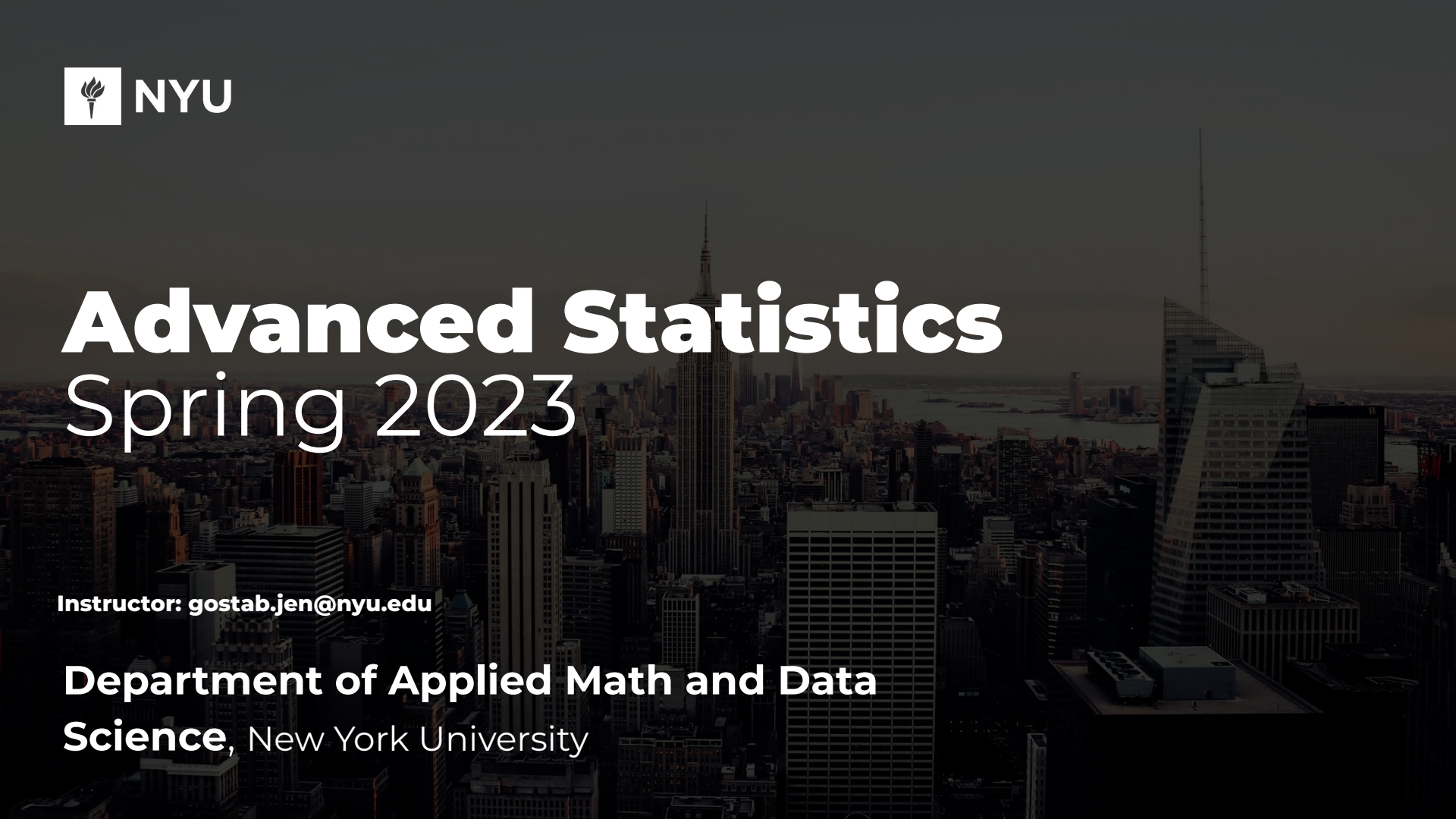


NYU

Advanced Statistics
Spring 2023

Instructor: gostab.jen@nyu.edu

Department of Applied Math and Data Science, New York University

Homework 2

- Spend time coding
- Write clear
- Week -8

NYU

Lab Assignment

Don’t forget to submit your lab activity to brightspace and ensure your full credit

NYU

Week 7-9 Plans

Review:

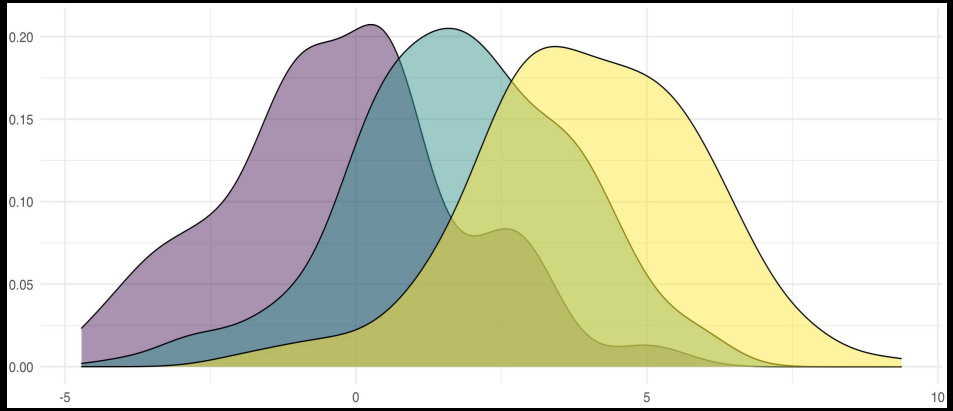
Multiple Regression

Logistic Regression

Intro to GLM-Generalized Linear Model

Practice for sample exam 2

Practice for R:
Regression in R



NYU

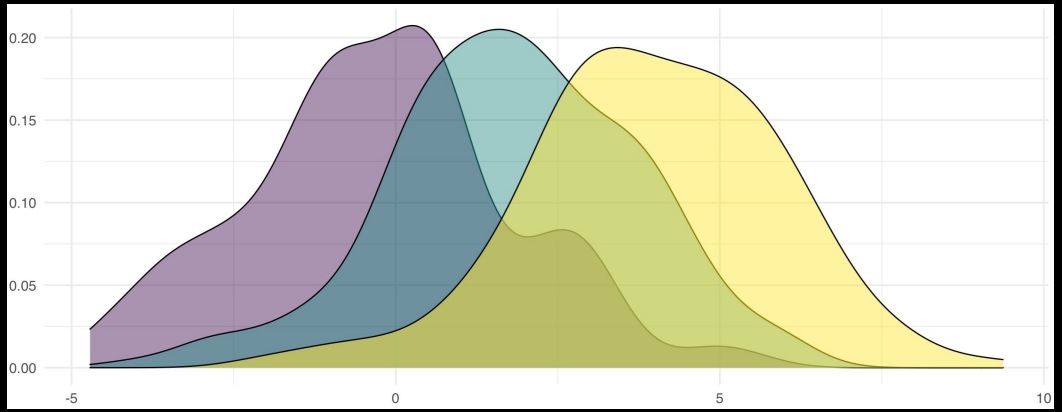
Multiple Regression

Review:

→Fundamental problem

Notations and math

Empirical Loss



NYU

Multiple Regression

We have some feature variables(IV)– A

We also observe outcomes(DV)- b

The most ideal way is to solve : $Ax = b$

such as $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & a_{17} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & a_{27} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & a_{37} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}$

Logistic Regression

- If the outcome is collected in binary form; there are some examples:

1) Federal President Election**(1 for win; 0 for lose)**
2) Survey response indicator variable **(1 for respondent; 0 for non-rep)**
3) Treatment assignment**(1 for receive treatment, 0 for placebo)**

Logistic Regression

- If the outcome is collected in binary form; there are some examples:

1) Federal President Election**(1 for win; 0 for lose)**
2) Survey response indicator variable **(1 for respondent; 0 for non-rep)**
3) Treatment assignment**(1 for receive treatment, 0 for placebo)**

Imagine selected students are required to complete a survey:

If given the data for which student responded to the survey

Think about what estimator we have for this variable *R*?

Logistic Regression

➔Survey response indicator variable **(1 for respondent; 0 for non-rep)**

Imagine selected students are required to complete a survey:

If given the data for which student responded to the survey

Think about what estimator we have for this variable *R*?

Can we get E(R)? Think about it.

Logistic Regression

➔Survey response indicator variable **(1 for respondent; 0 for non-rep)**

Imagine selected students are required to complete a survey:

If given the data for which student responded to the survey

Think about what estimator we have for this variable *R*?

Can we get E(R)? Think about it. **-Yes =0.6**

Can we get P(R=1) or P(R=1 | X) **This is a harder question.**

Logistic Regression

If we have a dataset as follows:

	R	P(R=1| X)	*x*
1	1	?	12
2	0	?	17
3	0	?	40
4	1	?	25
5	1	?	29

Logistic Regression

If we have a dataset as follows:

If we just need a good estimator for *E[R]*

Or

We want to precisely estimate the probability of responding to the survey for each unit in our sample

	R	P(R=1| X)	*x*
1	1	?	12
2	0	?	17
3	0	?	40
4	1	?	25
5	1	?	29

Logistic Regression

Calculate E[R] for this abbreviated data and Submit it

If we just need a good estimator for *E[R]*

Or

We want to precisely estimate the probability of responding to the survey for each unit in our sample

	R	P(R=1| X)	*x*
1	1	?	12
2	0	?	17
3	0	?	40
4	1	?	25
5	1	?	29