PSY 503: Foundations of Statistical Methods in Psychological Science

Introduction & Course Organization

September 3, 2025

Suyog Chandramouli

311 PSH (Princeton University)

Welcome to PSY 503!

About me

- Su – yog
  yog like in yogurt

- He/ him/ his

- I will be your instructor for
  - PSY 503: Fundamentals of Statistics
  - PSY 504: Advanced Statistical Methods
  - PSY 505: a seminar series on research methods/ statistical methods in psychology

- Statistical consulting / Research collaborations

My Journey

Bangalore, India

B.E. in Computer Science
Research Assistant (Indian Institute of Science)

2011

My Journey

Bangalore, India

B.E. in Computer Science
Research Assistant (Indian Institute of Science)

2011

My Journey

Bangalore, India

Bloomington, IN

Helsinki, Finland

Edmonton, Canada

(Princeton, NJ)

2025
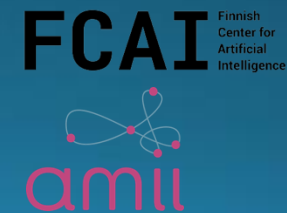
Research Interests
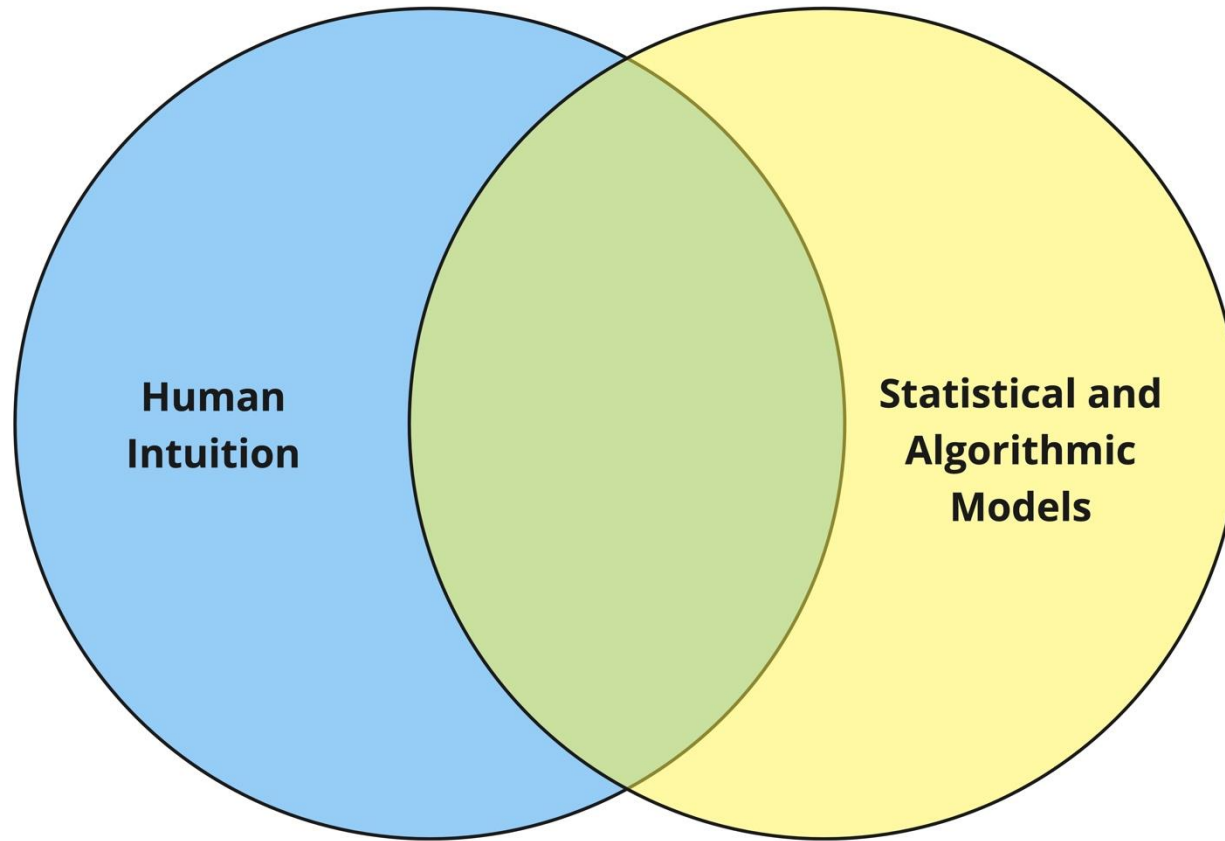
Research Interests
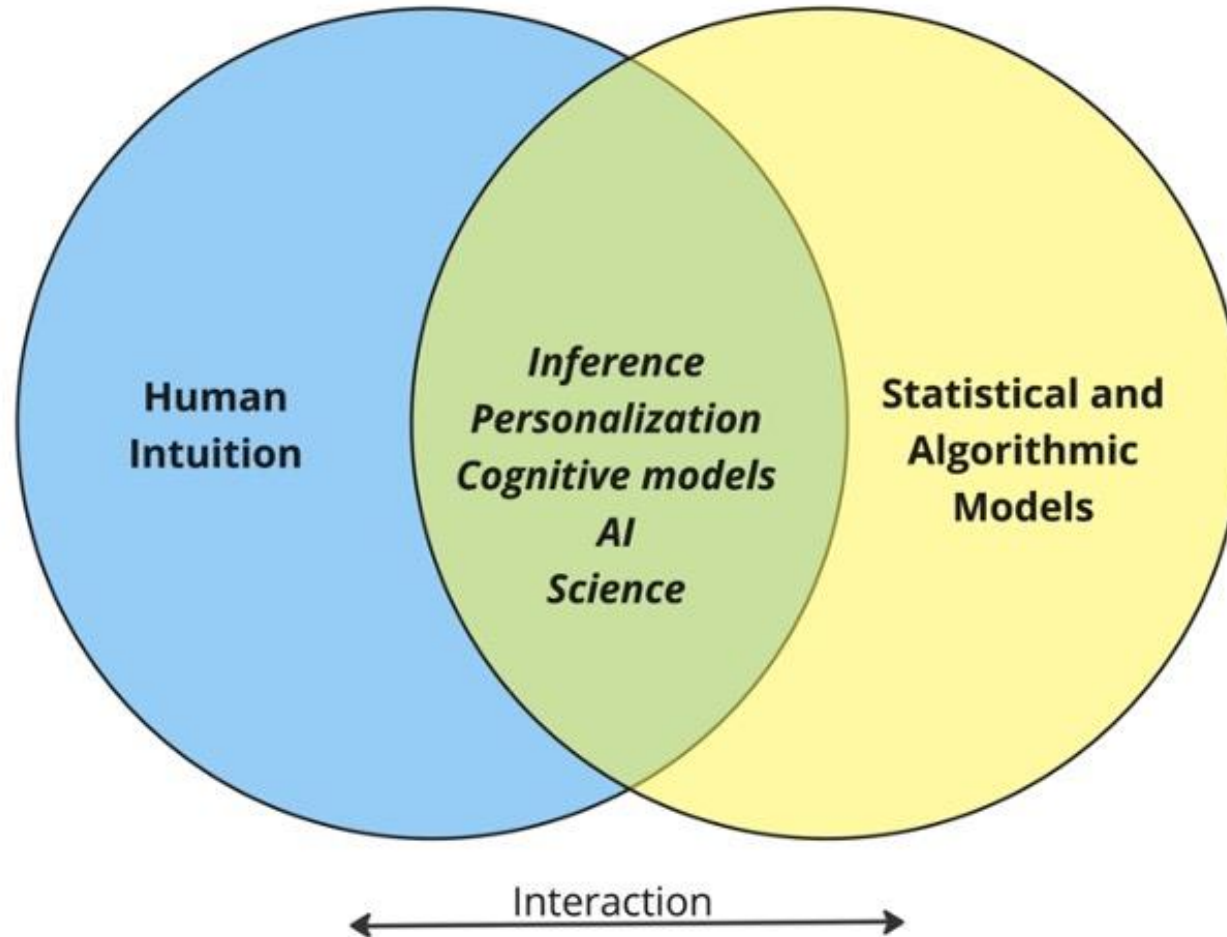
Ice breaker

- Name
- Lab
- Research interests
- Where feels like home to you?
- How was your summer?
- A boring fact about you?

Questionnaire Results

Experience with stats

How much experience do you have analyzing data?

6 responses

- Substantial
- Some
- Little
- No experience
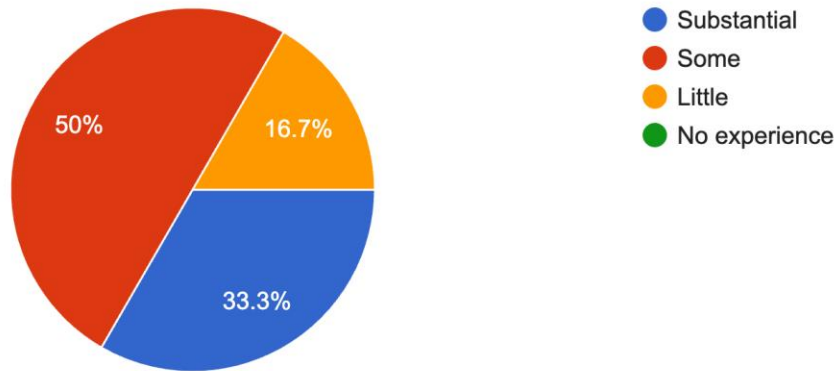
Little:
- "I have very little knowledge from > 5 years ago.. Don't feel confident.. And would hope for a refresher"
- Specific data analysis for undergrad theses and RA-ships
- Basic R

Some:
- T-tests s and ANOVAs
- Dabbled with R, MATLAB, SPSS, Jasp
- Using generative AI

Substantial:
- Data science major
- 3+stats courses

Experience with coding

How much experience do you have writing code?
6 responses

- Substantial
- Some
- Little
- None

66.7%
16.7%
16.7%
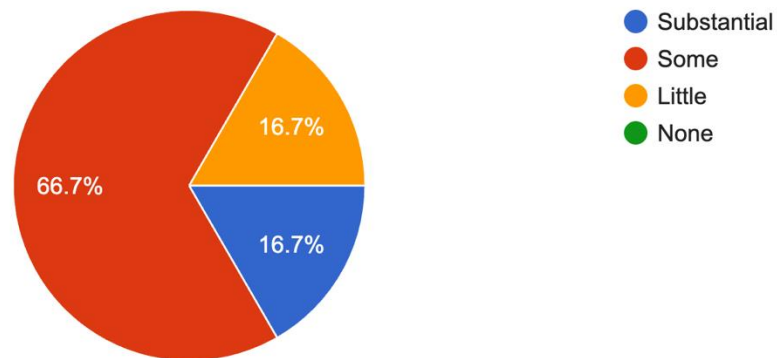
Some/ Little:
- R through courses
- Modified existing code
- With aid of generative AI

Substantial:
- Solid knowledge about Python, R, probabilistic programming languages
- Python (>3 yrs experience)

Feelings about math

How do you feel about math?

6 responses
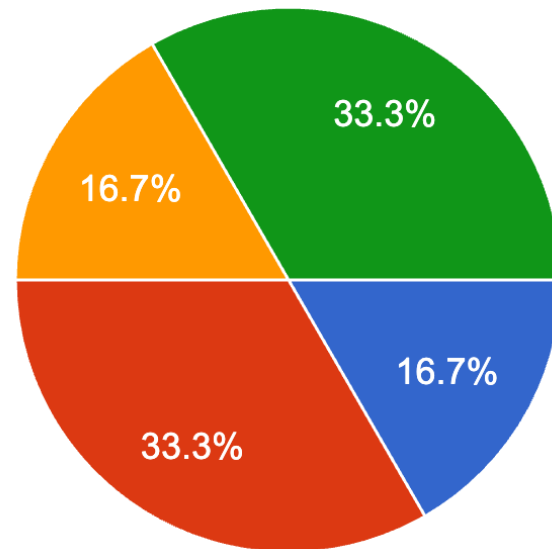
Stage 1: Regress each column of **X** on **Z**, $(X = Z\delta + \text{errors})$:

$$\hat{\delta} = (Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} X,$$

and save the predicted values:

$$\widehat{X} = Z\hat{\delta} = Z(Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} X = P_Z X.$$

Stage 2: Regress **Y** on the predicted values from the first stage:

$$Y = \widehat{X}\beta + \text{noise},$$

which gives

$$\beta_{2SLS} = (X^{\mathrm{T}} P_Z X)^{-1} X^{\mathrm{T}} P_Z Y.$$
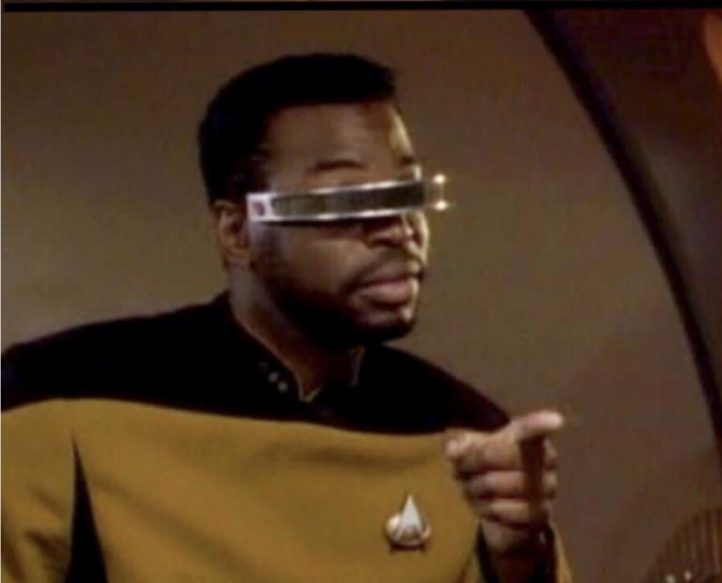
```r
model_2sls <- iv_robust(
  health ~ bed_net | treatment,
  data = bed_nets)
```

One-sample t test | Sign test / Signed rank test
Two-sample t test | Mann-Whitney U test
Paired t test | Wilcoxon signed rank
One-way ANOVA test | Kruskal Wallis tes
Pearson Corr / Linear Reg | Spearman Corr / Linea

Chi-square test / Exact test
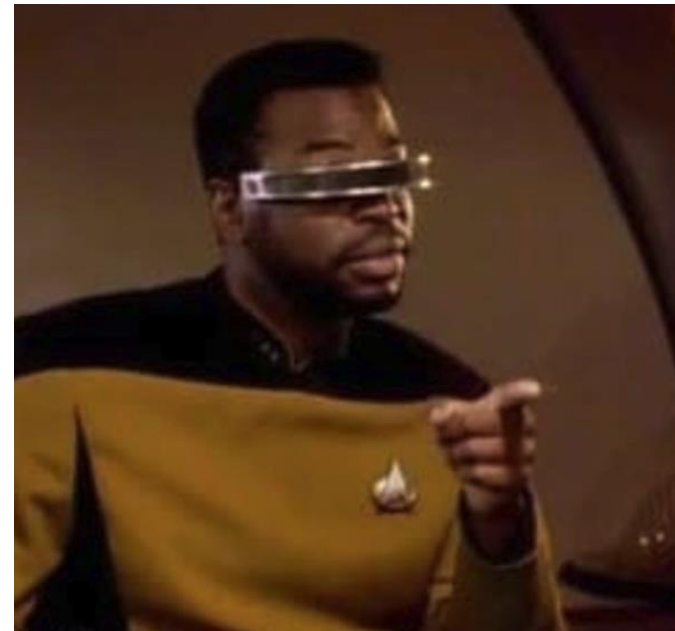Chi-square test / Fisher's exact test / Logistic regress
McNemar's test / Kappa statistic
Chi-square test / Fisher's exact test / Logistic regress
Logistic regression / Sensitivity & specificity / ROC

KM plot with Log-rank test
KM plot with Log-rank test
Cox regression

- Statistic
- Parameter
- Sum of squares
- Variance
- Standard deviation
- T-test
- Correlation
- Regression
- Experiment
- Random assignment
- Random sample
- Null hypothesis
- Alternative hypothesis
- Z-score
- Z distribution

- p < .05
- Random variable
- Quartiles
- Interquartile range
- Sampling distribution
- Randomization test
- Confidence interval
- Sampling variation
- Model
- Chi-square
- ANOVA
- Normal distribution
- Margin of error
- F-test
- Paired t-test

DATA = MODEL + ERROR

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$Y_i = \beta_0 + \varepsilon_i$$

Data Generating Process

DGP

M=58

Data

Sampling Distribution

Common statistical tests are linear models

Last updated: 28 June, 2019.Also check out the *Python version*!

	Common name	Built-in function in R	Equivalent linear model in R
Simple regression: lm(y ~ 1 + x)	**y is independent of x** P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	lm(y ~ 1) lm(signed_rank(y) ~ 1)
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y_1, y_2, paired=TRUE) wilcox.test(y_1, y_2, paired=TRUE)	lm(y_2 - y_1 ~ 1) lm(signed_rank(y_2 - y_1) ~ 1)
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	lm(y ~ 1 + x) lm(rank(y) ~ 1 + rank(x))
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y_1, y_2, var.equal=TRUE) t.test(y_1, y_2, var.equal=FALSE) wilcox.test(y_1, y_2)	lm(y ~ 1 + G_2)[A] gls(y ~ 1 + G_2, weights=...[B])[A] lm(signed_rank(y) ~ 1 + G_2)[A]

Different software/tools

For each program below, rate how proficient you are at using them
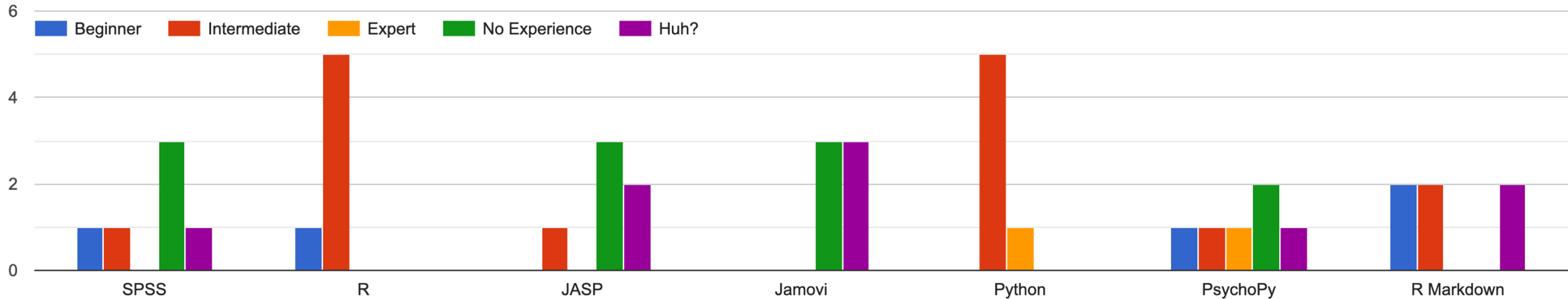
Different software/tools
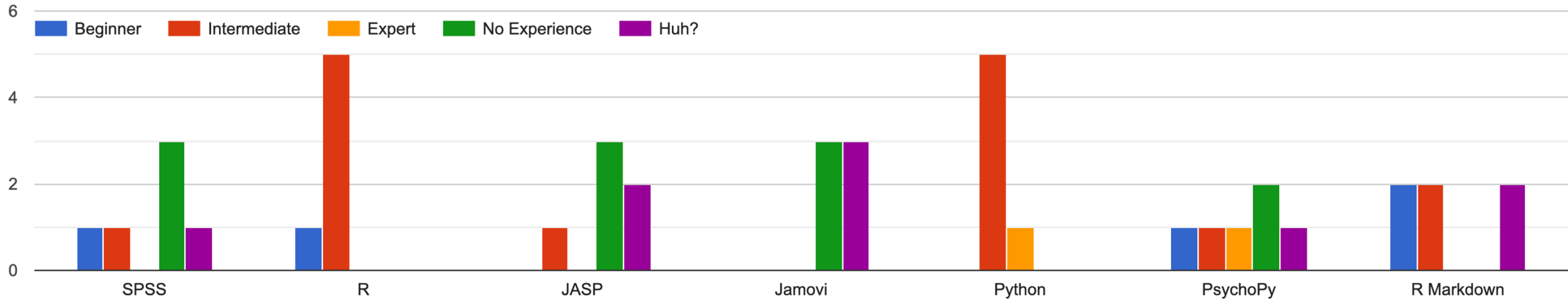
For each program below, rate how proficient you are at using them
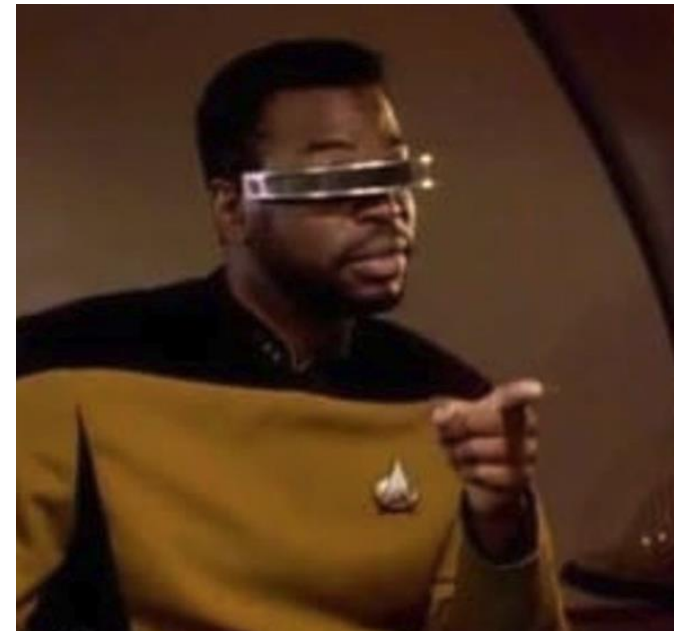
Topics you'd like to see discussed

- specific tools: R, SPSS, Python , and Matlab
- stats but w/o coding
- stats but w/o math
- stats with python only

- ML
- Evolutionary computational modeling
- Cognitive modeling – drift diffusion, RL

One-sample t test	Sign test / Signed rank test
Two-sample t test	Mann-Whitney U test
Paired t test	Wilcoxon signed rank
One-way ANOVA test	Kruskal Wallis tes
Pearson Corr / Linear Reg	Spearman Corr / Linea

- Chi-square test / Exact test
- Chi-square test / Fisher's exact test / Logistic regress
- McNemar's test / Kappa statistic
- Chi-square test / Fisher's exact test / Logistic regress
- Logistic regression / Sensitivity & specificity / ROC
- KM plot with Log-rank test
- KM plot with Log-rank test
- Cox regression

- Statistic
- Parameter
- Sum of squares
- Variance
- Standard deviation
- T-test
- Correlation
- Regression
- Experiment
- Random assignment
- Random sample
- Null hypothesis
- Alternative hypothesis
- Z-score
- Z distribution

- p < .05
- Random variable
- Quartiles
- Interquartile range
- Sampling distribution
- Randomization test
- Confidence interval
- Sampling variation
- Model
- Chi-square
- ANOVA
- Normal distribution
- Margin of error
- F-test
- Paired t-test

DATA = MODEL + ERROR

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$Y_i = \beta_0 + \varepsilon_i$$

Data Generating Process

DGP

M=58

Data

Sampling Distribution

The way I'm thinking

- We want both conceptual understanding and deployment skills (Lectures & Labs)

- Modeling
  - Understand traditional tests as special case of the General Linear model (GLM)
    - Regression all the way down

- Deployment
  - R, RMD/ Quarto
  - Github
  - Data manipulation, cleaning, visualization

R, Python

- Pandas was inspired by R
- Lots of effort to integrate these ecosystems
  (in data science contexts)
  - Reticulate
    - Call R from Python
    - Call Python from R
  - Quarto works with R & Python
  - Posit
    - Hadley Wickam, Wes McKinney
  - Tidyverse grammar in Python
    - Plotnine (ggplot)
    - Siuba (dplyr syntax)
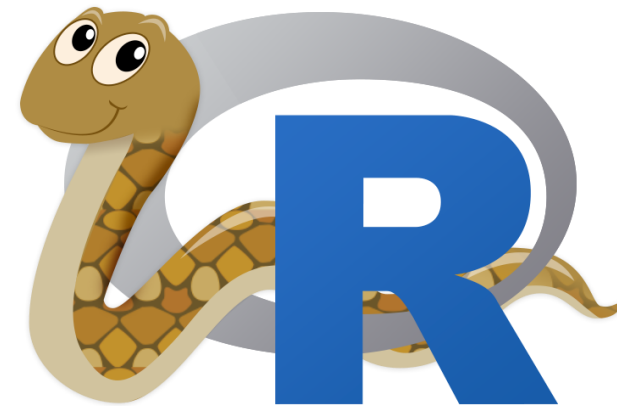
Course Materials (Books)

- **Main:** Learning statistics with R – Danielle Navarro
  - Extremely well-written
  - see Learning statistics with Python – Ethan Weed & Danielle Navarro

- Regression and other Stories – Andrew Gelman, Jennifer Hill, Aki Vehtari
  - Stories based on real-examples (rather than toy datasets)
  - Math supports the narrative
  - Art and practice of modeling

- Coursekata (optional)
  - Introductory level
  - Philosophical alignment (Data = Model + error, all tests are linear models, etc.)
  - Interactive and inline coding (folks with minimal coding experience loved it)

Course Materials (Others)

- Lecture, slides, and labs complement the reading
- Other articles, blog posts shared where relevant

Class format

- **Lectures (Monday)**
  - I'll highlight main concepts
  - My role is to introduce you to the basics of each topic and get you started with examples
  - It is your responsibility to dig deeper when you need and / or want to
    - Learning to be independent and thinking on your own is part of graduate school
- **Lab (Wednesday)**
  - Wednesday will be reserved for working on activities related to the topic
  - Also a time to seek clarification on topics from me or your classmates

Course Materials (Tools)

- R & R-Studio
  - R: programming language for data analysis
  - R-studio: more friendly and advanced programming environment

- Quarto / R-markdown
  - Quarto notebooks used in labs
  - open-source publishing system that allows you to embed code and explanatory text in the same document

- Github
  - create repositories to host project files
  - version control of code
  - enables collaborative work on projects

  We'll transition from primarily using Canvas to using Github for submissions

Choose your own adventure

- In-class presentations
  - go into a relevant rabbit hole of your interest (let's talk)
  - learn by teaching

- Canvas Discussions

- Project
  - Use own dataset or re-analyze existing data
  - Flexibility with statistical method used (let's talk)
  - ***Reproducible*** analysis submitted as a link to your Github / OSF
    - Reproducible Report
    - APA paper (using papaja)
    - Simulation-based power analysis
  - Will be discussed through the semester

Course Content

- See Syllabus and Course Schedule

- Canvas
  - Announcements
  - Discussions

- I will post slides link on Canvas / Google Docs before class so you can follow along

Next semester (Spring 2026)

- Advanced regression

- Mixed Models
  - Frequentist
  - Bayesian

- Other approaches based on your interests
  - Mediation/ Moderation
  - PCA
  - ..

Instructor contact

- e-mail: suyoghc@princeton.edu
- Twitter/X: suyoghc

- Anonymous feedback:
  https://www.admonymous.co/psy_503_feedback

Office Hours

- Monday: 1pm – 2:30 pm
- Wednesdays: 1pm – 2:30pm
- Please inform ahead of time

- Also by appointment

- PSH 325 or https://princeton.zoom.us/my/suyoghc

My statistical goals for you

- Read papers and understand what researchers are doing
  - Understand the logic and rationale behind statistical approaches
- Reason your way out of statistical jams
- NOT focused on performing low-level arithmetic or memorizing equations
- Analyze data with R
- Create reproducible documents
- **Obtain a skill set that generalizes outside academia**

Participation

- It is expected that you will come to class having read the articles assigned and prepared to discuss the material.

- Discussions in Canvas about the readings, statistics-related, and coding topics are also highly encouraged and count as participation.

- Attendance and participation are crucial for your success in the course.
  - If you are sick or are incapable of participating meaningfully in class, please stay home.
  - Please send me a note in case you are going to be absent for some reason! I will try my best to record in-class lecture for you (or record them at home). Same with late assignments.

Grades/ Assessments

- All assignments need to be turned in and complete in order to receive a passing grade

- Self-evaluation
  - Did you meet learning outcomes for the class? How well did you meet your personal goals?
  - Did you turn in all your assignments?
  - Based on my feedback, how well did you think you did on assignments?
  - How much time outside of class did you spend preparing for class?

- To me a letter grade of A would consist of finishing all the labs, engaging with the material at a sufficient level (ie., reading assigned articles, doing independent research outside class, asking questions during lecture), and generally being present in class.
  - If not, you can expect A-
  - A+ if I feel that you've gone above & beyond, or exceeded expectations

Final paper

- Reproducible statistical analysis
  - Find open data (OSF is a good place to start) in your field of interest
  - If you have published something or have a dataset, re-run in a reproducible manner
  - Very flexible with shaping it towards your interests and g
- Reproducible APA formatted paper using papaja
- Power analysis
  - Run simulations for a hypothetical replication attempt

Class environment

- I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, SES, etc.) To help accomplish this:

- If you have a name and/or set of pronouns that differ from those that appear in your official Princeton records, please let me know!

- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me. I want to be a resource for you. Remember that you can also submit anonymous feedback (which will lead to me making a general announcement to the class, if necessary to address your concerns).

- I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class (by anyone) that made you feel uncomfortable, please talk to me about it. (Again, anonymous feedback is always an option).

Questions ?

PSY 505 - Seminar in Advanced Research Methods

- 6 external speakers
- 1 internal speaker

- Relevant Topics (talks & tutorials) in the overlap of Stats & Psychology
- Designed to complement PSY 503 & PSY 504

- If you're enrolled, you're expected to attend the talk and the interactive workshop. Hot lunch served for all attendees
  - We sometimes brainstorm discussion questions ahead of time in a google doc.
  - I'll request visitors to share relevant (optional) readings if any.
  - If you'd like to meet 1:1, let me know. There may be slots available.

https://psych.princeton.edu/event-category/seminar-advanced-research-methods

Questions ?