Sociology 500: Applied Social Statistics
(Fall 2023)

Matthew J. Salganik Christina Pao Varun Satish

Last Edited: September 7, 2023

Matthew J. Salganik, Instructor

Christina Pao, Preceptor

Varun Satish, Preceptor

Class: 10:00am-11:30am Mondays and Wednesdays
Precept: 10:00am–12:00pm Thursdays

This is the first class in Sociology's two-course graduate statistics sequence. Starting from only basic math, we build up a foundation for linear regression and its application to causal inference. The course is focused on the tools needed to read and do research in the 2nd year empirical seminar, as well as the tools you will need to continue to learn throughout your career. The course draws examples from across the social sciences. Students can take it both as a first course in linear regression or as a deeper dive into regression than a typical undergraduate sequence.

1 The Basics

The first stop for anything you need as a member of the class should be the Canvas site.

1.1 Course Goals

This is the first course in the Department of Sociology's two-course graduate statistics sequence. Starting from only basic math, we build up a foundation for linear regression and its application to causal inference. Students will have the opportunity to learn the statistical and computational principles necessary to perform modern, flexible, and creative analysis of quantitative social data. By the end of this semester, you will be able on your way toward (1) engaging with modern, quantitative sociological research and (2) having a foundation for lifelong learning about quantitative methods.

Students enter this course with a huge range of goals related to modern, quantitative sociological research: some students seek to read and understand it, some seek to do it using existing methods, and some seek to develop new methods. Rather than ignoring this range of goals, we choose to embrace it. The course offers three "specializations", and your choice of specialization will impact your options for your SOC 504 replication project, your 2nd year empirical paper, and the kinds of research collaborations you can begin with faculty. The choice of specialization is completely up to you, and you can also switch as often as you like. To enable switching and create a cohesive experience, the lecture and precept will help prepare you for all three specializations. Each assignment will include some core activities and some activities that vary depending on your specialization. The specializations are:

- Sophisticated reader: You should choose this specialization if you want to do a basic replication project in Soc 504 and you do not want the option to use quantitative methods in your 2nd year empirical paper. In addition to core parts of the course, this specialization will focus on: 1) reading and writing about quantitative methods and 2) comparing quantitative methods to other methods. You will be responsible for acquiring the training you need to use other methods outside of this course. Choosing this specialization may make it difficult for your to collaborate with faculty doing research using quantitative methods.
- Sophisticated user: You should choose this specialization if you want to do a 2nd year empirical paper use existing quantitative methods in ways that they are commonly used by sociologists. In addition to core parts of the course, this specialization will focus on: 1) software engineering and 2) some mathematical and analytical reasoning. Choosing this specialization should be helpful if you want to collaborate with faculty doing research using quantitative methods.
- Sophisticated creator: You should choose this specialization if you want to modify existing qualitative methods to new contexts or develop entirely new quantitative methods. If this is your goal, you should expect to take additional courses beyond the required sequence in our department, such as the courses required for the Statistics and Machine Learning Graduate Certificate. If this is your goal you should also plan to do additional learning beyond your courses by participating in co-curricular activities such as reading groups, methodology seminars, and short courses. We would be happy to point you to these other opportunities, just tell us more about your interests. In addition to core parts of the course, this specialization will focus on: 1) additional foundational mathematical concepts 2) additional foundational computational concepts and 3) analytic reasoning. Choosing this specialization should be helpful if you want to collaborate with faculty doing research using quantitative methods.

Each of these goals is difficult to achieve at a high-level, and we expect you to strive for excellence no matter your goal. That said, we do expect that different specializations will require different amounts of time from you.

This course will require a lot of hard work from all of us; however, we have structured the class to provide you the maximal return on every hour of work you put in. As you read through this syllabus you will find numerous avenues for seeking help. If you are willing to put in the time, we are always happy to help. Please don't be shy about telling us where you need support.

1.2 Prerequisites

The most important prerequisite is a willingness to work hard on unfamiliar material. Learning statistical methods is like learning a new language, and it will take time and dedication to master its vocabulary, its grammar, and its idioms. However like studying languages, statistics and programming yield to daily practice and consistent effort.

We intentionally have no formal pre-requisites. Beyond high-school level algebra, it is helpful to have some familiarity with calculus (essentially knowing what derivatives and integrals are in principle even if you forget how to do the mechanics) and basic matrix operations (matrix multiplications and inverses). It will also be helpful if you have some experience with computer programming, specifically the R language. If these concepts are unfamiliar, you should review the materials from our Department of Sociology summer methods camp (`http://pusocmethodscamp.org/`).

Even if you have seen some of the materials in class before (e.g., you had an undergraduate class on linear regression), you will likely find a lot to learn here. If you are concerned that you may have already covered the material before, come talk to the instructor.

1.3 Preview of Course Structure

- Lectures (Monday/Wednesday 10:00AM-11:20AM):
  Lectures cover the core material and are organized into weeks. In each lecture you will be exposed to the main conceptual ideas of the week and have the opportunity to ask questions. Detailed slides are made available afterwards.

- Lab/Precepts (Thursday 10:00AM-12:00PM):
  We will often refer to this preceptor-led period as 'precept' but administratively it is a 'lab.' Regardless, the goal of this time is to learn the coding skills necessary to complete the problem sets. It also provides an opportunity to ask any followup questions about the material.

- Problem Sets (Due Wednesday 8PM):
  Learning comes through doing and the problem set is an opportunity to practice the material learned in the prior week. These problem sets will likely be the most time-consuming part of the class (self-reports in prior years suggest they take most people 5-10 hours per week) but they are also where the majority of the learning happens.

- Office Hours (5 hours per week, posted on Canvas)
  Across the three members of the instructional staff we will offer 5 office hours per week. Office hours with the instructor are primarily for discussing conceptual material while office hours with the preceptors should be used for assistance with the programming and/or problem sets. You can come to preceptor office hours without any specific questions and just sit and work on the problem sets if you like!

2 Materials

2.1 Computational Tools

The best way, and often the only way, to learn about data analysis and new statistical procedures is by doing. We will therefore make extensive use of a flexible (open-source and free) statistical software program called http://www.r-project.org, http://www.rstudio.com/, and a number of companion packages in the tidyverse style. Problem sets and the final exam will be completed in Quarto. You will learn how to program in this class, if you do not know already.

2.2 Readings

This class uses extremely detailed lecture slides (it is not uncommon to have 100+ slides in a week). We encourage you to think of these (and the lectures) as the primary reading in the class (although we will occasionally assign additional required reading). If you are someone who benefits a lot from reading material and you find the slides aren't working for you, come talk to me and I will help find a reading that is well situated to your background and interests.

Sometimes though it is just helpful to have a reference book. We include a few by topic below.

- Programming
  - Wickham, Çetinkaya-Rundel, and Grolemund. 2023. *R for Data Science, 2nd edition* (available online for everyone)
  - Healy, Kieran. 2018. *Data Visualization: A Practical Introduction.* Princeton University Press. (available online for everyone)
- Probability and Random Variables
  - Blitzstein and Hwang. 2019. *Introduction to Probability 2nd Edition* (available online through the library)
- Regression
  - Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press. (available online through JSTOR)
  - Aronow and Miller. 2019. *Foundations of Agnostic Statistics.* Cambridge University Press (available online through the library)
  - Fox, John. 2016 *Applied Regression Analysis and Generalized Linear Models. 3rd Edition.*
  - Shalizi, Cosma. Forthcoming. *Advanced Data Analysis from an Elementary Point of View.* Cambridge University Press (Note that this book is still being written and you can find draft PDFs on the linked page above.)
- Causal Inference
  - Hernán, Miguel A. and James M. Robins. 2023. *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC. (Note that this book is in press; you can find draft PDFs on the linked page above.)
  - Morgan, Stephen L, and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Second Edition.* Cambridge University Press. (available online through the library)

3 Assignments

There are three main types of assignments (summary here, details below):

1. **Preparing for and Attending the Class:** We expect you to come to class. Keep on top of the material and if there is something you don't understand, ask about it!
2. **Weekly problem sets:** Learning data analysis takes practice. The problem sets are described below.
3. **Final exam:** A cumulative take-home final exam will conclude the semester.

3.1 Preparing for the Class and Precept

Lectures and precept are most effective when you are following along and asking questions. The class is fairly small and will be at its best when we are using the opportunity to discuss the concepts. Please engage early and often. A key part of being able to engage is reviewing the notes before class so that you know what you do and don't know. Build in the time to do this review!

Reading statistical work (or lecture slides) can be challenging the first time you do it. There will be a temptation to skip over all the math—don't! The math is often where the action is in statistical work. Read carefully and go line by line seriously engaging with material: take notes, write down your impressions or confusions, talk with your classmates, and post questions on Ed (see below). All of your classes should be pushing your research forward and you will be better prepared to integrate this material into your work if you push for a deep level of understanding.

3.2 Problem Sets

Statistical methods are tools and it isn't very instructive to read a lot about hammers or watch someone else wield a hammer. You need to get your hands on a hammer or two. Thus, in this course, you will have problem sets on a weekly basis. The assignments will be a mix of analytic problems, computer simulations, and data analysis. Each assignment will include some core components, and some components geared to each specialization.

Format Assignments should be completed in Quarto which allows you to show both your answers and the code you used to arrive at them. If you haven't seen Quarto before, have a look at the Quarto website and materials on `http://pusocmethodscamp.org/`. We will use Gradescope for submitting and returning the assignments. Your wonderful preceptors will provide you with more detailed instructions before the first assignment is due.

Assessment Problem sets will be graded from 0-50 points. We also reserve the right to add bonus points for aesthetics including presentable graphs, clear code, nice formatting and well written answers. Solutions will be available directly after the problem set deadlines. Keeping up on the problem sets includes looking at the solutions key to make sure you understand what is happening!

Extensions and Late Policy You can have *one* no questions asked extension of one week on a problem set of your choosing. If you don't take an extension, we will drop your lowest grade (of any partially completed/submitted problem set). If all your problem sets are completed and with top-level grades (such that dropping the lowest wouldn't help you), we will add a comparable grade bonus to your final exam. Because we do not want to hold up the class we will not wait for everyone to submit their problem sets in order to post the solutions key. If you are turning your problem in late (with an initial extension or not) you are on your honor to *not look at the solutions* before submitting your work and you are required to explicitly write on your assignment that you have not looked at the solutions. If you are late without an extension (or exceed the one-week extension period given that you had an extension) your grade will drop on the problem set by 10 percentage points per day (5 points as the assignments are out of 50), with a maximum of 70 percentage points (35/50 points) taken off for your raw problem set score for turning it in late. You have until the beginning of reading period to submit a late problem set for partial credit but please don't wait that long—the material is very cumulative!

Extreme Circumstances I realize that the above extension policy may not be enough to deal with all situations. For example, if a student has a major illness this might disrupt weeks of school work. If an extreme disruption occurs, you need to reach out to me as quickly as possible. We will assess your situation and then in collaboration with your Director of Graduate Studies we will come up with a plan to get you up to speed as quickly as possible. If you do not reach out to me, then we cannot work with the Director of Graduate Studies to develop a plan.

Collaboration Policy: Unless otherwise stated, we encourage students to work together on the assignments, but you should write your own solutions (this includes code). That is, *no copy-and-paste* from other people's code. You would not copy-and-paste from someone's paper, and you should treat code the same way. However, we strongly suggest that you make a solo effort at all the problems before consulting others. It is easy to trick yourself into thinking you understand how to do something when you have someone else prompting you.

3.3 Final Exam

The final assessment of the class is a take-home exam. The exam is "open-book" in the sense that you can use the slides, your notes, books, and internet resources to answer the questions. However, the final exam must be completed *by yourself.* The exam will be available during the entire period allocated for take-home exams by the university. It will be approximately the length of a long problem set (although we caution that it might take longer if you are used to collaborating on the problem sets). We encourage you to start early. Before the final exam we will distribute a practice final that will help you in your preparations.

3.4 Grading

Final grades will be a weighted average of the final exam (30%), class participation (including Ed posting or getting involved during precept or class) (10%) and the weekly problem sets (60%).

3.5 ChatGPT and other AI tools

ChatGPT and other AI tools are potentially quite helpful for your research. It is not yet clear, however, the best way to integrate them into learning, if at all. For now, you are welcome to use ChatGPT (or other automated assistance tools) in this class if (1) you do not directly put the assignment question into ChatGPT and (2) you acknowledge their contribution and add a written description of how you are using them. Over the course of the semester, we will work together to try to figure out how ChatGPT and related tools are helping or hurting learning, and we will refine this policy as we go.

A few things to keep in mind:

- This class is designed to prepare you to conduct and evaluate research. At this point, it is not yet possible to put real research problems into ChatGPT and get reliable high-quality answers. Therefore, ChatGPT will not be able to do your research for you. That said, it is possible to take a research problem and break it into smaller parts and then use ChatGPT as a tool to help with some of those parts. Therefore, this is the skill we will allow you to explore in this class. You cannot directly put your assignments into ChatGPT, but you can break it down into smaller parts and use ChatGPT as a tool to help you with some of those parts.

Naturally, you—and not ChatGPT—are responsible for anything that you submit. Also, it is not yet clear if using ChatGPT in this way will help or hinder your learning.

- The assignments you will do in this class are a means to an end, not an end in themselves. They are designed to push you in ways that enable you to build new skills. Doing the assignment without the effort is not likely to build skills and will therefore have little value to you.
- ChatGPT is not perfect; far from it. You should expect that what you get from ChatGPT will be wrong occasionally. Programming is a setting where this might be OK, as described more in this post by Arvind Narayanan and Sayash Kapoor.
- If you are going to use ChatGPT extensively, it is important to understand how it works. I'd recommend this post "What Is ChatGPT Doing ... and Why Does It Work?"by Steve Wolfram as a starting point.

4 How to Learn in this Course

If you find this course challenging, you are not alone. Statistics can be challenging and we cover a lot of ground. However, I am confident that you can handle it. In this section of the syllabus I'm going to provide details on some of the forms of support that we offer in this class and pull back the curtain a bit on the pedagogical design.

> Your primary responsibilities in this class are to *work hard* and *communicate* with us about what you need. You can't learn if you aren't putting in the time. We can't help if we don't know there is a problem.

The course is designed to provide every tool we can think of to help you learn the material. If you are willing to put in the time, we want to ensure that time is used as effectively as possible.

4.1 Resources for Getting Help

There are a few main sources of support in the class.

1. Class and Precept
   We strongly encourage you to be an active participant in class and precept. Ask questions during class if you don't understand something that is happening. You don't even have to have a specific question: just raise your hand and let me know that you aren't following what is happening. I'm always happy to stop and go back.
2. Daily Feedback
   After every class I will pass around a note card and ask you to write down something about class. You can write something you liked or didn't like. Something you want to understand better or want to hear more about it. Maybe you want to know how a piece of material connects to the broader goals of the class. You can even just draw a smiley face. I will address the questions either in class or on Ed.

3. Readings and Slides
   If you are studying alone and you hit something you don't understand, your first instinct should be to study the readings and slides. There is a lot of material in the slides and they are intended to be reviewed multiple times, not just seen once during lecture.

4. Ed
   Ed is a classroom discussion board where you can post questions about the material. You will not be required to post, but the system is designed to get you help quickly and efficiently from classmates, the preceptors, and me. **Unless the question is of a personal nature or completely specific to you, you should not e-mail teaching staff**; instead, you should post your questions on Ed. The course staff will be monitoring the page, but we encourage you to help your classmates as well. A big part of why we use Ed is because reading other people's questions can be really helpful for bolstering your understanding of the material.

5. Preceptor Office Hours
   Each preceptor will offer two hours of office hours per week. You don't need to have a specific question and can just drop in. In past years students have used this as a convenient opportunity to gather to work collectively on the problem sets with someone to support you near by.

6. Instructor Office Hours
   I will have a dedicated office hour each week when you can drop by with questions about the conceptual material in the course.

7. Problem Set Solutions
   As soon as the problem sets are due the key is posted. I know it is really tempting to just turn your focus to the next problem set, but I highly encourage you to check the solutions each week. The class is extremely cumulative and it will help you out to lock down core concepts. Even if a concept in (for example) week 3 seems unrelated to week 4, it will definitely come back in a future work.

8. Final Exam Prep
   We will host a review session for the final exam sometime before the final exam period.

9. External Consulting Services
   Princeton offers numerous statistical consulting services. These should not be used for problem set help, but can be useful if you just need help understanding a broader concept.

This is a lot of resources but if you can think of something else that would be useful to you, we encourage you to come talk to us. Again, if you are willing to put in the time, we can get you a form of support that matches your needs.

4.2 How is the Course Designed?

At a high level, this course builds up the infrastructure of linear regression and the beginnings of causal inference from the basics of probability. The first six weeks focus on foundation elements: probability, random variables and the basics of statistical inference. The second four weeks covers linear regression and its variants. The final two weeks are devoted to causal identification and estimation.

Each week covers a specific topic with a series of closely connected lectures included in a common slide deck. These lectures are designed to fill you in on the core statistical ideas animating the week's topics. The lecture won't focus on code, it will focus on the underlying logic and the applications of the ideas to social science research.

Precepts will teach you the programming tools necessary to implement the things shown in lecture. The code shown in precept is very closely tied to what you will need to complete the problem sets.

The problem sets are where I expect the majority of the learning will be solidified. These assignments can be challenging and time consuming, but it is only through carefully engaging with the material that you will cement your understanding of it. If at the end of lecture you feel like you don't have a good handle on the material—that's to be expected. If after the problem set has been submitted you still feel uneasy with the material, you should come to office hours and talk to one of us about it.

Finally, there is a strong focus on the class on understanding why things work rather than just applying them. For example, we may program up our own functions for things `R` has built in functionality for. Why do we do that? By programming it ourselves or deriving a known result, we force ourselves to really understand the underlying mechanics. This not only improves our understanding of statistical analysis but it also helps learn new things in the future. Our goal isn't just for you to learn this material, it is prepare you to teach yourself new material in the future.

4.3 Advice from Prior Generations of Students

Each year I ask students to provide advice to future generations of students. Here is some advice from prior students responding to "What advice would you give to another student considering taking this course?" I think the advice is great and it may be helpful coming from other students.

- Be ready to spend a lot of time
- Ask questions if you don't know what's going on!
- Study hard, work hard, review the slides.
- Investing a considerable amount of time in getting familiar with R and its various tools will pay off in the long run!
- Go over the lecture slides each week. This can be hard when you feel like you're treading water and just staying afloat, but I wish I had done this regularly.
- It's challenging but very doable and rewarding if you put the time in. There are plenty of resources to take advantage of for help.
- It will be hard but you will learn so much.
- This course is very challenging but greatly contributed to my understanding of social statistics. If you're truly invested in the subject and willing to put in the work (more than you expect possibly), it will be one of the best courses you've taken.
- This is a course where you will learn a lot and spend most of your time doing the psets. I highly recommend office hours for clarification as lecture covers a lot of material.

4.4 A note to everyone that is not a first year PhD student in the Department of Sociology

This course is completely designed to provide training to first year PhD students in the Department of Sociology. That means that some of the topics covered—and the way that these topics are covered—may not be optimal for your particular backgrounds. This warning aside, previous generations of students from many different departments have found the course useful.

5 Course Outline

Readings below are optional reference if you want to have things written up in another way to follow along. They track as closely as I can manage to the course. Anything I really want to ensure you have read before you come to class, I'll tell you about in class.

Introduction and Probability

- Course Details, Outline and Requirements
- Core Ideas of Probability
- Marginal/Joint/Conditional Probability, Bayes' Rule, Independence

Optional Reading:

- Blitzstein and Hwang, Chapter 1 (Probability and counting)
- Aronow and Miller, Chapter 1

Random Variables

- Random Variables
- Marginal, joint, and conditional distributions
- Expectations, Conditional Expectations
- Covariance, correlation, and independence

Optional Reading:

- Blitzstein and Hwang, Chapter 2, 3-3.2 (random variables), 4-4.2 (expectation), 4.4-4.6 (indicator rv, LOTUS, variance), 5.1 - 5.4 (continuous random variables), 7.0-7.3 (joint distributions), Chapter 9 (Conditional expectation)
- Aronow and Miller, Chapter 2

Learning from Random Samples

- Populations, samples, estimation
- Point estimation
- Properties of estimators
- Interval Estimation

Optional Reading:

- Aronow and Miller Chapter 3.1-3.2.6 (IID Random variables and estimation), 3.4.1 (confidence intervals)

Testing and Regression

- Hypothesis testing
- Nonparametric regression
- Parametric models and linear regression
- Bias-variance tradeoff
- Regression as a predictive model

Optional Reading:

- Aronow and Miller, Chapter 3.4.2 (testing)
- Aronow and Miller, Chapter 4.1.1 (bivariate regression)

Simple Linear Regression

- Mechanics of Ordinary Least Squares
- Assumptions of the linear model
- Properties of least squares
- Inference with regression

Optional Reading

- Aronow and Miller 4.1.2 (OLS Regression)

Linear Regression with Two Regressors

- Mechanics of regression with two regressors
- Simpson's Paradox
- Omitted variables and multicollinearity
- Dummy variables, interactions, and polynomials

Multiple Linear Regression

- Matrix algebra and mechanics of multiple linear regression
- Classical Inference in a multiple linear regression model
- Agnostic regression
- Sandwich standard errors

Optional Reading:

- Aronow and Miller 4.1.3 (Regression with Matrix Algebra)
- Fox Chapter 9.1-9.4 skip 9.1.1-9.1.2 (Linear Models in Matrix Form)
- Fox Chapter 10 (Geometry of Regression)

What Can Go Wrong and How to Fix It

- Diagnostics with Residuals
- Unusual and Influential Data → Robust Estimation
- Nonlinearity → Generalized Additive Models

Optional Reading:

- Fox Chapters 11-13, 19
- Fox Chapter 19 Robust Regression
- Aronow and Miller Chapters 4.2-4.4 (Inference, Clustering, Nonlinearity)
- Angrist and Pishke Chapter 8 (Nonstandard Standard Error Issues)

Frameworks for Causal Inference

- Potential outcomes
- Nonparametric structural equation models

Optional Reading:

- Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86.3 (2021): 532-565.
- Pearl and Mackenzie. *The Book of Why.* 2018.
- Morgan and Winship Chapter 1: Causality and Empirical Research in the Social Sciences
- Morgan and Winship Chapter 13.1: Objections to Adoption of the Counterfactual Approach
- Angrist and Pishke Chapters 1-2
- Hernan and Robins (2016) Chapter 1: A definition of a causal effect
- Samii, Cyrus. 2016. "Causal Empiricism in quantitative research. *The Journal of Politics.*
- Morgan and Winship (2015) Chapter 2 (Potential Outcomes), Chapter 3 (Causal Graphs)

Causality With Measured Confounding

- The Assumption of No Unmeasured Confounding
- Choosing Conditioning Variables

Optional Reading:

- Angrist and Pishke Chapter 2 (The Experimental Ideal) Chapter 3 (Regression and Causality)
- Morgan and Winship Chapters 3-4 (Causal Graphs and Conditioning Estimators)
- Hernan and Robins Chapter 3 Observational Studies
- Elwert and Winship (2014) "Endogenous selection bias: The problem of conditioning on a collider variable" *Annual Review of Sociology*
- Morgan and Winship Chapter 11 Repeated Observations and the Estimation of Causal Effects
- Aronow, Peter M. and Cyrus Samii. 2016. "Does regression produce representative estimates of causal effects?" *American Journal of Political Science.*

Sensitivity Analysis

- Sensitivity Analysis

Optional Reading:

- Cinelli, Carlos, and Chad Hazlett. "Making sense of sensitivity: Extending omitted variable bias." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1 (2020): 39-67.

6 Inspirations

The development of that course was in turn influenced by a number of people particularly: Brandon Stewart, Matt Blackwell, Dalton Conley, Adam Glynn, Justin Grimmer, Jens Hainmueller, Erin Hartman, Chad Hazlett, Gary King, Kosuke Imai, Kevin Quinn, and Teppei Yamamoto. I am grateful to everyone who has contributed to these materials, directly or indirectly. I am also grateful to generations of past preceptors who have had a huge influence on the direction the class has gone including Clark Bernier, Emily Cantrell, Elisha Cohen, Max Fineman, Alex Kindel, Angela Li, Ian Lundberg, Shay O'Brien, Alejandro Schugurensky, Ziyao Tian, and Simone Zhang.