Unintended Consequences of Geographic Targeting

Jeff Larson, Surya Mattu and Julia Angwin

Abstract

For decades, The Princeton Review has prepared students for a battery of standardized tests for a price. In some cases, that price varies by ZIP code (or United States postal codes). The Princeton Review's website requests users enter their ZIP code before receiving a price for the individualized tutoring service. We at ProPublica analyzed the price variations for an online SAT tutoring service offered by The Princeton Review. The Princeton Review told ProPublica that the regional pricing differences for its "online tutoring package" were based on the "differential costs of running our business and the competitive attributes of the given market" and that any "differences in impact" were "incidental."

Results summary: We collected the price for The Princeton Review's "24-hr Online Tutoring," packages from each U.S. ZIP code and found that the prices varied by as much as $1,800. We compared the price in each ZIP code to the demographics and income of the ZIP code. Our analysis showed that Asians were disproportionately represented in ZIP codes that were quoted a higher price. As a result, Asians were 1.8 times as likely to be quoted a higher price than non-Asians. Our analysis also showed an increased likelihood of being quoted a higher price for ZIP codes with high median incomes.

Introduction

The Princeton Review offers these three levels of online individualized SAT tutoring on its website:

- SAT Private Level 24-hr Online Tutoring Package
- SAT Master Level 24-hr Online Tutoring Package
- SAT Premier Level 24-hr Online Tutoring Package

In order to purchase a package, a customer must first enter in a ZIP code. Students at Harvard University found that entering different ZIP codes resulted in different prices [1]. We improved on their original scraper, and in July and August 2015, we downloaded a list of prices for each package by ZIP code. We found that for each package those prices varied.

We sought to understand whether The Princeton Review's pricing system might disproportionately assign higher prices based on demographic characteristics, a phenomenon that has been found in other online pricing by ZIP code.

Background

All three of The Princeton Review's online SAT tutoring packages varied in price by geography. Depending on ZIP code, users saw one of three potential prices for the Private and Master packages. The most expensive package, Premier, displayed one of four possible prices by ZIP code.

The remainder of this paper will focus on the Premier package. For that package, prices ranged from $6,600 to $8,400, as shown in Table 1.

ZIP Code Count	Quoted Price
20,264	$6,600
5,055	$7,200
1,224	$7,800
349	$8,400

Table 1. Number of ZIP codes in the United States for each of the prices quoted for The Princeton Review's SAT Premier Level 24-hour Online Tutoring Package.

The Princeton Review's Premier prices are regionally distributed, with higher prices in California, the North East, most of Illinois and Wisconsin, Houston, Texas, and Cheyenne, Wyoming. (Figure 1).

Figure 1. The areas with the lowest price are shaded in tan ($6,600). The orange areas are the second highest price ($7,200). The dark orange areas are the third-highest price ($7,800). Red represents the highest price, which is not discernable at this scale, and includes most of New York City region.

Methods

We used census data to get a better understanding of the people who lived in these ZIP codes. We used the U.S. Census' ZIP Code Tabulation Areas (ZCTAs) [2], which are slightly different than the ZIP codes managed by the Postal Service. (For more see the Census' explanation of how ZCTAs are created.[3]) For this analysis, ZCTAs were the only viable geographic unit.

We used the following selected demographic fields from the 2013 5-year American Community Survey data [4] for our analysis.

1. Race, from Census Table B02001
2. Median Household Income in the Past 12 Months, from Census Table B19013
3. Educational Attainment, from Census Table S1501

The total number of ZCTAs used in this study was 26,892 of the 33,144 total ZCTAs. We ignored 6,252 ZCTAs because of large sampling errors or missing data.

The ZCTA demographic data allowed us to assess the relationships between characteristics of ZCTAs residents and The Princeton Review's price.

A logistic regression suggested candidate demographic variables for further analysis. We found that when controlling for factors like race and educational attainment, higher income ZCTAs had higher odds of being quoted a higher price by The Princeton Review. We also found that ZCTAs with higher percentages of Asian residents were more likely to be quoted one of the higher prices (For regression model results, see Figure 2).

We then conducted risk ratio analyses for Asians versus non-Asians and people in living above median-income ZCTAs versus below median-income ZCTAs.

Results

The combined ZCTA demographic and pricing data allowed us to generate a contingency table, as seen in Table 2.

A comparison of the rates at which Asians and non-Asians live in The Princeton Review's higher pricing ZIP codes resulted in an incidence rate ratio of 1.8, meaning that Asians are nearly twice as likely as non-Asians to be offered one of the higher price levels by The Princeton Review.

	High Price	Low Price
Asian	9,285,210	5,671,229
Non-Asian	100,107,183	190,378,831

Table 2. Asian vs. Prices Contingency Table.

Similarly, we generated a contingency table comparing ZCTAs with median household income above $73,632 to ZCTAs with median household income below that threshold. We chose $73,632 because it is one standard deviation away from the mean of the median incomes for ZCTAs.Table 3.

The incidence rate ratio was 2.05, which means that Americans living in areas with higher household income were twice as likely as those in low-income areas to be charged a higher price by The Princeton Review.

	High Price	Low Price
High Income	35,915,028	22,955,088
Low Income	73,477,365	173,094,972

Table 3. Income vs. Prices. Contingency Table

It could be argued that Asians are offered high prices for these courses because they live in wealthier areas. To test this proposition, we divided the ZCTAs in our analysis into categories: those with median household incomes above the U.S. national median income, and those below.

Our analysis showed, however, that even after isolating low median income ZCTAs, Asians had a disproportionate likelihood of being offered a high price by The Princeton Review, generating an incidence rate ratio of 2.2 as seen in Table 4. Isolating higher-income ZCTAs, somewhat narrows the incidence rate ratio to 1.4, as seen in Table 5.

Since the New York City metropolitan area is an outlier in the data, in that the entire area is assigned the higher price, we conducted a sensitivity analysis to see if removing the city's ZCTAs significantly changed our findings. The results remained consistent.

	High Price	Low Price
Asian	2,518,314	2,254,298
Non-Asian	37,450,836	120,330,880

Table 4: Contingency Table for Asians in areas with median incomes below the national median income of $53,046.

	High Price	Low Price
Asian	6,766,896	3,416,931
Non-Asian	62,656,347	70,047,951

Table 5: Contingency Table for Asians in areas with median incomes above the national median income of $53,046.

Discussion

Our analysis found that Asians were almost twice as likely as non-Asians to live in one of the areas being offered the higher price levels by The Princeton Review. And this effect remained consistent when stratifying ZCTAs by median income.

We also found that residents of relatively high income ZCTAs were much more likely to be offered one of The Princeton Review's higher price levels. A more comprehensive examination of the issues raised by this analysis can be found in a news article on ProPublica's website.[5]

Figure 2. A Logistic Regression Model for The Princeton Review's Pricing Scheme

.

	Dependent variable:
	higher price
Normalized Income	8.400***
	(0.214)
Asian Percent	6.388***
	(0.345)
Black Percent	0.522***
	(0.107)
Hispanic or Latino Percent	1.968***
	(0.083)
Bachelors Degree Percent	1.176***
	(0.150)
Constant	-3.403***
	(0.051)
Observations	26,892
Log Likelihood	-12,797.690
Akaike Inf. Crit.	25,607.390

*Note: White Percent is our control *p<0.1; **[<0.05; ***p<0.01*

Author Bios

Jeff Larson is the Data Editor at ProPublica. He is a winner of the Livingston Award for the 2011 series Redistricting: How Powerful Interests are Drawing You Out of a Vote. In 2011, he was a finalist for the Gerald Loeb Award for Distinguished Business and Financial Journalism. He is a graduate of the University of California, Santa Cruz.

Surya Mattu is a fellow at ProPublica and at Data & Society. He has worked as an engineer at Bell Labs and is a graduate from the New York University's Interactive Telecommunications Program. He has a degree from the University of Nottingham in the United Kingdom.

Julia Angwin is a senior reporter at ProPublica. From 2000 to 2013, she was a reporter at The Wall Street Journal, where she led a privacy investigative team that was a finalist for a Pulitzer Prize in Explanatory Reporting in 2011 and won a Gerald Loeb Award in 2010. In 2014, Julia was named reporter of the year by the Newswomen's Club of New York. In 2003, she was on a team of reporters at The Wall Street Journal that was awarded the Pulitzer Prize in Explanatory Reporting for coverage of corporate corruption. She is the author of two books, "Dragnet Nation: A Quest for Privacy, Security and Freedom in a World of Relentless Surveillance," (Holt, 2014) and "Stealing MySpace: The Battle to Control the Most Popular Website in America" (Random House, March 2009). She earned a B.A. in mathematics from the University of Chicago and an MBA from the Graduate School of Business at Columbia University.

References

1. Vafa K, Haigh C, Leung A, Yonack N. Price Discrimination in The Princeton Review's Online SAT Tutoring Service. *Technology Science*. 2015090101. September 1, 2015.
http://techscience.org/a/2015090101

2. U.S. Bureau of the Census' ZIP Code Tabulation Areas (ZCTAs).
https://www.census.gov/geo/reference/zctas.html

3. ZIP Code Tabulation Areas (ZCTAs). U.S. Bureau of the Census.
 http://www2.census.gov/geo/pdfs/education/brochures/ZCTAs.pdf

4. 2013 5-year American Community Survey Data. U.S. Bureau of the Census.
http://www.census.gov/data/developers/updates/acs-5-yr-summary-available-2009-2013.html

5. Julia Angwin, Surya Mattu, Jeff Larson "The Tiger Mom Tax: Asians are Nearly Twice as Likely to Get a Higher Price from Princeton Review," ProPublica.org, Sept. 1, 2015. http://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review