EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020
RAZ: Rebecca A. Zárate, MA

Overview

- Course Goals
- Statistics – What the Point?
  - The Purpose of Statistics
  - Limitations
- Think About It, Examples
- Extraordinary Claims
  - Poor Research
    - Andrew Wakefield
- Communication: Science and Society
  - Science and Society: Pop-Science
  - Open Access
- Data
- Statistical Software R

Course Goals

- Understand the two main classes of statistics: descriptive and inferential
- Be able to evaluate the quality of a research article, news story, headlines, or statistical finding
  - How was the study conducted?
  - How many participants?
  - How much of an effect does this have?
- Gain experience calculating statistics and running code in the statistical software R

Statistics – What's the Point?

STATS! - What is it good for!?

…absolutely everything*...

What's the point of statistics?

What do we use statistics for?

What good is it to provide the average value of something (like test scores or the average amount of lead in US water)?

Why do we care about the probability of something happening or not?

*with the understanding that statistics can be misrepresented for nefarious reasons or ignorance

Uses of Statistics

- Knowing the average…
  - … test score can give me an idea about how well the students are learning OR it can let me know how good or bad I'm doing as a teacher
  - … amount of lead in water gives us an idea of what is expected and more importantly what might be too high
- And probability is going to help us determine how likely or unlikely it is to see, for example, a city with lead levels 5x the national average

What's the point?

- Descriptive statistics provide helpful summary describing something in one number (but we lose some information by reducing like this)
  - Examples: Passer or batter ratings, GPA score
- Inferential statistics using sample data collected in an experiment or survey to make inferences about the population
  - Examples: How many homeless are there in the US? How many households lack internet service in the US?

What's the point?

- Assessing the Probability of Events can help reduce (or prepare for) risky situations
  - Example: What is the probability your new iPhone breaks in the next two years? Should you purchase the extended warranty?
- Identifying Important Relationships where it may be unethical to empirically test a hypothesis
  - Example: Does smoking cigarettes cause cancer– while controlling for other confounding variables?

But what about…

"Lies,
damned lies,
and statistics!"
– Mark Twain

Au Contraire, Mr. Twain…

Statistics are not lies nor do they lie.

Only the people who use statistics inappropriately or for their own agendas lie.

It's up to YOU, the budding bright mind that you are, to stay vigilant, ask question, look out for possible agendas… This will serve you well.

But Statistics Do Have Limitations

- Data and the statistics we decide to run on data help us understand but statistics…
  - Do not tell you the "correct answer"
  - Do not live in isolation (there is context to take into account)
  - Do not select and run themselves
  - Do not flush out every possible explanation
  - Can be misrepresented in a way to push an agenda
  - Are not interpreted the same way by everyone
  - Are only as good as the data you put in
    - Garbage in, Garbage Out

Garbage In, Garbage Out

"On two occasions I have been asked, "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" ... I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question."

— Charles Babbage,

Passages from the Life of a Philosopher

Think About It…

Think about it.

- Two students sitting side by side during a 50 question, multiple-choice test both got 5 questions wrong.
  - Were they cheating?
- These two students got the same 5 questions wrong.
  - Were they cheating?
- These two students chose the same wrong options for the same 5 questions.
  - Were they cheating?

What are the chances of each happening assuming they were NOT cheating?

Extraordinary Claims

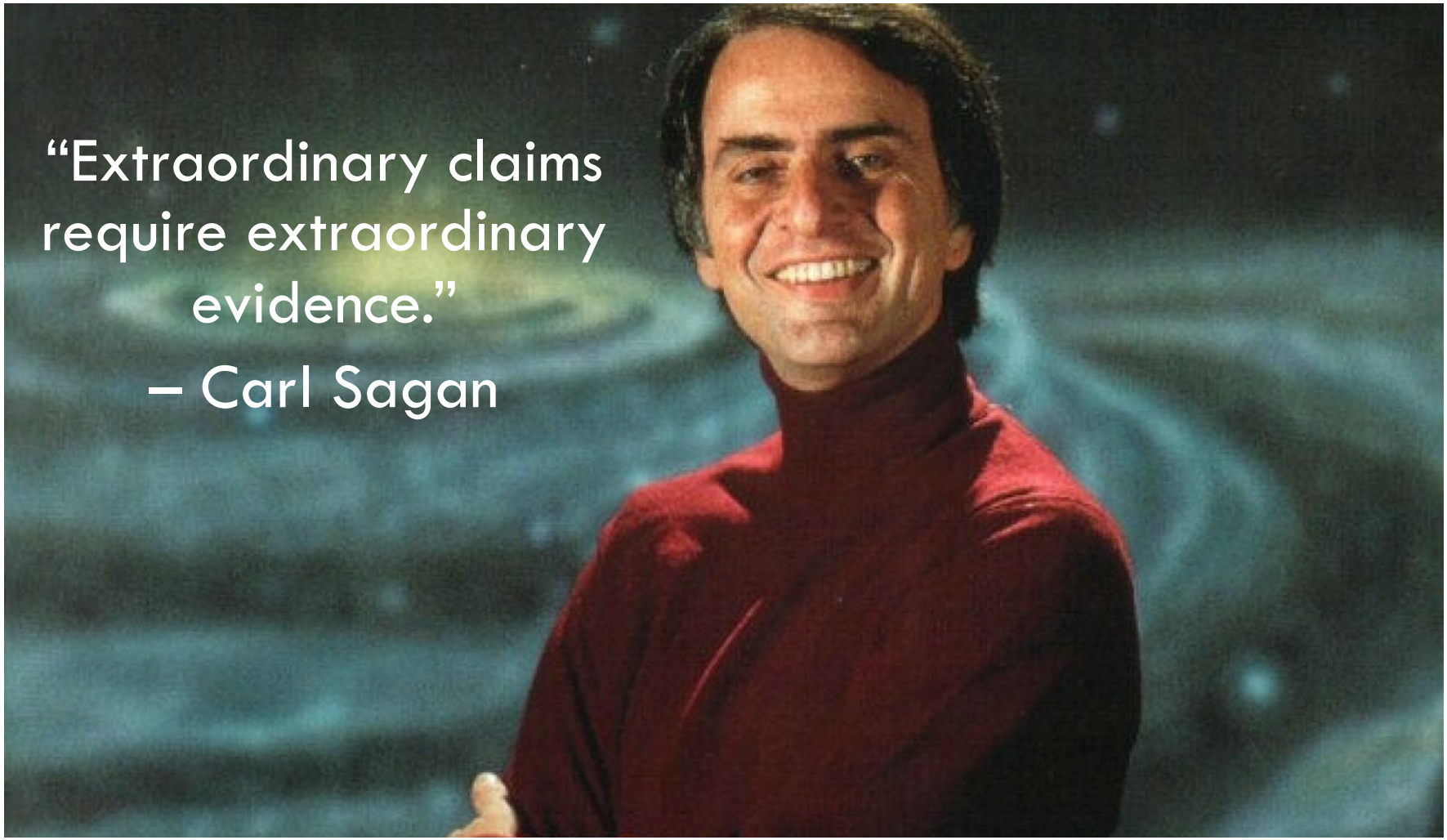
- "Amazing-New-Natural-Deodorant has over 9,000 5-Star reviews!"
  - Is it the best deodorant ever?!
  - What would be a more accurate thing to report?
  - Why might someone misrepresent data?
  - Does everyone love it?
    - Why or why not?
    - How do you know?
      - Do you know?

Extraordinary Claims

- "A new study finds that exposure to the household cleaning product, "Chemical-Sounding-Cleaner", increases your risk of a very specific type of cancer by 100x!"
  - Should you never buy that cleaner again?
  - What do you need to ask to better understand this claim?

Chemical-Sounding-Cleaner

Advice from the Great Carl Sagan



Extraordinary Claims

Extraordinary Claims

- Many times the media will publish what sounds like an extraordinary claim as fact
  - Ex. "New study shows drinking this soda *causes* cancer"
- In reality, many times the researchers and the study itself make no direct claims that something is the direct *cause* for something else.
  - Causation is difficult to establish
- Why then would the media (or even well respected research journals) publish such a claim?
  - Perhaps they have their own agenda?
  - Click bait, views, monetary gain?

Questions to Ask

- When you read a news or research article, you should always read it critically and ask some of the following questions:
  - Who funded it?
    - Do they have a vested interest in the project?
  - Who conducted it?
  - Sample size?
  - Research design?
  - Statistical analysis?
  - Data available?

Poor Research

- Andrew Wakefield…
  - Ever heard of him?
- Example of how one poorly conducted study can ripple out and affect society
- Published: 1998
- Retracted: 2010

Wakefield, 1998

"Illeal-lymphoid hyperplasia, non-specific colitis and pervasive developmental disorders in children"

Yes… This is the paper that started the MMR Vaccine - Autism Controversy.

In 1998, Andrew Wakefield along with 12 other authors published their article in the Lancet (the Vogue of journals).

In 2010 (12 years later!!), it was retracted.

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associa by the parents, with measles, mumps, and rub a vaccination in eight of the 12 children, with measl infection in one child, and otitis media in a . All 1 children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to a noid ul ration. Histology showed patchy chronic inflam tion i in 11 children and reactive ileal lymphoi perplasia in seven, but no granulomas. Be vioural disor s included autism (nine), disintegrativ sis (one), a possible postviral or vaccinal encephalitis o). There were no focal neurological ab malities and MRI and EEG tests were normal. Abno al laboratory results were significantly raised urinary ethylmal c acid compared with age-matched control (p=0.003), low haemoglobin in four children, a low s m IgA in four children.

Interp tation We ident associated gastrointestinal dis se and developmental regression in a group of previ ly normal chi n, which was generally associated in time with possible environmental triggers.

Lancet 1998; **351:** 637–41
See Commentary page

Inflammatory Bowel Disease Study Group, University Departments of Medicine and Histopathology (A J Wakefield FRCS, A Anthony MB, J Linnell PhD, A P Dhillon MRCPath, S E Davies MRCPath) **and the University Departments of Paediatric Gastroenterology** (S H Murch MB, D M Casson MRCP, M Malik MRCP, M A Thomson FRCP, J A Walker-Smith FRCP,), **Child and Adolescent Psychiatry** (M Berelowitz FRCPsych), **Neurology** (P Harvey FRCP), **and Radiology** (A Valentine FRCR), **Royal Free Hospital and School of Medicine, London NW3 2QG, UK**

Correspondence to: Dr A J Wakefield

Introduction

We saw several children who, after a p iod of apparent normality, lost acquired skills, includ g com unication. They all had gastrointestinal m ptoms, i luding abdominal pain, diarrhoea, and b ating and, in some cases, food intolerance. We d scribe the clinical fi dings, and gastrointestinal featur of these chi en.

Patients and meth ds

12 children, cons cutively r red to the department of paediatric gastr enterology with a hi y of a pervasive developmental d der with loss of acquired skills and intestinal symptoms (diarrh , abdominal pain, bloating and food intolerance), were inv gated. All children were admitted to the ward for 1 week, accomp ied by their parents.

Clinical investigations

We took histori including details of immunisations and exposure to infect us diseases, and assessed the children. In 11 cas the history was obtained by the senior clinician (JW-S). Neur logical and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental assessments included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known causes of childhood neurodegenerative disease. Urinary methylmalonic acid was measured in random urine samples from eight of the 12 children and 14 age-matched and sex-matched normal controls, by a modification of a technique described previously.[2] Chromatograms were scanned digitally on computer, to analyse the methylmalonic-acid zones from cases and controls. Urinary methylmalonic-acid concentrations in patients and controls were compared by a two-sample *t* test. Urinary creatinine was estimated by routine spectrophotometric assay.

Children were screened for antiendomyseal antibodies and boys were screened for fragile-X if this had not been done

RETRACTED

Wakefield, 1998

"Onset of behavioural symptoms was associated, by the parents, with measles, mumps, and rubella vaccination in 8 of the 12 children…

We identified associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers."

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associated by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities, ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation in 11 children and reactive ileal lymphonodular hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls ($p=0.003$), low haemoglobin in four children, and low serum IgA in four children.

Interpretation We identified associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

Lancet 1998; **351:** 637–41
See Commentary page

Inflammatory Bowel Disease Study Group, University Departments of Medicine and Histopathology (A J Wakefield FRCS, A Anthony MB, J Linnell PhD, A P Dhillon MRCPath, S E Davies MRCPath) **and the University Departments of Paediatric Gastroenterology** (S H Murch MB, D M Casson MRCP, M Malik MRCP, M A Thomson FRCP, J A Walker-Smith FRCP,), **Child and Adolescent Psychiatry** (M Berelowitz FRCPsych), **Neurology** (P Harvey FRCP), **and Radiology** (A Valentine FRCR), **Royal Free Hospital and School of Medicine, London NW3 2QG, UK**

Correspondence to: Dr A J Wakefield

Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and bloating and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features of these children.

Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (diarrhoea, abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for 1 week, accompanied by their parents.

Clinical investigations

We took histories including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental assessment included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known causes of childhood neurodegenerative disease. Urinary methylmalonic acid was measured in random urine samples from eight of the 12 children and 14 age-matched and sex-matched normal controls, by a modification of a technique described previously.[2] Chromatograms were scanned digitally on computer, to analyse the methylmalonic-acid zones from cases and controls. Urinary methylmalonic-acid concentrations in patients and controls were compared by a two-sample t test. Urinary creatinine was estimated by routine spectrophotometric assay.

Children were screened for antiendomyseal antibodies and boys were screened for fragile-X if this had not been done

RETRACTED

Andrew Wakefield, MD

- Financial conflicts of interest
  - Had a different vaccine patent for a "safer" vaccine
  - Was in cahoots with lawyers looking to sue for damages
- Professional and Research Misconduct
  - Used his own patients and forewent IRB protocols
  - Subjected patients to potentially unnecessary procedures, such as colon biopsies and lumbar punctures
- Poor Science to Begin With
  - Only 12 subjects; there were more authors!
  - Of those 12, their diagnoses (especially the regressive ASD) are highly questionable
  - Correlation does NOT equal causation

Communication: Science and Society

Communication: Science and Society

- One of the understated but crucially important part of statistics is the effective communication of those statistics
  - You could be the most brilliant biostatistician but if you cannot communicate your findings to your peers or the public, your work may have no real impact…
- But communication of statistics needn't fall only on the people doing the statistics. For example, journalism departments have started to realize the importance of understanding and communicating the numbers as well as creating engaging visualizations to help get the message across.

Communication: Science and Society

- "New cure for autism discovered!"
- "Mask do not help mitigate the spread of COVID-19."
- "Mask DO help mitigate the spread of COVID-19."

Pros and cons of pop science headlines?

News Headline vs. Actual Study

This is the Science News Headline

Science News *from research organizations*

COVID-19 causes 'hyperactivity' in blood-clotting cells

Date: June 30, 2020

Source: University of Utah Health

Summary: Changes in blood platelets triggered by COVID-19 could contribute to the onset of heart attacks, strokes, and other serious complications in some patients who have the disease, according to scientists. The researchers found that inflammatory proteins produced during infection significantly alter the function of platelets, making them 'hyperactive' and more prone to form dangerous and potentially deadly blood clots.

JUNE 23, 2020

Platelet Gene Expression and Function in COVID-19 Patients

Bhanu Kanth Manne, Frederik Denorme, Elizabeth A Middleton, Irina Portier, Jesse W Rowley, Chris J. Stubben, Aaron C. Petrey, Neal D. Tolley, Li Guo, Mark J. Cody, Andrew S. Weyrich, Christian Con Yost, Matthew T. Rondina, Robert A. Campbell

Blood blood.2020007214.

https://doi.org/10.1182/blood.2020007214 Article history

"These findings demonstrate that SARS-CoV-2 infection is associated with platelet hyperreactivity which may contribute to COVID-19 pathophysiology."

And this is what the paper actually said.

News Headline vs. Actual Study

- COVID may indeed be "associated with platelet hyperreactivity" and may indeed *cause* some of these changes… but more data is needed…
  - Sometimes there is some evidence for a finding, but many times it is limited. More studies with more participants usually need to be done.
- This does NOT completely discredit findings, but we must remember to take them with a grain of salt and always be willing to update our current understanding when new information and data become available
  - This is the essence of being a good researcher and a statistically literate member of society

Open it Up!

Open Access Explained

This is the first video to watch for your "Watch and React" assignments.

Data

What is Data?

What do we mean when we say "data"?

"We looked at the data and found x, y, z."

"The data show a, b, c."

"The data were collected through a randomized controlled design."

"Data is the merely the raw material of knowledge." - NYT

Data are…

- The measurements you've taken
- The responses you've gathered
- The value of something you measure
- The value of a variable

Etymology Moment!

- "The Latin word *data* is the plural of *datum*, "(thing) given," neuter past participle of *dare* "to give"."
  - Something that is "given", like a given fact

Next Up

- Before we dive into statistics we first need understand the different types of variables, how these variables are measured, and how variables can be used in research.

Variables

Statistical Software R

By Hand vs Statistical Software (R)

- People who do statistic *never* calculate anything by hand, though it is useful to do a few examples by hand so see how certain equations work.
  - In this class you will have some quizzes that will have you calculate some things. You have the option to do them by hand or use R. I will provide you with the ways to do it both ways.
- I *strongly* encourage you to use R whenever possible, practice helps.
  - You will have a group project at the end of the semester that has to be done in R…

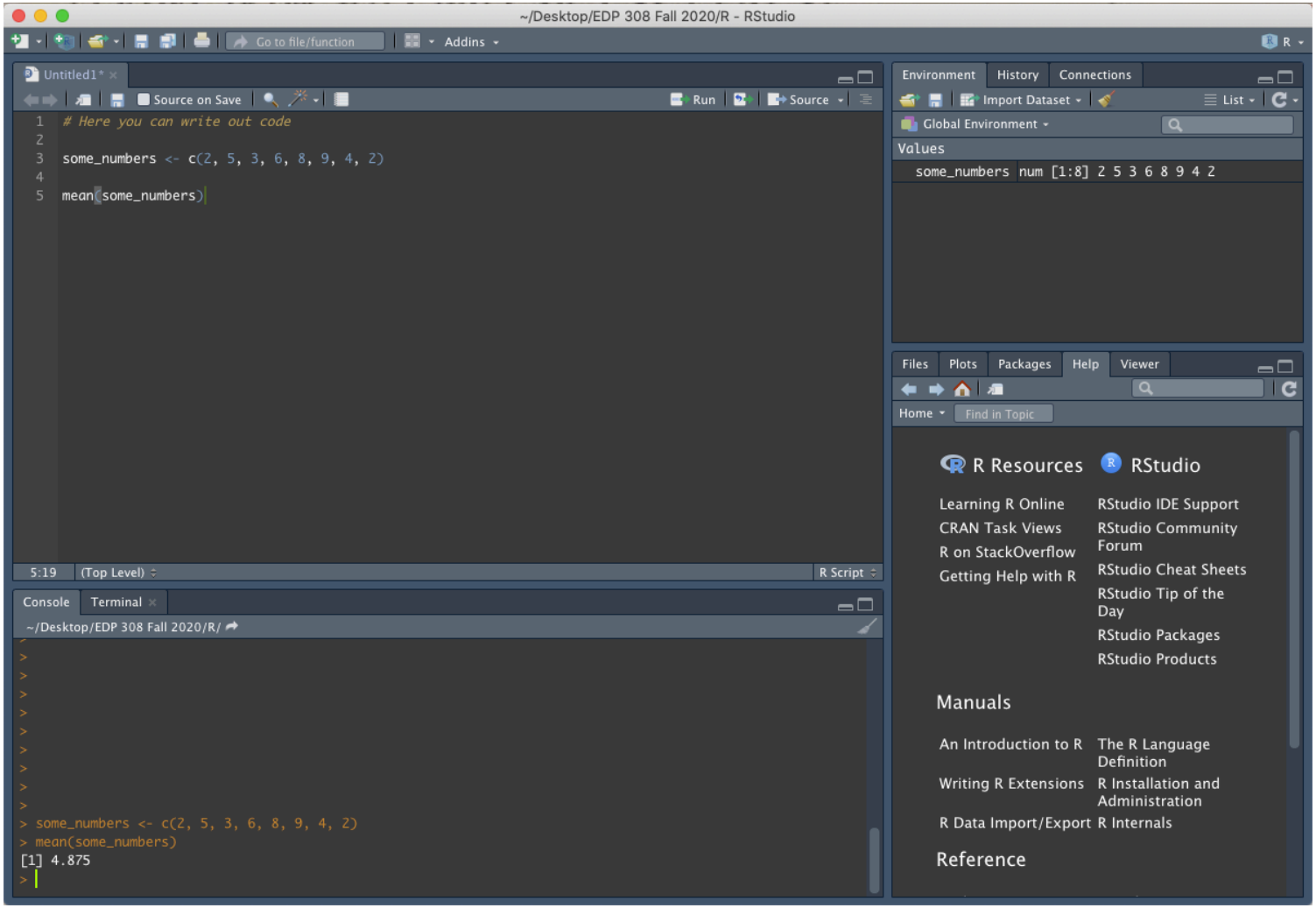
What is R?

- For this class, you will download the free software, R
- R is basically a giant, fancy calculator
- R is a completely free open sourced statistics software
  - Another great, free open sourced software is Python
  - Other statistics software include: SPSS, SAS, M-Plus, MATLAB and few others. However, these are propriety, meaning that cost money to purchase and/or use
    - Excel can do some statistical analysis but it's limited and propriety
- One aspect some people find intimidating about R when they start is that you have to do some coding, but it is this coding that allows you to have so much control

What is R Studio?

- You will also download R Studio
- R Studio is a nice environment to use R in that keeps things tidy
  - It's technically called a "integrated development environment" (IDE)
- It is much easier to use R inside of R Studio

R Studio


~/Desktop/EDP 308 Fall 2020/R - RStudio
Go to file/function
Addins
R
Untitled1*
Source on Save
Run
Source
1 # Here you can write out code
2
3 some_numbers <- c(2, 5, 3, 6, 8, 9, 4, 2)
4
5 mean(some_numbers)
5:19
(Top Level)
R Script
Console
Terminal
~/Desktop/EDP 308 Fall 2020/R/
> some_numbers <- c(2, 5, 3, 6, 8, 9, 4, 2)
> mean(some_numbers)
[1] 4.875
Environment
History
Connections
Import Dataset
List
Global Environment
Values
some_numbers num [1:8] 2 5 3 6 8 9 4 2
Files
Plots
Packages
Help
Viewer
Home
Find in Topic
R Resources
RStudio
Learning R Online
CRAN Task Views
R on StackOverflow
Getting Help with R
RStudio IDE Support
RStudio Community Forum
RStudio Cheat Sheets
RStudio Tip of the Day
RStudio Packages
RStudio Products
Manuals
An Introduction to R
Writing R Extensions
R Data Import/Export
The R Language Definition
R Installation and Administration
R Internals
Reference

Finding Help in R (Debugging)

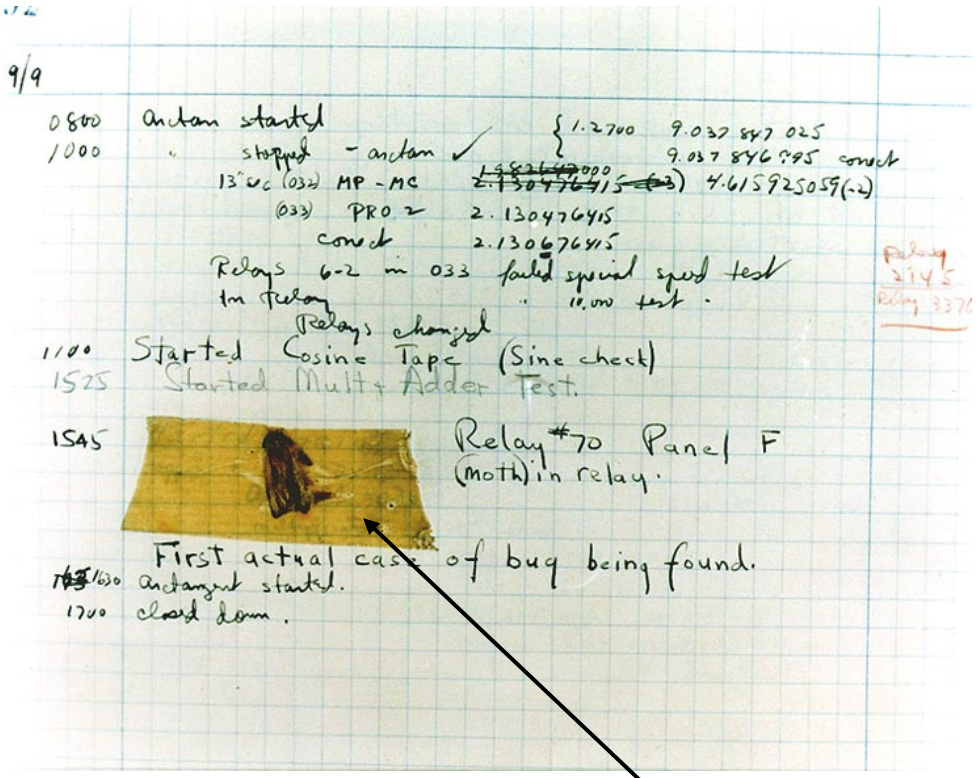
- Sometimes things don't work as they should. Sometimes R will throw back an error that might be cryptic and confusing. An invaluable skill in coding (and much of life, really) is being able to Google proficiently.
- Whenever R gives you an error, you could try copy pasting it into Google. Most likely someone else has had the same problem and asked for help on a forum. This can help you find the solution.
  - A very useful website that I use all the times:
    - https://stackoverflow.com/
  - Working together will also help you catch typos.
  - **And, of course, I am here to help you, too.**

Historical Moment: Grace Hopper

- **Grace Brewster Murray Hopper** (Dec. 9, 1906 – Jan. 1, 1992)
- American computer scientist and United States Navy Commodore (renamed to Rear Admiral, lower half in 1985)
- She was turned away from enlisting multiple times for various reasons such as being to old (34 years at the time), to small, and because her math skills were needed elsewhere. She enlisted in the Naval Reserves.

Historical Moment: Grace Hopper

- **Grace Brewster Murray Hopper** (Dec. 9, 1906 – Jan. 1, 1992)
- Fun Fact:
  - We "debug" code when it doesn't run or has errors
  - The reason it is called "debugging" is because an actual insect, a moth, found it's way into the Mark II (a computer at Harvard Hopper was working on) and caused some trouble.
  - Legend has it that Hopper stated they were "debugging" the system.



Why are you making us do this?!

- Maybe some of you think, "Oh, that's not for me. I don't do math or coding..." Says who?
  - Even if you never use R again in your life, I want to give you a small taste and the experience of doing real data analysis. By the end of this, you will have typed out and run some code all by yourself. And that is something to be very proud of. You will be ahead of the game and have done something not many people do (and some actively avoid).
    - I personally wish I would have been introduced to R in undergrad. I would have been so much further along. If you are thinking of graduate school, R can be an invaluable skill…
  - And if ever, for some other class or job, you need to analyze some data, you'll be able to fall back on some of the skills you learn here, like writing code, creating a visualization and interpreting the results.
- You CAN do this. I am going to help you throughout.