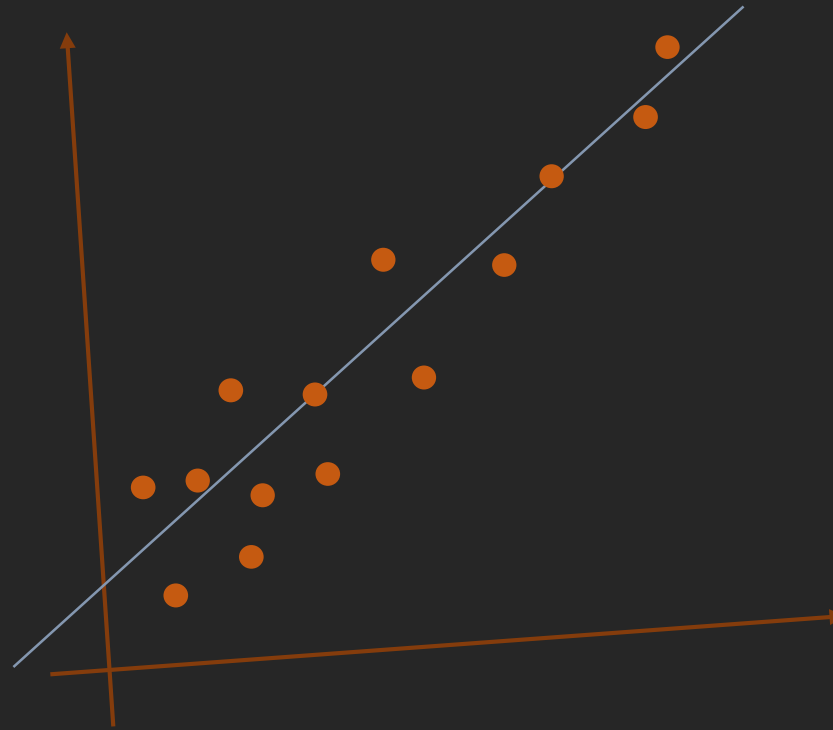# Applied Regression Analysis

STAT 4043 / STAT 5543

Introduction to the class

Pratyaydipta Rudra

Welcome to the class!

Outline

- Goal and expectation from this class.

- How the class will work.

- About quiz 1 and new student learning goal assessment.

Most important rule for this class

- **Please ask questions!**
  - In class, after class, office hours, email.
  - Please stop me if you need. I am happy to re-explain anything.

For every topic, I will try to answer the following 2 questions:

Why?

How?

# Why is applied regression important for you?

How can we make sure you learn the subject well to reach your goal?
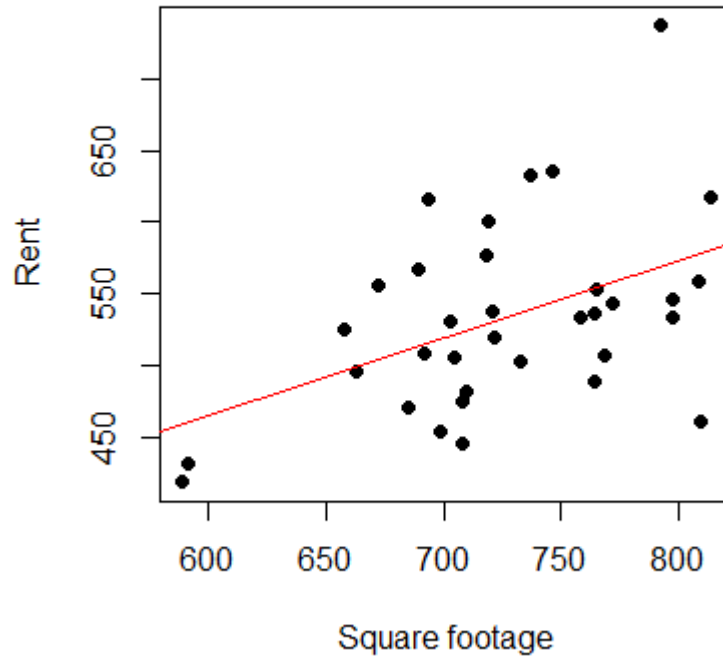
A diverse class

- This class consists of four different sections, two face-to-face, and two online.

- Overall, the class has students from very diverse academic backgrounds with vastly different goals.

What will we learn?

What do we intend to do with a regression model?
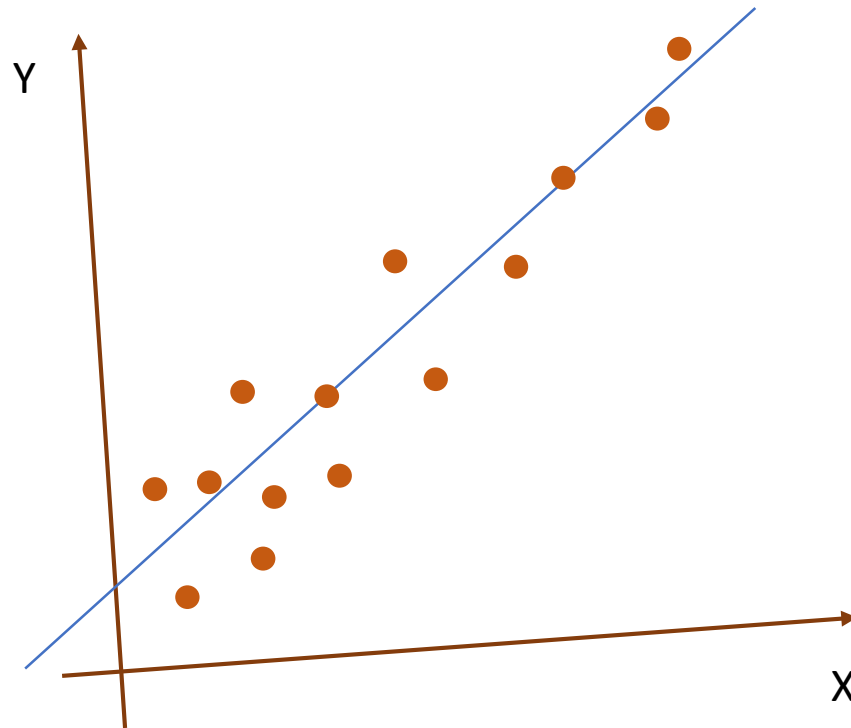
Example – relationship between square footage and rent of apartments (simple linear regression)

❖ Is there are relationship between the two variables?

❖ How strong is the relationship?

❖ Can we estimate how much the rent goes up per square foot, on average?

❖ Can we predict the rent for an apartment with a given square footage?

Multiple linear regression

- What are the other factors besides square footage that have an effect on the rent?

- If we collect data on these other variables, how can we update our model?

AGE: Age of the property.

SQFT: Square footage of unit.

SD: Amount of safety deposit.

UNTS: Number of units in the complex.
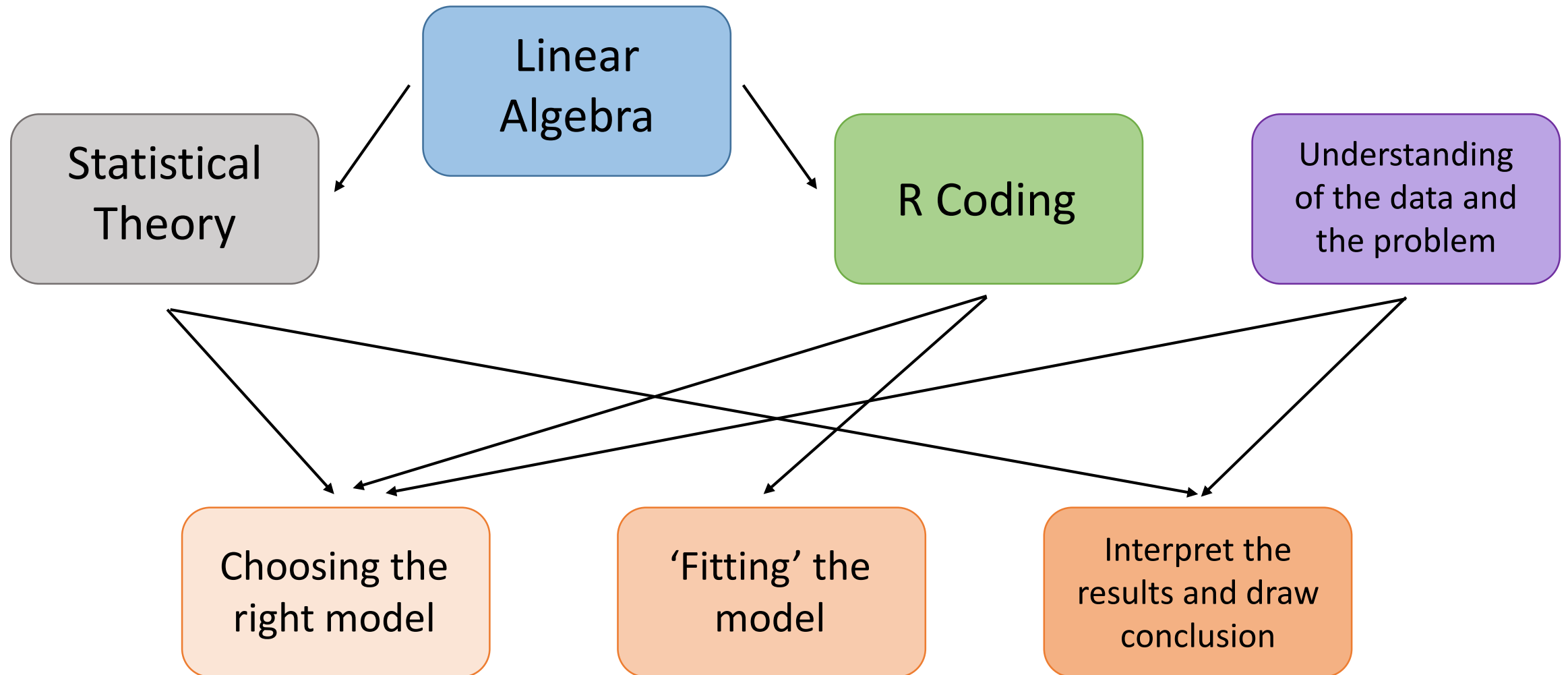
GAR: Presence of a garage (0: no, 1: yes).

SS: Presence of security system (0: no, 1: yes).

FIT: Presence of fitness facilities (0: no, 1: yes).

RENT: monthly rent.

AGE	SQFT	SD	UNTS	GAR	SS	FIT	RENT
7	692	150	408	0	1	0	508
7	765	100	334	0	1	1	553
8	764	150	170	0	1	1	488
13	808	100	533	0	1	1	558
7	685	100	264	0	0	0	471
7	710	100	296	0	0	0	481
5	718	100	240	0	1	1	577
6	672	100	420	0	0	1	556
4	746	100	410	1	1	1	636
4	792	100	404	1	1	1	737
8	797	150	252	0	1	1	546
7	708	100	276	0	1	0	445
8	797	150	252	0	0	1	533
6	813	100	416	0	0	0	617
7	708	100	536	0	1	1	475
16	658	100	188	1	1	1	525

Different tools to learn regression analysis

Linear Algebra

Statistical Theory

R Coding

Understanding of the data and the problem

Choosing the right model

'Fitting' the model

Interpret the results and draw conclusion

What you will not learn in class

- Basic probability and probability distributions
- Normal, t, F, χ^2 distributions
- Basic concepts of estimation and hypothesis testing; p-values and confidence intervals
- College algebra
- Basics of computer programming

Second week quiz will be based on pre-requisites.

Also…

- This is not a programming course. We will learn how to use R for regression analysis, but R programming is not going to be covered rigorously.
    - Some amount of self-learning is expected.

- This is not a linear algebra course. We will learn some linear algebra, only that much which is required to understand regression analysis.

Study resources

- Lecture slides, R codes and lecture videos

- Some notes on R programming

- The textbook is optional, I don't follow the book strictly.

- Practice problems

- Wednesday videos and Friday problem solving

Suggestions to do well in the course

- Stay with the course, don't wait till homeworks are assigned or till the day before the exam. Go back to the materials taught in class every week.

- Use all the resources (e.g. R notes).

- It is not impossible to get lost during a lecture. Please stop me and ask right away.

- Solve the practice problems.

- When assignments/exams are returned, figure out your mistakes and try to solve the problem again on your own.

- Use the internet for questions related to R-coding (a guide is posted).

Questions, suggestions from you?

Assessment

- Assignments: 25%
- Short quizzes: 15% - **Some in-class, some online**
- Midterm Exams: 40%
- Project: 20%

- See syllabus for more details and a tentative schedule.
- **Online students need to arrange proctoring for the exams.**

Homeworks and practice problems

- 5 graded assignments (one additional assignment for graduate students).

- Practice problems: meant to prepare you for the graded items.

- I can help you as much as you need on practice problems. Please solve them!

- Help on graded assignments will be limited.

Computer usage

- Installing R to your computer is required. You will need to use R throughout the course.

- I encourage you to bring a laptop to the class.

- Get accustomed to Canvas.
  - Slides, notes, assignments (usually on Fridays), grades will be posted to Canvas.

Conduct

- Attendance will be taken for bookkeeping purposes. Not used for grading.

- Please limit in class computer usage to academic activities.

- Maintaining academic integrity is important.
  - Feel free to discuss homework assignments, but you must write up your own assignment and write your own codes.
  - No collaboration is allowed for other forms of assessments.

Usage of artificial intelligence

- Not allowed for solving assignments.

- It can sometimes be an interesting learning tool, but I don't recommend using it.

- Except perhaps for debugging codes...

fortune.com
ChatGPT went from accurately answering a simple math problem 98% of the time to just 2%

Questions?

Tasks for you before the next class

- Read syllabus and the class materials guide posted under the syllabus module. Bring any questions for me.
- Install R and Rstudio to your laptop (guide posted) before the next class.

About quiz 1

Quiz 1 is a syllabus quiz plus new student survey

- Please read the syllabus first before you attempt it.
- Second part of it has some questions about yourself. You need to answer all questions (except the last one) to get credit.

Example: My answers

- Preferred Name for the class: Dr Rudra

- Department: Statistics (Previously Biostatistics, Biostatistics, Statistics and Statistics)

My answers

- Goals and expectations from this course?
  - Students will be able to develop basic understanding of regression models.
  - Given a real data, students will be able to identify the right model for the data, analyze the data and interpret the results.

- What statistics course have you taken before?
  - …

My answers

- Have you taken any linear algebra course?
    - Yes

- Have you used statistical software R before?
    - Yes

My answers

- Tell me something about yourself
  - I completed my masters in statistics from Indian Statistical Institute and came to the US for my PhD, which I completed in 2015 from UNC Chapel Hill. Then I completed 3 years of post doc from UC Denver.

  - I took statistics in high school because that was the only alternative to biology (which I hated). Now my research area is biostatistics!

  - Outside academia, I love to play tennis and photograph wildlife.

Questions?