Economics 2222A

Lecture Thu Sep 10

Professor Nirav Mehta

University of Western Ontario

Fall 2026

Question: Is it more effective to take notes on the laptop or by hand?

Question: Is it more effective to take notes on the laptop or by hand?

- 66 students were assigned to watch a TED talk
  - 33 were randomly assigned to take notes using a laptop
  - 33 were randomly assigned to take notes by hand

Question: Is it more effective to take notes on the laptop or by hand?

- 66 students were assigned to watch a TED talk
  - 33 were randomly assigned to take notes using a laptop
  - 33 were randomly assigned to take notes by hand
- Each group was then tested on the content of the TED talk for both *factual* and *conceptual* questions

Question: Is it more effective to take notes on the laptop or by hand?

- 66 students were assigned to watch a TED talk
  - 33 were randomly assigned to take notes using a laptop
  - 33 were randomly assigned to take notes by hand
- Each group was then tested on the content of the TED talk for both *factual* and *conceptual* questions
- Researchers then computed the average score on the test, broken down by whether the questions were factual or conceptual

Two kinds of questions

1. What would you *expect to see* if the **truth** was
   - there's no effect of note-taking medium on performance for either type of question?

Two kinds of questions

1. What would you *expect to see* if the **truth** was
   - there's no effect of note-taking medium on performance for either type of question?
   - taking notes by laptop is very effective for factual questions?

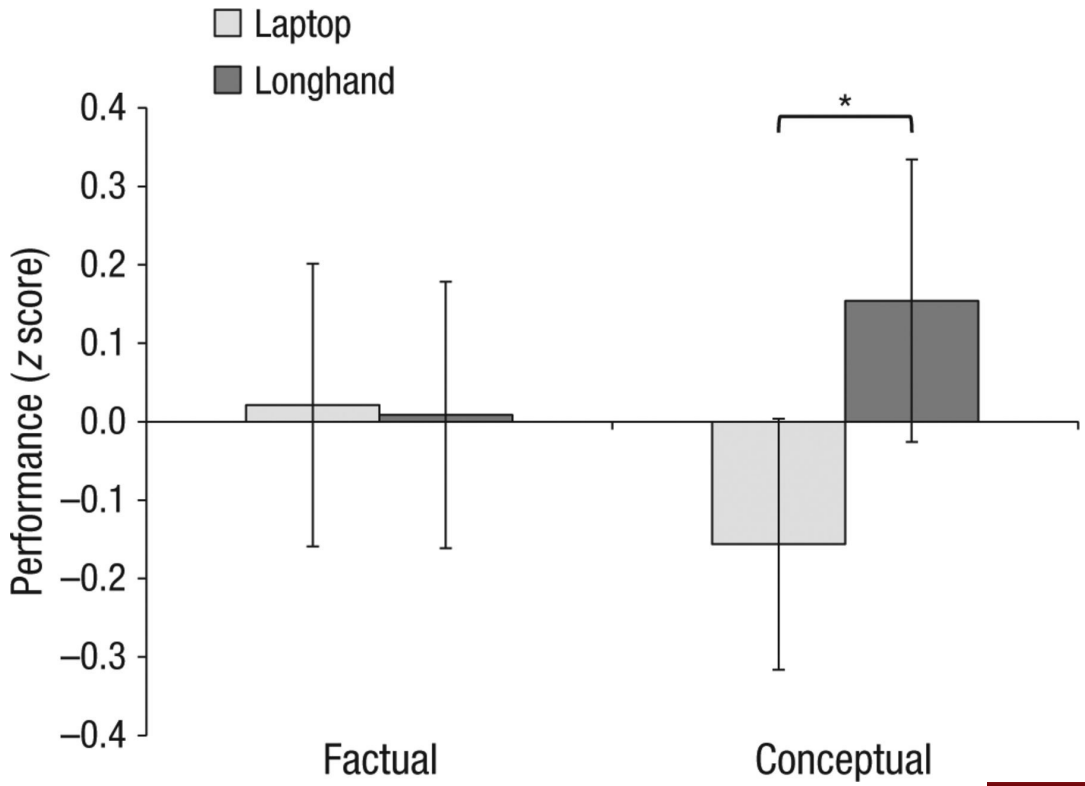
Two kinds of questions

1. What would you *expect to see* if the **truth** was
   - there's no effect of note-taking medium on performance for either type of question?
   - taking notes by laptop is very effective for factual questions?
2. What might you *say about the truth* if you **saw**
   - the average score on conceptual questions was *twice as high* when students took notes by hand?

Two kinds of questions

1. What would you *expect to see* if the **truth** was
   - there's no effect of note-taking medium on performance for either type of question?
   - taking notes by laptop is very effective for factual questions?
2. What might you *say about the truth* if you **saw**
   - the average score on conceptual questions was *twice as high* when students took notes by hand?
   - *slightly higher* average scores on conceptual questions for those who took notes by hand?

Fig. 1. Mean z-scored performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 1).



Mueller P A , and Oppenheimer D M Psychological Science 2014;0956797614524581

Concepts used to think about this study's results

- Mean
  - e.g., "the average score on factual questions was 0.01"

Concepts used to think about this study's results

- Mean
  - e.g., "the average score on factual questions was 0.01"
- Conditional mean
  - e.g., "the average score on conceptual questions was 0.15 when taking notes by hand and -0.15 when using laptops"

Concepts used to think about this study's results

- Mean
  - e.g., "the average score on factual questions was 0.01"
- Conditional mean
  - e.g., "the average score on conceptual questions was 0.15 when taking notes by hand and -0.15 when using laptops"
- Probability statement
  - e.g., "there's a 2% chance the difference in average scores could be due to random chance"

Concepts used to think about this study's results

- Mean
  - e.g., "the average score on factual questions was 0.01"
- Conditional mean
  - e.g., "the average score on conceptual questions was 0.15 when taking notes by hand and -0.15 when using laptops"
- Probability statement
  - e.g., "there's a 2% chance the difference in average scores could be due to random chance"
- Hypothesis testing
  - e.g., "the results of the experiment lead us to reject at the 5% significance level that note-taking medium has no effect on conceptual topics"

Concepts used to think about this study's results

- Mean
  - e.g., "the average score on factual questions was 0.01"
- Conditional mean
  - e.g., "the average score on conceptual questions was 0.15 when taking notes by hand and -0.15 when using laptops"
- Probability statement
  - e.g., "there's a 2% chance the difference in average scores could be due to random chance"
- Hypothesis testing
  - e.g., "the results of the experiment lead us to reject at the 5% significance level that note-taking medium has no effect on conceptual topics"

You will be familiar with *all* of these concepts by the end of this course!

About me

My name is Nirav Mehta.

- I was born in the USA.
- I did a BA in Economics and International Relations, MSc in Economics, and a PhD in Economics
- I came to Western in 2011, after completing my PhD, and I love this course!!

About me

My name is Nirav Mehta.

- I was born in the USA.
- I did a BA in Economics and International Relations, MSc in Economics, and a PhD in Economics
- I came to Western in 2011, after completing my PhD, and I love this course!!

I am an applied microeconomist, whose research focus is mostly on education and health economics.

- I use probability and statistics every day in my research.
- Example research question:
  - How responsive are physicians to changes in reimbursement rates for treatments?
  - Process: Get data on payment schedules, physician behavior, and patient outcomes, and estimate models of physician behavior to inform public policy (e.g., OHIP reimbursement rates)!

About this course

- The course outline is available on OWL Brightspace at `https://westernu.brightspace.com/`
- Make sure you review it very carefully
- Here are some highlights

About this course

Contact info

- Nirav Mehta
- Office: SSC 4011
- Email: nirav.mehta@uwo.ca
- Office hours: Tuesdays and Thursdays 10:15-11:15am, in person at SSC 4011, or by appointment

About this course

Contact info

- Nirav Mehta
- Office: SSC 4011
- Email: nirav.mehta@uwo.ca
- Office hours: Tuesdays and Thursdays 10:15-11:15am, in person at SSC 4011, or by appointment

Class times and locations

- Section 001 T: 11:30am-12:30pm; TH: 11:30am-1:30pm in SSC 3024
- Section 002 T: 2:30-3:30pm; TH: 1:30-3:30pm in AHB 1B02

Textbook

There is **no required textbook**: my lectures are designed to be self-contained.

- You are expected to attend all lectures; lecture notes will not be distributed.

Textbook

There is **no required textbook**: my lectures are designed to be self-contained.

- You are expected to attend all lectures; lecture notes will not be distributed.

If you want a conventional textbook, use Newbold, P., Carlson, W., and B. Thorne. Statistics for Business and Economics (Eighth edition). (NCT8)

Options:

1. The NCT8 eBook can be purchased for $102, or rented for 180 days for $68, via the Campus Bookstore.
2. Second-hand or earlier editions (e.g., NCT7) are fine.
3. There are several copies on reserve at the Weldon Library.

I will provide chapter references in case you'd like to follow along in the textbook.

Grading

There are graded homeworks, in-class quizzes, and a final exam.

Grading

There are graded homeworks, in-class quizzes, and a final exam.

- **Homeworks:**
  - 6 graded homeworks, which will cover recent material
  - Posted on the course website ahead of their due dates (see the course outline); submitted online by 11pm on the due date
  - Late homeworks are forfeited
  - The lowest homework score is dropped

Grading

There are graded homeworks, in-class quizzes, and a final exam.

- **Homeworks:**
  - 6 graded homeworks, which will cover recent material
  - Posted on the course website ahead of their due dates (see the course outline); submitted online by 11pm on the due date
  - Late homeworks are forfeited
  - The lowest homework score is dropped
- **Quizzes**
  - 6 quizzes, written during regularly scheduled lecture time on Tuesdays spread across the term (dates on the course outline)
  - Quizzes will typically focus on material from the preceding one to two weeks
  - The lowest two quiz scores are dropped
  - No make-up quizzes

Grading

There are graded homeworks, in-class quizzes, and a final exam.

- **Homeworks:**
  - 6 graded homeworks, which will cover recent material
  - Posted on the course website ahead of their due dates (see the course outline); submitted online by 11pm on the due date
  - Late homeworks are forfeited
  - The lowest homework score is dropped
- **Quizzes**
  - 6 quizzes, written during regularly scheduled lecture time on Tuesdays spread across the term (dates on the course outline)
  - Quizzes will typically focus on material from the preceding one to two weeks
  - The lowest two quiz scores are dropped
  - No make-up quizzes
- **Final exam**
  - Cumulative. The date and time will be set later this semester by the Registrar's office.

Grading

The course grade is computed using the following weights:

Homework:	10% (best 5 $\times$ 2% per homework)
Quizzes:	40% (best 4 $\times$ 10% per quiz)
Final exam:	50%

Grading

The course grade is computed using the following weights:

Homework:	10% (best 5 $\times$ 2% per homework)
Quizzes:	40% (best 4 $\times$ 10% per quiz)
Final exam:	50%

Regrading:

- You may submit a quiz for a regrade within one week of when it has been handed back in class but note that I will scrutinize the entire quiz during the regrade.
- This means your grade may go up, go down, or stay the same

Conditions for passing the course

- You must pass the final exam (i.e., receive a mark of at least 50% on it) to pass the course.
  - A student who does not pass the final exam will receive a course grade of no more than 48%.
- Writing at least four of the six quizzes is a condition of writing the final exam.
  - A student who misses more than two quizzes without approved academic consideration may be prohibited from writing the final exam.

Missed assessments

Flexibility is already built into the grading scheme:

- A missed homework or quiz simply becomes one of your dropped scores; no further action is needed.

Missed assessments

Flexibility is already built into the grading scheme:

- A missed homework or quiz simply becomes one of your dropped scores; no further action is needed.
- If circumstances beyond your control cause you to miss more homeworks or quizzes than would already be dropped, contact me and your academic counsellor as soon as possible (no later than 48 hours after the missed assessment).

Missed assessments

Flexibility is already built into the grading scheme:

- A missed homework or quiz simply becomes one of your dropped scores; no further action is needed.
- If circumstances beyond your control cause you to miss more homeworks or quizzes than would already be dropped, contact me and your academic counsellor as soon as possible (no later than 48 hours after the missed assessment).
- See the course outline for the full academic consideration policy.

Introduction

This lecture

- Why study probability and statistics?
- Describing data
  - Graphically
  - Numerically: Descriptive statistics

Introduction

What is statistics?

- We use *statistics* to draw inferences about the real world in a systematic fashion
  - *Probability* provides the foundation for statistical analysis

Introduction

What is statistics?

- We use *statistics* to draw inferences about the real world in a systematic fashion
  - *Probability* provides the foundation for statistical analysis

What can we use statistics for?

- Inferences in the social sciences (effect of government programs, like subsidies to farmers or school choice programs)

Introduction

What is statistics?

- We use *statistics* to draw inferences about the real world in a systematic fashion
  - *Probability* provides the foundation for statistical analysis

What can we use statistics for?

- Inferences in the social sciences (effect of government programs, like subsidies to farmers or school choice programs)
- Forecasting (weather, stock market)

Introduction

What is statistics?

- We use *statistics* to draw inferences about the real world in a systematic fashion
  - *Probability* provides the foundation for statistical analysis

What can we use statistics for?

- Inferences in the social sciences (effect of government programs, like subsidies to farmers or school choice programs)
- Forecasting (weather, stock market)
- Anything where we have a model of some process where we don't know what's going to happen with certainty

Note on this course

- In this course we are not studying micro- or macroeconomics.

Note on this course

- In this course we are not studying micro- or macroeconomics.
- This is often a very abstract course, but I will try to provide examples along the way

Note on this course

- In this course we are not studying micro- or macroeconomics.
- This is often a very abstract course, but I will try to provide examples along the way
- It is key that you practice to understand the concepts, because they will reappear over and over again!
  - This is good if you nail the concept. Bad if you delay.
- Lectures will almost always be at the board (today's an exception...)!

Preliminaries

Definition (Population)

complete list of items of interest

Definition (Sample)

a subset of the population.

At the basis is something deep:

- We want to know the truth about the *population*, using data from a *sample*
- Key idea: we can't typically see the entire population, so use the sample.

Example: Major choice

- Suppose you're considering applying to different social science majors (e.g., economics, sociology, criminology, ...)
- One consideration might be earnings post graduation
  - Average earnings for graduates from different majors could inform you about this
- So suppose you'd like to learn the "true" value of average earnings for grads from different programs
  - But... the average earnings for any one person aren't the same from year to year
  - and the average earnings of any two people likely differ, even if they've graduated from the same program
  - Also, even if we could measure earnings for all prior grads, the population contains people who haven't even graduated yet, like you!
- How can you use sample earnings data from prior grads to inform your choice of program?

First, a definition...

Definition (Statistic)

a numerical summary of the data in your sample.

Example (Earnings)

How would we determine the average earnings for college graduates, by major (the population of interest)?

We might conduct a labor market survey or a census sampling some fraction of the population. Statistics Canada does this, as does the US Census Bureau, via the American Community Survey. At the end of this lecture, we will look at actual earnings data from the 2021 American Community Survey.

Tools we need in this course

Definition (Descriptive statistics)

Numbers that describe our sample.

All the data in a round of the American Community Survey constitute a sample of the population of all people in America.

We can then compute descriptive statistics summarizing various aspects of interest about our sample.

Tools we need in this course, cont.

Definition (Inferential statistics)

Link sample statistics to the population via inferences.

In our example, all the earnings for people in all Social Science majors are the population.

Tools we need in this course, cont.

Definition (Inferential statistics)

Link sample statistics to the population via inferences.

In our example, all the earnings for people in all Social Science majors are the population.
In general we seek to make two types of inferences:

1. *Use the sample to get "estimates" for population parameters (i.e., average earnings by major)*

Tools we need in this course, cont.

Definition (Inferential statistics)

Link sample statistics to the population via inferences.

In our example, all the earnings for people in all Social Science majors are the population.
In general we seek to make two types of inferences:

1. *Use the sample to get "estimates" for population parameters (i.e., average earnings by major)*
2. *Use the sample to test claims ("hypothesis testing") about the population (e.g., "economists earn more than sociologists")*

Probability

- We need *probability* because we can't know things for sure, because we're not seeing all the events we're interested in
  - some come from the future, some come by finding every last person who's graduated with a criminology degree, etc.

Probability

- We need *probability* because we can't know things for sure, because we're not seeing all the events we're interested in
  - some come from the future, some come by finding every last person who's graduated with a criminology degree, etc.
- Probability distributions are our model of the world

Probability

- We need *probability* because we can't know things for sure, because we're not seeing all the events we're interested in
  - some come from the future, some come by finding every last person who's graduated with a criminology degree, etc.
- Probability distributions are our model of the world
- We use this model of the world to infer things about the population, that is, the "truth".

Parameters

Definition (Parameter)

underlying value of object of interest in the population, i.e., the truth

Parameters

Definition (Parameter)

underlying value of object of interest in the population, i.e., the truth

The *statistic* is the sample analogue of the population parameter.

Parameters

Definition (Parameter)

underlying value of object of interest in the population, i.e., the truth

The *statistic* is the sample analogue of the population parameter. We use the sample that we observe to estimate population parameters.

- This is our best "guess" about the population parameter (average earnings by major in the population)
- Note: sample variables will be in Latin letters (e.g., a, b), population variables will be in Greek (e.g., α, β).

$\rightarrow$ BOARD

(projector off)

Relationships between variables

First, let's look at some descriptive statistics on earnings from the American Community Survey (ACS).

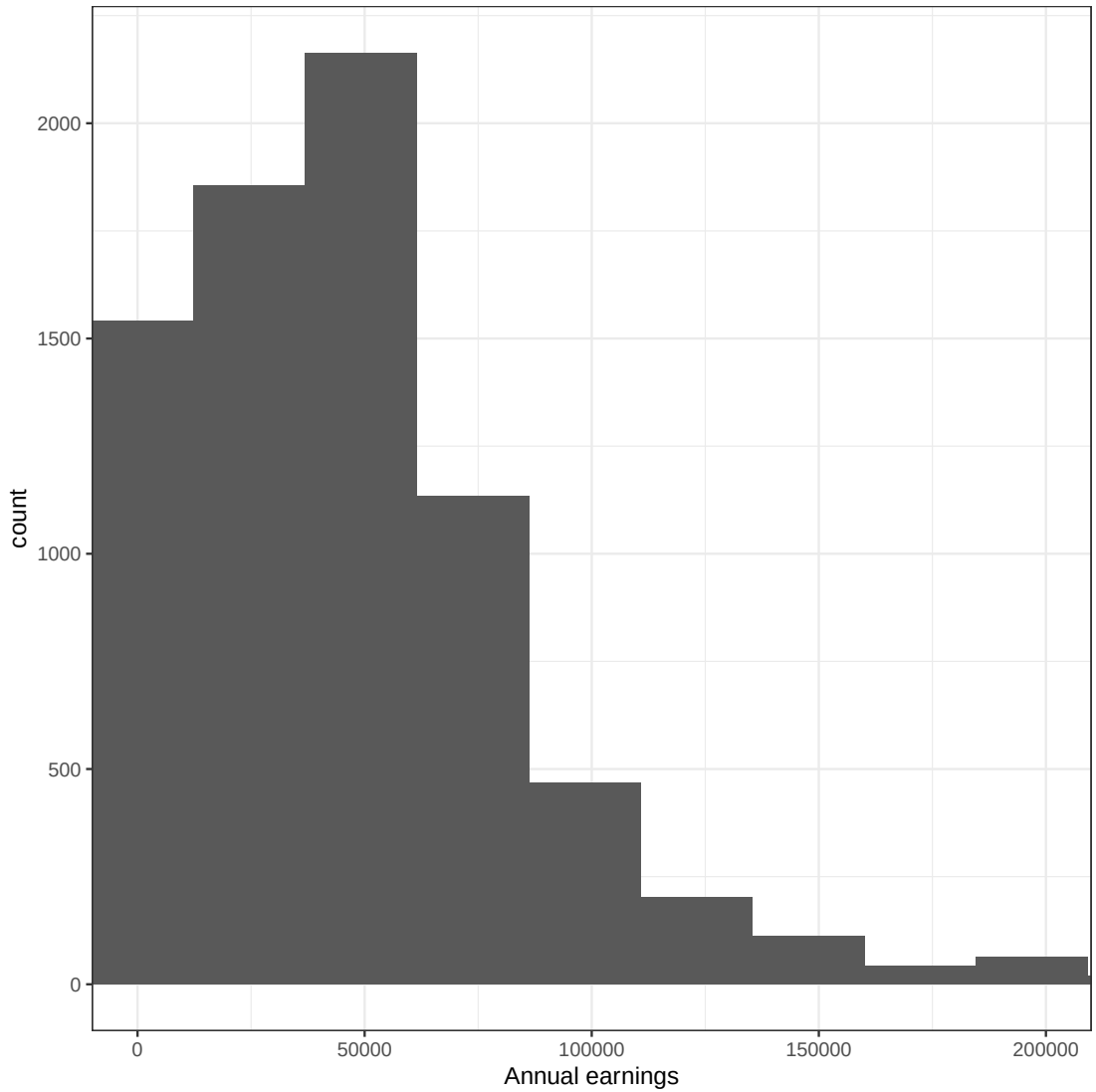
- We will focus on people who graduated with Social Science degrees sampled in the 2021 American Community Survey
- We'll restrict our sample to people aged 23-30 to get a good idea of earnings for people comparable to you post-grad

Then, we'll look at some graphs showing relationships between variables.

Means, medians, and standard deviations of earnings by Social Science major from ACS sample

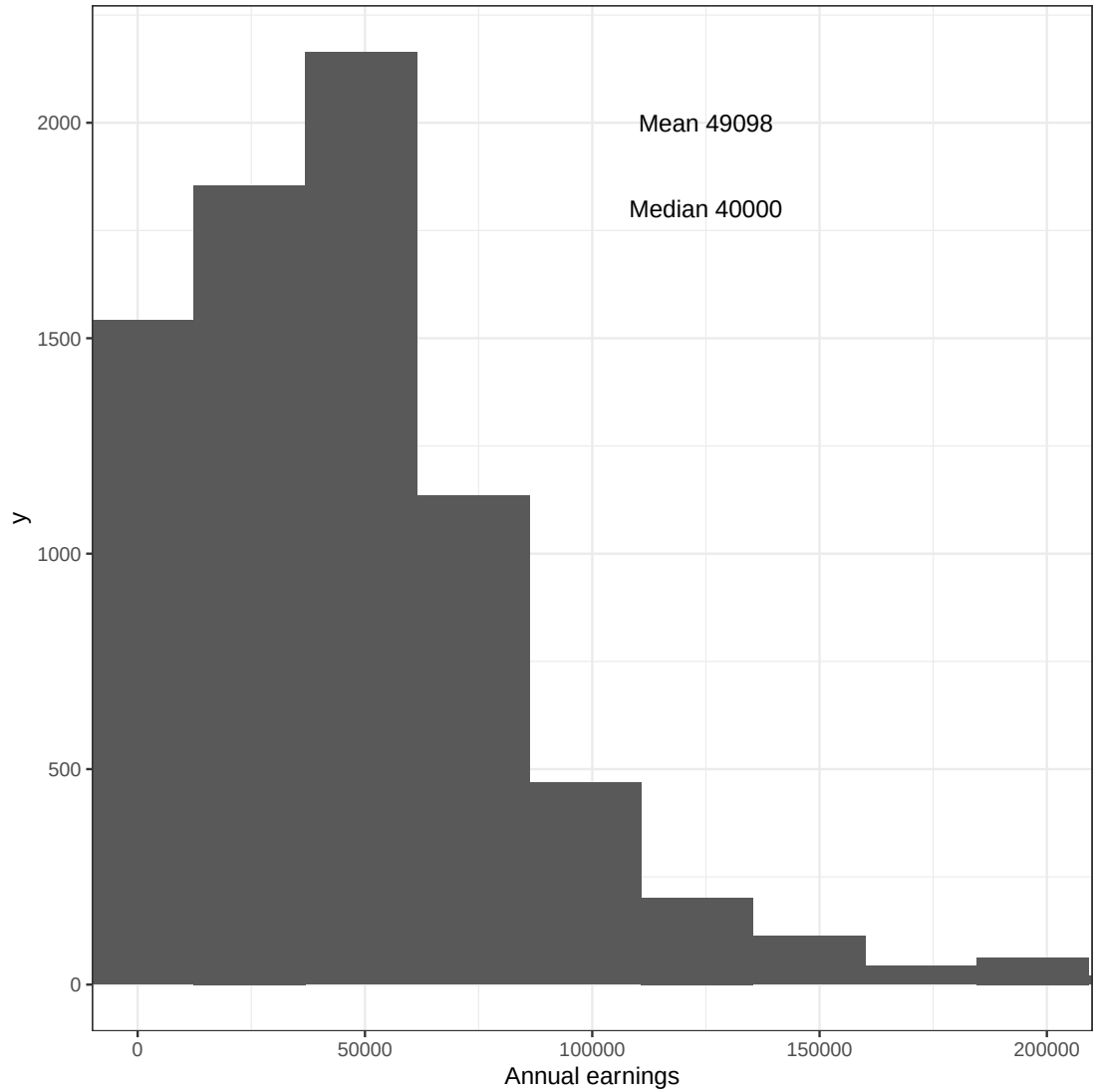
Major	Mean	Median	SD
CRIMINOLOGY	39386.76	39900.00	24387.40
ECONOMICS	61913.39	50000.00	64653.94
GENERAL SOCIAL SCIENCES	35262.59	32000.00	29486.61
GEOGRAPHY	42466.20	39000.00	40582.30
INTERDISCIPLINARY SOCIAL SCIENCES	37330.98	35000.00	28732.86
INTERNATIONAL RELATIONS	46474.55	42000.00	38131.46
MISCELLANEOUS SOCIAL SCIENCES	47128.70	44500.00	30459.25
POLITICAL SCIENCE AND GOVERNMENT	49768.97	42000.00	48149.74
SOCIOLOGY	37074.59	32500.00	38829.00

Histogram of earnings for ACS sample

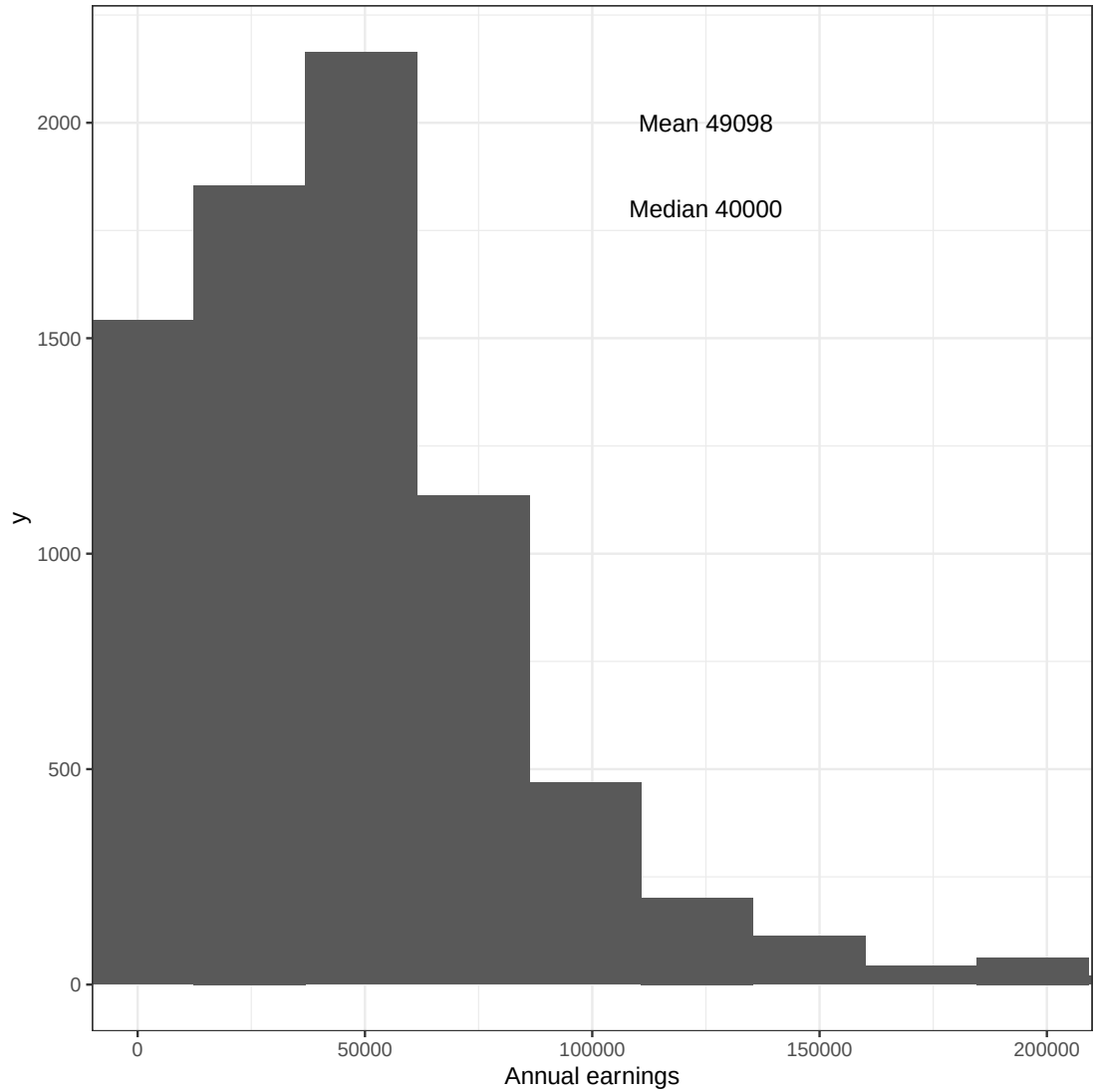


Is this skewed? Let's look at how the mean and median compare

Histogram


Mean 49098
Median 40000
2000
1500
1000
500
0
y
0
50000
100000
150000
200000
Annual earnings

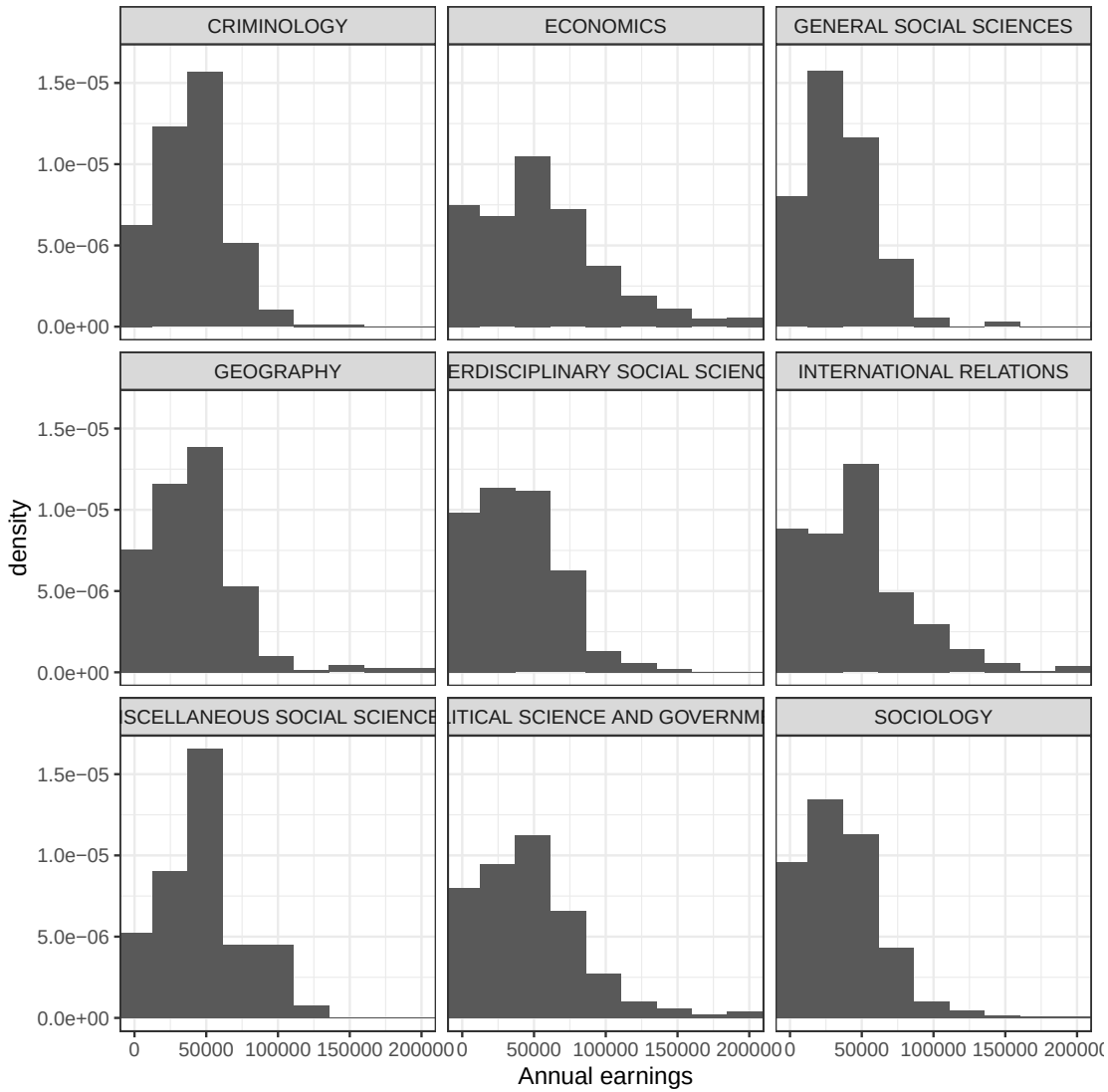
Histogram



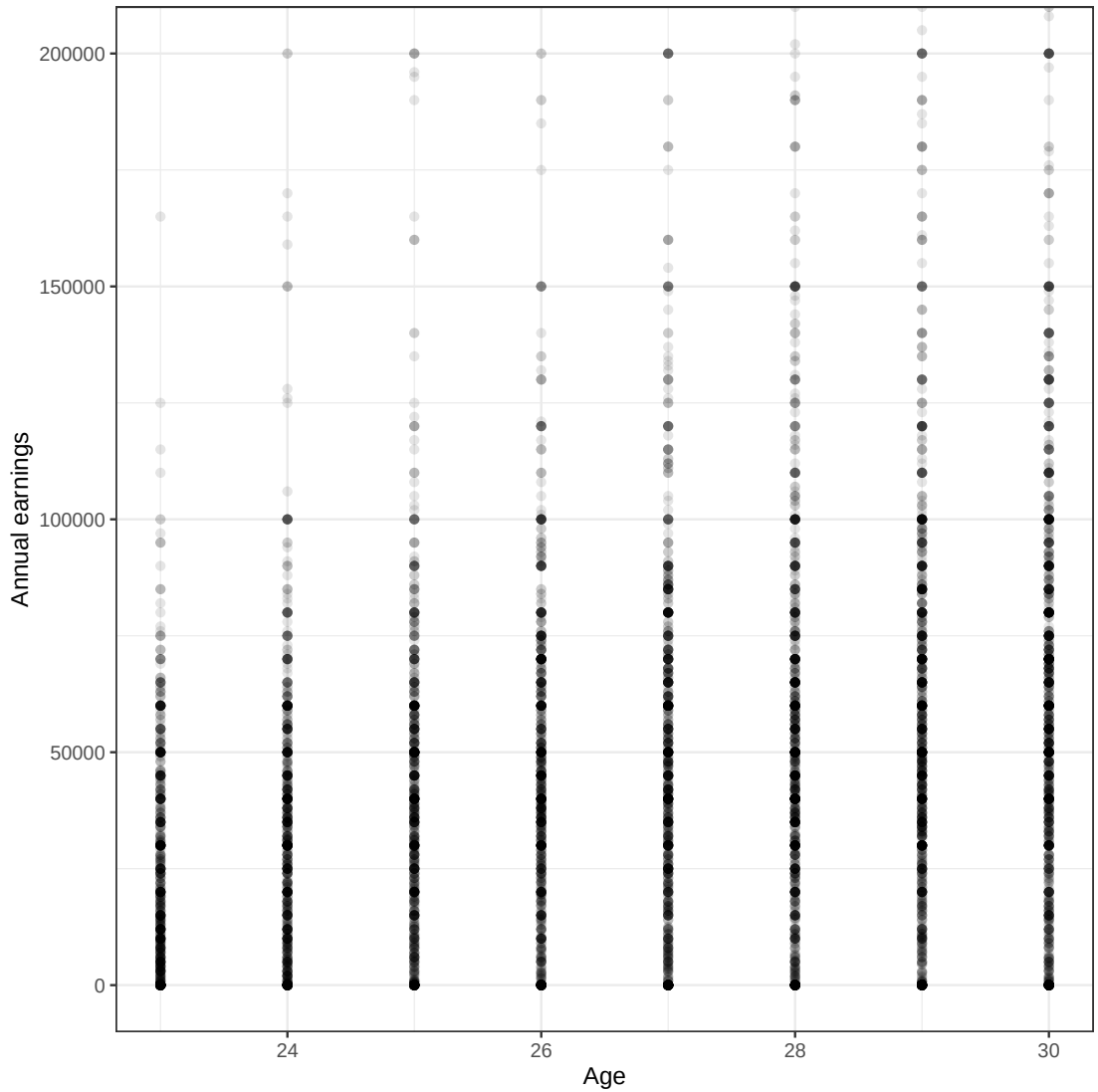
The mean is well above the median (see the annotations), so earnings are *right skewed*

Conditional histogram

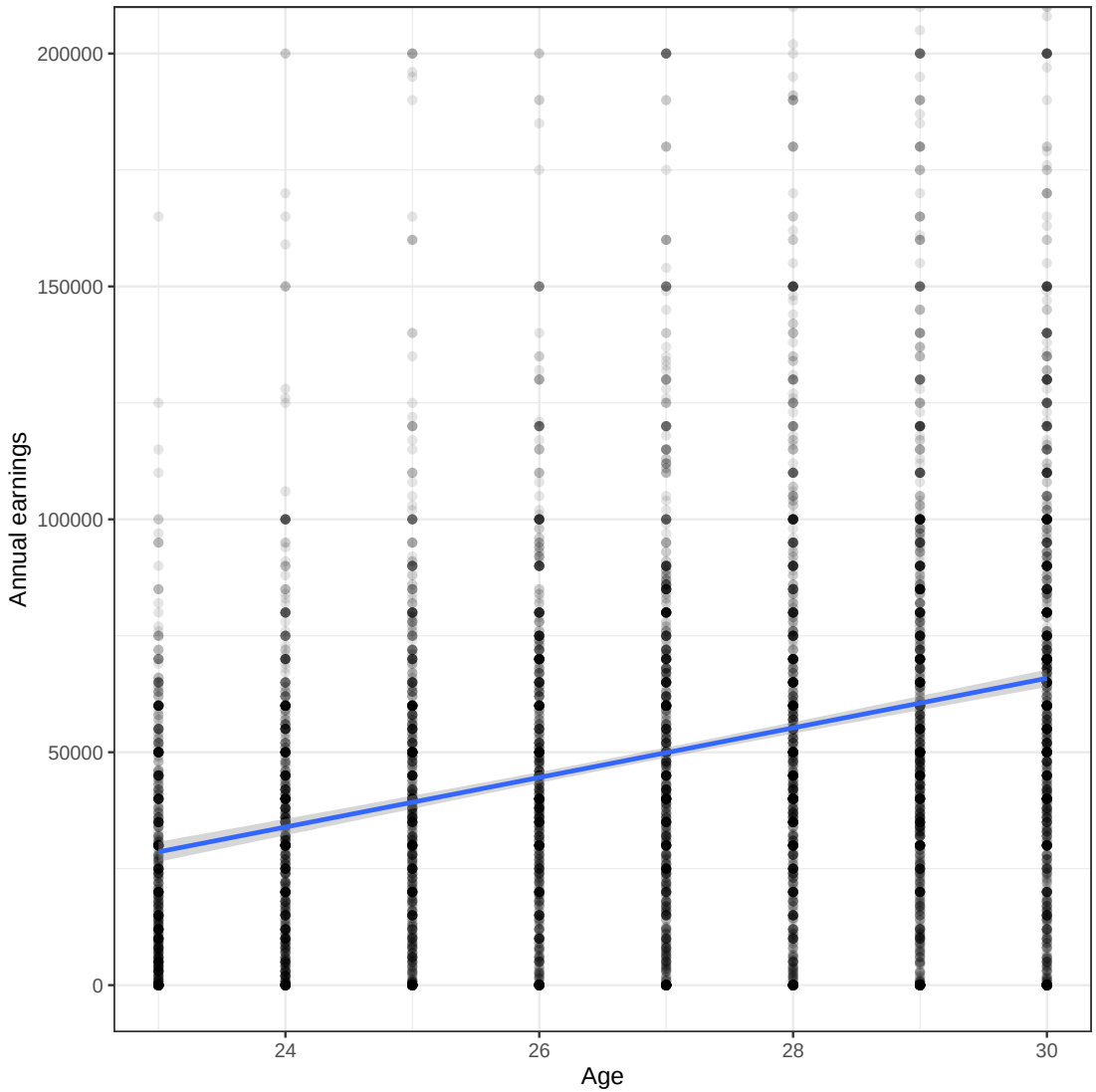
What do earnings by social science major look like?



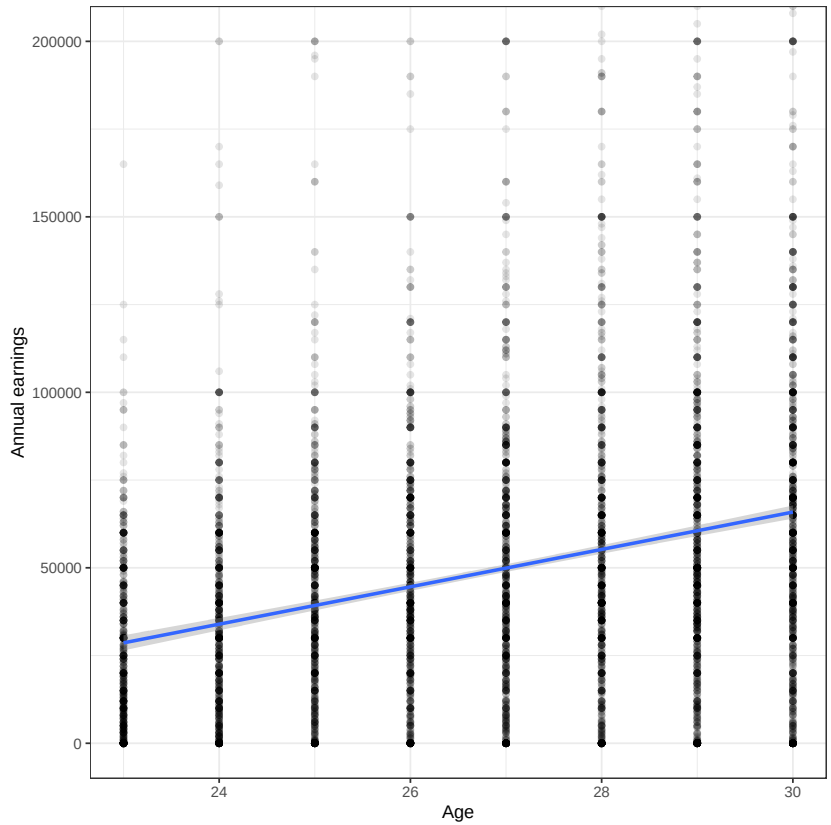
Scatterplots: Earnings by age



Scatterplots: Earnings by age + line of best fit



Line of best fit, covariance, and correlation



- Cov(earnings, age) $= s_{\text{earnings, age}} = 27{,}681$
- Cor(earnings, age) $= r_{\text{earnings, age}} = 0.24$