THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Introduction to predictive analytics

Lecture 1

STA 371G

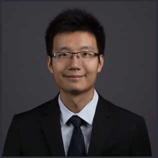
Course goals

- Use regression and time series analysis to build predictive models
- Build decision trees to help make decisions under uncertainty
- Utilize simulations to forecast outputs based on uncertain inputs
- Given a new business situation, select an appropriate analysis, carry it out, and effectively communicate the results
- This is a practical course!

About the course staff

- Instructor: **Brian Lukoff, Ph.D.**
  - Office hours: M/W 11 AM - 12 PM in CBA 3.440
  - Contact: `brian.lukoff@utexas.edu` or 415-652-8853
- TAs:
  - Office hours: W/Th/F 2-4 PM in CBA 4.304A
  - Help session: F 11 AM - 12 PM in the ModLab

Jared Fisher


Han Jiang

Sai Ravi

Who am I?

- **Educator:** Previously taught at Harvard University and Boston University
- **Entrepreneur:** Currently co-founder and CEO of Perusall; formerly co-founder and CEO of Learning Catalytics (acquired by Pearson)
- **Engineer/statistician:** Software engineering/analytics background

1. Find someone who...

2. Course logistics

3. Let’s do some statistics, yo

For each box on your bingo card, find someone who matches the description in the box. You must use a different person for each box.

The winner will be crowned the STA 371G Bingo Champion$^{\text{TM}}$.

1. Find someone who...
2. Course logistics
3. Let’s do some statistics, yo

Canvas

- Access at `canvas.utexas.edu`
- This is your home base for the course
- Make sure you can log in and are enrolled in STA 371G in Canvas

Class participation

- We will use Learning Catalytics so you can get practice of the concepts during class
- Buy online ($12) at `learningcatalytics.com` or use for free if you bought for another class (you may still have access from STA 309)
- Grading based on completeness only; answer 75% of the questions for full credit
- Bring a laptop, smartphone, or tablet to every class
- A note about devices in class

Reading assignments

- No textbook is required for this course
- We will use Perusall for reading assignments
- Access for free at `perusall.com`
- Use Perusall to ask and answer your classmates questions and have discussions in the text
- This will help you better understand the text and will help me gear class time to what topics you are having the most trouble with
- Reading assignments are due by 7 PM; grading is based on effort and thoughtfulness of your questions and comments

Statistical computing

- We will use R for statistical analysis throughout the course
- This is industrial-strength, state-of-the-art, and free software for statistical computing
- We will access R through RStudio, a graphical interface for R
- Download R and RStudio at `rstudio.com`

Homework

- Regular homework assignments during the semester
- Submit in QUEST by 11:59 PM on the due date
- Why homework?

Exams

- Two midterm exams and a cumulative final exam
- All are given outside of class times; you'll sign up for a time that is convenient for you
- Tests are in the ModLab; you'll have access to R during every exam
- Your final exam will overwrite your lowest midterm grade if it helps your overall grade

Team project

- One team project
- You will pick a data set (or create one, e.g. through a survey) and apply regression techniques (we'll learn about this!) to build a predictive model

Grading

Class participation	**5%**
Reading assignments	**5%**
Homework	**15%**
Team project	**15%**
Midterm 1	**20%**
Midterm 2	**20%**
Final Exam	**20%**

1. Find someone who...

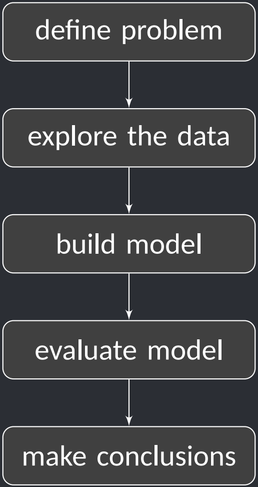
2. Course logistics

3. Let’s do some statistics, yo

Purpose of a model

- **Make a prediction** about one variable based on the others
- **Understand the relationships** between the variables

Data analysis process



define problem

explore the data

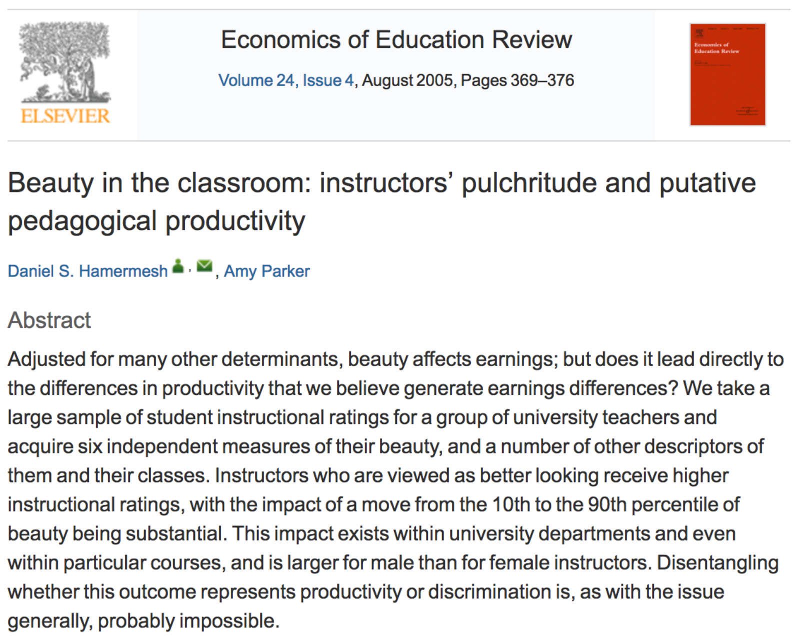
build model

evaluate model

make conclusions

Define the problem

What personal characteristics about an instructor do you think are predictive of the scores they receive on student evaluations?



Economics of Education Review

Volume 24, Issue 4, August 2005, Pages 369–376

Beauty in the classroom: instructors’ pulchritude and putative pedagogical productivity

Daniel S. Hamermesh, Amy Parker

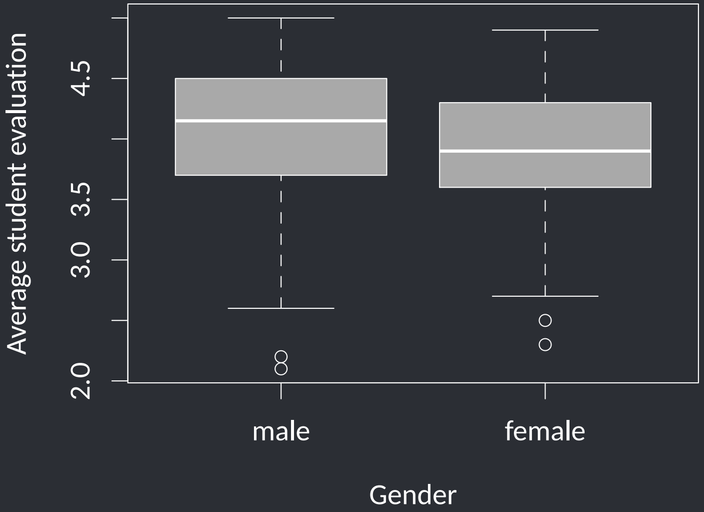
Abstract

Adjusted for many other determinants, beauty affects earnings; but does it lead directly to the differences in productivity that we believe generate earnings differences? We take a large sample of student instructional ratings for a group of university teachers and acquire six independent measures of their beauty, and a number of other descriptors of them and their classes. Instructors who are viewed as better looking receive higher instructional ratings, with the impact of a move from the 10th to the 90th percentile of beauty being substantial. This impact exists within university departments and even within particular courses, and is larger for male than for female instructors. Disentangling whether this outcome represents productivity or discrimination is, as with the issue generally, probably impossible.

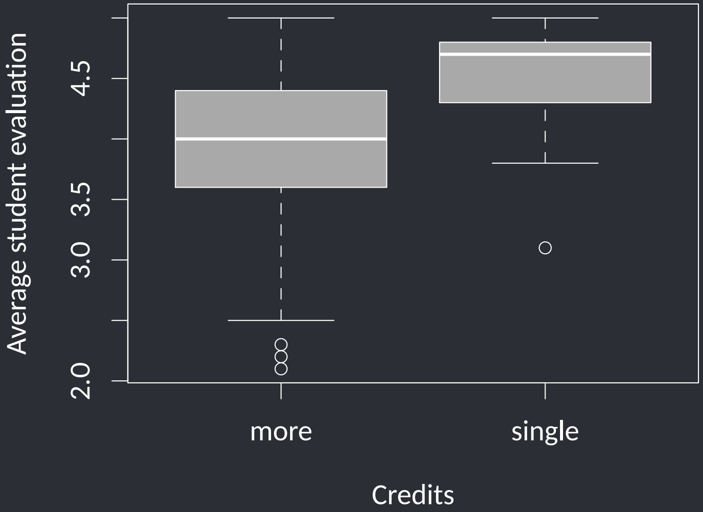
Hamermesh & Parker (2004) data set

- Student evaluations of $N = 463$ instructors at UT Austin, 2000-2002
- For each instructor:
  - **beauty**: average score from a six-student panel)
  - **gender**: male or female
  - **credits**: single- or multi-credit course
  - **age**: age of instructor
  - (and more…)

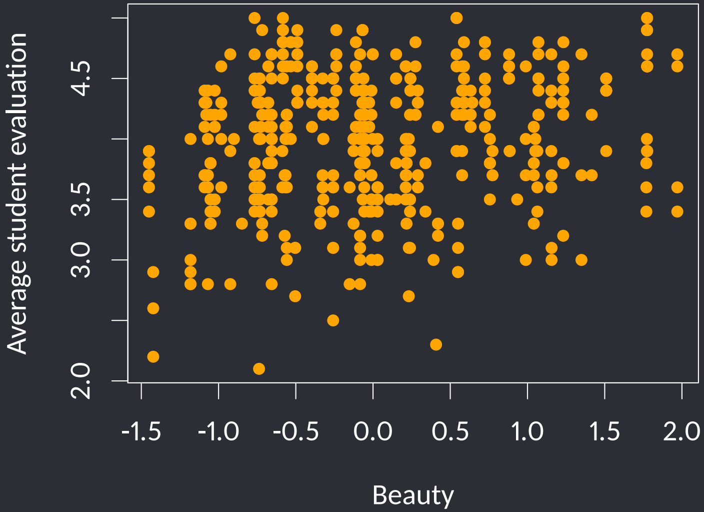
Explore the data



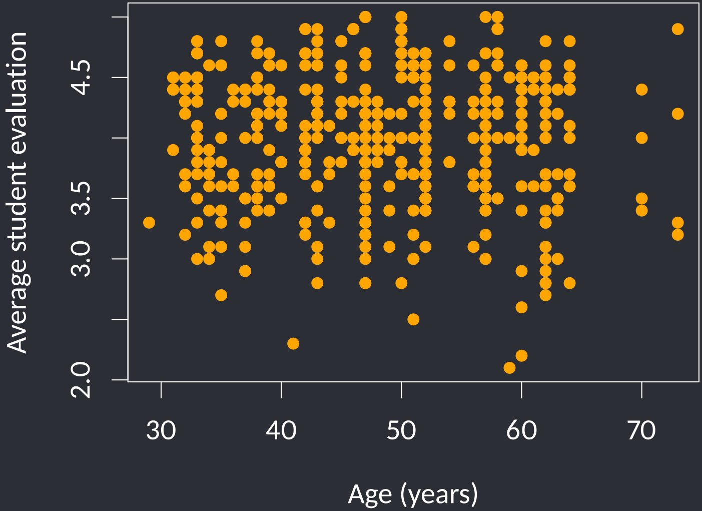
Explore the data



Explore the data



Explore the data



Build the model

A regression model lets us create a model that incorporates all of these relationships to best predict evaluation scores:

$$\widehat{\text{eval}} = 4.13 + 0.16 \cdot \text{beauty} - 0.2 \cdot \text{female} + 0.58 \cdot \text{credits} + 0 \cdot \text{age}$$

Build the model

A regression model lets us create a model that incorporates all of these relationships to best predict evaluation scores:

$$\widehat{\text{eval}} = 4.13 + 0.16 \cdot \text{beauty} - 0.2 \cdot \text{female} + 0.58 \cdot \text{credits} + 0 \cdot \text{age}$$

We predict a 40-year-old female, with a beauty score of 2, teaching a multi-credit course would get an evaluation score of

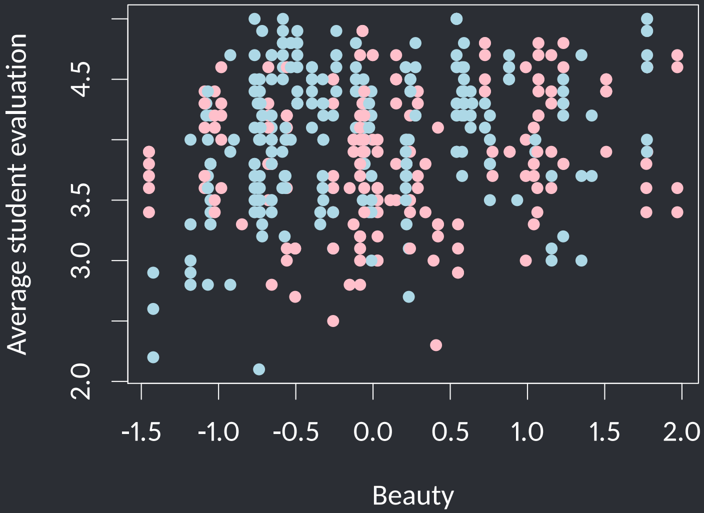
$$\widehat{\text{eval}} = 4.13 + 0.16 \cdot 2 - 0.2 \cdot 1 + 0.58 \cdot 0 = 4.18.$$

Evaluate the model

How could you evaluate the quality of this model?

Can we do better?

Do you see a different pattern between men and women?



Six for the weekend

1. Read the syllabus
2. Install R and RStudio on your computer
3. Make sure you can log in to Canvas
4. Purchase (or see if you already have access to) Learning Catalytics
5. Create a Perusall account
6. Bring a device to class on Monday