INFO251 – Applied Machine Learning

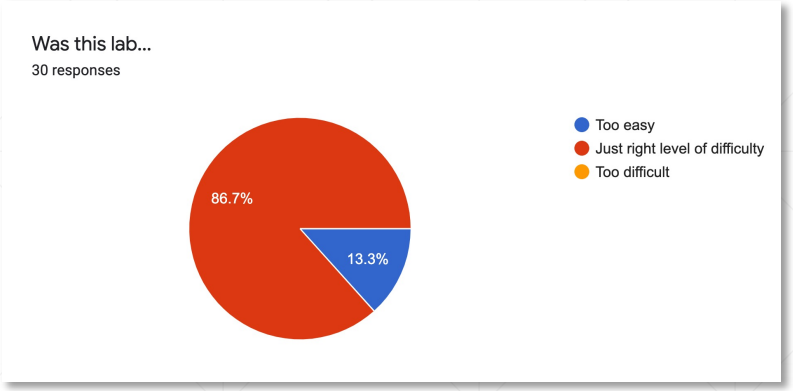
Lab 2

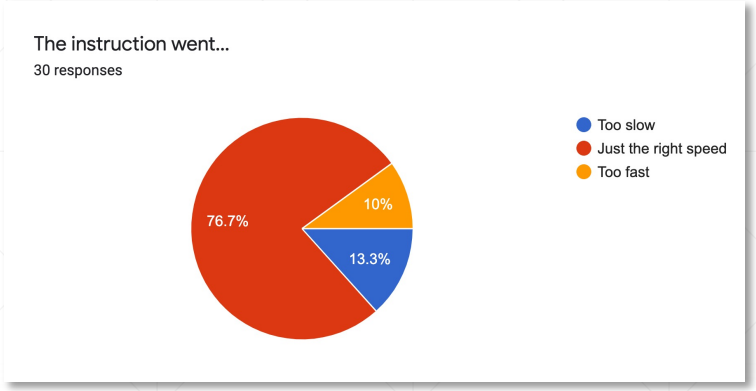
Emily Aiken

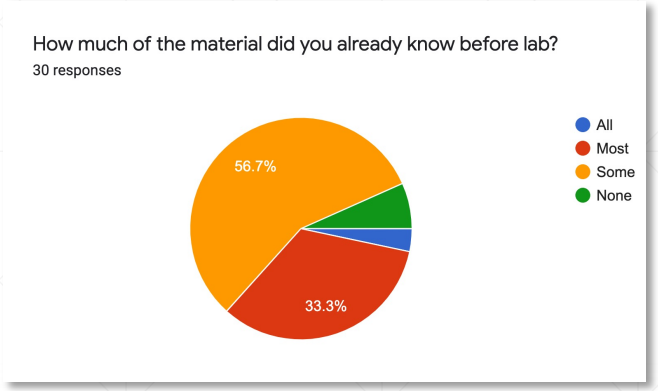
Announcements

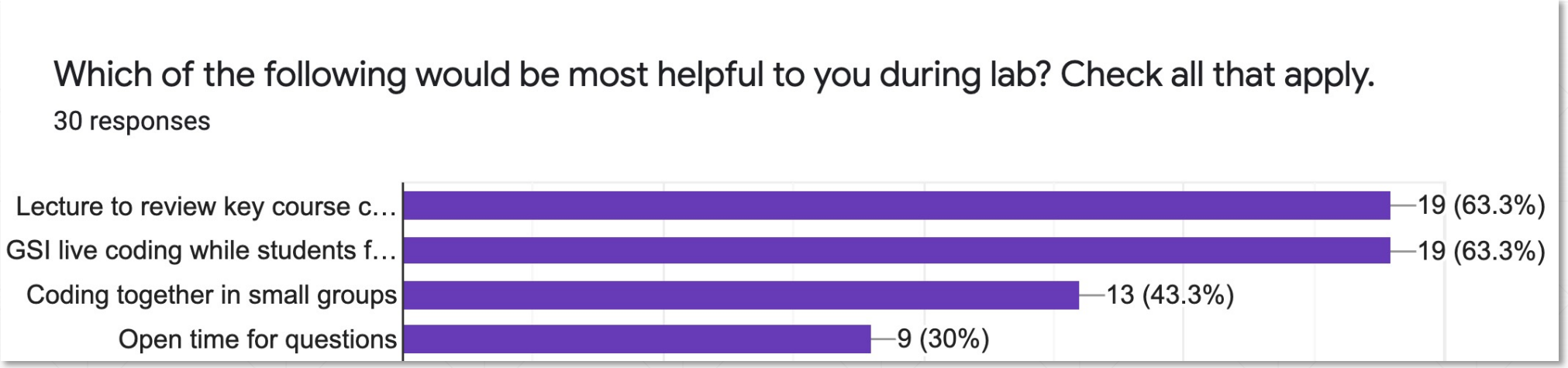
- **Problem Set 1** solutions are posted
- **Problem Set 2** due Monday February 7
- **Add/drop deadline** on Friday, February 18
- **Lecture**: In person starting next week, SH rooms 202 and 210
- **GSI office hours**: Hybrid starting this week, SH room 6a or on zoom

Feedback


Was this lab...
30 responses
Too easy
Just right level of difficulty
Too difficult
86.7%
13.3%


The instruction went...
30 responses
Too slow
Just the right speed
Too fast
76.7%
10%
13.3%


How much of the material did you already know before lab?
30 responses
All
Most
Some
None
56.7%
33.3%


Which of the following would be most helpful to you during lab? Check all that apply.
30 responses
Lecture to review key course c…
19 (63.3%)
GSI live coding while students f…
19 (63.3%)
Coding together in small groups
13 (43.3%)
Open time for questions
9 (30%)

Today: Regression and Impact Evaluation

- Univariate regression
- Multivariate regression
  - Dummy variables
  - Interaction terms
- Difference-in-differences

Today's tools: `pandas` and `statsmodels`

Today's Data: Minimum Wage and Employment

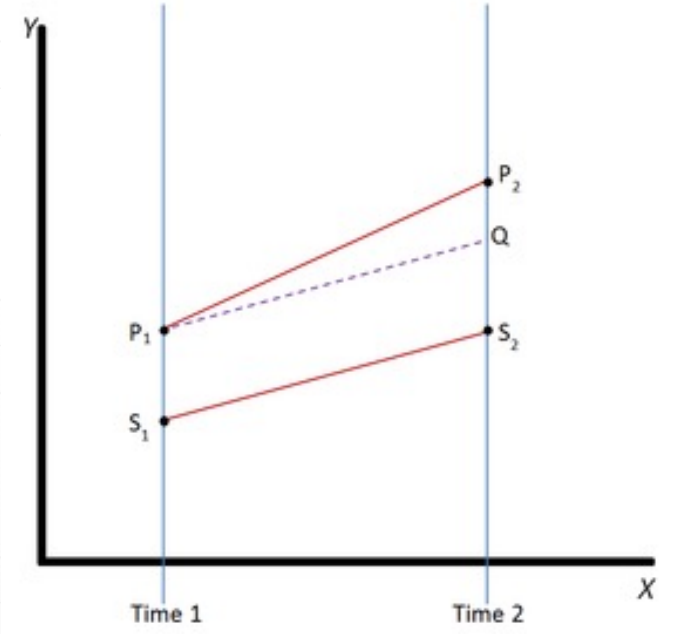
- **Source**: Card and Krueger (1994). Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review* Vol. 90 No. 5.
- April 1992: NJ raises the minimum wage from $4.25/hour to $5.05/hour.
- Card and Krueger conducted a survey with 410 fast food restaurants in NJ and Eastern PA, once in March 1992 and once in December 1992.
  - Number of employees
  - Region, chain, hours of operation, number of registers, other covariates

Data Types

Person	Test Score	Has Siblings	Marital Status	Education
A	740	1	Married	12
B	632	1	Never married	11
C	676	0	Divorced	10
D	572	1	Widow/widower	11
E	580	0	Married	8
F	738	1	Married	7
G	710	0	Divorced	9

Difference-in-differences

- Compare **changes in outcomes** in the treatment group and the control group
- Assumptions:
  - “Parallel trends”: Trends in pre-treatment outcomes are the same in treatment and control groups
  - Composition of treatment and control groups stable across time
  - No spillover effects



Difference-in-differences

- $Y = B_0 + B_1*Time + B_2*Treatment + B_3*(Time*Treatment) + B_4*Covariates$