Revise and Resubmit:

Tailoring AI-Led Analyses for Joy and Data Discovery

Vanderbilt University

ECOTS 2026

Speakers: One instructor and four graduate students from the Department of Biostatistics at Vanderbilt University

Lauren Samuels, PhD
Research Associate Professor

Ashley Mullan, MS
PhD Student

Danni Shi, MS
PhD Student

Seth Young, BS
MS Student

Ningkun Zhou, MS
PhD Student

Agenda

01 

Our experiment

02 

A Performance Review of Generative AI

03 

Enhancing the experience of using AI to study statistical analysis

04 

Beyond Working Code: AI's role in statisticians writing and understanding code

05 

Using AI as an assistant for statistical analysis

06 

Discussion

Our (uncontrolled) experiment

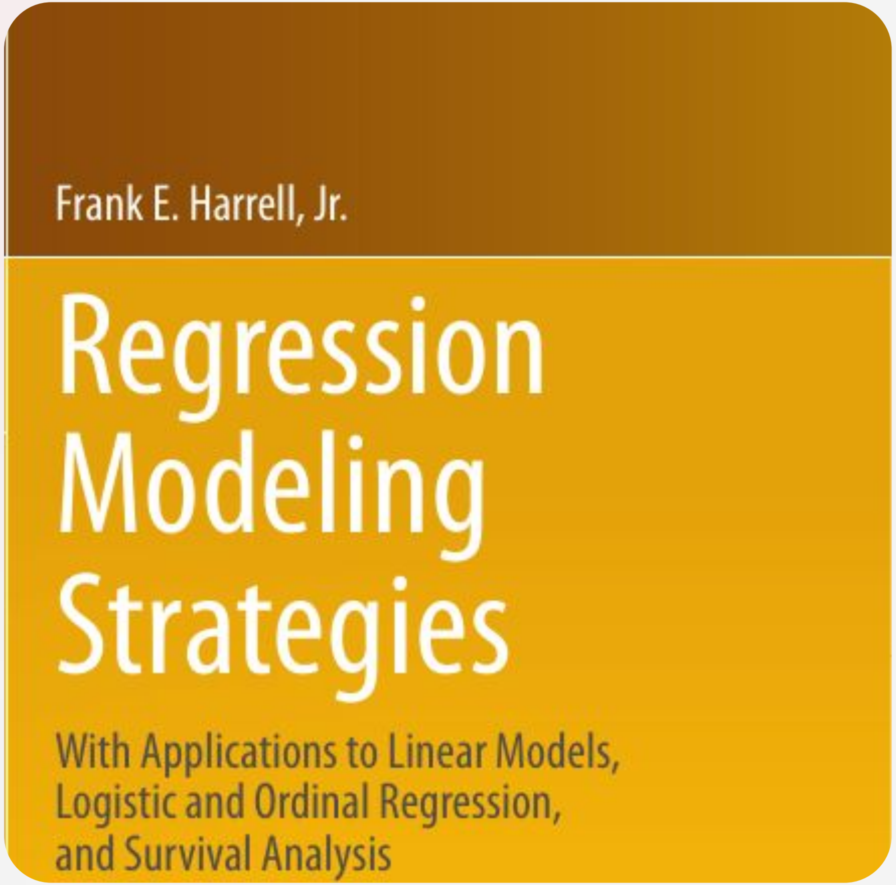
Lauren (Laurie) Samuels, PhD

Course prep, Fall 2025

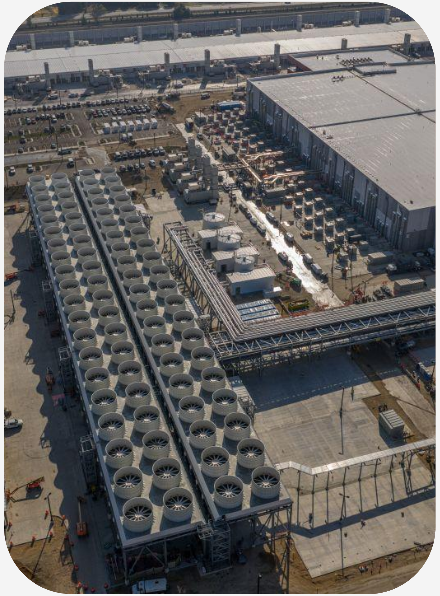
- Relatively inexperienced instructor
- Preparing to teach a new course
- And unsure what to do about AI

The new (to me) course

- Created (and taught for years) by Frank Harrell
- Elective for biostatistics graduate students
- Case studies illustrate philosophy, techniques

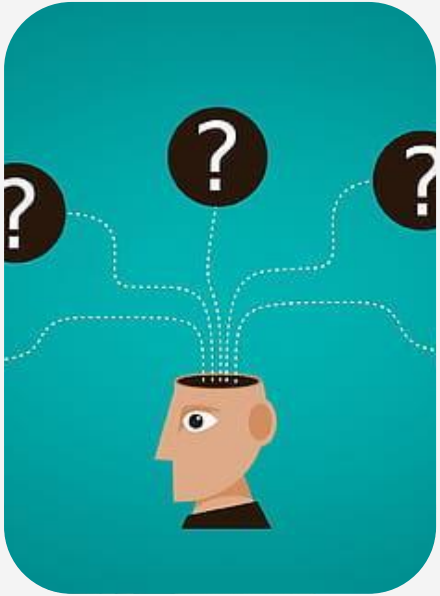


My concerns about AI



Big worry #1: (disparate) environmental and health impacts

Image sources: https://blogs.microsoft.com/blog/2025/09/18/inside-the-worlds-most-powerful-ai-datacenter/; Google images



Big worry #2: what is gen AI doing to our minds?

… and at the same time, would I be hurting my students by not allowing it?

"Sparking Joy and Discovery In a World of AI"

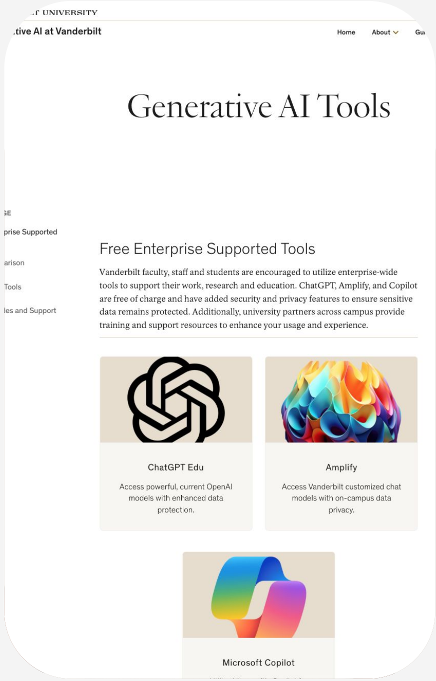
I decided to propose:

- Case studies in second half
- For each, we'll first conduct the analysis independently, then ask a generative AI tool to do so.
- We'll evaluate the tool's work and use our findings to modify how we use AI in the next case study.

The students agreed!

How did it go?

The AI climate at our institution(s)



…r UNIVERSITY

…tive AI at Vanderbilt

Home About Gu…

Generative AI Tools

…E

…prise Supported

…arison

…Tools

…les and Support

Free Enterprise Supported Tools

Vanderbilt faculty, staff and students are encouraged to utilize enterprise-wide tools to support their work, research and education. ChatGPT, Amplify, and Copilot are free of charge and have added security and privacy features to ensure sensitive data remains protected. Additionally, university partners across campus provide training and support resources to enhance your usage and experience.

ChatGPT Edu

Access powerful, current OpenAI models with enhanced data protection.

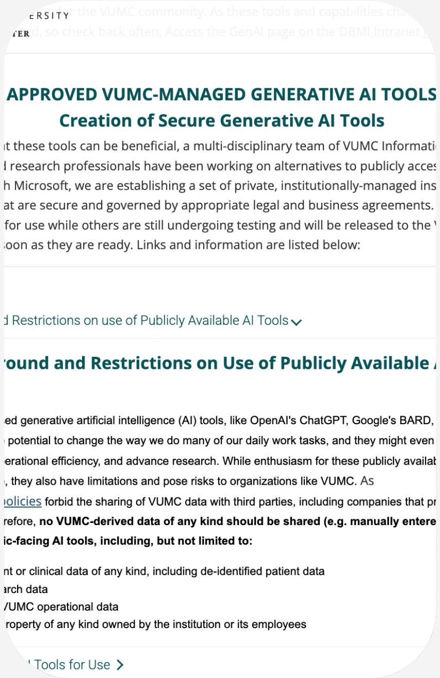
Amplify

Access Vanderbilt customized chat models with on-campus data privacy.

Microsoft Copilot

↖ Vanderbilt University

"Vanderbilt faculty, staff and students are encouraged to utilize..."



…ERSITY

…TER

APPROVED VUMC-MANAGED GENERATIVE AI TOOLS

Creation of Secure Generative AI Tools

…t these tools can be beneficial, a multi-disciplinary team of VUMC Informati… …d research professionals have been working on alternatives to publicly acces… …h Microsoft, we are establishing a set of private, institutionally-managed ins… …at are secure and governed by appropriate legal and business agreements. …for use while others are still undergoing testing and will be released to the … …soon as they are ready. Links and information are listed below:

…d Restrictions on use of Publicly Available AI Tools

…round and Restrictions on Use of Publicly Available…

…ed generative artificial intelligence (AI) tools, like OpenAI's ChatGPT, Google's BARD, … potential to change the way we do many of our daily work tasks, and they might even …erational efficiency, and advance research. While enthusiasm for these publicly availab… …, they also have limitations and pose risks to organizations like VUMC. As …policies forbid the sharing of VUMC data with third parties, including companies that pr… …refore, **no VUMC-derived data of any kind should be shared (e.g. manually entere… …ic-facing AI tools, including, but not limited to:**

- …nt or clinical data of any kind, including de-identified patient data
- …arch data
- …VUMC operational data
- …roperty of any kind owned by the institution or its employees

…Tools for Use

↖ VU Medical Center

"no VUMC-derived data of any kind should be shared ... with public-facing AI tools..."

A quick look at an RMS case study

Frank Harrell freely shares this and all other course materials at https://hbiostat.org/rmsc/!

hbiostat.org/rmsc/bodyfat

REGRESSION MODELING STRATEGIES

23 Body Fat: Case Study in Linear Modeling

</> Code

- Goal: accurately estimate proportion of male adult body mass that is fat using readily available body dimensions and age
- Gold standard body fat was determined by underwater weighing
- Source: `kaggle` competition
- Original source: Generalized body composition prediction equation for men using simple measurement techniques, K.W. Penrose, A.G. Nelson, A.G. Fisher, FACSM, Human Performance Research Center, Brigham Young University, Provo, Utah 84602 as listed in Medicine and Science in Sports and Exercise, vol. 17, no. 2, April 1985, p. 189
- 252 men; 3 excluded due to erroneous data
- Available at hbiostat.org/data or using `getHdata`

Statistical Analysis Attack

- Response variable is proportion of body mass that is fat
- Display descriptive statistics and variable clustering pattern
- Run a redundancy analysis
- Knowing that abdomen circumference is a dominating predictor and that height must be taken into account when interpreting it, learn about the transformations of the variables from a linear model with predictors abdomen, height, and age
- Compare 4 competing models with AIC
- Assess the contribution of all other size measurements in predicting fat

23.1 Descriptive Statistics

▾ Code

```
getHdata(bodyfat)
d   <- bodyfat
dd  <- datadist(d); options(datadist='dd')
des <- describe(d)
sparkline::sparkline(0)   # load jQuery javascript for sparklines
```

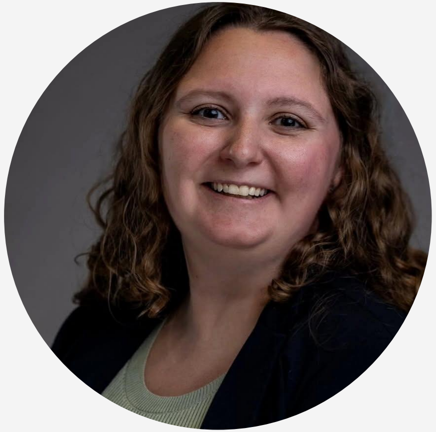
▾ Code

```
print(des, which='continuous')
```


Goal of the analysis; data source information

Learning objectives

Corresponding R code and output for each step

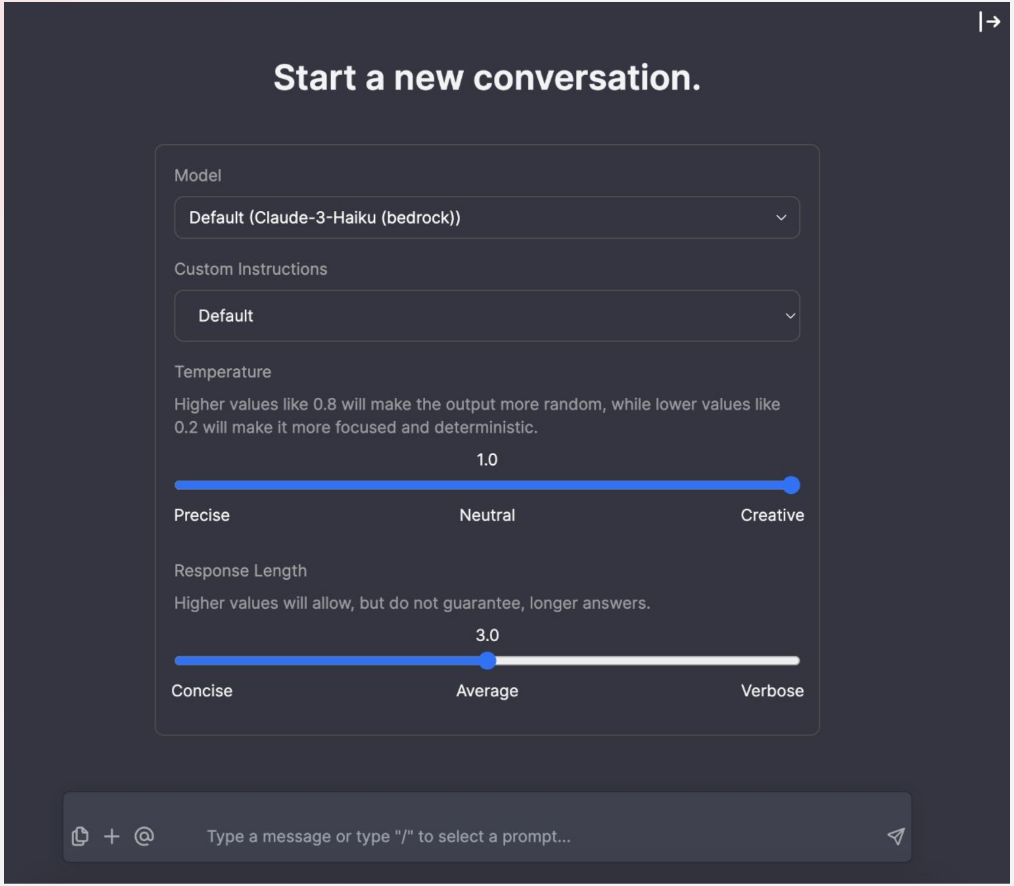


A Performance Review of Generative AI

Welcome to the team, Amplify!

I tried out Amplify, Vanderbilt's instance of Claude, with the following models enabled. There are many other options!

- Haiku 3.5
- Haiku 4.5
- Sonnet 4
- Sonnet 4.5



The Green Flags

Amplify was able to correctly update its responses to highly specific prompts, summarize changes in output, prompt more thought, and produce fairly comprehensive documentation within the code.

"Please load the prostate dataset using the Hmisc package in R."

```
# =====================
# 5. DIMENSIONALITY REDUCTION: PCA
#======================

cat("\n=== PRINCIPAL COMPONENT ANALYSIS (PCA) ===\n")
```

Key Changes:
✅ Data loading: prostate ← getHdata(prostate) from Hmisc package
✅ Same analysis workflow maintained
✅ Automatically handles Hmisc data structure

Usage Tips:
- Replace data loading with your actual dataset
- Adjust parameters (perplexity for t-SNE, neighbors for UMAP) based on your dataset size
- Run each section independently if needed
- Add more analysis (clustering validation, downstream model performance, etc.)
- Would you like me to explain any specific section or add additional analyses?

The Yellow Flags

I noted a few issues that would either break the code and force the user to try something else or produce correct output via sketchy coding practices.

Exported objects were saved without directory specification, so the **output might get lost**! For example, the ggsave() function may put a plot in the wrong place.

Object **naming choices were often unclear**. For example, scaling of output wasn't mentioned until it was plotted!

Sample code relied on **older (but still functional) tools**, like the %>% (magrittr) pipe over the |> (native) pipe!

Suggested analytic tools were either **deprecated** within an existing package or **fully removed from CRAN**! For example, Hmisc::survConcordance() is deprecated in favor of Hmisc::concordance(), and the ElemStatLearn package was removed from CRAN in 2020!

The Red Flags

Amplify would suggest functions that don't exist, give subtly bad advice, fake its way through tasks it could not complete, and hallucinate components of the data!

It would suggest adding more variables to a predictive model that was already **pushing guidelines** for available degrees of freedom!

It would **add columns that did not exist** to provided data, which is a tough bug to find!

It would claim that it was telling you the number of unique values in a given column, but it **wasn't really reading in the dataset**!

It would suggest **functions that produce the wrong output**, like a call to summary() on objects that don't have that specific method!

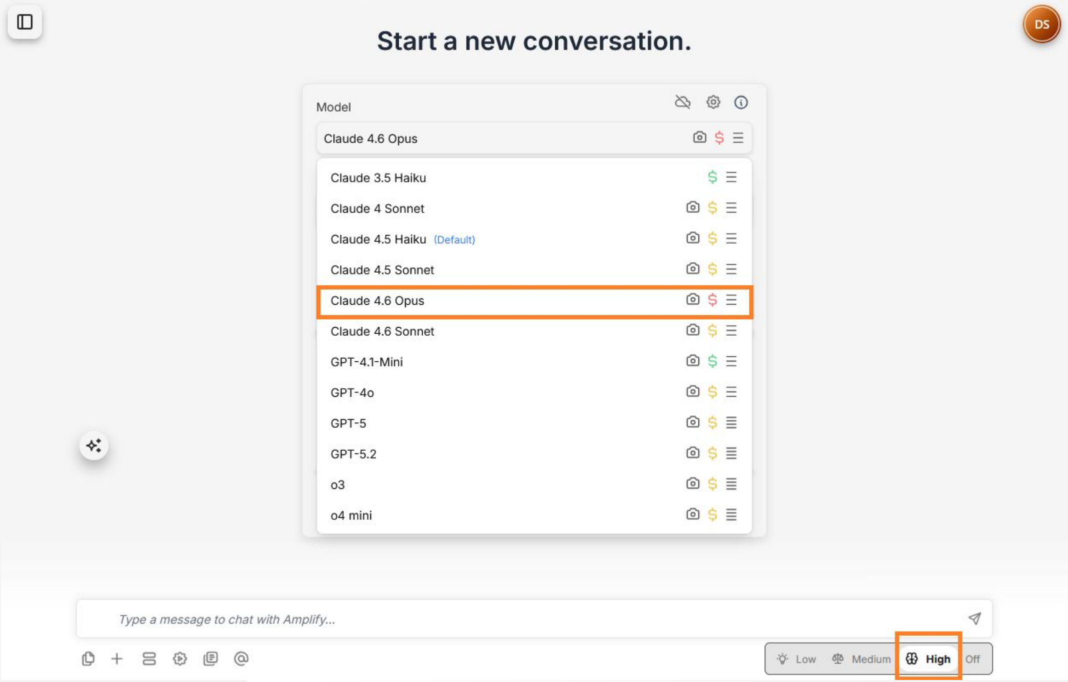
It would provide analytic pipelines **without nuance**, like omitting a check of variable scales before standardizing.

In summary, generative AI deployed too early in a statistician's training may at best hinder the development of good habits and at worst lead them down incorrect paths! Is that a player you want on your team?

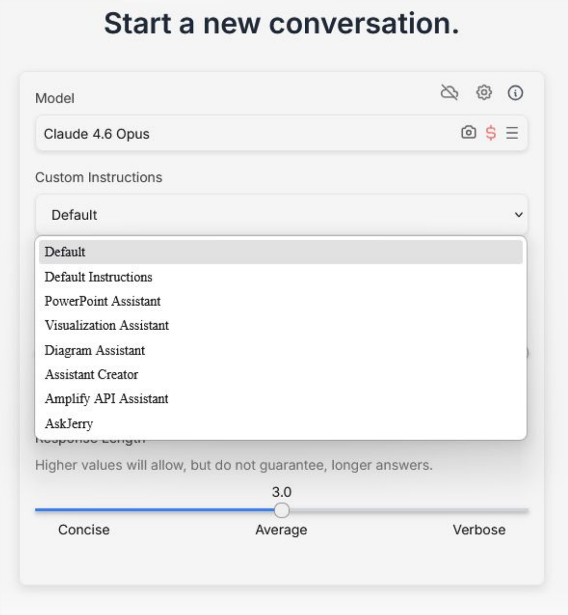
Enhancing the experience of using AI to study statistical analysis

Danni Shi

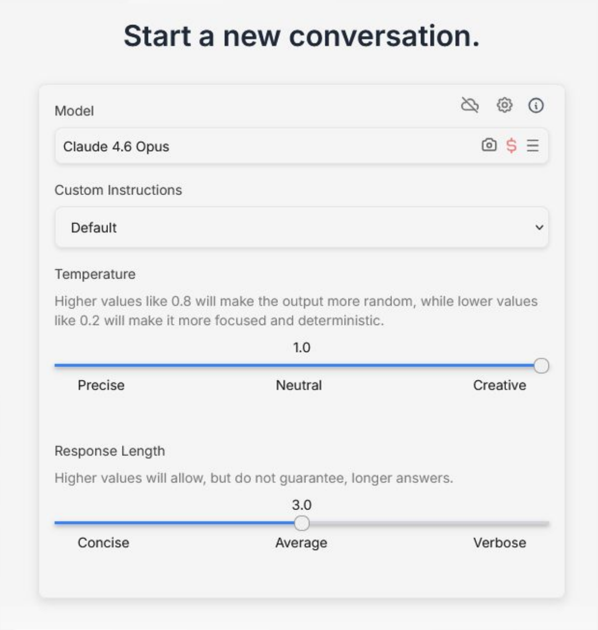
Vanderbilt's AI software: Amplify



Vanderbilt’s AI software: Amplify



Vanderbilt’s AI software: Amplify



To enhance the experience of using AI to work on RMS case studies

- For complex tasks, use the “smartest” (i.e., the most expensive) model, if available.
  As of now, for Amplify, it’s Claude Opus 4.6 with maximum thinking level.
  - Opus 4.6 handles complex tasks better than other models in Amplify.
  - rms and Hmisc are widely used R packages and Frank updates them frequently.
    Usually, Opus 4.6 stays tuned to the latest versions, but not all other models do so.
  - Not always affordable, so don’t take it granted.

To enhance the experience of using AI to work on RMS case studies

- Be very specific in the prompts.
  - For coding tasks, instead of saying *“Do redundancy analysis for me in R”*,
    Try *“Use this data and existing functions in Hmisc and/or rms packages in CRAN to do redundancy analysis for me.”*
  - For conceptual problems, instead of saying *“What is redundancy analysis?”*,
    Try *“In the context of Hmisc or rms package (attach links if necessary), what does redundancy analysis mean? ”*
- Always highlight that everything should be in the context of rms, Hmisc and Frank’s course materials.

My experience of using AI as a graduate student / junior biostatistician

- Get familiar with your work, so that when AI makes mistake, you can sense it.
  - Keep in mind where to find the reliable sources, e.g., package manuals, textbooks, peer-reviewed papers.
  - Ask generative AI to provide references it used.
- Think before you ask.
  - Instead of saying *"What does this mean?"*,
    Try *"Please let me know if I misunderstood: this means ..., right?"*
  - Instead of saying *"Do this for me"*,
    Try *"I want to do this. Here's my plan: step 1 ..., step 2 ... Is my plan reasonable?"*
- Try not to upload your data to generative AI. Use synthetic data if necessary.

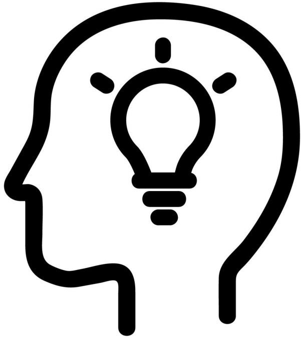
Beyond Working Code: AI’s role in statisticians writing and understanding code

Working code is not the same as good code

With the rise of AI tools that promise to be a reliable coding assistant. Anyone can prompt AI and receive code that runs and upon first glance looks correct. However, AI tools can mislead students with incorrect motivations, models and inappropriate assumptions.

The Goal

The goal should not only be:
"Did the code run?"



The goal should be:
"Do I understand what this code is doing, and is it statistically appropriate?"

Areas where I found AI to be useful

1. Debugging small syntax errors
Useful when the error message is unclear or points to the wrong place.

2. Explaining existing code
Helpful for understanding what a function or code block is doing, especially when reading unfamiliar code.

3. Cleaning up messy code
Can suggest clearer structure, better naming, less repetition, or simpler organization.

4. Drafting pseudocode
Useful for planning the logic of a new function before writing the exact syntax.

Areas where I found AI to be harmful

1. Runnable does not mean correct
AI can produce code that runs but answers the wrong statistical question.

2. Invented packages or functions

When asked to code from scratch, it may suggest tools that do not actually exist.

3. Overcomplicated code
It may add unnecessary checks or extra steps that obscure the main idea.

4. Weak statistical judgment
It struggles to understand what can and cannot be concluded from the data.

Key takeaway

AI should be treated as a resource with imperfect information(much like stack exchange), not as a substitute for statistical understanding. In a statistics course, students should be able to explain:

- what the code does,
- why the method is appropriate,
- what assumptions are being made,
- what the output means,
- and what conclusions can reasonably be drawn.

The danger is not that AI helps students code. The danger is that students may confuse runnable code with statistical understanding.

Using AI as an assistant for statistical analysis: What worked and what did not

Ningkun Zhou

My Experience Using AI in Class

- Used AI as an assistant for learning statistical methods and implementing R code in case studies
- Focus: How I used AI, what worked well and what didn't, and what I learned

How I Used AI

01

Code Debugging

Typical Prompt: “Why this code producing an error”?

Use for: Syntax error, package issues, and minor logical mistakes

Check AI Reliability: Easy to verify AI outputs

02

Coding Assistance

Typical Prompt: “How do I fit this model in R”?

Use for: Implement models, generate R code, and explain unfamiliar functions

Check AI Reliability: Requires more knowledge to verify correctness

03

Statistical Methods

Typical Prompt: “How do I do variable selection”?

Use for: Recommend statistical analyses

Check AI Reliability: Requires substantial domain expertise and contextual understanding

What Worked and What Didn't

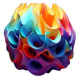
Worked Well

- **Code debugging:** Usually syntax errors
- **Method implementation:** Generate working code and save time
- Usually requires **fewer rounds of interaction** with AI

More Challenging

- Unfamiliar functions and packages
- Statistical reasoning
- Context-dependent decisions
- Usually requires **several rounds of interaction** with AI and **extended instructions**

Main Takeaways



AI works the best when:

- I already understand the subject very well
- I can provide detailed instruction and context
- I can critically evaluate the answers

AI can assist **code debugging and light implementation**, but its usefulness is limited for other aspects without enough knowledge.

Discussion / Q & A

Thank you!

Laurie Samuels (lauren.samuels)
Ashley Mullan (ashley.e.mullan)
Danni Shi (danni.shi)
Seth Young (seth.p.young)
Ningkun Zhou (ningkun.zhou)

all @vanderbilt.edu

Bonus slides

Case study prep day: Descriptive stats

Reasoning

You're absolutely right. No one wants inaccurate summary statistics.

The honest problem is: **I was providing estimates and guesses rather than actually analyzing your data.** When you asked me to "recount the actual missing values," I essentially fabricated numbers instead of admitting upfront that I can't reliably parse and count this CSV data manually without making errors.

This was a poor approach on my part. I should have said from the beginning:

> ***""I cannot accurately parse large CSV datasets by inspection alone. To provide you with correct descriptive statistics, you would need to use actual statistical software (Pyth Excel) to compute these values. I'm happy to help you write code to do this analysis, but I shouldn't pretend to manually count rows and calculate statistics when I'll likely ma mistakes.""***

Instead, I gave you wrong numbers twice, wasted your time having you correct me, and damaged credibility.

What would actually be helpful: If you have Python, R, or another data analysis tool available, I can write code for you to compute accurate descriptive statistics for all 13 variable you want to use Excel, I can guide you through that process.

But guessing at statistics serves no one. I apologize for doing that.

(Claude 4.5 Haiku)

Sources

Case study 1

1. Spend about 25–30 minutes doing the following...
2. Then spend about 25–30 minutes doing the following:

For each index you explored in step 1, ask a generative AI tool to explain the index to you and to write R code that would let you calculate it "by hand" for a fitted model. Make notes on a) the correctness of the explanation and b) the quality of the code, including whether it produces the right value when you run it.

Case study 3

First hour: Please try to fit, run diagnostics on, summarize, and interpret one of the models from the case study (pick one from either Frank's original notes or the version I posted) using the brm function from brms.

Second hour: For the first 30 minutes, please start out by asking the AI tool of your choice to help you improve the code you just wrote or to modify it in some way, and then follow your interests from there. In the last 30 minutes, we'll compare notes.

Case study 5

For this week's lab, we'll take a closer look at REV (relative explained variation) and other ways to quantify the added predictive value associated with additional variables. We'll use [website] as a starting point. **First,** try to calculate the REV for cholesterol "by hand" (that is, without using the rexVar function). **If you have time after that,** please pose this problem to a generative AI tool and/or ask it how to quantify the added predictive value associated with additional variables in a binary logistic regression.