How do Start Ratings Affect Customer Choice

Team ERLT
Hanxi Gu, Bingyao Lei, Yujie Li, Changbo He

---

1. Introduction

In today's digital age, customer reviews play a crucial role in shaping consumer behavior. The rise of e-commerce and social media provide customers the opportunity to access a plethora of information about products and services. Customer reviews can provide valuable insights into the quality, features, and usability of a product or service, thus can help customers make informed decisions about what and which to buy. Increasingly, consumers are turning to reviews from other consumers to inform their purchasing decisions. According to the biggest online review management company Podium, 93% of American consumers have made daily buying decisions based on an online review, merely 2% have never made a purchase considering reviews (Podium.com. 2017). Positive reviews can increase sales and improve a company's reputation, while negative reviews can lead to lost sales and damage to a company's brand image. Customer review is such a powerful and complex tool, under its big umbrella star rating technique is one of the most popular forms. A star rating is a visual representation of customer feedback and satisfaction with a product or service, usually based on a scale of 1 to 5 stars. In this study, we specifically look at the star rating aspect of customer review. We employ a field experiment to find out how star rating can affect customers on choosing and evaluating products.

Many extant studies have shown good rating reviews are correlated with higher sales of products. In a recent study done by Li et al.(2022), they compare the sales of a product in one generation to the sales of the previous generation and control for customer review ratings. They find that the previous generation valence, i.e. average star rating, has a positive impact on current generation sales. The impact becomes even higher when the standard deviation of current generation ratings is high and the current generation valence is high. In another study done by Chevalier et al.(2003), they examine the customer reviews and sales of books on Amazon.com and BarnesandNoble.com. Using the rank of the book on the website to measure the sale, they find that an improvement in the book's review rating leads to an increase of sales at the site. Moreover, a high-star review helps the book climb up the rank at the site faster. However, a one-star review, or bad review, has a greater impact than a five-star review, or good review. Dellarocas et al.(2007) argues in their paper that customer reviews should be considered as an attribute of the product itself, their star rating should be utilized as a vital measurement number just like the sales of the previous period when forecasting the sales of a product. Studies mentioned above all show that good rating could improve the demand of a product, thus indirectly implying the idea that good rating could increase customer's perception of value within a product.

2. Experimental Design

In order to measure the causal effect of ratings and reviews on consumers' perceived value and product choice, we conducted a framed field experiment on USC campus. In our experiment, participants were invited to browse a website (e506.netlify.app), which included two homogeneous Bluetooth speakers - product A and product B - along with detailed information and reviews of the products. After browsing the website, participants were asked to fill out a questionnaire assessing their valuation and selection of product A and B.

Sample

Between April 4th and April 14th, 2023, we conducted the experiment with a sample of 107 participants who were all USC students, comprising 51 males, 55 females, and one person preferred not to say. Participants were engaged via posters and personal invitations.

The first method for recruiting participants was posting posters in teaching buildings on the USC campus. Buildings were randomly assigned to four experimental groups, including one control group and four treatment groups (see Graph 1).

The second method was randomly inviting students on campus. Participants were randomly assigned to one of the four experimental groups by using a true random number generator (random.org).

Questionnaire

The questionnaire consisted of four parts. The first part required participants to estimate the likely price range of product A, as well as its precise price. The second part was the same estimation for product B. The third part asked participants to select which of the two products they would prefer to purchase. The final part comprised background questions such as gender, age, degree/academic class, and online shopping frequency.

Treatment and control groups

To simulate the comparison process when online shopping, we presented participants in each experimental group with two products, A and B, each with two settings - setting 1 and setting 2. In setting 1, the product was presented with a 5-star rating and all positive reviews. In contrast, setting 2 presented the product with a lower rating, including both positive and negative reviews.

The control group of the experiment was presented with both product A and product B at setting 2. In contrast, treatment group 1 was presented with both products at setting 1. Treatment group 2 was presented with product A at setting 1 and product B at setting 2, while treatment group 3 was presented with product A at setting 2 and product B at setting 1. Table 1 shows the settings for each experimental group.

Hypotheses

In our study design, we assumed that the preference for product A and B would be the same in both control group and treatment group 1. For treatment groups 2 and 3, we acknowledged that

while the 5-star ratings could be perceived as manipulated, they still represented better ratings than the real reviews. Therefore, we were uncertain about the impact of such ratings on participants' perceptions and preferences.

This paper will focus on comparing the predicted price of product A in two groups, with product B held constant in the same setting. Specifically, we will compare treatment group 2 (A+, B-) with the control group (A-, B-), and treatment group 1 (A+, B+) with treatment group 3 (A-, B+). We will also compare the price of product B controlling product A at the same setting. Specifically, we will compare treatment group 3 (A-, B+) with the control group (A-, B-), and treatment group 1 (A+, B+) with treatment group 3 (A+, B-).

Therefore, we have formulated two null hypotheses (H_0 and $H_0^{'}$), and the corresponding alternative hypotheses (H_1 and H_2):

H_0*: There is no difference in the average predicted price of product A between the two groups when controlling for the setting of product B.*

H_1*: There is a significant difference in the average predicted price of product A between the two groups when controlling for the setting of product B.*

$H_0^{'}$*: There is no difference in the average predicted price of product B between the two groups when controlling for the setting of product A.*

H_2*: There is a significant difference in the average predicted price of product A between the two groups when controlling for the setting of product B.*

Threats to internal and external validity

Based on our study design, there are several potential threats to both internal and external validity that should be considered. The potential threat to internal validity is related to the recruitment method via poster posting. The SUTVA would be threatened if two or more acquaintances scan the same poster and answer questions together.

Furthermore, the fact that all participants in our study are USC students may undermine the external validity of our study. In addition, the experimental environment cannot accurately reflect online shopping in the real world.

3. Analysis

In this section, we will first present the variables in the analysis and assess the balance of covariates across the control and three treatment groups using a balance table. Next, we will estimate the average treatment effects on the outcome variables and conduct further statistical testing. Finally, we will estimate the heterogeneous treatment effects to measure the effect of a treatment or intervention that may vary across different subgroups of individuals.

Variables

Table 2 provides an overview of all the variables included in our study and their corresponding encoding mechanism. It allows readers to understand measurement scales and categories of each variable easily, which is essential for interpreting the results.

Balance Table

It is important to assess whether the treatment and control groups are balanced on key baseline characteristics. Table 3 is the balance table for the experiment, comparing the mean values and p-values of these characteristics between the control group and the three treatment groups.

As shown in Table 3, the treatment and control groups are balanced on several key baseline characteristics. Specifically, there are no significant differences in the mean values of gender, age, and online shopping frequency between the control group and the three treatment groups. It suggests that the randomization process succeeds in assigning participants to different groups and balancing the groups on these characteristics.

However, there is a significant difference in the mean value of the degree/academic class between the treatment and control groups, with treatment group 2 having a higher mean value than the other groups. It indicates that there may be some residual confounding due to differences in the educational level of participants. Therefore, we will include degree as a covariate in the subsequent analysis to account for this difference and control its potential effect on the outcomes.

Summary Statistics

To further understand the distribution of variables, we report relevant statistics in this section. As shown in Table 4, the means of the baseline characteristics do not vary much across groups.

We also examined the distribution of the preference variable among the four groups in Graph 2. All groups except the control group have the expected outcomes as the preference corresponds with the review settings. But in the control group, where product A and product B have the same all-star-level review setting, 56.5% of the participants indicated they preferred product A. It suggests internal invalidity may exist in our experiment as we did not set the reviews of both products equal when we designed the review website for the study.

Main Analysis

In this section, we aim to investigate the average treatment effect of including different star-level reviews across groups. We will first examine product A and compare the average treatment effect (ATE) between the control group and treatment group 2, and between treatment group 1 and treatment group 3 since the setting of product B is the same in these groups. Moving onto product B, we will compare the ATE between the control group and treatment group 3, as well as between treatment group 1 and treatment group 2. Lastly, we will investigate the product preference and compare the ATE between these four group combinations.

It is worth noting that in our experiment, we do not need to use the local average treatment effect (LATE) because our treatments are randomly assigned. It means that the potential outcomes for each group are independent of treatment assignment.

The basic main regression equation is listed below, where Y stands for the outcome variables, including the estimated price of A, the estimated price of B, and preference, X stands for the control or treatment group, is the constant term, β_1 is the estimated coefficient, and ε is the error term:

$$Y = \alpha + \beta_1 X + \varepsilon$$

However, as the balance table indicates, we need to include degree as a covariate in the regression to control for its potential effect on the outcomes. Thus, we run another regression as well:

$$Y = \alpha + \beta_1 X + \beta_2 Degree + \varepsilon$$

As shown in the first two columns of Table 5, our treatment involves increasing the star rating of product A while setting the star rating of product B close to reality or at a 5-star level. The results show that the treatment has a positive effect on the estimated price of product A. Specifically, the estimated price of A increases by 18% over the control group mean in treatment group 2, and by 0.7% over the treatment group 3 mean in treatment group 1. When taking the degree variable into account, the estimated price of A increases by 19% and 1.3%, respectively. However, these differences are not statistically significant at conventional levels with p-values of 0.22, 0.15, 0.96, and 0.92, respectively. Similarly, the treatment does not have a statistically significant effect on the estimated price of product B across groups. It suggests that we cannot reject the null hypothesis, and the star-rating manipulation does not significantly affect the perceived value of both products.

Table 5 also provides insight into the preference between products A and B. We find that when the star ratings of product A are held constant across groups, an increase in the star rating of product B results in a significant shift in customer preference towards product B, with a p-value of 0.002.

In summary, these findings imply that even though star-rating manipulation has a limited impact on the estimated price of the products, it still plays an important role in shaping customer preference for the product with higher star-rating reviews.

Distribution of Outcome

For further analysis, we examine the distribution of outcome variables across different groups. Graphs 3 and 4 illustrate the probabilistic distribution of product A and B prices. We observe that when we compare the price probabilities of the same group combinations as in the main analysis section, the result aligns with the previous conclusion, customers tend to estimate a higher price for the product with higher ratings.

Heterogeneous Treatment Effects

We now turn our attention to examining heterogeneous treatment effects across different subgroups. We included various demographic variables in the regression, such as gender, age, degree, and online shopping frequency, to explore potential differences in the treatment effect. The heterogeneous treatment effects regression equation is listed below, where Y stands for the outcome variables, including the estimated price of A, the estimated price of B, and preference, X stands for the control or treatment group, D stands for the demographic variables, X*D stands for the interaction term of group variable and demographic variable, ε is the constant term, β_i are the estimated coefficients, and is the error term:

$$Y = \alpha + \beta_1 X + \beta_2 D + \beta_3 (X * D) + \varepsilon$$

Table 6 shows that all coefficients related to gender are insignificant. Similar results are obtained for all other demographic variables, except for the coefficient of the group variable in the equation of price of B, which is significant when considering the online shopping frequency and its interaction term. Specifically, the coefficient suggests a 73.8% increase in the estimation of the price of B over the mean in treatment group 2 when increasing the star rating of product B, by setting product A to remain at a 5-star rating, with a p-value of 0.033. As the coefficient for the interaction term is positive, it means that individuals with higher online shopping frequency tend to perceive less value in the 5-star products as they might be more aware of the manipulation of the good reviews. Regarding preference, there is no significant influence from demographic variables.

The findings from Graph 5 verify that individuals with a higher frequency of online shopping tend to assign a higher value to products with a rating closer to reality. Interestingly, we also find that respondents who do not have online shopping at all also show the same value prescription as the respondents who have high online shopping frequency.

In summary, these results suggest that demographic variables do not significantly influence treatment effects, except for the significant influence of online shopping frequency on the estimated price of product B.

4. Discussion

Result expectation

We expect that review manipulations have significant effects on customer's perceptions and decision-making. However, the results are not significant enough to reject our null hypotheses. Even so, the reviews are still significantly influential in shaping customer preference.

Internal validity

In our experiment, all the treatment and control groups are assigned randomly. Other than the potential threats to internal validity we mentioned in the Experimental Design section, other threats also appear during the experiment. The balance table shows that the imbalance distribution

of the degree variable may harm the accuracy of potential outcomes. Graph 2 also suggests that the review manipulation process when we created the page view website may lead to inherent bias, which harms the internal validity.

External validity

When it comes to external validity, due to a relatively small scale of participants, our results may not perfectly match the real market environment. By taking the standard deviation of the price, which is $17.78, and taking $5 as minimum detectable effect, the sample size should be at least 397 observations. However, in our experiment, we managed to get 107 valid observations, which is too small compared with the expected size. Therefore, our experiment cannot reflect the real situation in the market.

Further improvements

For further research, a larger scope of subjects is needed. If the research aims to investigate the student market, then we should investigate more than one school to get the results. Our experiment only includes one specific product, bluetooth speaker, which may not fully represent how the rating system could influence people's decision making in real life. There is a possibility that participants know little and do not even care about one specific product. To make up for this, a bucket of goods and services with different reviews should be provided, instead of only one product. In that way, the ATE can be more accurate and significant when we have the overall effects of products that people are likely to care about.

5. Conclusion

In conclusion, this research study assessed the impact of star-rating manipulation on consumers' perceived value and preference between two similar products. The balance table showed successful randomization and balance of control and treatment groups on most key baseline characteristics, except for degree/academic class. Summary statistics revealed significant variation in price expectations and internal invalidity due to unequal reviews for the control group. The main analysis found that increasing the star rating of product A resulted in a positive but statistically insignificant effect on the estimated price of A, while it still plays an important role in shaping customer preference for the product with higher star-rating reviews significantly. The heterogeneous treatment effects analysis showed a significant shift in preference for product B when its star rating increased, but only for groups where product A had a lower star rating. Overall, the study highlights that customers show preferences towards the products with higher ratings.

Appendix

Table 1: Experimental Groups

Control group	**Treatment group 1**	**Treatment group 2**	**Treatment group 3**
A- (Setting 2)	A+ (Setting 1)	A+ (Setting 1)	A- (Setting 2)
B- (Setting 2)	B+ (Setting 1)	B- (Setting 2)	B+ (Setting 1)

Notes. "+" represents setting 1 (a 5-star rating with all positive reviews); "-" represents setting 2 (a lower rating with both positive and negative reviews).

Table 2: Variables Table

	Group	Preference	Gender	Degree/ Academic class	Online shopping frequency
0	Control Group A: Setting 2 B: Setting 2	Neither A nor B	Female	Freshman	No online shopping at all
1	Treatment Group 1 A: Setting 1 B: Setting 1	A	Male	Sophomore	1-2 times a month
2	Treatment Group 2 A: Setting 1 B: Setting 2	B	Prefer not to say	Junior	3-4 times a month
3	Treatment Group 3 A: Setting 2 B: Setting 1	No preference between A and B		Senior	5-8 times a month
4				Graduate	More than 8 times a month
5				Other	
n	107	107	107	107	107

Notes. Other variables including the probability of product price in each subrange, the product price, and the age of the respondent are not shown in the above table as they are not dummy variables. For the difference between each setting, please refer to the Experimental Design section.

Table 3: Balance Table

	CG	T1	T2	T3	CG vs T1	CG vs T2	CG vs T3	T1 vs T2	T1 vs T3	T2 vs T3	Joint test
Gender	0.609	0.469	0.429	0.500	0.314	0.208	0.500	0.760	0.832	0.640	0.657
Age	22.174	22.938	22.929	22.417	0.104	0.105	0.664	0.976	0.178	0.166	0.184
Degree	3.043	3.500	3.893	3.208	0.193	0.009**	0.700	0.068	0.360	0.016*	0.046*
Online shopping frequency	1.826	1.844	1.857	2.083	0.946	0.916	0.350	0.960	0.351	0.435	0.775
n	23	32	28	24							

Notes. This table presents the average characteristics of respondents in the control and treatment groups of the experiment. Column 1 to 4 refer to different control and treatment groups. The subsequent columns present *p*-values corresponding to the test of equality between each pair of groups. The final column presents the result of a joint test of equality across all groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Summary Table

	CG					*T1*				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
A~$[0,20]	23	9.609	14.494	0	56	32	15.219	21.277	0	74
A~$[21,30]	23	41.913	30.148	0	90	32	30.500	22.000	0	82
A~$[31,40]	23	33.261	27.810	0	90	32	37.375	26.753	1	100
A>$40	23	15.217	24.391	0	74	32	16.906	24.245	0	90
Price of A	23	33.646	9.640	19.9	60	32	35.868	15.778	20	100
B~$[0,20]	23	18.435	27.502	0	100	32	13.500	18.327	0	60
B~$[21,30]	23	36.391	28.022	0	90	32	30.000	22.098	0	100
B~$[31,40]	23	32.696	24.941	0	85	32	37.438	24.257	0	100
B>$40	23	12.478	16.175	0	65	32	19.063	22.782	0	90
Price of B	23	31.303	9.374	19	55	32	40.356	25.502	20	150
Preference	23	1.345	0.775	0	3	32	1.781	0.751	1	3
Gender	23	0.609	0.499	0	1	32	0.469	0.507	0	1
Age	23	22.174	2.167	18	24	32	22.938	1.243	20	24
Degree	23	3.043	1.66	0	5	32	3.500	1.016	0	5
OS Freq.	23	1.826	0.937	1	4	32	1.844	0.954	0	4

	T2					*T3*				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
A~$[0,20]	28	7.071	10.794	0	40	24	10.625	23.050	0	100
A~$[21,30]	28	33.107	26.853	0	100	24	35.333	21.571	0	100
A~$[31,40]	28	39.143	24.786	0	100	24	33.042	28.645	0	100
A>$40	28	20.679	26.469	0	100	24	21.000	30.146	0	100
Price of A	28	39.908	22.595	18	129	24	35.624	17.131	15	100
B~$[0,20]	28	10.821	14.189	0	57	24	9.167	21.709	0	100
B~$[21,30]	28	49.143	27.490	5	100	24	35.708	33.050	0	100
B~$[31,40]	28	29.500	21.560	0	80	24	25.625	22.687	0	67
B>$40	28	10.536	17.013	0	60	24	29.500	35.490	0	100
Price of B	28	31.092	9.685	15	60	24	36.416	17.661	5	100
Preference	28	1.500	0.882	0	3	24	2.042	0.624	1	3
Gender	28	0.429	0.504	0	1	24	0.500	0.590	0	2
Age	28	22.930	0.979	21	24	24	22.417	1.613	19	24
Degree	28	3.893	0.497	2	5	24	3.208	1.351	0	4
OS Freq.	28	1.857	1.113	0	4	24	2.083	0.929	0	4

Notes. The data of the first to fourth row and the sixth to ninth row are absolute values ranging from 0 to 100 as they are the probability measurement of the product price falls within each price range.

Table 5: Main Regression Table

	CG vs T2 Y=Price of A		*T1 vs T3* Y=Price of A		*CG vs T3* Y=Price of B		*T1 vs T2* Y=Price of B	
X	3.131	3.975	-0.122	-0.233	1.704	1.677	-9.264	-6.922
	(2.529)	(2.723)	(2.210)	(2.245)	(1.384)	(1.400)	(5.119)	(5.151)
Degree		-1.987		-0.761		0.490		-5.962
		(2.334)		(1.918)		(1.475)		(3.110)
Constant	33.646***	39.695***	35.991***	38.763***	31.303***	29.810***	49.621***	68.145***
	(3.748)	(8.036)	(4.651)	(8.419)	(2.966)	(5.397)	(7.930)	(12.388)
n	51	51	56	56	47	47	60	60

Y=Preference	*CG vs T2*		*T1 vs T3*		*CG vs T3*		*T1 vs T2*	
X	0.076	0.096	0.130	0.147	0.231***	0.228***	-0.281	-0.281
	(0.118)	(0.127)	(0.094)	(0.094)	(0.068)	(0.069)	(0.211)	(0.219)
Degree		-0.047		0.112		0.057		-1.47e
		(0.109)		(0.081)		(0.072)		(0.132)
Constant	1.348***	1.493***	1.651***	1.242***	1.348***	1.173***	2.063***	2.063***
	(0.174)	(0.376)	(0.199)	(0.354)	(0.146)	(0.265)	(0.326)	(0.526)
n	51	51	56	56	47	47	60	60

Notes. * p < 0.05, ** p < 0.01, *** p < 0.001

Table 6: Heterogeneous Treatment Effects Table

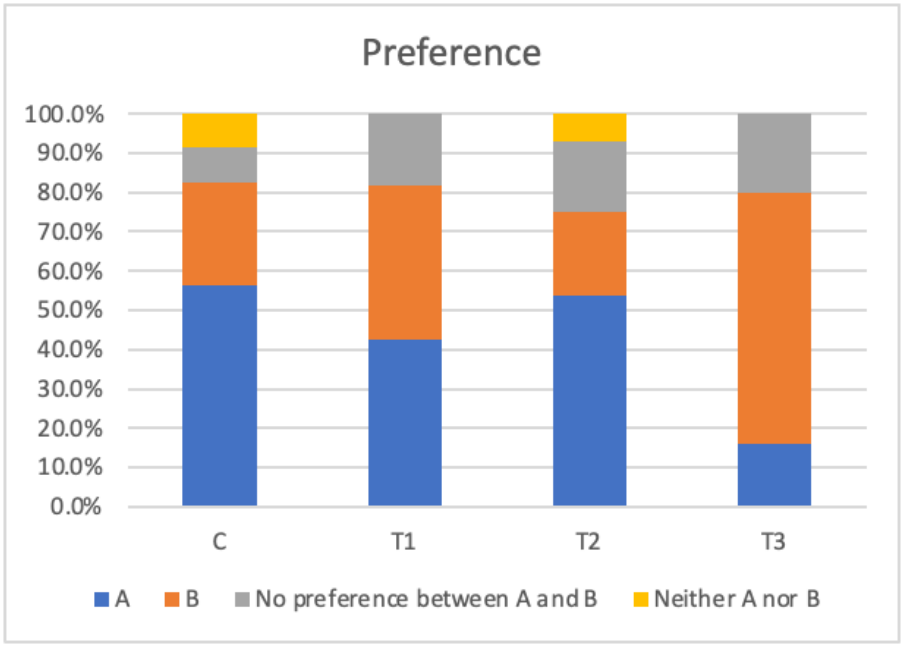
	CG vs T2 Y=Price of A		*T1 vs T3* Y=Price of A		*CG vs T3* Y=Price of B		*T1 vs T2* Y=Price of B	
X	1.733	1.144	0.428	1.845	1.019	2.239	-12.394	-22.950*
	(3.802)	(5.441)	(3.015)	(5.152)	(2.006)	(3.342)	(6.983)	(10.511)
Gender	-5.506		4.822		3.604		-10.706	
	(7.796)		(9.303)		(5.992)		(16.137)	
X*Gender	2.104		-1.212		1.631		6.950	
	(5.228)		(4.157)		(2.593)		(10.443)	
OS Freq.		-2.719		-3.383		-0.609		-14.459
		(4.157)		(4.877)		(3.305)		(8.079)
X*OS Fre.		1.093		-0.709		-0.231		7.420
		(2.610)		(2.351)		(1.548)		(4.985)
Constant	37.000***	38.611***	33.749***	41.568***	29.109***	32.415***	54.511***	76.285***
	(6.083)	(8.492)	(6.447)	(10.232)	(4.675)	(6.752)	(10.941)	(16.838)
n	51	51	56	56	47	47	60	60

Notes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

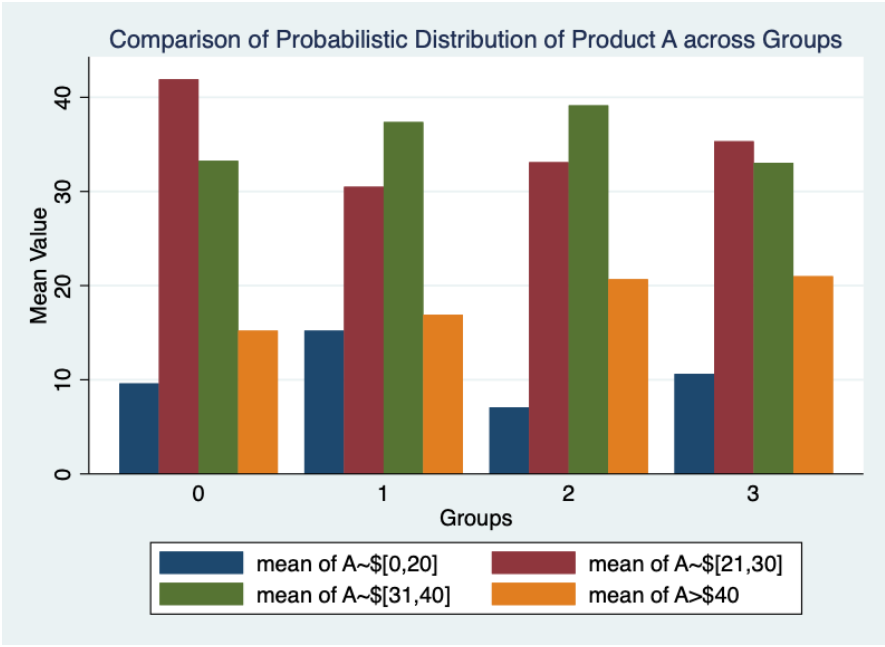
Graph 1

Building	Group
A4 Kaprielian Hall, KAP-1	4
A4 Kaprielian Hall, KAP-2	1
B3 Drama Center, DRC	2
B4 Ray R. Irani Hall, RRI	3
B5 Biegler Hall, BHE	1
B5 Olin Hall, OHE	2
B6 Salvatori Computer Science Center, SAL	2
C2 McClintock Building, MCC	4
C4 Grace Ford Salvatori Hall, GFS-1	4
C4 Grace Ford Salvatori Hall, GFS-2	3
C6 Watt Hall, WAH	1
C7 Fisher Museum of Art (Harris Hall)	1
D6 Leventhal School of Accounting, ACC	4
D6 Zumberge Hall, ZHS-1	3
D6 Zumberge Hall, ZHS-2	2
E4 Booth Hall, BMH	3
E4 Social Sciences Building, SOS	1
E4 Taper Hall, THH-1	4
E4 Taper Hall, THH-2	3
E7 Musick Law Building, LAW	2

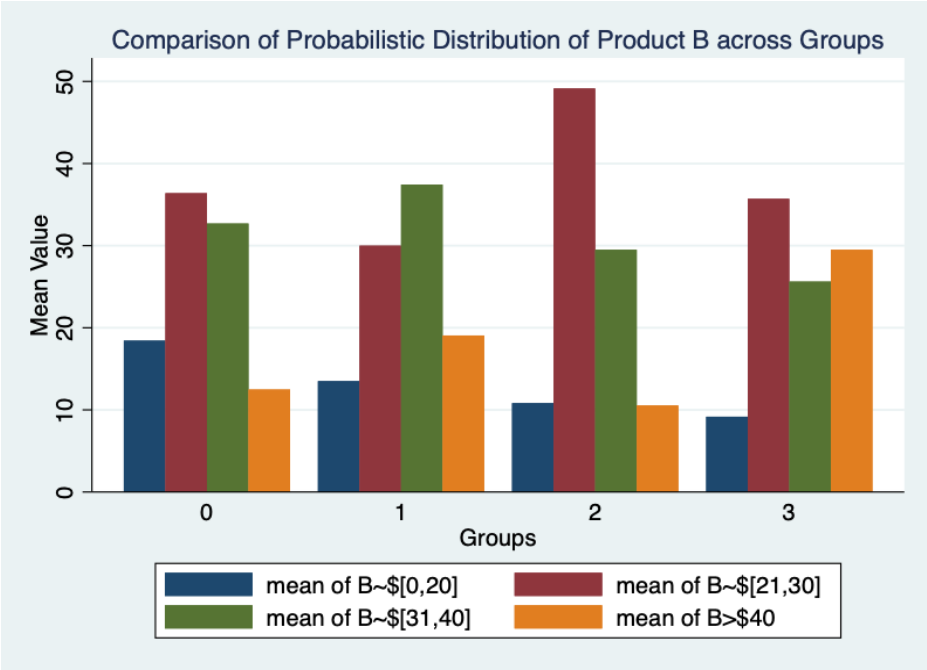
Graph 2



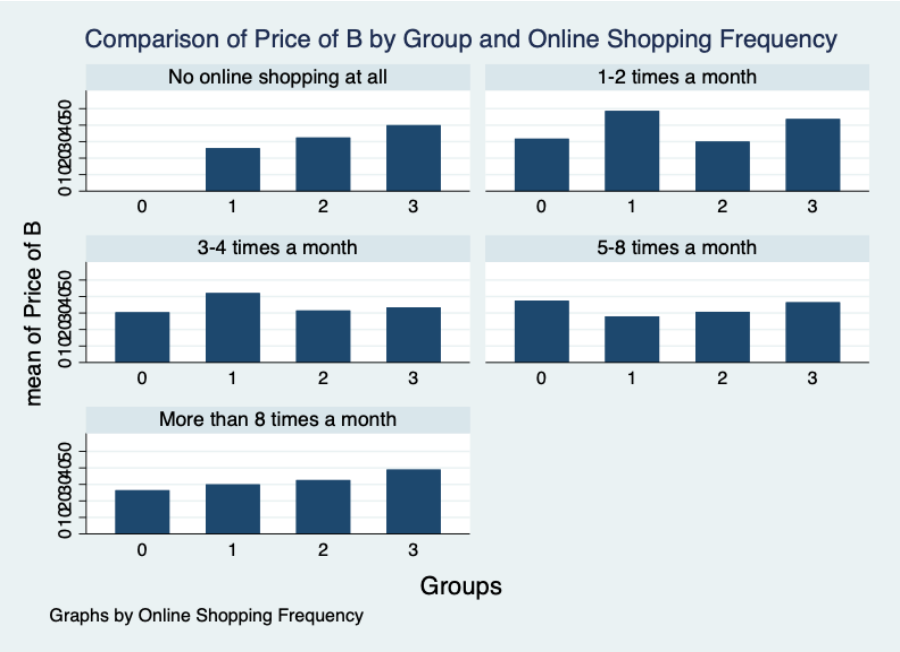
Graph 3



Graph 4



Graph 5



Reference

Chevalier, Judith A., and Dina Mayzlin. “The Effect of Word of Mouth on Sales: Online Book Reviews.” NBER Working Paper Series, 2003, p. 10148, https://doi.org/10.3386/w10148.

Dellarocas, Chrysanthos, et al. “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures.” Journal of Interactive Marketing, vol. 21, no. 4, 2007, pp. 23–45, https://doi.org/10.1002/dir.20087.

Li, Linyi, et al. “History Matters: The Impact of Online Customer Reviews Across Product Generations.” Management Science, vol. 68, no. 5, 2022, pp. 3878–903, https://doi.org/10.1287/mnsc.2021.4061.

Podium 2017 state of online reviews. Podium. (n.d.). Retrieved April 23, 2023, https://learn.podium.com/rs/841-BRM-380/images/Podium-2017-State-of-Online-Reviews.pdf.