ABV Effect on Beer Ratings

Lab 2: Moonsoo Kim, Saurabh Pattarkine, Ben Michaels

Introduction

The 2022 United States beer market was valued at 115 billion dollar with a 6% dollar growth in the craft beer section. While beer consumption overall has decreased, craft beer market share and annual production continues to increase. In addition, overall brewery counts continue to increase annually yet the number of annual breweries closing are rapidly approaching the number opening [1].

This changing yet promising market has been shown to be influenced by beer rating applications which allow users to rate consumed beers via their mobile device or via a web site. The number of users and reviews from these apps has increased annually and a positive correlation between sale price and ratings has been demonstrated in previous literature[2],[3].

This study aims to explore the relationship between ABV (X) and overall consumer review scores (Y). Understanding this relationship is essential for breweries to tailor their product offerings to meet consumer expectations, potentially leading to increased sales and market share. While there are many factors that play into the overall rating of beer, ABV is a measurable and controllable component of the brew. Brewers can manipulate and plan for ABV during the brewing process by modifying ingredients, process, and/or temperatures. The models constructed in this study will shed light on the nuanced relationship between ABV and consumer ratings, offering actionable insights for breweries seeking to enhance the success of their products.

Brief Description of the Data

Our data was downloaded from the "Beer Profile and Rating Set" hosted on Kaggle with credits given to the Beer Tasting Profiles Dataset by sp1222 and the 1.5 Million Beer Reviews by Tanya Cashorali. The units of observation are unique beers, each contributing data on ABV, beer style, brewery details, and consumer ratings. An exact start and end date of the reviews was not provided and was created from a dataset uploaded in 2017. The data source was BeerAdvocate user reviews.

Operationalization of Key Concepts:

The independent variable (X) in our study is the ABV of beer, operationalized as a flexible function, allowing for potential nonlinear relationships. The dependent variable (Y) is the overall consumer review score, a comprehensive metric of a beer's quality based on consumer evaluations. Alternative variables, such as beer style or brewery size, are considered as potential covariates to address potential confounding.

Explanation of Key Modeling Decisions:

All exploration and model building was performed on a 30% subsample of the data. The remaining 70%, which totaled 2238 beers, was used to generate the statistics in this report. Our next step was to eliminate beers with ABVs greater than 15.0 (n=14). This decision was made due to our microbreweries inexperience and lack of equipment to brew beers above this threshold. The brewing process is drastically different to produce beers at this ABV and as we were unable to produce a beer at this ABV, we excluded them from the analysis.

[1] Brewers Association. "National Beer Sales and Production Data" (2022)

[2] Pozner, J.-E., DeSoucey, M., Verhaal, J. C.,and Sikavica, K. (2022). Watered Down: Market Growth, Authenticity, and Evaluation in Craft Beer. Organization Studies, 43(3), 321-345.

[3] Michis, A. A. (2023). Hedonic decomposition of beer prices: Consumer ratings and quantity discounts. International Journal of Food and Agribusiness Marketing, 35(4), 443-462.

A significant p-value (5.487e-11) in the ANOVA suggests that the quadratic model has a statistically better fit. The linear model (AIC=1135.040) and quadratic model (AIC=1093.886) suggest that the quadratic is a better fit in terms of the trade-off between goodness of fit and model complexity.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	952.00	182.41				
2	951.00	174.34	1.00	8.07	44.01	0.00

	df	AIC
linear_model	3.00	1135.04
quadratic_model	4.00	1093.89

Figure one shows the relationship between ABV and overall review. The relationship suggests that overall review initially increases as the ABV increases to about 10, then begins to decrease at higher ABVs. Based on the quadratic form of ABV vs overall reviews, we created a derived variable (ABVSquared) by squaring the ABV value.

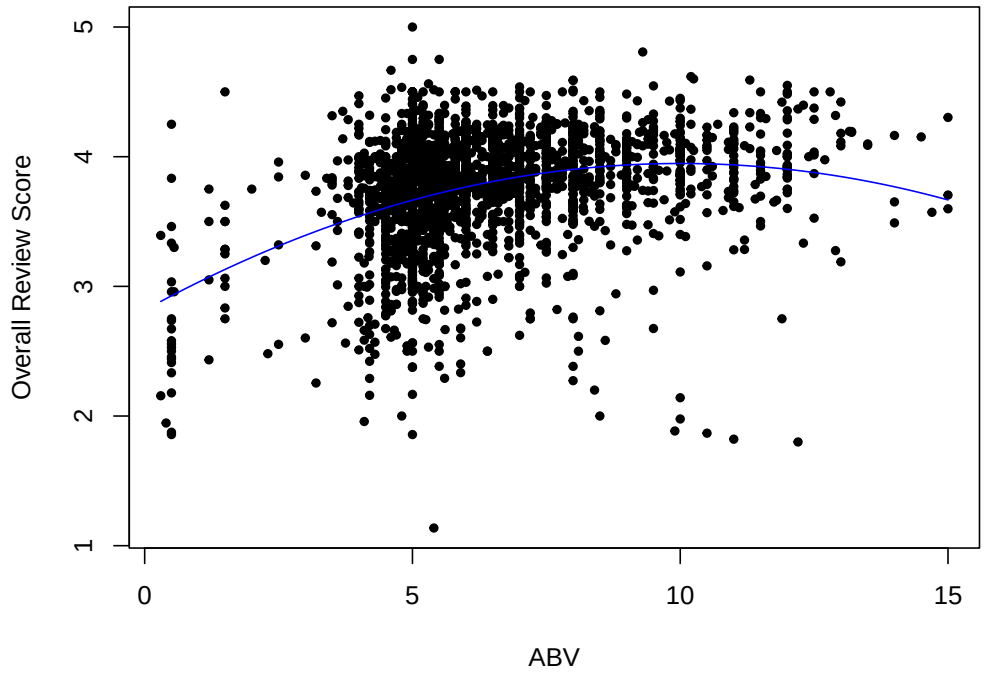

Figure 1: ABV vs Overall Review

Our basic model consists of the following regression form:

$$\widehat{OverallReviewAvg} = \beta_0 + \beta_1 \cdot (ABV) + \beta_2 \cdot (ABV^2)$$

Figure 2 shows calculated correlation coefficients between variables and the Overall Review score:

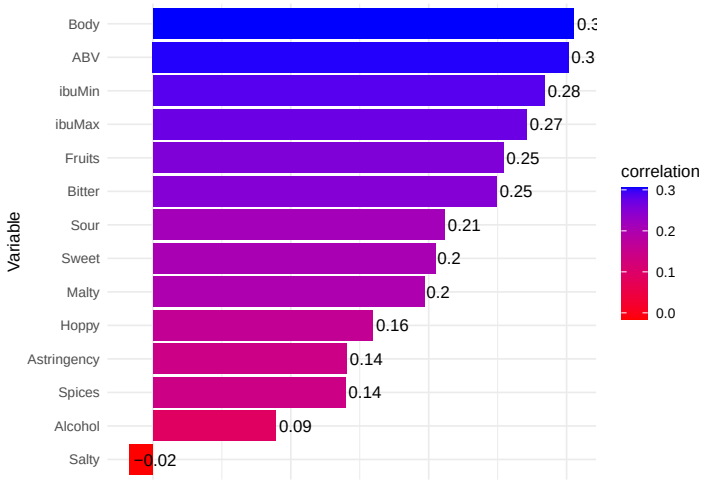

Figure 2: Correlation with Overall Review

Table 1: Estimated Regressions

	Output Variable: Average Overall Review		
	review_overall		
	(1)	(2)	(3)
ABV	0.23***	0.21***	0.16***
	(0.02)	(0.02)	(0.02)
ABVSquared	−0.01***	−0.01***	−0.01***
	(0.001)	(0.001)	(0.001)
Body			0.003***
			(0.0004)
Bitter			0.001**
			(0.0005)
Fruits			0.003***
			(0.0003)
ibuMin		0.01***	0.01***
		(0.001)	(0.001)
ibuMax		−0.0001	−0.002*
		(0.001)	(0.001)
Constant	2.82***	2.80***	2.77***
	(0.06)	(0.06)	(0.06)
Observations	2,229	2,229	2,229
R^2	0.13	0.15	0.23
Residual Std. Error	0.41 (df = 2226)	0.41 (df = 2224)	0.39 (df = 2221)

Note: HC_1 robust standard errors in parentheses. Body, Bitter, and Fruits variables are review averages for those components. *ibuMin* and *ibuMax* are assigned based on the style of beer.

Results

Table 1 shows the results of three represented regressions. For all models, ABV was highly statistically significant. The R2 value increased as more variables were included in each model. As an example using the basic model, if our micro brewery produced three beers with ABVs of 4,9,and 15 percent, we could expect average ratings of 3.57, 4.08, and 4.02 respectively. The additional variables we included could also be considered to brew a beer to maximize consumer ratings. Using the third model, if a brewery were to try and maximize reviews, they would want an ABV in the 9-11 range, use ingredients that increased body, bitter, and fruit scores, and stay closer to the minimum IBU range for the style of beer. Figure 3 shows the predicted relationship between overall reviews and ABV.

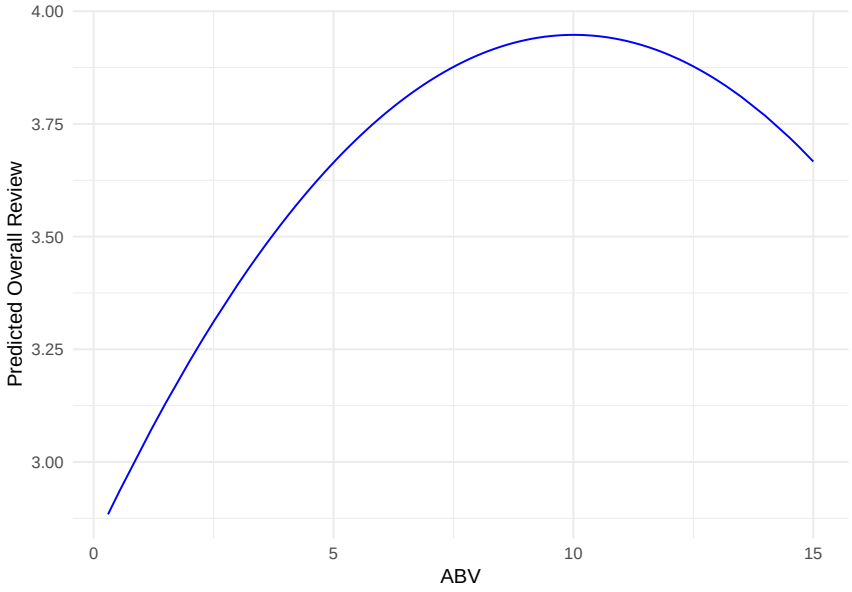

Figure 3: ABV vs Predicted Overall Review

Limitations

Consistent regression estimates require an assumption of independent and identically distributed (IID) observations. As beer is distributed and consumed geographically, there is the possibility of geographic clustering based on groups of users rating a beer either higher or lower due to geographic preferences.

In addition, beer reviews occur over time and could be influenced by prior reviews which are viewable by a user before they submit their own review. The pharmacodynamics of ethanol create the possibility of a time based effect where reviews increase (or potentially decrease) as more is consumed. This translates to the possibility of the last, or first, beer being rated higher depending on the effect of ethanol on the user's body. The dataset did not contain information above the collection time frame and if we were to action this data, we would want a more recent dataset that would reflect current trends and taste preferences.

There is also the possibility that omitted variables exist in our data. For example, marketing, bottle/can design, hype, or tradition could all influence the beer ratings left by users.

Last, we did not address the quantity of beer reviews in our analysis. This could be a separate analysis where the impact of number of reviews is quantified relative to the overall rating.

Conclusion

In conclusion, this study demonstrates the relationship between ABV and consumer ratings in the beer market. By addressing the research question, breweries can make informed decisions about product design, contributing to enhanced product success. This can help them build and enhance the product not only with respect to taste but also other factors like feel, texture, aroma etc. The broader implications of the findings are discussed, emphasizing the practical relevance for breweries aiming to navigate the competitive landscape successfully.