must be well-lit and quiet.
- You must be seated at a desk that is empty apart from a computer and any permitted material for the exam. A cell phone may be at hand for emergency communication and picture taking (see below), but may not normally be used during the exam.
- You must have set up a camera that clearly shows your desk, your computer screen, and yourself. The video feed of this camera needs to be visible throughout the exam and may not be turned off during the exam time. The camera may be either a second camera or the only camera used in the video feed. A camera embedded in a laptop screen (which does not show the screen, desk or your hands) is not sufficient by itself.
- You may mute your microphone by default, but be ready to unmute it when asked.
- Your computer sound should be turned on, in case there is an urgent message from the instructor.
- You will receive the exam paper electronically, e.g., via the chat function of Zoom. You may view the paper on your computer and write solutions on your own paper. These solution should then be photographed and the image files emailed to the instructor, horst@sjtu.edu.cn.
- The photographs of the answers must be taken before the end of the exam and should be emailed as soon as feasible. Allowances for network connections will be made as necessary. However, no solutions may be photographed after the exam time has passed.

## Class Participation

Another part of the course grade is based on "class participation." This will encompass a range of factors, including taking part in the regular polls that will appear in the lectures, following the online video procedures (e.g., with regard to keeping the camera turned on and entering your name in western transliteration) and other issues that may arise during the term.

In general, it will be easy to get full marks for this component, assuming a reasonable level of cooperation and engagement in the class.




Garry Trudeau, *Doonesbury*, April 27, 2008, https://www.gocomics.com/doonesbury/2008/04/27

## Honor Code policy

Please familiarize yourself with JI's Honor Code, found at

https://www.ji.sjtu.edu.cn/academics/academic-integrity/honor-code/.

In addition to the general rules for examinations etc., the following specific policy applies to the coursework:

**Coursework and Term Projects: External Sources, Collaboration and Piazza**

Students within each assignment group, are allowed and supposed to collaborate in a completely unrestricted manner. However, it is prohibited to discuss assignment problems or to share solutions with members of other groups in any way.

Instead, if you encounter problems in solving an exercise, you are explicitly encouraged to post a question on Piazza. **Posting a reasonable question on Piazza is never a violation of the Honor Code.** However, do not share full and complete solutions to current assignments on Piazza (or on any other public medium). Such sharing will very likely constitute a violation of the Honor Code. Please use reasonable judgment!

The above comments regarding collaboration and the use of Piazza in coursework also apply, *mutatis mutandis*, to term projects.

It may happen that you find the solution of a homework problem in some outside source (book, internet site, etc.). In that case you are not allowed to just copy the solution; this is considered a violation of the Honor Code.

The correct way of using external sources for homework problems is to understand the contents of your source and then to write down the solution in your own words and without referring back to the source the solution of the problem. Your solution should differ in style significantly from the published solution. If in doubt, cite the source that you used.

On the other hand, for term projects, you should **always** cite all sources that you use, both in-text and in the bibliography.

Lastly, Section 5 of the Honor Code is fully enforced: any violation of the Honor Code by a term project or an assignment group will cause all group members to be sanctioned equally.

All members of the Teaching Team will be happy to answer any any questions regarding the application of the Honor Code.

## Course grade components

Below is a summary of the components of the course grade:

| Grade Component | Points |
|---|---|
| Course Outcome Quizzes | 7 |
| Completing the Peer Evaluation for the Term Project | 1 |
| Completing the Peer Evaluation for the Coursework | 1 |
| Completing the Course Evaluation | 1 |
| Class Participation | 5 |
| Midterm Exam | 25 |
| Final Exam | 30 |
| Online Modules | 15 |
| Term Project | 15 |
| Total | 100 |

The course will be graded on a letter scale (A-F), with a certain number of points corresponding to each letter grade.

The grading scale will *usually* be based on the top approximately 6-12% of students receiving a grade of A+, with the following grades determined by (mostly) fixed point increments.

Apart from this normalization, the grade distribution is up to you! If (for example) all students obtain many points in the exams, I am happy to see everyone receive a grade of A. Students are primarily evaluated with respect to a fixed point scale, not with respect to each other.

## And now, enjoy the course...

Please don't hesitate to let any member of the teaching team know if you have any questions about course policies or content. Also, don't hesitate to contact us if you need help or support, either academically or otherwise. We are there to help!