Constraints, Incentives, And Consumer Behavior Portfolio

Keean Kawai

UC Berkeley

Economics & Data Science Student

keeankawai@berkeley.edu

Table of Contents

Consumer Behavior Survey Modeling & Analysis | Overview

Timeline: March 2025 – May 2025

Context

- Designed and administered a structured consumer behavior survey to **60 SMCCD students**
- Collected original data on **monthly income, discretionary spending, savings behavior, and financial literacy**

Objectives

- Analyze the relationship between **income levels and total discretionary spending**
- Identify **spending allocation patterns** across income quartiles by category
- Evaluate how **financial literacy correlates with savings outcomes**

Methods

- Performed extensive **data cleaning, transformation, and feature engineering** in Excel
- Constructed income specifications **including and excluding zero income observations** for robustness
- Conducted **income vs. discretionary spending analysis** and **two sample t tests**
- Calculated **category level spending allocation by income quartile** using rounded monthly income
- Applied **boxplot and distributional analysis** to compare savings across self rated financial literacy groups

Survey Dataset Structure and Variables

Preview of Entire Dataset

Student ID	Monthly Income Estimated to 100	Food Estimated to 10's	Entertainment	Technology	Clothing	Skincare	Self-Rated Financial Literacy Score (
1	700	250	150	0	120	40	5
2	1000	200	50	80	150	20	3
3	1100	310	200	200	200	0	1
4	3000	500	0	110	0	0	4
5	0	200	300	80	400	10	3
6	0	110	70	0	90	50	3
7	800	300	150	0	300	80	3
8	1200	280	90	100	100	120	4

Total Discretionary Spending	Savings	Income Quartile	Food Share	Entertainment	Technology	Clothing	Skincare	Literacy Group (Low [1,2], Mid
560	140	Q2	0.446428571	0.267857143	0	0.2142857	0.07142857	High
500	500	Q3	0.4	0.1	0.16	0.3	0.04	Mid
910	190	Q3	0.340659341	0.21978022	0.21978022	0.2197802	0	Low
610	2390	Q4	0.819672131	0	0.180327869	0	0	High
990	-990	Q1	0.202020202	0.303030303	0.080808081	0.4040404	0.01010101	Mid
320	-320	Q1	0.34375	0.21875	0	0.28125	0.15625	Mid
830	-30	Q2	0.361445783	0.180722892	0	0.3614458	0.09638554	Mid
690	510	Q3	0.405797101	0.130434783	0.144927536	0.1449275	0.17391304	High
680	-180	Q2	0.279411765	0.073529412	0	0.5735294	0.07352941	Low

****All variables cleaned, standardized, and derived from original survey responses.***

Methodology and Experimental Design

Survey Design (Data Collection)

- Designed and deployed a structured consumer behavior survey to **60 SMCCD students**
- Collected standardized inputs on:
  - **Monthly income** (including zero income observations)
  - **Category level discretionary spending** (Food, Technology, Entertainment, Clothing)
  - **Savings behavior**
  - **Self Rated Financial Literacy** (grouped for comparative analysis)
- Ensured consistent units and comparable responses by using a fixed time window (**monthly**) across all questions

Data Preparation and Feature Engineering

- Executed comprehensive **data cleaning and validation** to remove nonresponses (N/A) and enforce numeric formatting
- Standardized income by **rounding monthly income to the nearest hundred** to reduce reporting noise and support stable grouping
- Engineered core analytical variables:
  - **Total Discretionary Spending** = sum of category expenditures
  - **Spending Allocation Shares** = category spend divided by total discretionary spend (used for composition analysis)
  - **Income Quartiles** based on monthly income for segmentation and distributional comparison

Analytical Strategy (What Was Tested)

- **Income vs. Total Discretionary Spending**
  - Modeled the relationship between income and total discretionary spending using:
  - A baseline specification **excluding zero income observations**
  - A robustness specification **including zero income observations**
    - This dual approach tests whether results are driven by edge cases or persist across the full sample

Two Sample Hypothesis Testing (t Test)

- Conducted **two sample t tests** to evaluate whether mean discretionary spending differs across income defined groups
- Reported statistical significance and used results to support or reject observed spending gaps

Quartile Based Allocation Analysis

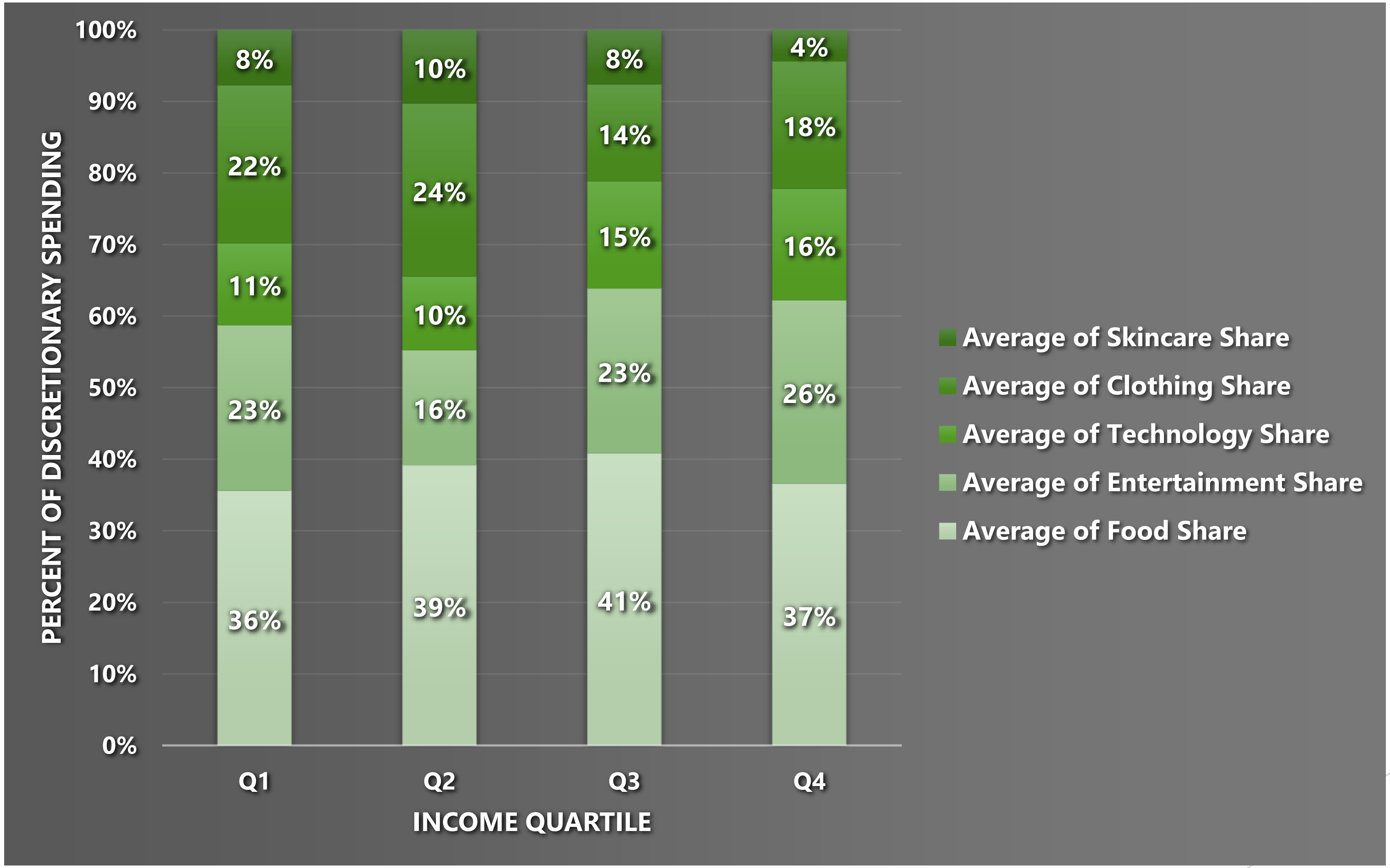
- Segmented students into **income quartiles**
- Computed average discretionary spending by category within each quartile to estimate how consumption composition shifts as constraints relax
- Interpreted category sensitivity as a proxy for **income elasticity** in spending behavior

Savings Distribution by Financial Literacy (Boxplots)

- Grouped respondents by **self rated financial literacy**
- Used **boxplot distribution analysis** to compare savings dispersion, medians, and outliers across groups
- Focused on distributional differences rather than only means to capture behavioral heterogeneity

Allocation of Discretionary Spending by Income Quartile

Normalized Bar Chart Modeling



Income Quartile Definition

Q1: Monthly income below $200
Q2: $200 to $800
Q3: $800 to $1,285
Q4: $1,285 and above

Allocation of Discretionary Spending by Income Quartile

Normalized Bar Chart Analysis

Key Observations

Food spending dominates across all income quartiles, accounting for roughly **36% to 41% of discretionary budgets**

- This matters because **food** represents a **near necessity good**, so even as income rises, its **budget share** remains relatively stable

Technology and entertainment shares increase with income, rising from **11% to 16% and 23% to 26% respectively from Q1 to Q4**

- This pattern exists because higher income relaxes constraints, allowing spending to shift toward higher elasticity, non-essential categories

Clothing share declines in middle income groups before rebounding at the top quartile, **indicating non-linear consumption behavior**

- Essentially, mid income students prioritize substitution and budgeting, while higher income students re expand discretionary choice

Skincare spending remains a small but persistent category, with the **highest share in lower income quartiles**

- This suggests skincare may function closer to a routine maintenance good rather than a luxury good for students

Analytical Insight

These allocation shifts matter because **consumer choice theory** predicts that as income constraints loosen, spending reallocates toward categories with higher perceived **marginal utility**

By using income quartiles rather than raw income, the **analysis highlights relative financial position**, which better explains compositional spending changes than absolute income alone

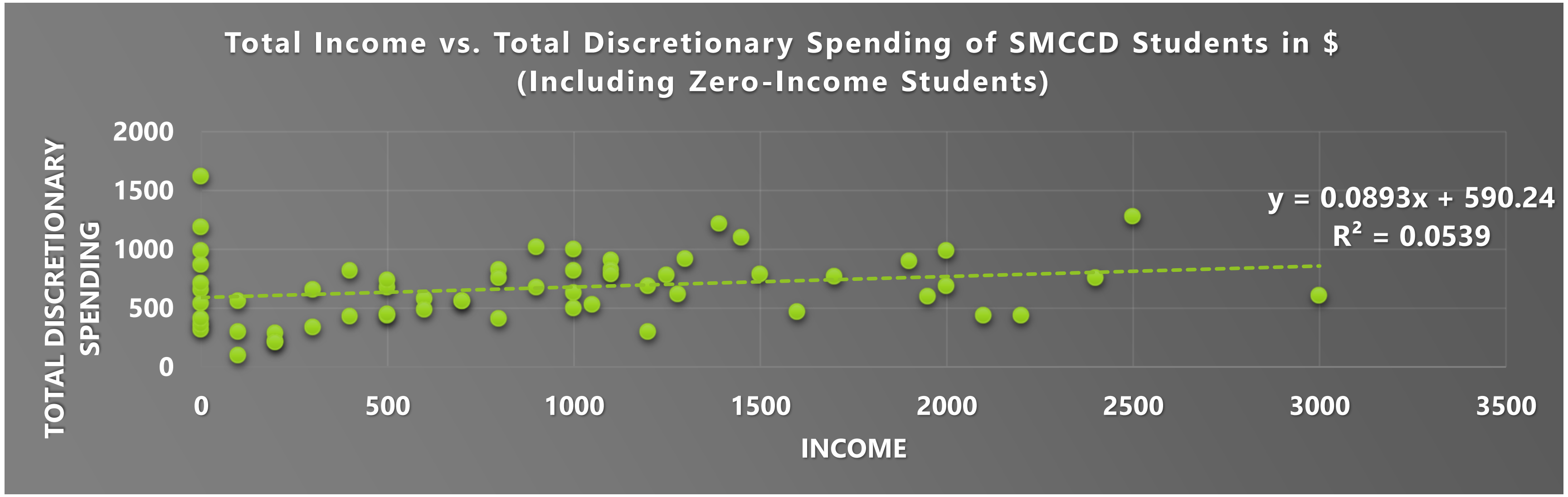
Why This Matters for Analysis

This chart demonstrates **category level income elasticity**, not just total spending differences

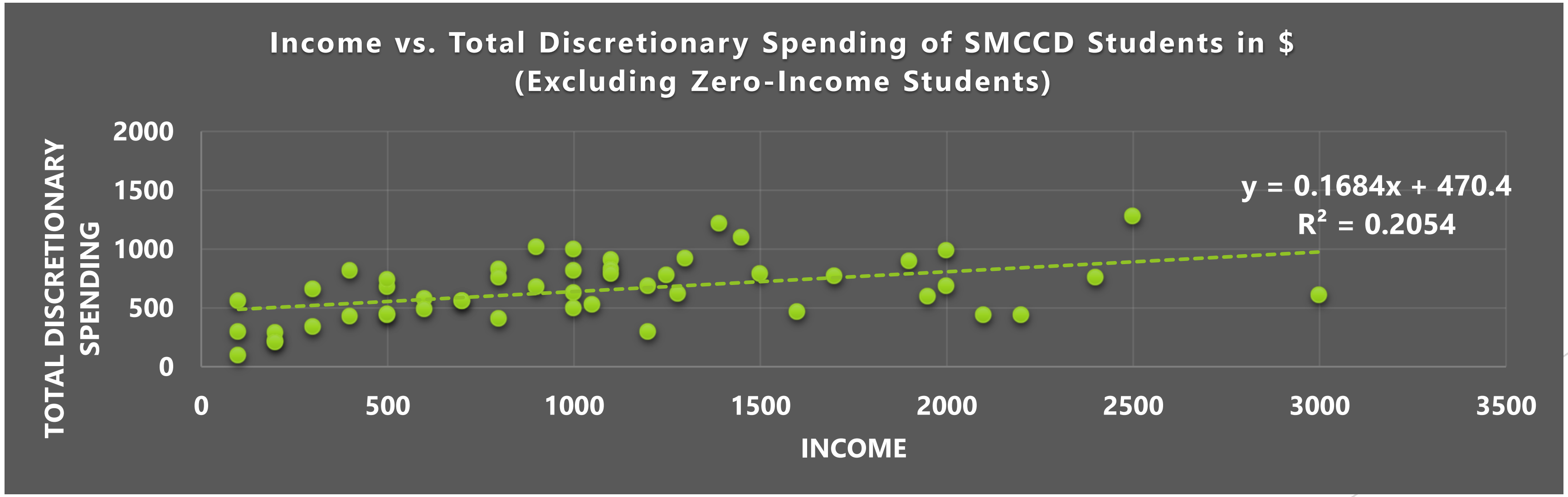
It shows that income affects ***what students spend on*****, not only *****how much they spend***, which is critical for consumer behavior and **market segmentation analysis**

Income vs. Total Discretionary Spending

Robustness Analysis Including and Excluding Zero-Income Students



N=60



N=50

Income vs. Total Discretionary Spending

Analytical Results and Distributional Context

Descriptive Statistics (Context)

Full Sample (Including Zero-Income Students)

- *Income*: Mean $889.5 | Median $800 | SD $751.6
- *Discretionary Spending*: Mean $669.7 | Median $660 | SD $289.1

Excluding Zero-Income Students

- *Income*: Mean $1,067.4 | Median $1,000 | SD $697.4
- *Discretionary Spending*: Mean $650.2 | Median $645 | SD $259.2

Key Regression Results

•Including Zero-Income Students

- *Line of best fit*: *Spending = 0.0893 × Income + 590.24*
- R^2 = **0.0539**
- Interpretation: Income explains **~5% of variation** in discretionary spending

•Excluding Zero-Income Students

- *Line of best fit*: *Spending = 0.1684 × Income + 470.4*
- R^2 = **0.2054**
- Interpretation: Income explains **~21% of variation** in discretionary spending among earners

Analytical Setup

•Evaluates the relationship between monthly income and total discretionary spending

•Two specifications used to test robustness:

- Full sample, including 10 zero-income students (**n = 60**)
- Income-earning subsample excluding zero-income students (**n = 50**)

Descriptive statistics reported to contextualize variance and model fit

Interpretation

•Income displays substantially higher variance than discretionary spending, which limits linear explanatory power

•Discretionary spending remains relatively stable even as income increases, indicating budget smoothing behavior

•Removing zero-income students nearly quadruples explanatory power, confirming strong sample composition effects

•Once income becomes a binding constraint, discretionary spending responds more predictably to income changes

Why This Matters

•Confirms that income affects spending conditionally, not uniformly

•Demonstrates disciplined use of robustness checks, distributional analysis, and economic reasoning

•Strengthens confidence that observed patterns reflect behavior, not data artifacts

Hypothesis Testing: Income Group Differences in Discretionary Spending

Two-Sample t-Test Assuming Unequal Variances

Analytical Setup

•Objective: Test whether *mean discretionary* spending differs across income groups

•Test used: *Two-sample t-test assuming unequal variances*

- This choice matters because spending variance differs across groups, and assuming equal variance would bias inference

•Hypotheses:

- Null hypothesis (H_0): Mean discretionary spending is equal across income groups
- Alternative hypothesis (H_1): Mean discretionary spending differs across income groups

Test Results

•Group 1 Mean: **$242.78**

•Group 2 Mean: **$374.00**

•Sample sizes: $n_1 = 18$, $n_2 = 25$

•Degrees of freedom: **40**

•t-statistic: **−0.5731**

•Two-tailed p-value: **0.5698**

•One-tailed p-value: **0.2849**

•Critical value (two-tailed, $\alpha = 0.05$): **2.0211**

Decision

•Fail to reject the null hypothesis at conventional significance levels

•Observed mean difference is ***not statistically significant***

Interpretation

•Although higher income students exhibit higher average discretionary spending, the difference is not statistically distinguishable from zero

•This occurs because within-group spending variance is large relative to between-group mean differences

•Essentially, income explains some variation in spending levels, but not enough to generate statistically significant group separation in this sample

•This reinforces earlier regression results showing moderate to low explanatory power, even when positive trends are present

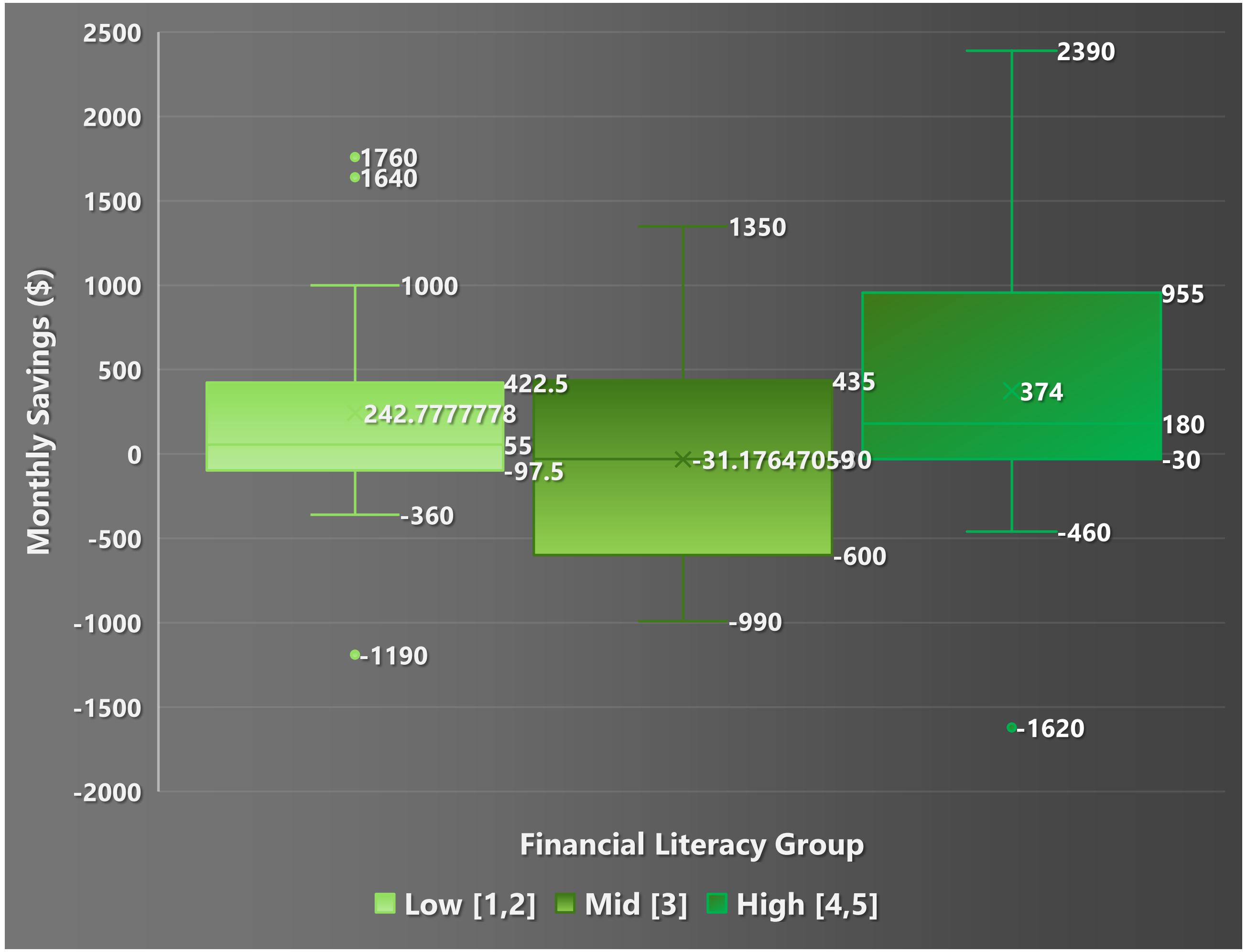
Why This Matters

•Confirms that observed income-spending relationships are *directional rather than definitive*

•Demonstrates correct use of ***inferential statistics***, not just descriptive or visual analysis

•Shows analytical discipline by distinguishing economic intuition from statistical evidence

Saving Distributions by Financial Literacy Group
Boxplot Distribution Modeling



Analytical Setup

Respondents grouped by self-rated financial literacy:

- *Low literacy*: ratings 1–2
- *Mid literacy*: rating 3
- *High literacy*: ratings 4–5

Outcome variable: monthly savings, including positive savings and debt

Visualization method: boxplot distribution analysis to capture medians, dispersion, and outliers

Saving Distributions by Financial Literacy Group
Boxplot Distribution Analysis

Key Distributional Results

•**Low Financial Literacy**

- Mean: ~$243 | Median: $55
- Q1: −$97.5 | Q3: $422.5
- Interpretation: Positive average driven by upper-tail outliers, while typical students save little

•**Mid Financial Literacy**

- Mean: ~−$31 | Median: ~−$31
- Q1: −$600 | Q3: $435
- Interpretation: Largest downside risk and weakest central tendency, indicating unstable savings behavior

•**High Financial Literacy**

- Mean: ~$374 | Median: $180
- Q1: −$30 | Q3: $955
- Interpretation: Higher typical savings and a strong upper tail, despite occasional negative outcomes

Comparative Analysis

•Median savings increase monotonically with financial literacy, suggesting literacy improves **typical financial outcomes**, not just exceptional cases

•The mid-literacy group performs worse than both low and high groups, indicating that partial knowledge may increase risk-taking without sufficient budgeting discipline

•Higher financial literacy compresses downside risk while expanding upside potential, shifting the entire savings distribution rightward

Why This Matters

•Mean-based comparisons would understate behavioral differences; **distributional analysis reveals structural shifts** across groups

•Financial literacy appears to function as a **risk management mechanism**, reducing the likelihood of sustained negative savings

•Results imply that improving literacy can materially affect financial resilience, even when income levels are similar

•This complements earlier income analyses by showing that **behavioral factors**, not just financial constraints, shape outcomes

Key Insights and Economic Interpretation

Consumer Behavior Under Income and Knowledge Constraints

Core Findings

•**Income is a meaningful but incomplete predictor of discretionary spending**

- Regression results show positive income effects, but low to moderate R^2 values indicate substantial behavioral heterogeneity
- Spending levels respond to income only once income becomes a binding constraint

•**Spending composition shifts more clearly than total spending**

- Higher income is associated with increased allocation toward technology and entertainment
- Essential categories such as food maintain stable budget shares across income levels

•**Financial literacy materially affects savings outcomes**

- Higher literacy groups exhibit higher median savings, stronger upper-tail outcomes, and reduced downside risk
- Distributional analysis reveals behavioral differences masked by mean-only comparisons

Economic Interpretation

•Consumer behavior reflects a **constraint-driven optimization process**, where individuals allocate limited resources to maximize perceived utility
•Income primarily determines *what consumers can choose*, while financial literacy determines *how efficiently they choose*
•Observed budget smoothing and weak linear income effects are consistent with consumer choice and behavioral finance theory
•The interaction of income constraints and knowledge gaps explains why similar income levels produce divergent financial outcomes

Strategic and Market Implications

•**Consumer segmentation requires both income and behavioral dimensions**

- Income alone underpredicts purchasing behavior and financial resilience
- Literacy and decision quality act as hidden segmentation variables

•**Product positioning and pricing strategies** should account for heterogeneous willingness to spend across categories
•**Financial education and tooling** represent high-leverage interventions that can shift consumer outcomes without income changes
•**Consulting and analytics insight**: robust conclusions require segmentation, robustness checks, and distributional analysis rather than single-point estimates

Why This Project Stands Out

•Demonstrates end-to-end **data analytics workflow**: survey design, data cleaning, feature engineering, regression analysis, hypothesis testing, and distributional modeling
•Applies **economic theory to real consumer data** with clear behavioral interpretation
•Translates quantitative findings into **actionable market and strategy insights**