

# Analysis of class surveys

## Part 1: Is self-reported statistics proficiency associated with the number of PSTAT courses completed?

Justin Zhou,Ella Yang,Lucy Cao,Cecilia Jiang,Janice Jiang,Wendy Zhu

2025-10-19

### Table of contents

<b>1</b>	<b>Executive summary</b>	<b>1</b>
<b>2</b>	<b>Data description</b>	<b>2</b>
<b>3</b>	<b>Question of interest</b>	<b>2</b>
<b>4</b>	<b>Findings (Q1)</b>	<b>2</b>
4.1	Proficiency composition . . . . .	3
4.2	PSTAT courses by proficiency (boxplot) . . . . .	4
4.3	Average PSTAT courses (mean $\pm$ SD) . . . . .	5
4.4	Average PSTAT courses (trend) . . . . .	6

### 1 Executive summary

We analyzed anonymized intake-survey responses from two sections to answer one question: **Is self-reported statistics proficiency associated with the number of PSTAT courses completed?** Using descriptive plots and a one-way ANOVA with Tukey's post-hoc test (implemented in `Question_1.Rmd`), we find evidence of a statistically significant association: respondents reporting higher statistics proficiency have, on average, completed more PSTAT courses. Because the responses are a **sample** of offered-seat students, we emphasize clear visuals and center-focused summaries appropriate for modest (n).

## 2 Data description

- **How obtained.** Google Forms intake survey; exported and shared as two section files with identifiers removed plus a metadata file.
- **Sample.** Respondents from two course sections; treated as a **sample** from all offered-seat students (not necessarily a census).
- **Measurements.** Demographics; self-assessed **proficiency** and **comfort** (Likert 1–5); indicators for prior coursework; and project preferences. For Q1, we use `stat.prof` (statistics proficiency; ordered beginner → intermediate → advanced) and `pstat_courses_total` (sum of PSTAT course indicators per student).

## 3 Question of interest

1. **Q1.** Does self-reported statistics proficiency relate to the number of PSTAT courses completed?

## 4 Findings (Q1)

**Question.** Is self-reported statistics proficiency associated with the number of PSTAT courses completed?

**What we did.** We summarized the distribution of proficiency levels (Beginner, Intermediate, Advanced) and compared total PSTAT coursework across groups using descriptive plots (composition chart, boxplot, and bar chart of means with uncertainty). We then ran a one-way ANOVA and Tukey’s HSD post-hoc comparisons.

## 4.1 Proficiency composition

**Distribution of Students by Statistics Proficiency Level**

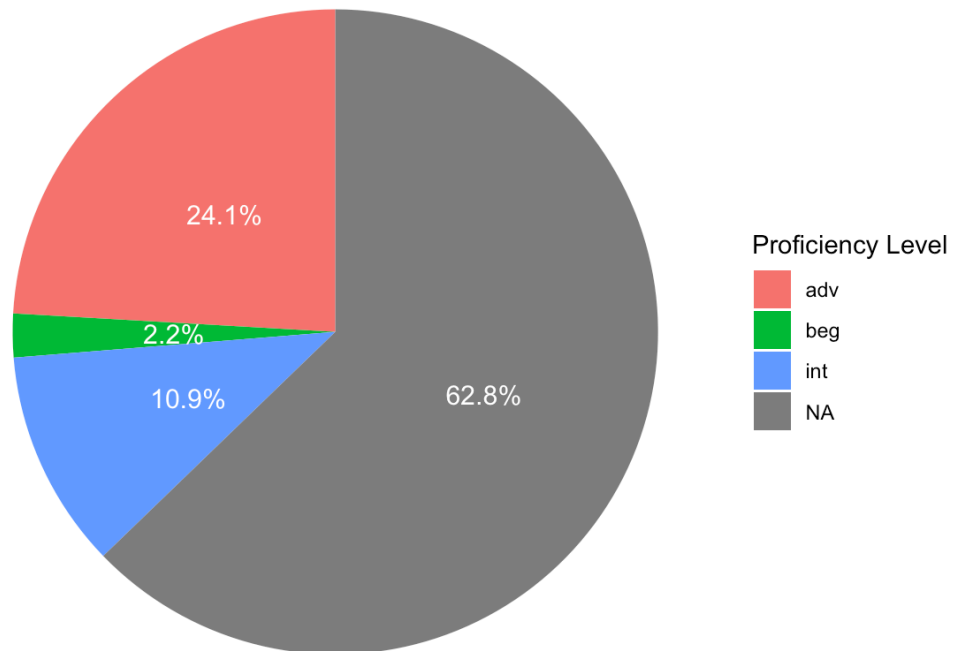


Figure 1: Distribution of Students by Statistics Proficiency Level

## 4.2 PSTAT courses by proficiency (boxplot)

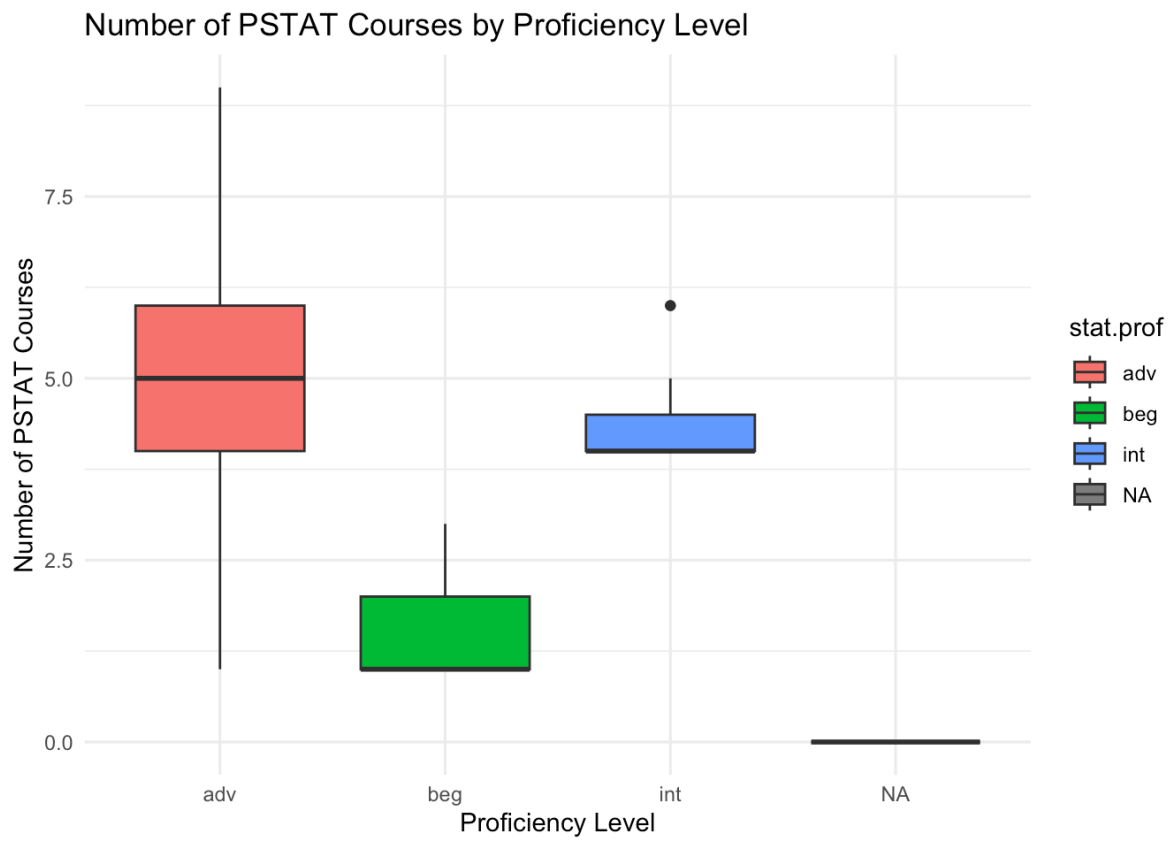


Figure 2: Number of PSTAT Courses by Proficiency Level

### 4.3 Average PSTAT courses (mean $\pm$ SD)

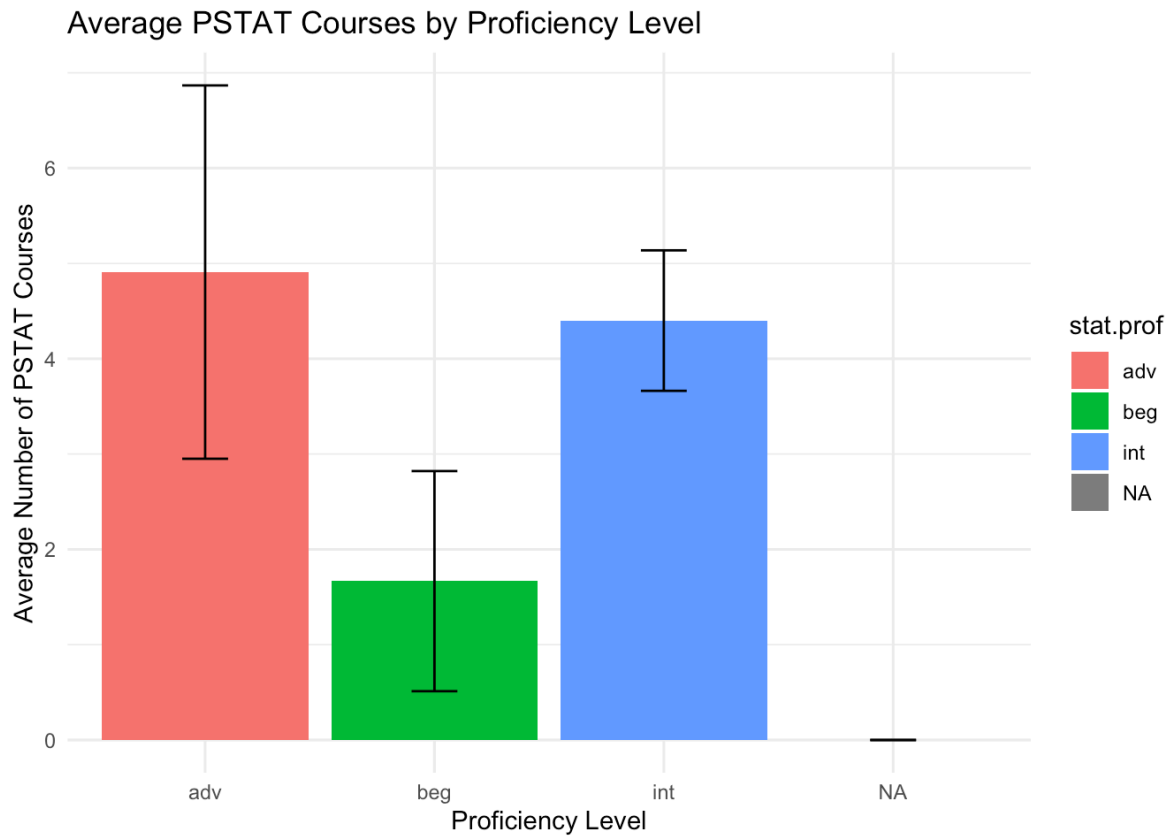


Figure 3: Average PSTAT Courses by Proficiency Level

#### 4.4 Average PSTAT courses (trend)

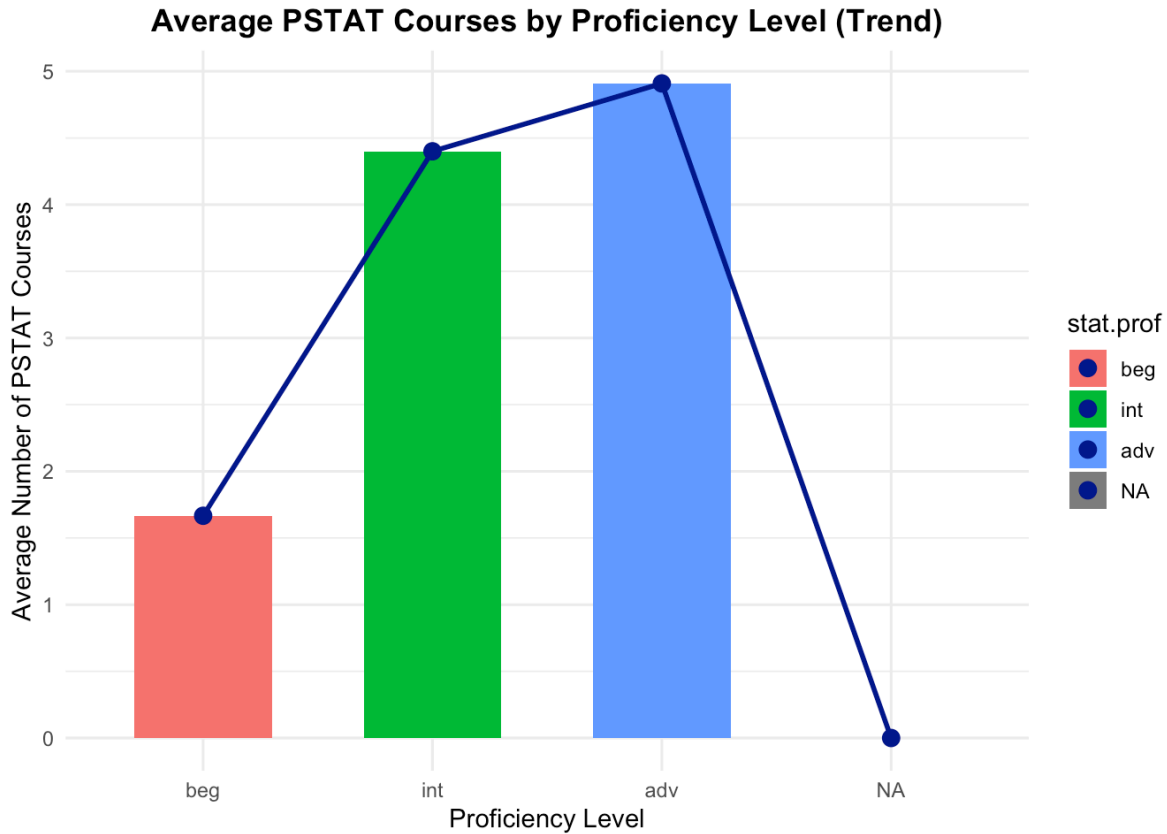


Figure 4: Average PSTAT Courses by Proficiency Level (Trend)

**Results.** - Visual summaries show an upward trend: students reporting higher statistics proficiency tend to have taken more PSTAT courses. - One-way ANOVA:  $p = 0.008$  (less than  $\alpha = 0.05$ ), indicating a statistically significant difference in coursework totals across proficiency levels. - Tukey's HSD pairwise comparisons: - Advanced vs Beginner:  $p = 0.0062$  (significant). - Intermediate vs Beginner:  $p = 0.033$  (significant). - Intermediate vs Advanced:  $p = 0.59$  (not significant).

**Interpretation.** Within this sample, greater exposure to PSTAT coursework is associated with higher self-reported statistics proficiency, especially when comparing Beginners to Intermediate and Advanced students. We treat these results as descriptive evidence from the sample of offered-seat students and avoid over-interpreting extremes; the focus is on clear, center-based summaries supported by standard inference.