Critical assessment of ChatGPT data analysis "skills"

Feedback

Eva Cantoni, Anna van Es

University of Geneva

Umeå University, June 9th, 2025

Motivation

UNIGE Statement on artificial intelligence (28.05.2025, evolving):

> *The University of Geneva (UNIGE) supports the development and use of artificial intelligence (AI). Above all, it sees it as an opportunity, based on the principle that the risks associated with it can be minimized through a responsible approach.*

Principles for the use of generative artificial intelligence :

- Responsibility
- Legality
- Academic integrity and transparency
- Economy and ecology

Motivation

Teachers interest on two questions:

- How to implement University of Geneva guidelines
- What are the “skills” of ChatGPT in terms of data analysis

Experimental setting

Setting

- Compulsory course «Modern Flexible Regression» of the master in statistics (90-120 ECTS). Students take this course either at the 2nd or the 4th semester.
- Spring 2024: 12 students. Spring 2025: 20 students.

Goal

- Explore the potential and pitfalls of ChatGPT to provide practical experience in generative AI-assisted data analysis.

Project goals

The primary goal of the project is to empower Master of Statistics students to explore the potential of ChatGPT for comprehensive data analysis. The project will involve leveraging ChatGPT's capabilities to extract insights, generate hypotheses, and derive meaningful conclusions from a given dataset. Through this exploration, students will gain hands-on experience in integrating generative artificial intelligence (genAI) into statistical analysis workflows.

The project provides an opportunity for students to bridge the gap between traditional statistical methods and emerging AI technologies, fostering a well-rounded skill set crucial for today's data-driven landscape.

Datasets

A dataset provided for each group (3-4 students), together with a scientific question of interest.

- Real data sets available on R packages, modified, reduced size
- Simulated datasets, inspired by real situations, including issues of outliers, missingness not at random, collinearity, etc.

An example: A study of hospital visit frequency among chronic patients (available information: HospitalVisits, Age, BMI, DietType, SmokingStatus, ExerciseHours, Height, Weight, Healthcare). Research question: Design a preventative healthcare campaign targeting high-risk individuals who frequently visit hospitals. Choose one important risk factor to focus your campaign on.

Instructions

The general instruction is to use ChatGPT to answer the scientific question of interest based on the provided data, with the methodological and statistical tools of the course and with the R software.

Either the free version or the commercial version of ChatGPT were allowed, with the need to declare which version has been used. Which version is used had no impact on the grade.

Deliverables

Each group participating in the project is expected to hand in:

- The history of the conversation with ChatGPT (maximum of 10 queries, unlimited).
- A critical assessment on the ChatGPT analysis (maximum of 3 pages, minimum font size 12).
- Scientific argumentation (1 page)
- Reflection on the group collaboration (1 page)
- The slides of the oral presentation to follow, to be discussed by a group of peers
- The R code with output, if relevant to motivate the assessment of the model proposed by ChatGPT.

Learning goals

At the end of the project, students will be able to:

- Apply statistical techniques to identify patterns and trends within the dataset
- Formulate meaningful questions and hypotheses to properly analyze a data set
- **Integrate ChatGPT into the statistical analysis process to enhance data interpretation**
- **Evaluate the strengths and limitations of using generative AI for data-driven insights**
- Collaborate with peers to improve the overall quality of the analysis
- Communicate findings by delivering an effective oral presentation
- Manage time and resources effectively throughout the analysis process

Assessment and grading

Evaluation based on the quality of analysis, effective use of ChatGPT, collaboration with peers, and the clarity and persuasiveness of the oral presentation.

Bonus (to add to the grade of the final written exam), as the sum of a group grade (up to 0.5) based on the written document and an individual grade (up to 0.25) based on the oral presentation.

(Group) grade that counts for 30%, final exam counts for 70% .

- 10% for the scientific argumentation
- 50% for the critical assessment
- 10% for the reflection on the collaboration
- 10% for the overall quality of the report
- 20% for the oral presentation

Teachers's feedback 2024

- Positive acceptance of the project by students (voluntary basis, all participated).
- Various strategies adopted by students (e.g., novice vs. experienced statisticians), resulting in more or less informed approaches.
- Data analysis task too easy for ChatGPT
- Some disappointment with the quality of the critical evaluation provided by the students: most of the comments were very general. (Generated by ChatGPT?)

Critical assessment details 2025

The critical assessment should address the following:

> *Assess the relevance, strengths, and weaknesses of ChatGPT's responses. More specifically, discuss the accuracy and coherence of ChatGPT's statistical choices, the validity of the applied methodologies in relation to the course content, the robustness of the interpretations, any potential biases, errors, or inconsistencies in the generated response. Discuss also whether ChatGPT's analysis aligns with statistical and methodological best practices, and identify any limitations or inconsistencies.*

Teachers's feedback 2025

- Students creativity (lay persons vs educated person, genAI vs master student, ...)
- More challenging task for ChatGPT (and students as well)
- Importance of good prompts
- ChatGPT performance:
  - struggles with variables types, e.g. categorical variables, if not precisely instructed
  - In general, missingness issue not properly dealt with
  - In the health visit example: 1) impressive reasoning on "modifiable" variables 2) bias: strongly advocating to target BMI (based on the large existing literature) even though other variables where more appropriate (statistical and practical significance) based on the data at hand

Research project 2025 (voluntary basis)

Prof. G. Molinari (TECFA, UNIGE) : Research in the Computer-Supported Collaborative Learning (CSCL) field.

The aim of the research is to study how students in groups use ChatGPT to analyze and interpret data. We will focus on how the use of ChatGPT influences group dynamics and on the students' experience of working with ChatGPT in groups.

Collected data:

1. Recording of the online work sessions with ChatGPT
2. Interactions with ChatGPT
3. Responses to two questionnaires (pre and post)
4. Access to the project deliverable

Interdisciplinary research project

Prof. G. Molinari (TECFA, UNIGE) with 6 other professors from different fields (social sciences, translation and interpreting, computer science, psychology)

Research project on

The transformation of skills through Language Models as seen by Swiss academic institutions

Selected references

Al Labadi, L. and Ly, A. (2025).

Enhancing statistics education through Project-Based Learning (PBL) and the emergence of ChatGPT.

Teaching Statistics.

Schwarz, J. (2025).

The use of generative AI in statistical data analysis and its impact on teaching statistics at universities of applied sciences.

Teaching Statistics.

Tu, X., Zou, J., Su, W. J., and Zhang, L. (2023).

What should data science education do with large language models.

arXiv preprint arXiv:2307.02792, 3.

THANK YOU FOR YOUR ATTENTION

www.shutterstock.com · 641809102