

YData: An Introduction to Data Science

Lecture 01: Introduction

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University
Spring 2021

Credit: data8.org





Elena Khusainova



John Lafferty



Ben Evans
**Computational
infrastructure**



Yousef Mustafa
TF



Fangming Xu
TF

Schedule and files:

<http://ydata123.org/sp21/>

Course info, grades, etc:

<https://canvas.yale.edu>

We will send out a link to a **survey** to collect data that may be used as examples in class. The responses will likely be made publicly available so that we can work with the data during class. None of the questions are required, so feel free to skip any questions you do not want to answer. Please only submit your responses once.

YData website: <http://ydata123.org>

YData Piazza site: <https://piazza.com/yale/spring2020/sds123>

What is Data Science?

Drawing useful conclusions from data using computation

- **Exploration**

- Identifying patterns in information
- Uses visualizations

- **Inference**

- Quantifying whether those patterns are reliable
- Uses randomization

- **Prediction**

- Making informed guesses
- Uses machine learning

YData Seminar Courses

- Data science is driven by applications
- Every data-driven subject brings new challenges
- YData seminars are small, independent courses taught by Yale faculty who are excited to share their expertise
- We encourage you to consider enrolling in one of the YData seminars offered this semester

Currently available YData Seminar Courses

S&DS 172: YData: **Science for Political Campaigns**

Instructor: J. Kalla

S&DS 173: YData: **Analysis of Baseball Data**

Instructor: E. Meyers

S&DS 174: YData: **Statistics in the Media**

Instructor: A. Donoghue

S&DS 175: YData: **Measuring Culture**

Instructor: D. Karell

S&DS 176: YData: **Humanities Data Mining**

Instructors: DeRose/Duhaime

S&DS 177: YData: **COVID-19 Behavioral Impacts**

Instructor: Y. Yan

Course Structure

- Three lectures per week
- Weekly homework assignments
(lowest homework grade will be dropped at end of semester)
First homework due by 11:59 PM on Thursday, January 23
- Weekly practice exercises (ungraded)
- Three projects
- Drop-in office hours (see Canvas)
- Midterm during lecture hour on Wednesday, March 4, 2020
- **Final exam scheduled on Friday, May 1 (2 PM)**

Details can be found at <https://canvas.yale.edu>

Course grade

Your overall course score will be determined as a weighted average of weekly homework (25% of final grade), projects (25%), midterm exam (20%), and final exam (30%)

A letter grade will be assigned based on:

A: 93 - 100	A-: 90 - 93	B+: 87 - 90	B: 83 - 87	B-: 80 - 83
C+: 77 - 80	C: 73 - 77	C-: 70 - 73	D: 60 - 70	F: Below 60

Honors: 90-100	High Pass: 80-90	Pass: 70-80	Fail: Below 70
-----------------------	-------------------------	--------------------	-----------------------

If adjustments are made to this grading scale, it will only be in a direction that will result in a higher letter grade.

Computational and Inferential Thinking: The Foundations of Data Science

By Ani Adhikari and John DeNero ([Adhikari and DeNero, 2018](#))

Freely available at <https://www.inferentialthinking.com>

- Ask a friend
- Ask on Piazza
Participation on Piazza through asking and answering questions is *strongly* encouraged
- Come to office hours (see Canvas for days and times)

Asking questions is encouraged

- Discuss questions with each other (except on exams)
- Submit homework individually, but discuss with others (don't share written solutions or code)
- Submit projects individually or with one partner (only undergraduates may work with a partner)

The limits of collaboration

- Don't share solutions with each other (except project partners)
- Copying or other dishonesty will result in failing the course

Data science can be used for investigating many different types of questions.

For example, **do left-handers die younger than right-handers?**

Do left-handers die younger than right-handers?

Psychological Bulletin
1991, Vol. 109, No. 1, 90-106

Copyright 1991 by the American Psychological Association, Inc.
0033-2909/91/\$3.00

Left-Handedness: A Marker for Decreased Survival Fitness

Stanley Coren
University of British Columbia
Vancouver, British Columbia, Canada

Diane F. Halpern
California State University, San Bernardino

Life span studies have shown that the population percentage of left-handers diminishes steadily, so that they are drastically underrepresented in the oldest age groups. Data are reviewed that indicate that this population trend is due to the reduced longevity of left-handers. Some of the elevated risk for sinistrals is apparently due to environmental factors that elevate their accident susceptibility. Further evidence suggests that left-handedness may be a marker for birth stress related neuropathy, developmental delays and irregularities, and deficiencies in the immune system due to the intrauterine hormonal environment. Some statistical and physiological factors that may cause left-handedness to be selectively associated with earlier mortality are also presented.

***According to a study from 1991, the answer is “Yes”**

Do left-handers die younger than right-handers?

Psychological Bulletin
1991, Vol. 109, No. 1, 90-106

Copyright 1991 by the American Psychological Association, Inc.
0033-2909/91/\$3.00

Left-Handedness: A Marker for Decreased Survival Fitness

Stanley Coren
University of British Columbia
Vancouver, British Columbia, Canada

Diane F. Halpern
California State University, San Bernardino

Life span studies have shown that the population percentage of left-handers diminishes steadily, so that they are drastically underrepresented in the oldest age groups. Data are reviewed that indicate that this population trend is due to the reduced longevity of left-handers. Some of the elevated risk for sinistrals is apparently due to environmental factors that elevate their accident susceptibility. Further evidence suggests that left-handedness may be a marker for birth stress related neuropathy, developmental delays and irregularities, and deficiencies in the immune system due to the intrauterine hormonal environment. Some statistical and physiological factors that may cause left-handedness to be selectively associated with earlier mortality are also presented.

***According to a study from 1991, the answer is “Yes”**

- Average age of death for left-handers: 66

Do left-handers die younger than right-handers?

Psychological Bulletin
1991, Vol. 109, No. 1, 90-106

Copyright 1991 by the American Psychological Association, Inc.
0033-2909/91/\$3.00

Left-Handedness: A Marker for Decreased Survival Fitness

Stanley Coren
University of British Columbia
Vancouver, British Columbia, Canada

Diane F. Halpern
California State University, San Bernardino

Life span studies have shown that the population percentage of left-handers diminishes steadily, so that they are drastically underrepresented in the oldest age groups. Data are reviewed that indicate that this population trend is due to the reduced longevity of left-handers. Some of the elevated risk for sinistrals is apparently due to environmental factors that elevate their accident susceptibility. Further evidence suggests that left-handedness may be a marker for birth stress related neuropathy, developmental delays and irregularities, and deficiencies in the immune system due to the intrauterine hormonal environment. Some statistical and physiological factors that may cause left-handedness to be selectively associated with earlier mortality are also presented.

***According to a study from 1991, the answer is “Yes”**

- Average age of death for left-handers: 66
- Average age of death for right-handers: 75

*Example found in “Seeing Through Statistics” by J. Utts

Think & Share

The results of the study addressing the question “Do left-handers die younger than right-handers?” suggest

- A** Yes, left-handed people do die younger than right-handed people. Period.
- B** Yes, the study suggests that left-handed people do die younger than right-handed people, but the study should be repeated in exactly the same manner to see if the same results appear.
- C** No, there is no evidence whatsoever that left-handed people die younger than right-handed people.
- D** I need to know more information about how the study was performed before I can decide.

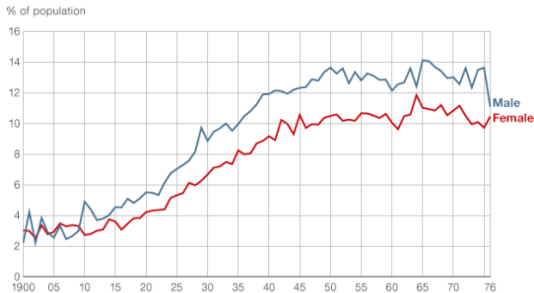
Do left-handers die younger than right-handers?

- Left-handers: 66 vs. Right-handers: 75
- Flaws in the study?

Do left-handers die younger than right-handers?

- Left-handers: 66 vs. Right-handers: 75
- Flaws in the study?
- Did not take into account that in the early part of the 20th century, many children were forced to write with their right hands reducing the population of naturally left-handed people in the older group

Left handedness 1900-1976



Source: Chris McManus Right Hand, Left Hand

Do left-handers die younger than right-handers?

- Left-handers: 66 vs. Right-handers: 75
- Flaws in the study?
- Did not take into account that in the early part of the 20th century, many children were forced to write with their right hands reducing the population of naturally left-handed people in the older group
- Ideas for improvement?

Do left-handers die younger than right-handers?

- Left-handers: 66 vs. Right-handers: 75
- Flaws in the study?
- Did not take into account that in the early part of the 20th century, many children were forced to write with their right hands reducing the population of naturally left-handed people in the older group
- Ideas for improvement?
- A prospective study - follow group of current left and right handers until death

Do left-handers die younger than right-handers?

Where the data come from matters.

Example

(DEMO)

Adhikari, A. and DeNero, J. (2018), “Computational and Inferential Thinking,” Gitbook.