BIOS6606: LectureAug26 Statistics For Basic Sciences Introduction

Kathleen Torkko, PhD, MSPH, MS
Associate Professor
Department of Pathology
Department of Biostatistics and Bioinformatics
Course Director, Main Lecturer

Image: www.dailymotion.com/video/x5i72x4_introduction_fun

Section on Statistical Education

TL04

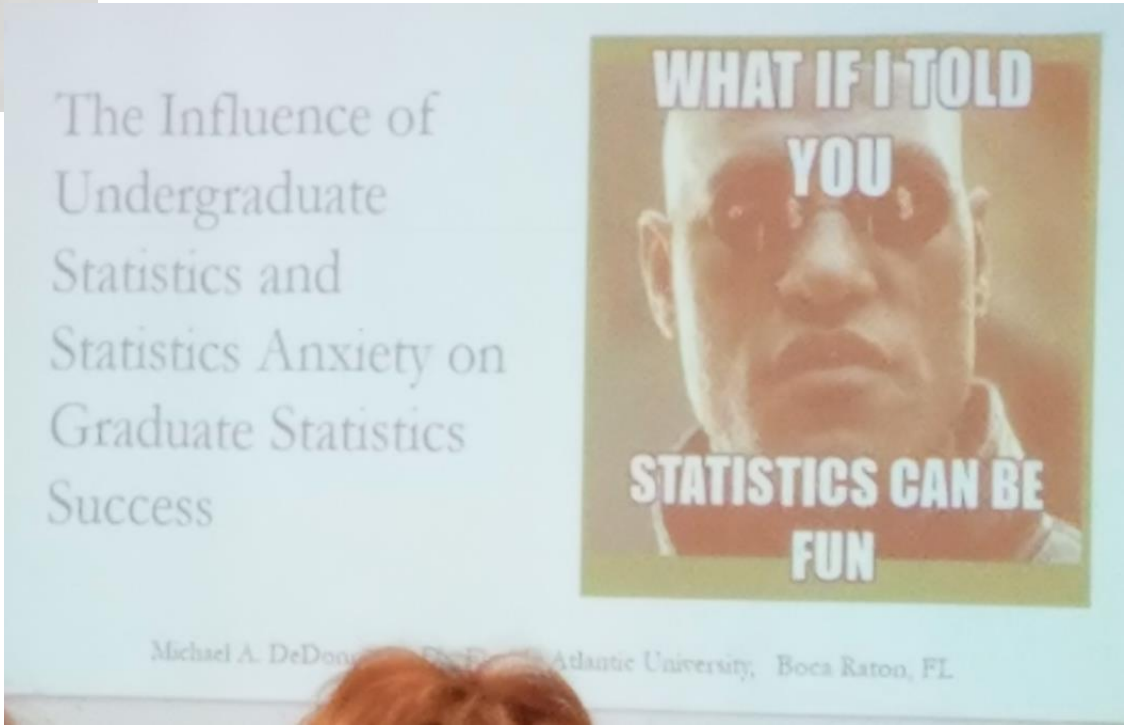
Why Do Students Hate Statistics?

Michael DeDonno, Florida Atlantic University

We will consider the behavioral and cognitive factors that seem to affect a student's desire and interest to take statistics courses.

SOLD OUT

Joint Statistical Meeting
Baltimore, August 2017



Goals of this Course

- The goal is not to turn you into statisticians, but into basic scientists who appreciate the need for good statistical practices
    - You should feel comfortable performing some statistical analyses on your own, but also realize when it is necessary to seek professional statistical help
- Specific goals/aims will be given with each lecture
- A main goal is to get you to *think* differently – like a statistician

    A data analysis is much more than putting data into a statistical program and pushing enter to get a p-value

Aims of this Course

- Give you a good foundation in classical basic statistics
- Use basic research data whenever possible
- Include time for practice exercises and discussion
- Study the practice and application of statistics within the context of reducing bias
    - Doing the best (most honest; least biased) analysis that fits the data and the research question

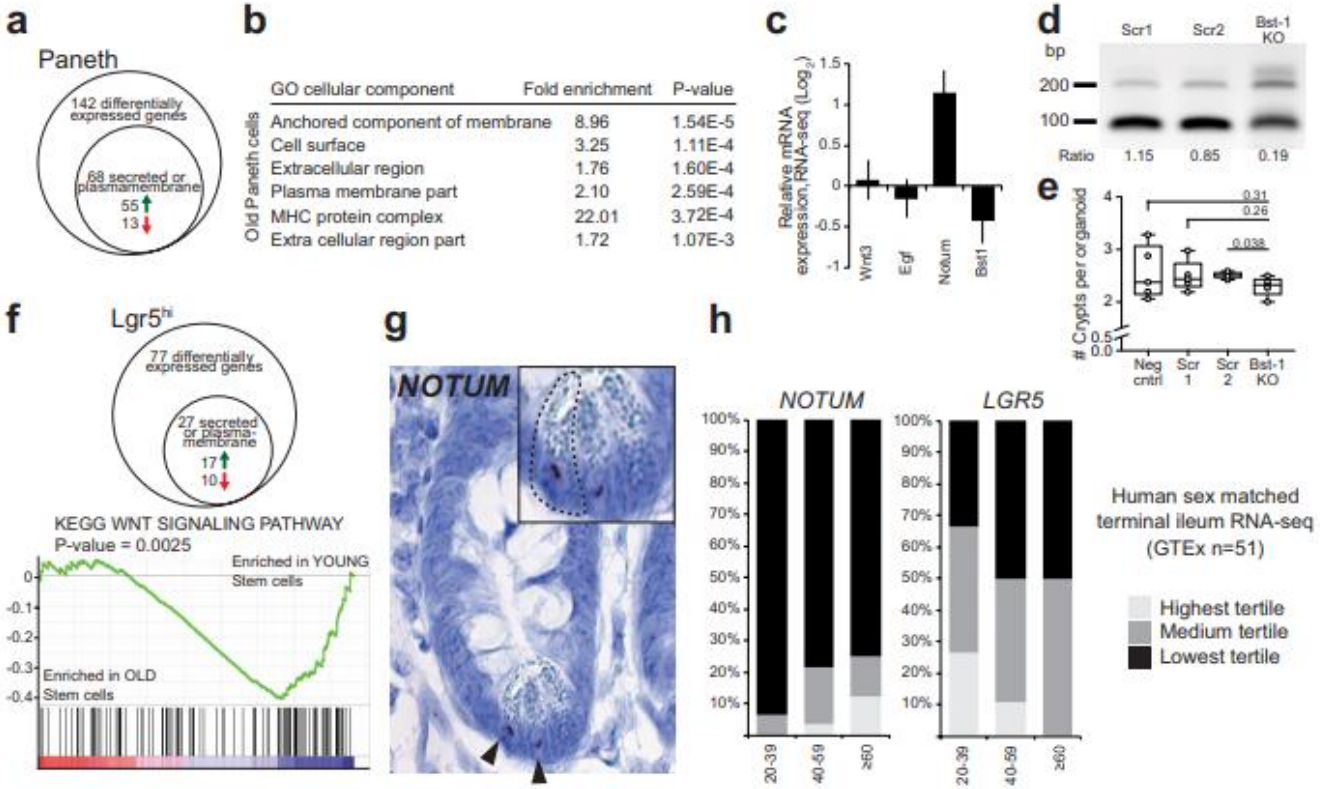
Why do basic scientists need to study statistics?

Notum produced by Paneth cells attenuates regeneration of aged intestinal epithelium

NATURE | VOL 571 | 18 JULY 2019

Nalle Pentinmikko[1], Sharif Iqbal[1,14], Miyeko Mana[2,14], Simon Andersson[1], Armand B. Cognetta III[3], Radu M. Suciu[3], Jatin Roper[4], Kalle Luopajärvi[1], Eino Markelin[1], Swetha Gopalakrishnan[1], Olli-Pekka Smolander[1], Santiago Naranjo[2], Tuure Saarinen[5,6], Anne Juuti[6], Kirsi Pietiläinen[5], Petri Auvinen[1], Ari Ristimäki[7], Nitin Gupta[8], Tuomas Tammela[9], Tyler Jacks[2,10], David M. Sabatini[2,10,11], Benjamin F. Cravatt[3], Ömer H. Yilmaz[2] & Pekka Katajisto[1,12,13]*

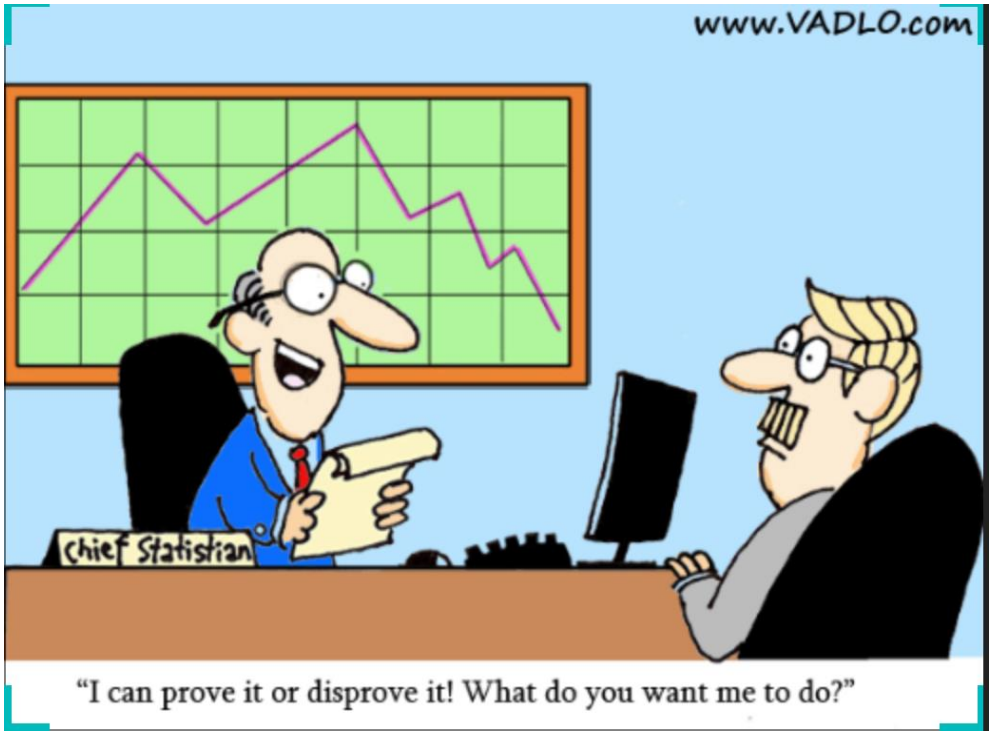
Extended Data Fig. 2 | Characterization of gene expression in old Paneth and ISCs.



Gels from PCR

gene set enrichment analysis

RNAscope for *In Situ* Hybridization



"There are three kinds of lies: lies, damned lies and statistics"

- Mark Twain or Benjamin Disraeli?

Statistics: The only science that enables different experts using the same figures to draw different conclusions.

American humorist Evan Esar (1899-1995) in his *Comic Dictionary*

Crowdsourced research: Many hands make tight work

Raphael Silberzahn & Eric L. Uhlmann

Nature 08 October 2015

ONE DATA SET, MANY ANALYSTS

Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question (about football players' skin colour and red cards).

“ It is easy to lie with statistics, but it is easier to lie without them”

Frederick Mosteller

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”

H.G. Wells

“The fact is that, despite its mathematical base, statistics is as much an art as it is a science. A great many manipulations and even distortions are possible within the bounds of propriety. “

Darrell Huff, How to lie with statistics

“Treat statistics as a science, not a recipe.”

- Biostatistician Andrew Vickers

TWO BROAD TYPES OF STATISTICS

MATHEMATICAL STATISTICS

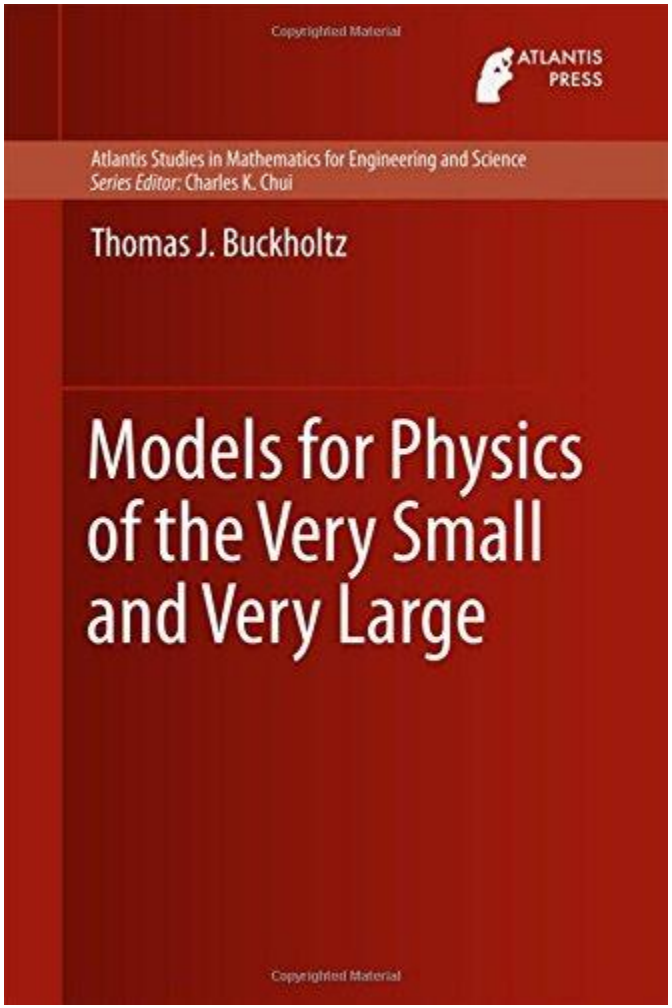
- ***Very*** mathematical
  - Calculus
  - Matrix algebra
- Focus on proofs & derivations
- Development of new statistical methods

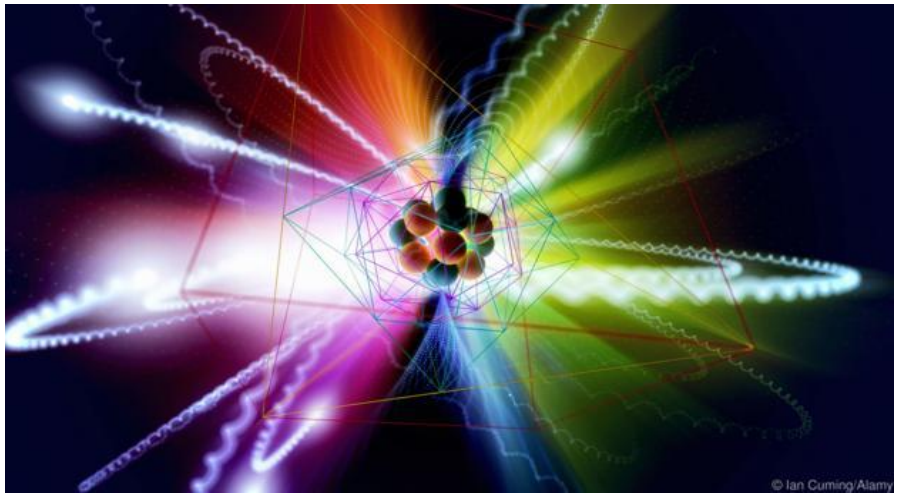
"Real" statisticians

APPLIED STATISTICS

- Can have far less emphasis on mathematics

  …but some math is good for you!
- Emphasis is on:
  - Conceptual understanding
  - Data analysis (i.e., *application* of statistical methods)
  - Correct use of statistical software (GraphPad Prism)

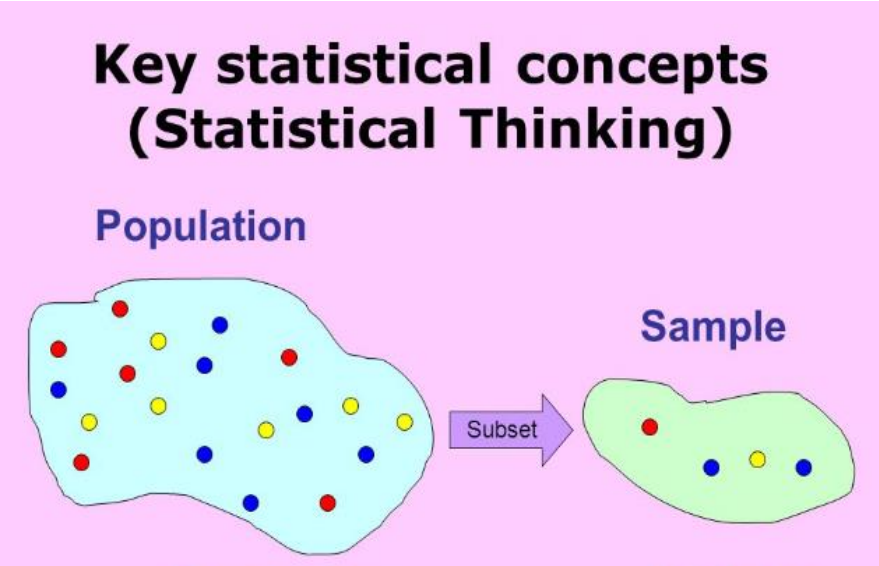




We will not be talking about the “–omics” or big data

A Basic Concept about the Statistical Tests We Will Discuss this Term

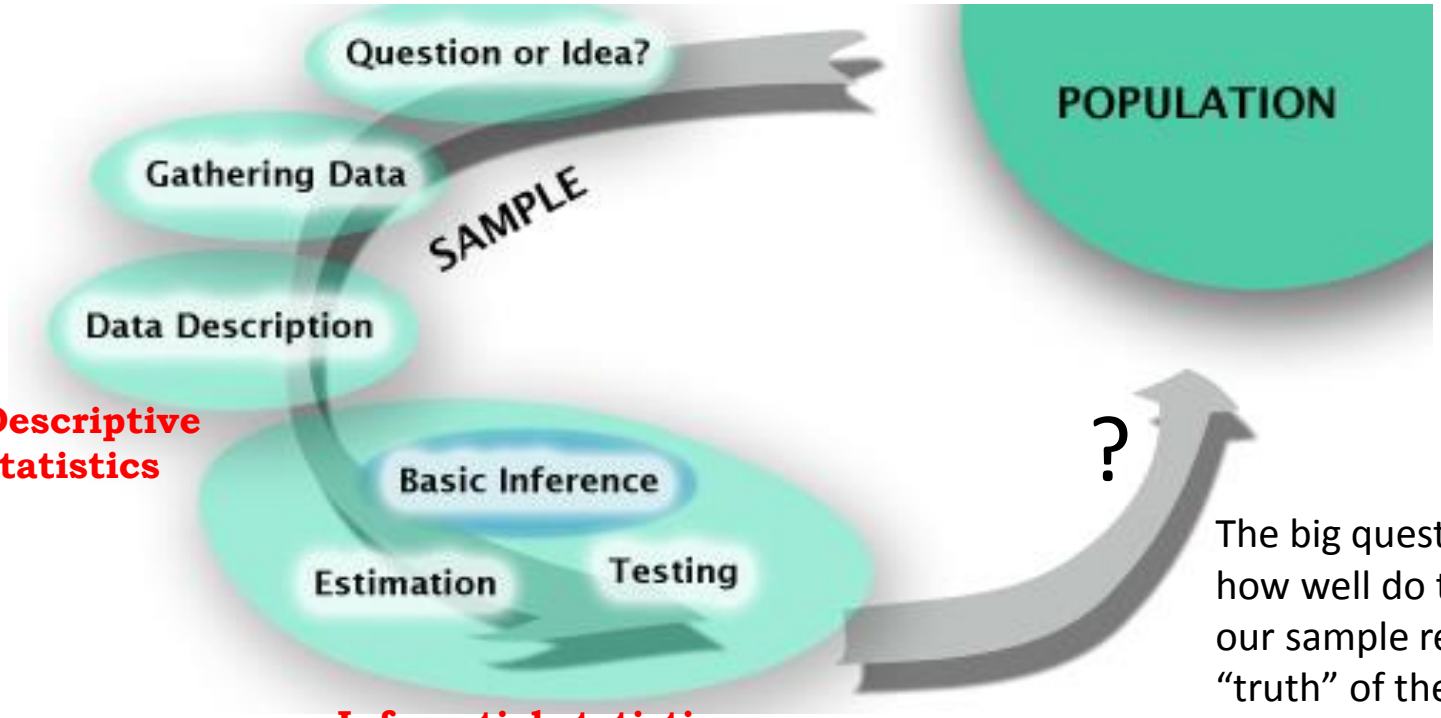
Scientists work with samples taken from larger populations



Women with breast cancer
Women with breast cancer that metastasize
Prostate cancer cells
Prostate cancer cells from high grade disease
Neurons from the brain of a mouse

Image: slideplayer.com/slide/7536181/

The Research Process and Statistics



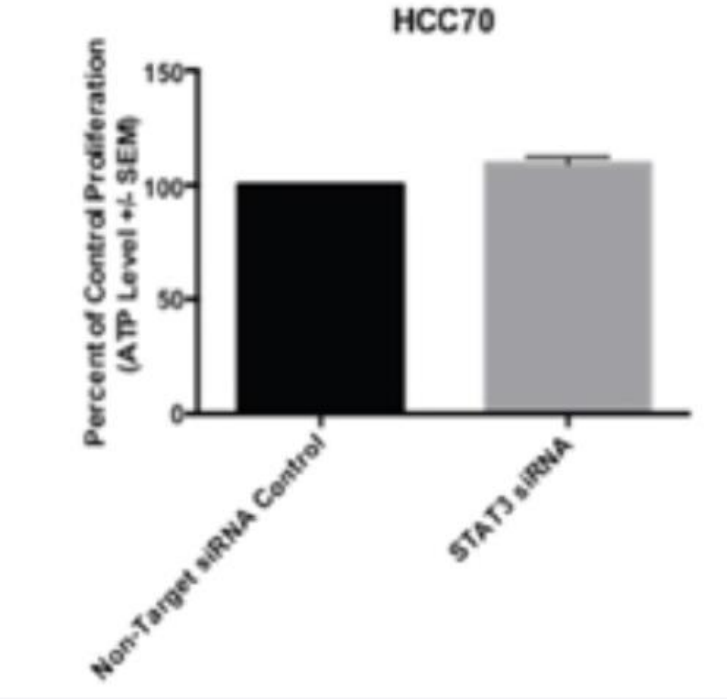
The big question is:
how well do the results from our sample represent the "truth" of the population?

- Start with a question (hypothesis) about a population
- Gather data from a subset of that population, known as a sample
- Describe that data using the various methods
- Use the data from the sample to make inferences about the larger population
- Determine a measure of effect (estimation) and test a statistical hypothesis

Statistics is the science of collecting, describing, analyzing, and making inference from data

Descriptive Statistics in Science

Genomic regulation of invasion by STAT3 in triple negative breast cancer.

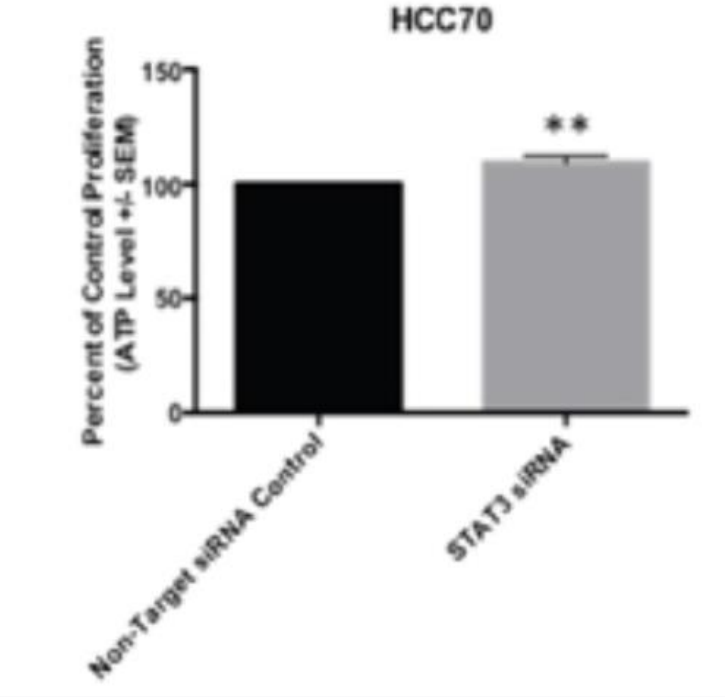


HCC70 proliferation measure after siRNA mediated STAT3 knockdown for 96 hrs.

McDaniel JM. Oncotarget. 2016 Dec 24. doi: 10.18632/oncotarget.14153. [Epub ahead of print]

Descriptive and Inferential Statistics in Science

Genomic regulation of invasion by STAT3 in triple negative breast cancer.



HCC70 proliferation measure after siRNA mediated STAT3 knockdown for 96 hrs. Knockdown of STAT3 did not result in reduced proliferation, but increased proliferation by 10% (**t-test, p-value=0.004).

McDaniel JM. Oncotarget. 2016 Dec 24. doi: 10.18632/oncotarget.14153. [Epub ahead of print]

NATURE | NEWS

Science joins push to screen statistics in papers

New policy follows efforts by other journals to bolster standards of data analysis.

Richard Van Noorden

03 July 2014

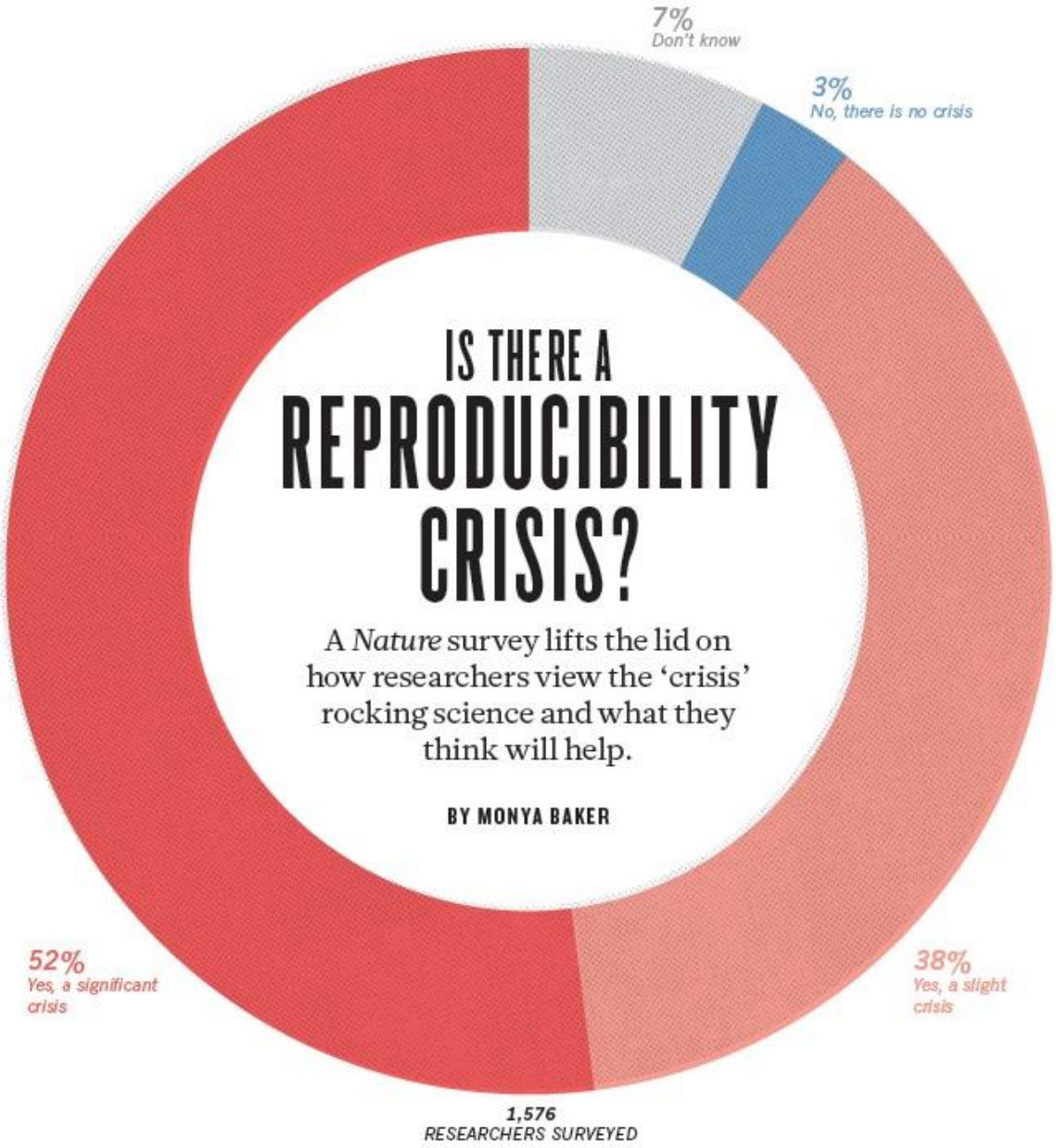
Rights & Permissions

Liu Ming/Alamy

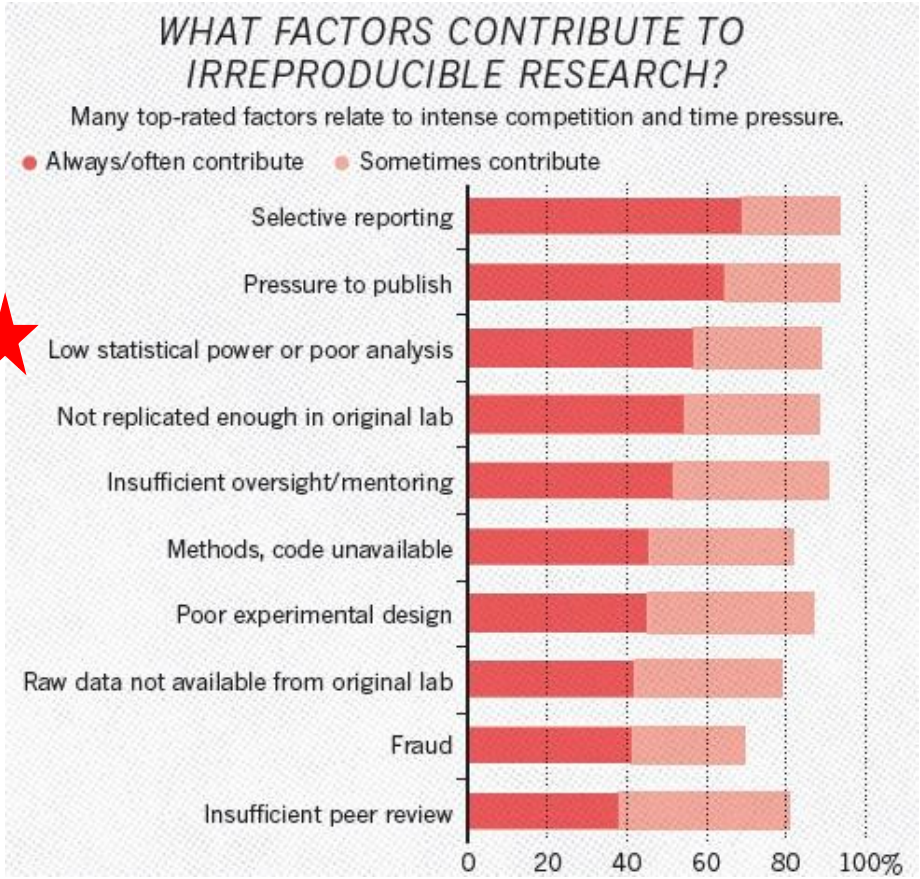
Professional statisticians will be asked to pick apart selected manuscripts during the peer review process at *Science*.

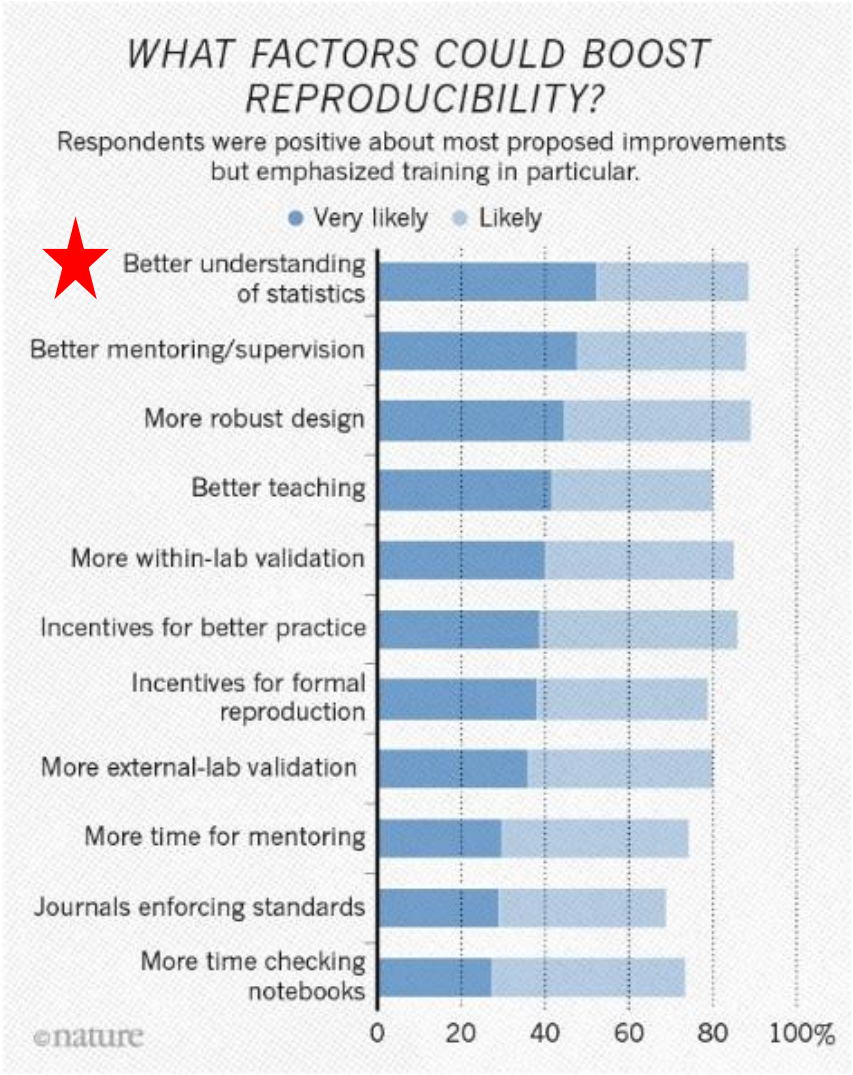
- The journal *Science* is adding an extra round of statistical checks to its peer-review process. The policy follows similar efforts from other journals, after widespread concern that basic mistakes in data analysis are contributing to the irreproducibility of many published research findings

- Professional scientists are expected to know how to analyze data, but statistical errors are alarmingly common in published research



Nature 533, 452–454 (26 May 2016)





Nature 533, 452–454 (26 May 2016)

Fifteen common mistakes encountered in clinical research

1. Failure to carefully examine the literature for similar, prior research
2. Failure to critically assess the prior literature
3. Failure to specify the inclusion and exclusion criteria for your subjects
4. Failure to determine and report the error of your measurement methods
5. Failure to specify the exact statistical assumptions made in the analysis
6. Failure to perform sample size analysis before the study begins
7. Failure to implement adequate bias control measures
8. Failure to write and stick to a detailed time line
9. Failure to vigorously recruit and retain subjects
10. Failure to have a detailed, written and vetted protocol
11. Failure to examine for normality of the data
12. Failure to report missing data and use of an intention to treat analysis
13. Failure to perform and report power calculations
14. Failure to point out the weaknesses of your own study
15. Failure to understand and use correct scientific language

Clark GT et al Journal of Prosthodontic Research 55 (2011) 1–6

Checklist to Improve Reproducibility

Corresponding Author Name: _____

Manuscript Number: _____

Reporting Checklist For Life Sciences Articles

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Figure legends

☐ Check here to confirm that the following information is available in all relevant figure legends (or Methods section if too long):

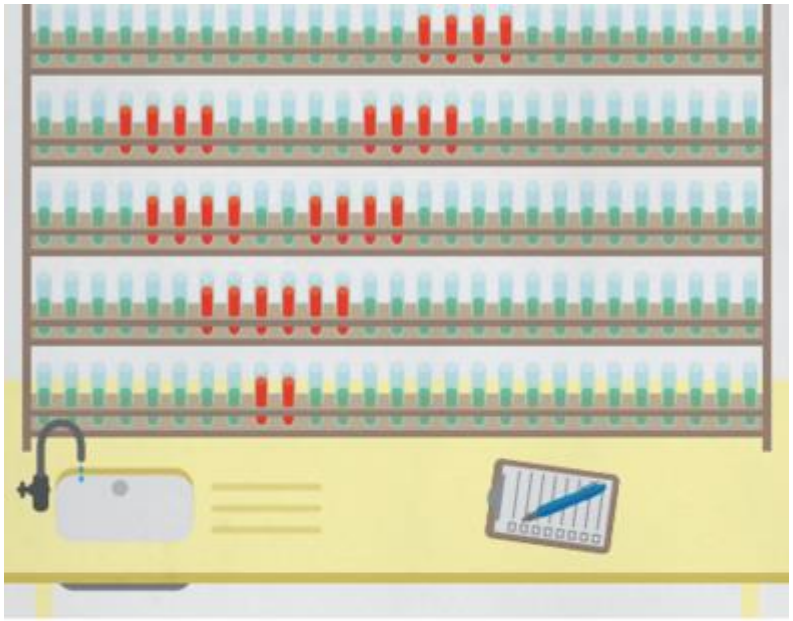
- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represent **technical or biological replicates** (including how many animals, litters, culture, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**: (For small sample sizes (n<5) descriptive statistics are not appropriate, instead plot individual data points)
  - very common tests, such as *t*-test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - **statistical test results**, e.g., ***P* values**;
  - definition of '**center values**' as **median** or **mean**;
  - definition of **error bars** as **s.d.** or **s.e.m.** or **c.i.**

This checklist will not be published. Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the Methods section for statistics, reagents and animal models. Below, provide the page number or section and paragraph number (e.g. "Page 5" or "Methods, 'reagents' subsection, paragraph 2").

npg nature publishing group | reporting checklist for life sciences articles

Statistics and general methods	Reported in section/paragraph or page #:
1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give section/paragraph or page #)	
For animal studies, include a statement about sample size estimate even if no statistical methods were used.	
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established? (Give section/paragraph or page #)	
3. If a method of randomization was used to determine how samples/ animals were allocated to experimental groups and processed, describe it. (Give section/paragraph or page #)	
For animal studies, include a statement about randomization even if no randomization was used.	
4. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding. (Give section/paragraph or page #)	
For animal studies, include a statement about blinding even if no blinding was done.	
5. For every figure, are statistical tests justified as appropriate?	
Do the data meet the assumptions of the tests (e.g., normal distribution)?	
Is there an estimate of variation within each group of data?	
Is the variance similar between the groups that are being statistically compared? (Give section/paragraph or page #)	

(Continues on following page)



NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

Nature. 2014 Jan 30;505(7485):612-3

Preclinical research, *especially work that uses animal models*, seems to be the area that is currently most susceptible to lack of reproducibility.

RESEARCH ARTICLE PLOS ONE | DOI:10.1371/journal.pone.0166733 November 17, 2016

The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research

Marc T. Avey[1,2]*, David Moher[1,3], Katrina J. Sullivan[1], Dean Fergusson[1], Gilly Griffin[1], Jeremy M. Grimshaw[1,4], Brian Hutton[1,3], Manoj M. Lalu[1,7], Malcolm Macleod[5], John Marshall[6], Shirley H. J. Mei[7], Michael Rudnicki[7], Duncan J. Stewart[7,8], Alexis F. Turgeon[9,10], Lauralyn McIntyre[1,11], Canadian Critical Care Translational Biology Group[¶]

To help assess bias, numerous basic *in vivo* experimental parameters are often not explicitly reported or relegated to supplemental material sections of high-profile biomedical science journals. The parameters include:

housing conditions
randomization
blinding
sample sizes
data handling and analysis

Differences from last year

New TAs
More basic science data
No multivariate analyses
Midterm later in term

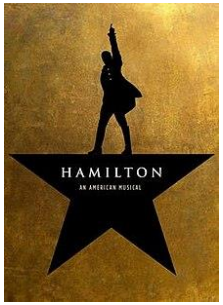


Image: www.motivationalmemo.com/what-a-difference-it-makes-to-be-different/

Some lessons learned from teaching this class so far

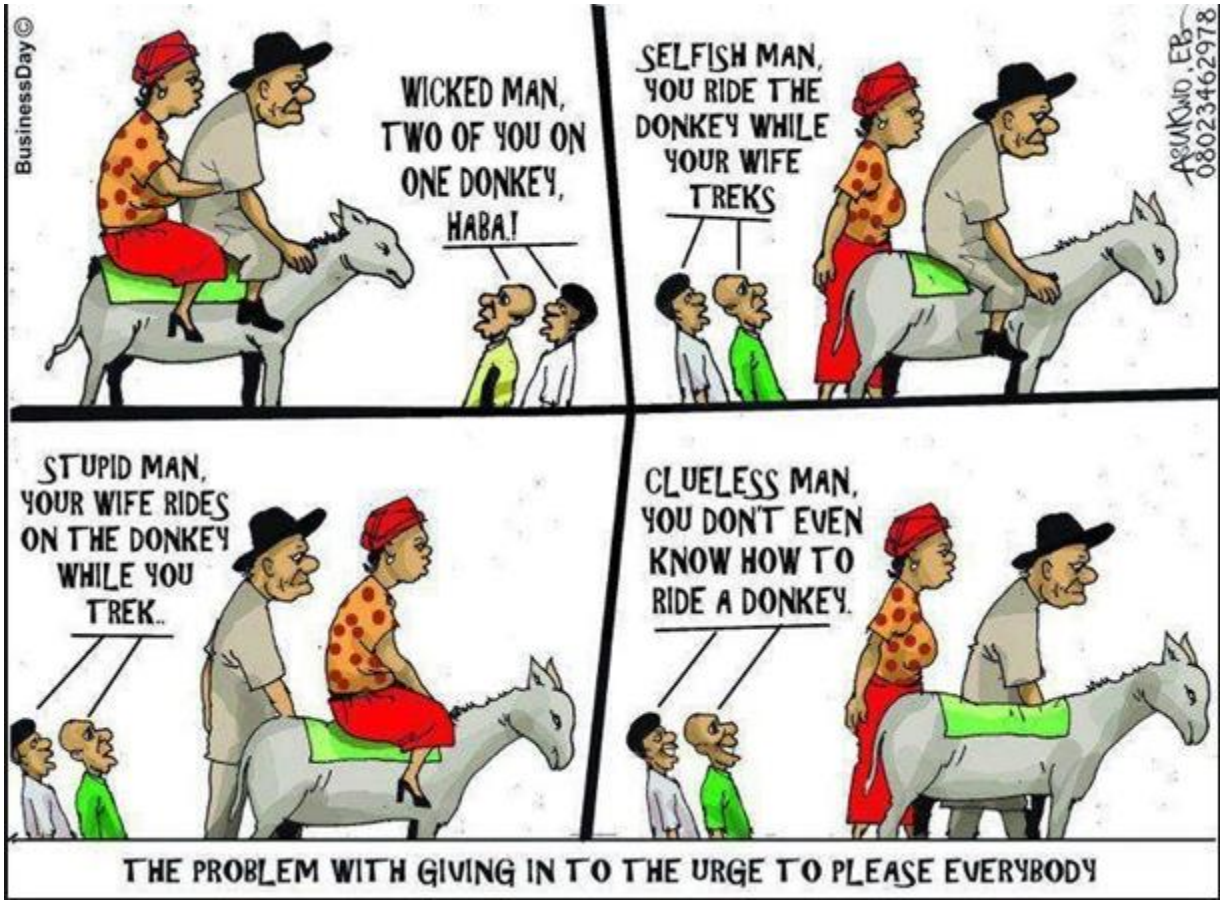


Image: www.pinterest.com/pin/310959549251550313/


SUGGESTION
BOX
BAG

Teaching Points

- My lecture slides are dense
- Questions are encouraged – a dialog is often the best way to teach
  - "Suggestion box"
- Review sessions are for you
  - Provide questions ahead of time, ask as many as you want during the session
  - Work through examples – not turned in for grades
- ? In class practice – bring your laptop or share
  - I will post dataset on Canvas the day before class so download before class
- I like to give background and history of tests to provide context
- Working in a group for homework is encouraged – just make sure you understand
- Take advantage of office hours, both TAs' and mine

Fun Fact Indicator: Things that you will not be part of any exam

Yes, it is true the color of **Hippo's milk** is bright **pink**. The reason why it is **pink** is that **hippo** secretes two kind of unique acids called "Hipposudoric acid" and "Norhipposudoric acid". The two acids got their names from the word Hippopotamus. Jan 24, 2013

Why Hippos Milk is Pink — 10 Random Facts About Hippos : Facts List
factslist.net/2013/01/why-hippos-milk-is-pink-10-random-facts-about-hippos/

Image: www.dumpaday.com/random-pictures/random-fun-facts-15-pics/

Kathleen Torkko

PhD in Analytic Health Sciences: Epidemiology Track

Variation in vitamin D receptor polymorphisms and prostate cancer; differences by race/ethnicity

MSPH

Screening for chlamydia in adolescent females

MS in Marine Aquaculture

Growth and maturation of *Laminaria japonica* in Tokyo Bay

The Japanese Journal of

PHYCOLOGY

CONTENTS

Course Mechanics
Contact Information and Office hours

Contact will be through Canvas, class time, office hours, e-mail.
We will try to respond with in 24 hours, but realize that given our busy schedules, we may not always be able to do so.

Dr. Torkko (kathleen.torkko@ucdenver.edu)

Office: RC1-South Rm L18-5113

Office hours: TBD

come to my office

Randy Jin (xin.2.jin@cuanschutz)

Office hours and location: TBD

Software

Software

www.graphpad.com

GraphPad Prism

Overview

GraphPad Prism, available for both Windows and Mac computers, combines scientific graphing, comprehensive curve fitting (nonlinear regression), understandable statistics, and data organization.

GraphPad Prism was originally designed for experimental biologists in medical schools and drug companies, especially those in pharmacology and physiology. Prism is now used much more broadly by all kinds of biologists, as well as social and physical scientists. More than 200,000 scientists in over 110 countries rely on Prism to analyze, graph and present their scientific data. It is also widely used by undergraduate and graduate students.

> While it won't replace a heavy-duty statistics program, Prism lets you easily perform basic statistical tests commonly used by laboratory and clinical researchers. Prism offers t tests, nonparametric comparisons, one-, two- and three-way ANOVA, analysis of contingency tables, and survival analysis. Analysis choices are presented in clear language that avoids unnecessary statistical jargon.

How to get GraphPad Prism

$0/semester for a student license will be paid by the Department of Biostatistics and Bioinformatics

In the assignment, there is a place to indicate if you need a license.

I will need your e-mail address to activate the license.

You will be able to install Prism on only **one** computer.

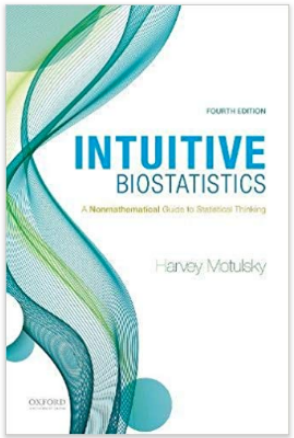
Textbook (optional)

Image: bookcheetah.com/blog/the-3-phases-of-textbook-reading/

Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking 4th Edition

by Harvey Motulsky (Author)

★★★★★ 1 customer review



Paperback $31.62 - $63.22 | Other Sellers from $31.62

Rent $31.62

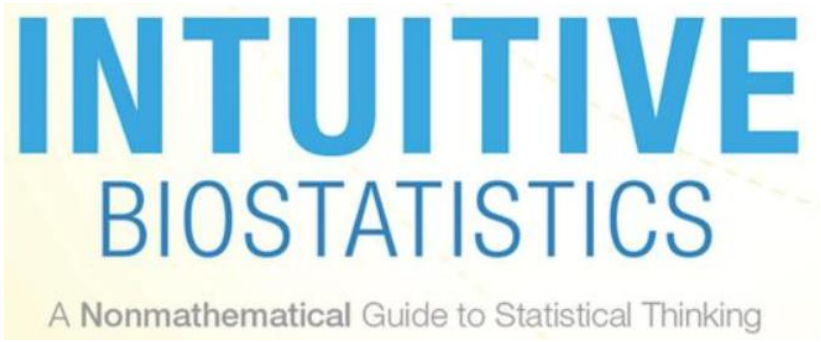
Buy new **$63.22**

In Stock.
Ships from and sold by Amazon.com. Gift-wrap available.

Coupon ☐ Save an extra $3.88 when you apply this coupon. Details

Want it Monday, Aug. 20? Order within **16 hrs 37 mins** and choose **One-Day Shipping** at checkout. Details

List Price: ~~$69.95~~
Save: $6.73 (10%)
17 New from $58.23
FREE Shipping.
Deliver to Denver 80201
Qty: 1
Add to Cart
Buy Now



Grades

Copyright © 2002 Newspaper Enterprise Association, Inc.

Course Mechanics
Evaluation

- Exams (two-thirds of grade)
  - Midterm (in class): one-third
  - Final (in class and possible take-home portion): one-third
  - In-class is multiple choice and short answer
    - You will be allowed a study guide - 2 pages of notes
- Assignments due for most lectures (one-third of grade)
  - One assignment per week and due the following week
  - If assignment is only reading, a quiz *may be* given
  - 1-3 hours a week per 1 credit hour allowed
    - Try for closer to 1 hour
  - Submit to Canvas with your name saved in the file name (i.e., SmithJane_AssignmentWeek02)
  - TA's will be responsible for grading the assignments
    - Address questions to them first
    - Instructors will resolve disputes with me as the final arbiter

Assignments submitted late without good reason will have 25% deducted from the score

Course Mechanics

- 24 lectures
- 5 Review sessions
- 2 in-class tests (midterm and final) with take home component (?)
- One office hour session each week for instructor
- One office hour session each week run by TA
- Attendance is optional (but you are still responsible for content)
- Attendance will be taken at each class
  - Will not be used to determine grade but may add a few extra credit points

Course Schedule (subject to change)

BIOS 6606: COURSE SCHEDULE FALL 2019 (ver 16AUG2019) *SUBJECT TO CHANGE*

Week01 08/26 LECTURE - INTRODUCTION
08/28 LECTURE –Probability, sampling error, and bias
Week02 09/02 LABOR DAY
09/04 LECTURE - PRISM tutorial, good data practices
Week 03 09/09 LECTURE - DESCRIPTIVE STATISTICS 1: Types of data, measures of central tendency and variability
09/11 LECTURE - DESCRIPTIVE STATISTICS 2: Good graphing practices
Week04 09/16 LECTURE – INFERENTIAL STATISTICS TEST ASSUMPTIONS 1 – Assumption of normality
09/18 LECTURE - INFERENTIAL STATISTICS TEST ASSUMPTIONS 2 – Other test assumptions (independence, equal variances, etc.)
Week05 09/23 LECTURE - CONCEPTS OF INDEPENDENCE IN BASIC SCIENCES – Identifying experimental units
09/25 LECTURE – CONCEPTS FOR INFERENTIAL STATISTICS – The real meaning and use of the p-value
Week06 09/30 LECTURE – SCIENTIFIC RIGOR - Repeatibility, Replicability, Reproducibility
10/02 LECTURE - Introduction to Big Data in Biomedicine; Guide to other courses
Week07 10/07 LECTURE - INFERENTIAL STATISTICS 1– Comparing one or two independent groups of continuous data (parametric and non-parametric tests)
10/09 LECTURE - INFERENTIAL STATISTICS 2 – Comparing one or two independent groups of categorical data
Week08 10/14 LECTURE – INFERENTIAL STATISTICS 3 – Comparing two groups of related data (Continuous and categorical)
10/16 REVIEW SESSION 1: Tests comparing one or two groups of data: t-tests, z-score, chi-square, etc.
Week09 10/21 LECTURE - INFERENTIAL STATISTICS 4 - Comparing >2 independent groups of continuous or categorical data (ANOVA, chi-square, etc)
10/24 LECTURE - INFERENTIAL STATISTICS 5- Comparing >2 groups of related continuous or categorical data (repeated measures ANOVA, etc.)
Week10 10/28 REVIEW SESSION 2: ANOVA, repeated measures ANOVA plus student questions for midterm
10/30 **MIDTERM**
Week11 11/04 LECTURE - BASICS OF STATISTICS: power and sample size, multiple comparisons
11/06 LECTURE - STATISTICS USING THE FANTASTIC 2X2 TABLE: Risks and odds, Odds ratio, Relative risk, sensitivity, specificity, positive and negative predictive values
Week12 11/11 LECTURE - INFERENTIAL STATISTICS 6 – Comparing two groups of continuous data using correlation and simple linear regression
11/13 REVIEW SESSION 3: simple linear regression, correlation
Week13 11/18 LECTURE - INFERENTIAL STATISTICS 7 – Survival Analyses (Kaplan-Meier, time to event data)
11/20 LECTURE - STATISTICS FOR OTHER PROCEDURES: ROC, Student requests
Week14 11/25 REVIEW SESSION 4: Survival, ROC
11/27 LECTURE - PRE-THANKSGIVING FUN: more practice using data sets from published papers
Week15 12/02 LECTURE - SOME BASIC SCIENCE SPECIFIC ISSUES: Choosing the experimental unit redux, using small sample sizes
12/04 LECTURE – GOOD STATISTICAL PRACTICES: Choosing the right test and writing up statistical results
Week16 12/09 REVIEW SESSION 5: Free form – bring your questions
12/11 **FINAL**

Review sessions

Exams

Practicalities

Lecture slides will be posted *usually* the day before class by 6 pm on Canvas.

Assignments will be once a week, assigned on Wednesdays after class
The questions will cover the lecture topics for that week

Assignments will be due (uploaded to Canvas) the following Wednesday before class (by 10:30 am)

Canvas Organization

Files organized by categories then by week

Lectures (folder) *pdf files of lectures*
- Week01Aug26Aug28 (folder)
  - LectureAug26 (pdf file)
  - LectureAug28 (pdf file)
- Week02LaborDaySep04 (folder)

Assignments (folder) *mostly Word and pdf files; an occasional Excel file*
- Week01Aug26Aug28 (folder)
  - Week01AssignmentDueAug27 (word file)
  - <specific paper(s) used in assignment (pdf file)>
- Week02LaborDaySep04 (folder)

Canvas Organization – cont.

Datasets (folder) *mostly Excel files*
- Week01Aug26Aug28 (folder)
  - LectureAug26Datasets (folder) *Examples used in class*
  - LectureAug26Datasets (folder) *Examples used in class*
  - Week01AssignmentDatasets (folder)

Papers (folder) *mostly pdf files*
- Week01Aug26Aug28 (folder)
- Week02LaborDaySep04 (folder)

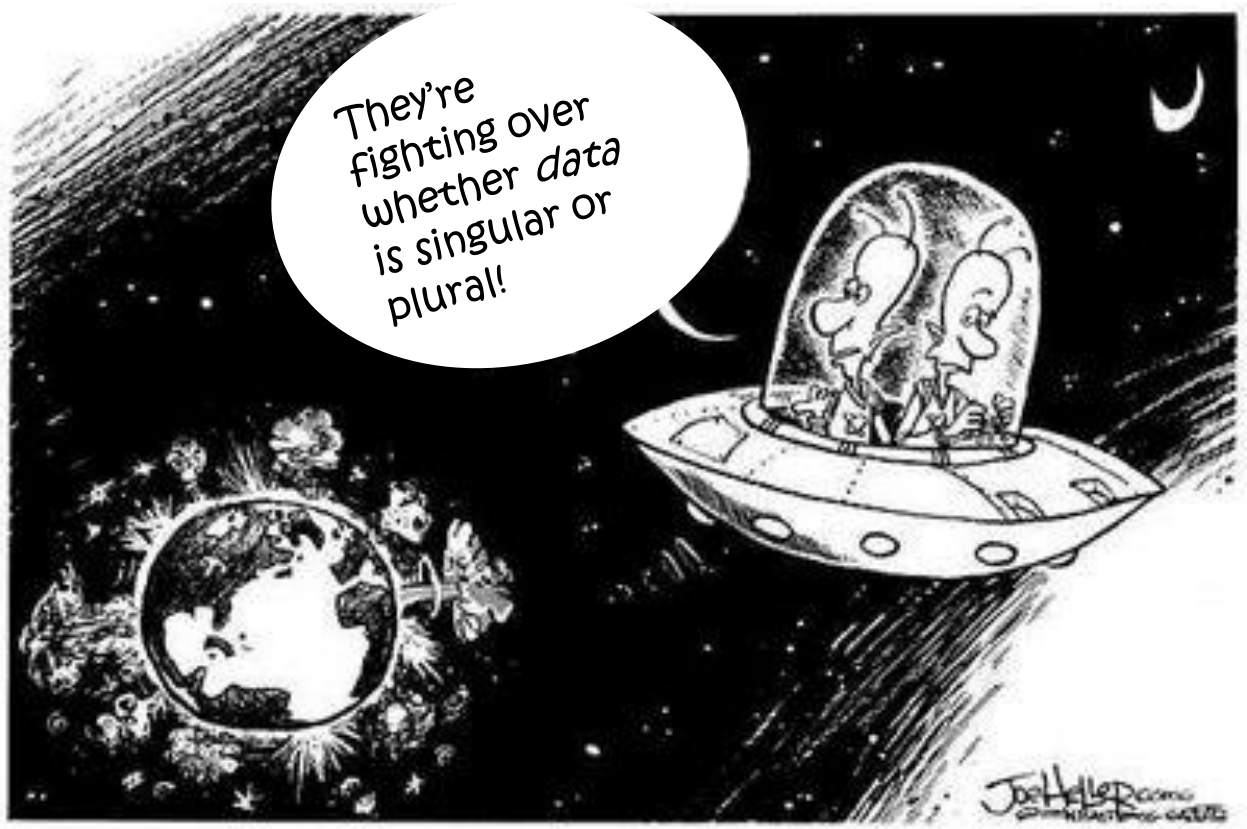
Syllabus (Folder)
- Syllabus, Course schedule, Other information as needed

The Big Picture

- Think
  - A data analysis is much more than putting data into a statistical program and pushing enter to get a p-value

- Shades of gray
  - "Truth" vs. estimation (AKA inference)
  - Statistical significance vs. clinical/biological significance
  - Choosing the right test for the data and research question
    - Reducing bias
  - Determining if data are normally distributed or homoscedastic

This course is more like a collection of short stories than a novel



Source: http://quantumenergyprocess.files.wordpress.com/2013/04/religion-war-cartoon-02.jpg

ASSIGNMENT

Week01 Assignment: Due Tomorrow Aug 27

Reading (see Canvas)

Save file in the following format (LastnameFirstinit_WeekXXAssignment)
TorkkoK_Week01Assignment

EMAIL TO ME (kathleen.torkko@cuanschutz.edu)

- Complete questionnaire
- Tell me which graduate program you belong to
- Tell me what you expect from the class (e.g., specific types of analysis)
- Give me your e-mail
- Upload to Canvas by 8 pm 8/27 (so I can talk about it in class on Wednesday)
- Give me some times that would work for TA office hours on Monday and Tuesday and my office hours on Wednesday afternoon

There are no right or wrong answers. For this assignment only, full points will be awarded for providing the information on time.


RANDOM.ORG (Randomness and Integrity Services Ltd) [IE] https://www.random.org/
Home
Games
Numbers
Lists & More
Drawings
Web Tools
Statistics
Testimonials
Learn More
Login
Lottery Quick Pick
Keno Quick Pick
Coin Flipper
Dice Roller
Playing Card Shuffler
Birdie Fund Generator
Search RANDOM.ORG
Search
True Random Number Service
Do you own
Check out our app!
What's this fuss about true randomness?
Home
Games
Numbers
Lists & More
Drawings
Web Tools
Statistics
Testimonials
Learn More
Login
RANDOM.ORG
Search RANDOM.ORG
Search
True Random Number Service
Do you own an iOS or Android device? Check out our app!
Coin Flipper
This form allows you to flip virtual coins. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.
Flip 1 virtual coin(s) of type Australian $1
Flip Coin(s)
Reset Form

Home Games Numbers Lists & More Drawings Web Tools Statistics Testimonials Learn More Login

RANDOM.ORG

Search RANDOM.ORG Search

True Random Number Service

Do you own an iOS or Android device? Check out our app!

Coin Flipper

You flipped 1 coin of type Australian $1:

Timestamp: 2018-08-19 22:16:15 UTC

Flip Again Go Back

Then do 2 coins and record in the excel file the number:

both heads	n
first heads second tails	n
first tails second heads	n
both tails	n

RANDOM.ORG

Do you own an iOS or Android device? Check out our app!

Dice Roller

This form allows you to roll virtual dice. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.

Roll 1 virtual dice.

Roll Dice Reset Form

	Full bag
Color	n
Blue	13
Orange	10
Green	7
Yellow	5
Red	10
Brown	8
Total	53