**Do labels impact wine price perceptions?**

INFO 241 Final Project Report

Romit Barua, Aiste Cechaviciute, Benjamin Fell, Gautham Koorma

**Table of Contents**

[**1. Introduction**](#_4n88hgn7ycbw) **2**

[**2. Experimental Design**](#_y4p9z9xl4mi9) **3**

[Experiment Overview](#_7noo7q3dsiex) 3

[Recruitment](#_kwn2e14bp2eo) 5

[Power Analysis and Sample Size](#_adjb03min134) 6

[Recruitment process](#_8pz3ldwtffow) 7

[Randomization](#_s4lg02ggz2zn) 8

[Delivery of Treatment](#_ymmj1hqa79lq) 9

[Inclusion Criteria](#_3676ssf42xbo) 9

[Exclusion Criteria](#_dzjg8fkgg19m) 9

[**3. Analysis**](#_rnbcgt3mm38k) **9**

[Data](#_7t661uf87wqr) 9

[Models and Findings](#_oxyv0irzmffw) 13

[**4. Learnings**](#_8emgxzell8j9) **19**

[**5. Future Work**](#_ujphrkya0afb) **19**

[**6. Conclusion**](#_tmqbdcaiqsu0) **20**

[**7. References**](#_2t8awagxxm6o) **20**

[**8. Appendix**](#_1g3decw0y9jc) **21**

[1. Treatment A: Wine labels containing limited information about the wine in a printed version.](#_9jxhwojdwbqd) 21

[2. Treatment B: Bottles were presented to subjects](#_pj2duwe5tzsv) 21

[3. Survey: Hunger level question](#_cdiumddaj5rh) 22

[4. Survey: Wine expertise question](#_e2sa7d72fvj) 22

[5. Survey: Wine consumption question](#_ysfitv5umjh6) 23

[6. Survey: Wine adventurousness question](#_6l7d2v29fb2n) 23

# 

# 1. Introduction

The United States is one of the top wine producing countries in the world. Throughout 2021, the country produced 24.1 million hectoliters of wine, and sales grossed $78.4 billion. Within such a large market, bringing new wines or maintaining a competitive edge is crucially important to wine-makers and sellers. Our study aims to aid wine-makers, sellers, and distributors by providing some insight into what elements of their product impact consumers' perception of quality and expense. In our experiment, we focused on understanding whether seeing the information on the wine label, or seeing the entire bottle itself, aided consumers in correctly estimating a wine's listed price.

Given the size of the wine industry in the United States, it is unsurprising that previous research similar to this topic has been conducted. In *The Color of Odors* [(Morrot)](#_fx43t5j4a9n) , researchers found that the smell and taste of wine could be affected by changing the color of the wine. In the Journal of Wine Economics, Robin Goldstein researched the relationship between consumer enjoyment of wine and its price. He found that the relationship was not always positively correlated among individuals without wine training [(Goldstein)](#_xi42lfaf50v4). There have also been studies regarding wine competitions. Robert Hogson found that “...winning a Gold medal is greatly influenced by chance alone” [(Hodgson)](#_1qtb9j2k8c32).

All of this previous research indicates that average consumers (consumers who have not received wine sommelier training), have a hard time distinguishing between the quality and price of wines just by tasting the wine. For this reason, we chose to create a simplified price-estimation task, where we asked participants to taste differently priced wines of the same type, and match it to the list price of the wine. Based on the existing research, we expected this to still be a challenging task to complete based on tasting the wine alone. In order to understand the impact of seeing additional information that can be found on the wine bottle, our study introduces incremental information to participants, because we assume that the more information is presented to participants, the better participants will perform the task. This is because the wine labels and bottles themselves convey information through aesthetics, brand, cork variety, and vintage.

We hypothesized that being provided with additional information such as a wine-card or the actual wine bottle itself would improve participants' accuracy in matching the wine to the correct price as opposed to allowing them to taste the wines without seeing this information.

# 2. **Experimental Design**

## **Experiment Overview**

To test the hypothesis that seeing wine labels impacts the perception of wine prices, we designed a lab experiment that aims to measure people’s ability to distinguish between wines of different prices when they have access to different levels of information about each wine. The following figure shows a timeline of all interactions with the experiment participants:



Figure 1: Timeline of all interactions with the experiment participants

For the experiment, we selected four different Pinot Noir wines from France. This decision was taken to minimize potential bias that could be introduced by using different types of wines. These bottles were also selected and purchased from a specialty shop rather than a grocery store to minimize the chances of subjects being familiar with any of the wines beforehand. The table below lists the wines along with the labels that were assigned to them for the experiment.

| **Label** | **Price (masked)** | **Wine Name (available to treatment groups only)** |
| --- | --- | --- |
| A | $6.99 | [Loudenotte Cuvée Réservée Pinot Noir](https://www.vivino.com/US/en/loudenotte-cuvee-reservee-pinot-noir/w/7511227) |
| B | $64.00 | [Santenay “Premier Cru” “Les Gravières”](https://shopbanquet.com/northberkeleyimports/products/remoissenet-pere-et-fils-santenay-1er-cru-les-gravieres-2019/61565fa721559c2dd3724d44) |
| C | $16.00 | [Cave De Gortona Pinot Noir](http://www.northberkeleyimports.com/wordpress/wp-content/uploads/2019/04/GORTONA_pinot.pdf) |
| D | $36.00 | [Bourgogne Pinot Noir](http://www.northberkeleyimports.com/wordpress/wp-content/uploads/2016/01/GAVIGNET_bourgpn1.pdf) |

Table 1: Wines with their corresponding letter labels

On the day of the experiment, participants arrived at a lobby and checked-in with a member of our team. We asked the participants to self-report some information detailed further below following which the participants were randomly assigned to one of three groups:

* **Control:** This group could only taste the four wines and did not have access to any additional information about the wines during the wine-tasting
* **Treatment A:** This group was provided a printed wine card that listed the name of the wine, region of France, type of wine and alcohol percentage while tasting the wines.
* **Treatment B:** This group was allowed to see and interact with all four bottles of wine while tasting the wines.

Participants self-reported the following information through an online form (screenshots provided in Appendix 3 to 6) presented to them on a tablet during the check-in process. This information is used as covariates in our analysis section:

* **Race/ethnicity:** Multiple choice categorical
* **Gender identity:** Multiple choice categorical
* **Age:** Continuous variable
* **Hunger level:** self reported hunger level prior to wine tasting (categorical)
* **Wine expertise:** self reported familiarity with wine (categorical)
* **Wine consumption:** number of days per week they consume wine (continuous)
* **Wine adventurousness:** how often they try new wines (categorical)

After completing the check-in process and being assigned to a group, each participant was individually taken to a wine tasting room to ensure they were isolated and not influenced by other participants. In the wine tasting room, two team members prepared the experimental setup based on which group each participant belonged to and followed a script for each group to ensure that information about the experiment was delivered consistently. During the tasting, each participant was presented with four glasses of wine labeled A, B, C, and D. The labels were used in a multiple choice wine price matching question each participant had to answer. The question was presented to them on an online form on a tablet and is provided for reference in the figure below.



Figure 2: Wine price matching question

The responses collected in the form were used to compute the **outcome variables** for our analysis. The primary outcome variable is the **percentage of correctly matched wine-price pairs** in their response. In the analysis section, we describe further how these percentages are computed for use in regression analysis.

Based on our experimental setup, we have **three potential outcomes:**

**Y(0):** Percentage of correctly matched wine-price pairs if the subject is not exposed to any additional information (neither wine-card nor bottles)

**Y(1):** Percentage of correctly matched wine-price pairs if the subject is exposed to a wine card containing each wine’s name, region, type, and alcohol percentage.

**Y(2):** Percentage of correctly matched wine-price pairs if the subject is exposed to each wine’s bottle.

## **Recruitment**

Given that we had budget and time constraints, we designed a lab experiment that focused primarily on the population in our extended networks in the Berkeley area, most of whom were fellow students at UC Berkeley and their partners. Arranging a large scale wine tasting experiment with people from around the Bay Area presented operational challenges that we wanted to avoid. The downside of this approach is that we did not have random selection from the population - we could only perform random assignment to the treatment and control groups. The implications of this choice is that we cannot guarantee external validity of our results. In other words, our results would not generalize well outside the Berkeley population.

### **Power Analysis and Sample Size**

Before starting our recruitment process, we conducted a power analysis to understand the impact of different ATE values in our experiment and ultimately define the target sample size that would be needed in order to find statistical significance in case the average treatment effects are what we expect them to be.

During our power analysis, we considered only one control and one treatment group. In the actual experiment, we split treatment into categories A and B based on feedback from the initial power analysis. We ran three power simulations using ATE values of 20%, 10%, and 5% percentage accuracy respectively. We also assumed that the direction of the ATE will be positive i.e. the treatment group will have a higher percentage accuracy compared to the control group. For the control group and treatment groups, we added a standard deviation of 15% and 10% respectively. We selected the ATE values and other parameters in the simulation based on our review of prior research in the area while also considering reasonable bounds for simulating a wide enough spread in the ATE values.

As shown in the plot below, we find that in the 20% and 10% ATE scenarios, we see relatively high power, ranging from 65% to 100%. On the other hand, if we obtain smaller ATE’s in our experiment, we will not be able to conclusively reject the null hypothesis that seeing the wine bottle has no impact on the ability to predict wine bottle prices.

The following figure shows the power as a function of the sample size for all three ATE scenarios:

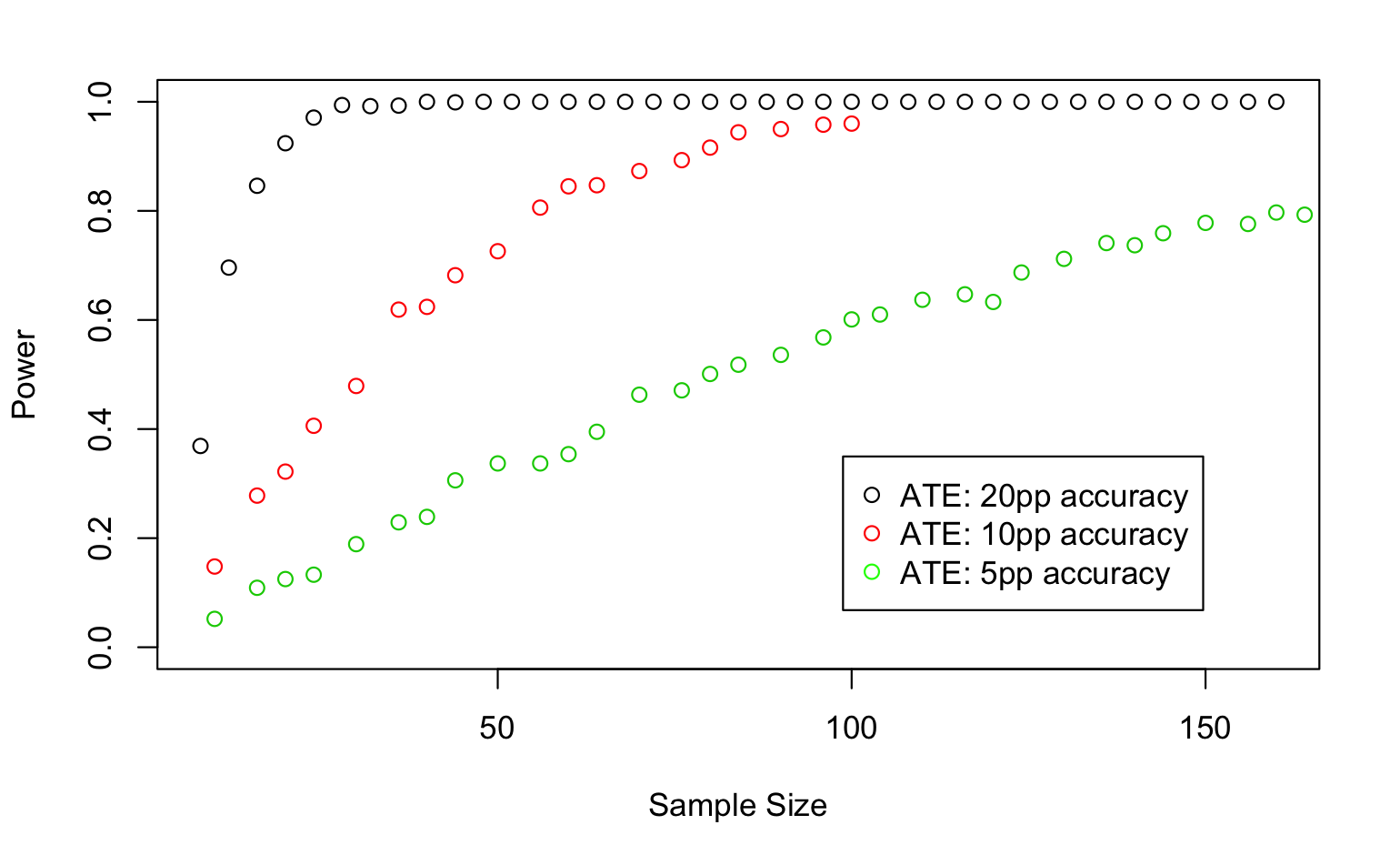


Figure 3: Power Analysis: We would have good chances (at least 70%) of obtaining statistical significant effects if the ATE is at least 10pp and the sample size is over 50 subjects.

We therefore defined a target sample size of 50 subjects to be sufficient to reject the null hypothesis in case the ATE is at least 10pp.

### **Recruitment process**

After obtaining a target sample size of 50 individuals, we created and sent an initial interest form through UC Berkeley slack channels and email lists on 11/15/23 to create some exposure and interest in the experiment within the UC Berkeley community. With this survey we aimed to understand the number of potential participants that would be interested in this study, obtain their contact information so we can send them follow-up notifications, and define the ideal schedule for the experiment that would maximize conversion among the group of interested participants. 23 participants expressed interest from the initial survey - many of whom mentioned that they would bring friends and relatives. The ideal schedule for conducting the experiment was decided to be 11/18/23 from 5-9 pm. Word of mouth messaging among the participants and others in the Berkeley area helped us grow the sample size to the final number of 47 participants.

During the day before the experiment (11/17/23), we sent notifications to all responders to confirm or update their attendance to the event. To provide additional incentives for those that showed up, we organized a social event after the experiment with free drinks and snacks. Only two of the respondents did not show up for the experiment after this follow-up confirmation notification. The following figure summarizes our recruitment process and flow of participants:

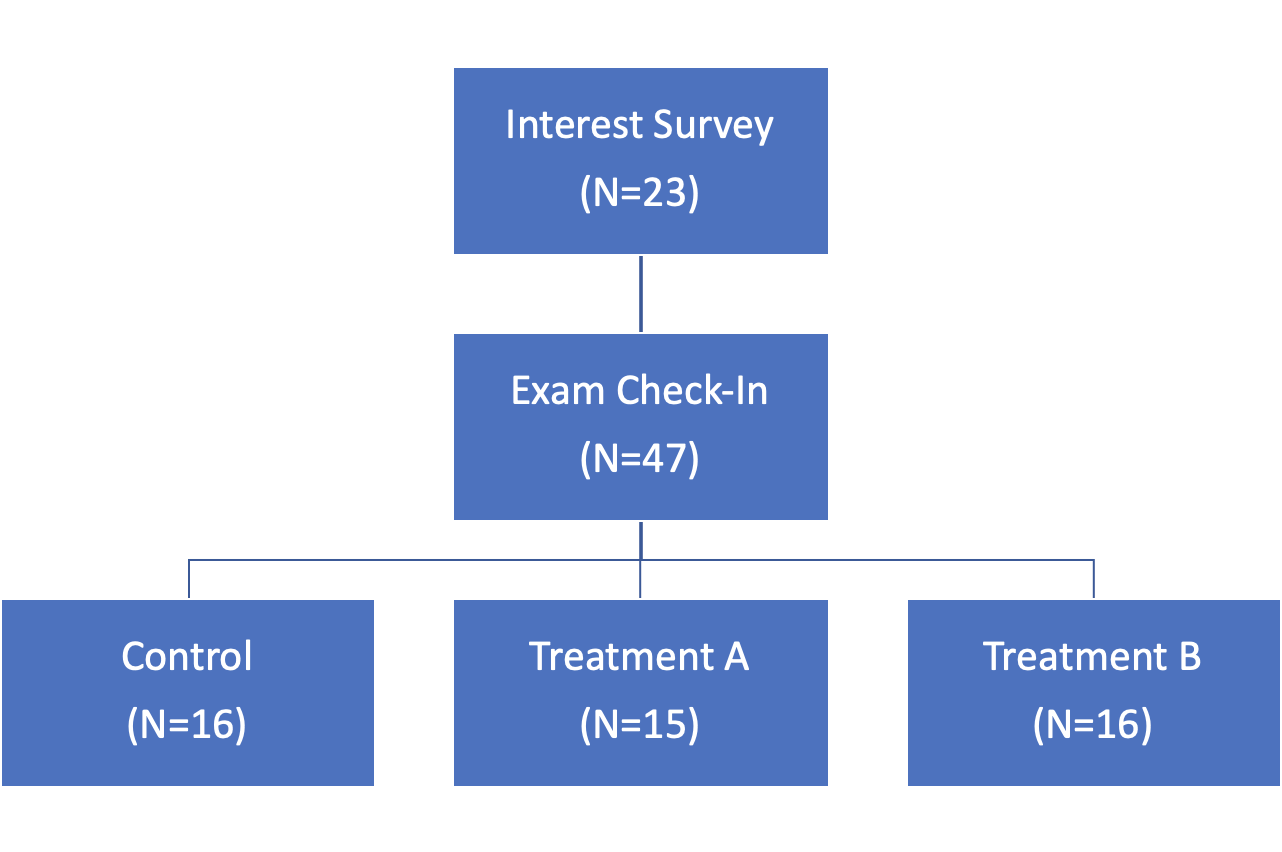


Figure 4: Flow of observations measured through the recruitment and randomization processes.

## **Randomization**

One of the main challenges in our experimental design was randomization. Because we were recruiting participants through a survey several days before the event, we anticipated that not only many of them were not going to show up, but also that word of mouth was going to bring unexpected participants during the event. This meant that if we performed randomization on the potential participants in our surveys we would risk group imbalance if there was attrition. We were also concerned that if we randomly assign the participants when they arrived at check-in time, we would risk creating group imbalance in this scenario as well as introduce potential challenges in experiment administration.

To address these concerns, we decided to implement a pre-randomization process as follows:

1. We created a list of numbers from 1 to 70.
2. We randomly assigned each number to either Control, Treatment A, or Treatment B.
3. We checked if the treatment assignment was roughly balanced at sample sizes of 30, 50, and 70 to ensure that we had sufficient participants in each group for each of these turnout rates.
4. After we obtained a balanced random list at all 3 sample sizes, we assigned participants to each number incrementally starting at the number 1 as they arrived and checked-in for the experiment.

We believe that this approach was reasonable because it **eliminates attrition** while still maintaining random assignment with a guarantee that the groups are on average pretty balanced. Another concern that we had was that the order of participants could introduce some bias in the results. For example, more motivated or experienced wine tasters could show up before inexperienced wine tasters. To address this potential bias, we introduced covariates to control for some of these potential sources of bias and we also ran spillover analysis to test if the ATE of the first half of participants was different from the ATE of the second half of participants. This test is detailed further in the analysis section below and did not show significant differences across covariates, meaning that participants with an early check-in were similar to those with a late check-in.

## **Delivery of Treatment**

Since this was a lab experiment under direct supervision by the researchers, it was straightforward to guarantee delivery of treatment to subjects. Before a participant entered the wine-tasting room, the researchers made sure to set up the wine-tasting room appropriately. The same amount of wine was measured and poured in new recyclable glasses, glasses were ordered in the same way for every subject, and exactly the same instructions were given to each participant in each group. In addition, the wine-card and bottles presented to the respective treatment groups were placed in the same position for each participant. Given these measures, we are confident that we were able to **guarantee compliance** and ensure that the experiment was conducted with the same environment and conditions within each group.

## **Inclusion Criteria**

For obvious reasons, we only encouraged and accepted participation for people over the age of 21 who were legally allowed to drink alcohol. No other inclusion criteria was enforced.

## **Exclusion Criteria**

Right after each participant submitted their response to the wine-price matching question, one of the researchers asked them if they knew or recognized any of the four wines or bottles presented. If any of the subjects knew at least one of the wines, they would automatically be excluded from the dataset. Fortunately, none of the participants were excluded for this reason.

# 3. Analysis

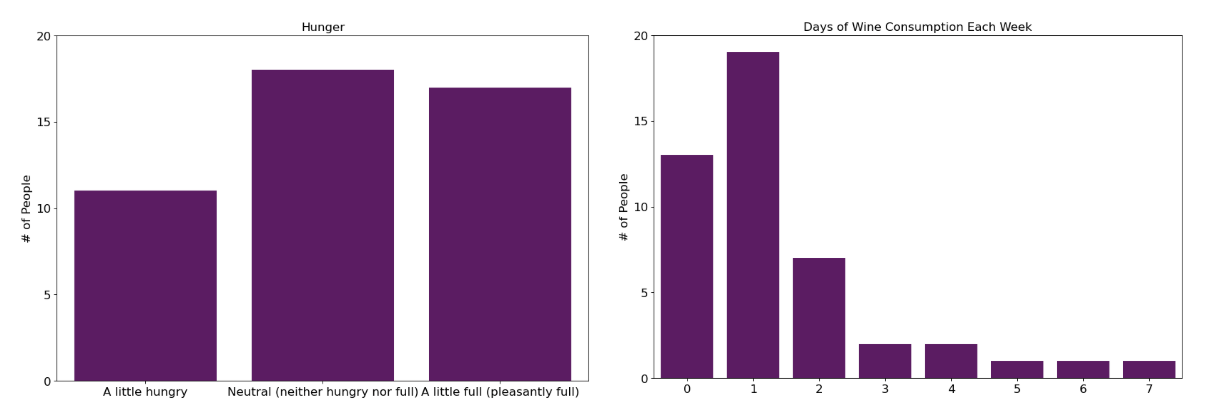
### **Data**

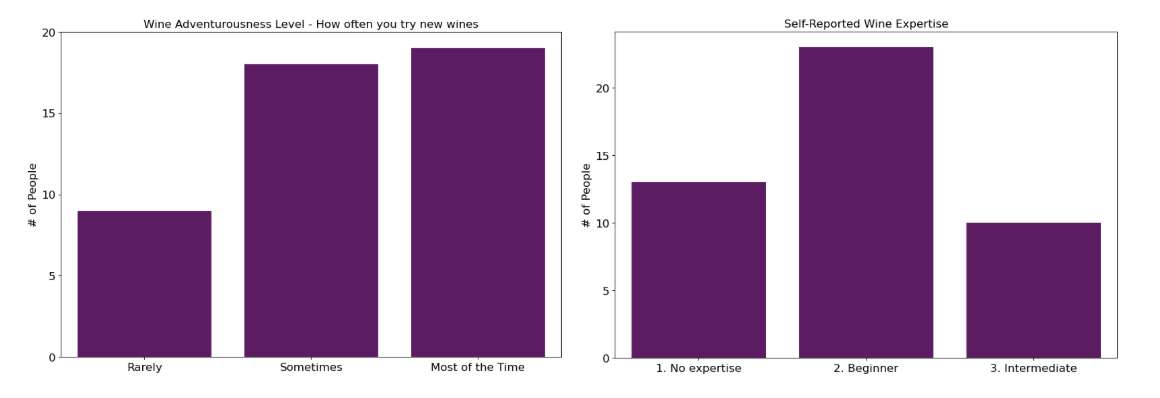
As outlined above, our **data collection process** on the day of the experiment involved the following steps:

* First, participants who arrived at the lobby were greeted by a member from our team who added them to a list that had arrival positions randomized for assignment into control, treatment A, or treatment B.
* After a participant was checked in and assigned to a group, they were handed a tablet while they waited in the lobby and were asked to self-report data including the 7 covariates mentioned above.
* Finally, in the room where the wine tasting was conducted, a participant was also presented with a tablet where they recorded their responses for the wine to price matching.

The data was collected using Google Sheets and was used to generate a data frame for analysis that had 47 rows representing the observations for each participant.

Figure 5 below shows the distribution of the 7 **covariates** we collected from the participants. Collected information includes age, race, and gender. Further, based on covariates in prior research, we also collected information on the subject’s level of hunger, wine expertise, amount of weekly wine consumption and wine adventurousness (how frequently subjects try new wine). Individual distribution plots of the covariates we collected are presented below.





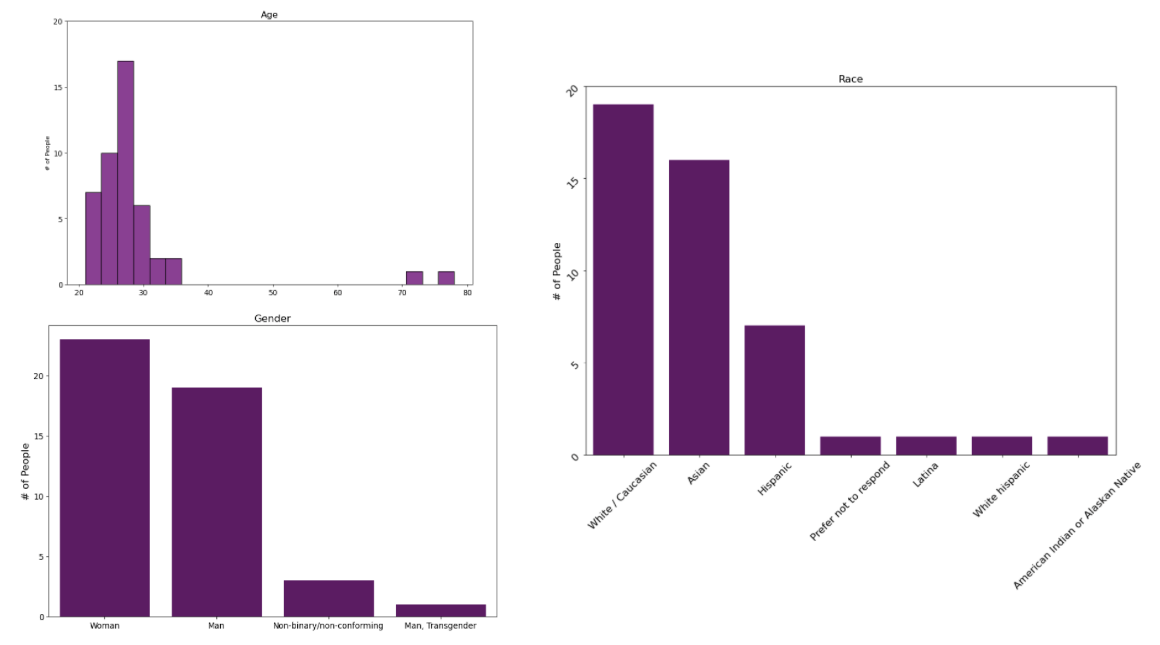


Figure 5: Distribution of the 7 covariates collected

The **outcome variable** used in our analysis is the percentage accuracy of correctly matched wine-price pairs. Due to the degrees of freedom, subjects could answer with correct percentages of 0%, 25%, 50% and 100%. We expected that this outcome variable would vary based on the treatment assignment and mapped it to the **operational space** of the models we ran to test if the outcome was impacted by treatment in a statistically significant manner.

Figure 6 shows the distribution of wine price guesses for each wine used in the experiment. The tall bars associated with the $6.99 and $16.00 wines on the left as well as the tall bars associated with the $36.00 and $64.00 wines on the right form loose clusters which suggests that on average, people are able to separate the lower priced wines from the higher priced wines with reasonable accuracy.

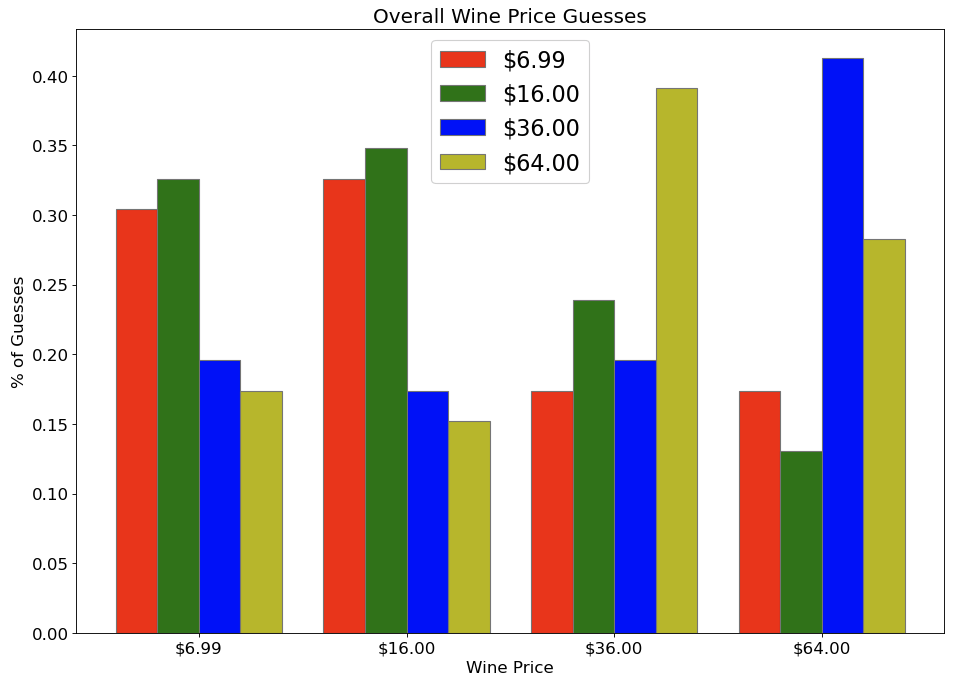


Figure 6: Distribution of wine price guesses for each wine bottle

Figure 7 shows the distribution of wine price guesses in the control and the two treatment groups in the experiment. No group was better overall at the price-matching task, and there were no statistically significant differences of group performance.

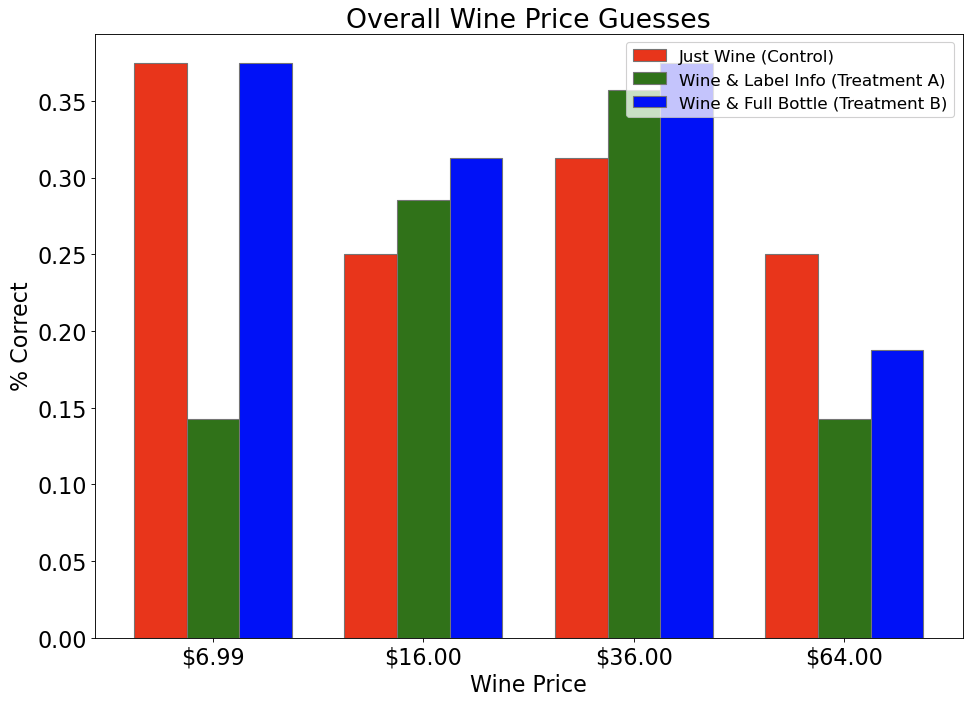


Figure 7: Distribution of wine price guesses in the control and two treatment groups

### **Models and Findings**

Prior to running the full analysis on our results, we first checked to confirm that we had no issues in our data. This included checking for covariate balance, heterogeneous treatment effects, and spillover.

To ensure that we didn’t have any issues with covariate imbalance, for each treatment we generated two regressions. One where we regressed a subjects’ treatment assignment onto a constant and one where we regressed the subjects’ treatment assignment onto all of the covariates. The constant model finds the distribution of treatment assignments based on the random assignment procedure. The full-covariate model shows whether any of the covariates are predictive of which treatment group a subject would be assigned to. We found that none of the covariates had statistically significant coefficients, showing that none of the individual covariates were predictive of treatment assignment. Further, for each treatment, we ran an ANOVA between the constant model and full-covariate model. As shown below, we see that the ANOVA returned statistically insignificant results for both treatment A and treatment B, showing that our covariates are balanced.



Figure 8: ANOVA between the constant model and full-covariate model (covariate balance check)

We additionally checked to ensure that there was no heterogeneous treatment effect. To do so, we selected the covariates that we thought might be potential heterogeneous treatment effects, specifically the self-reported wine expertise and gender.

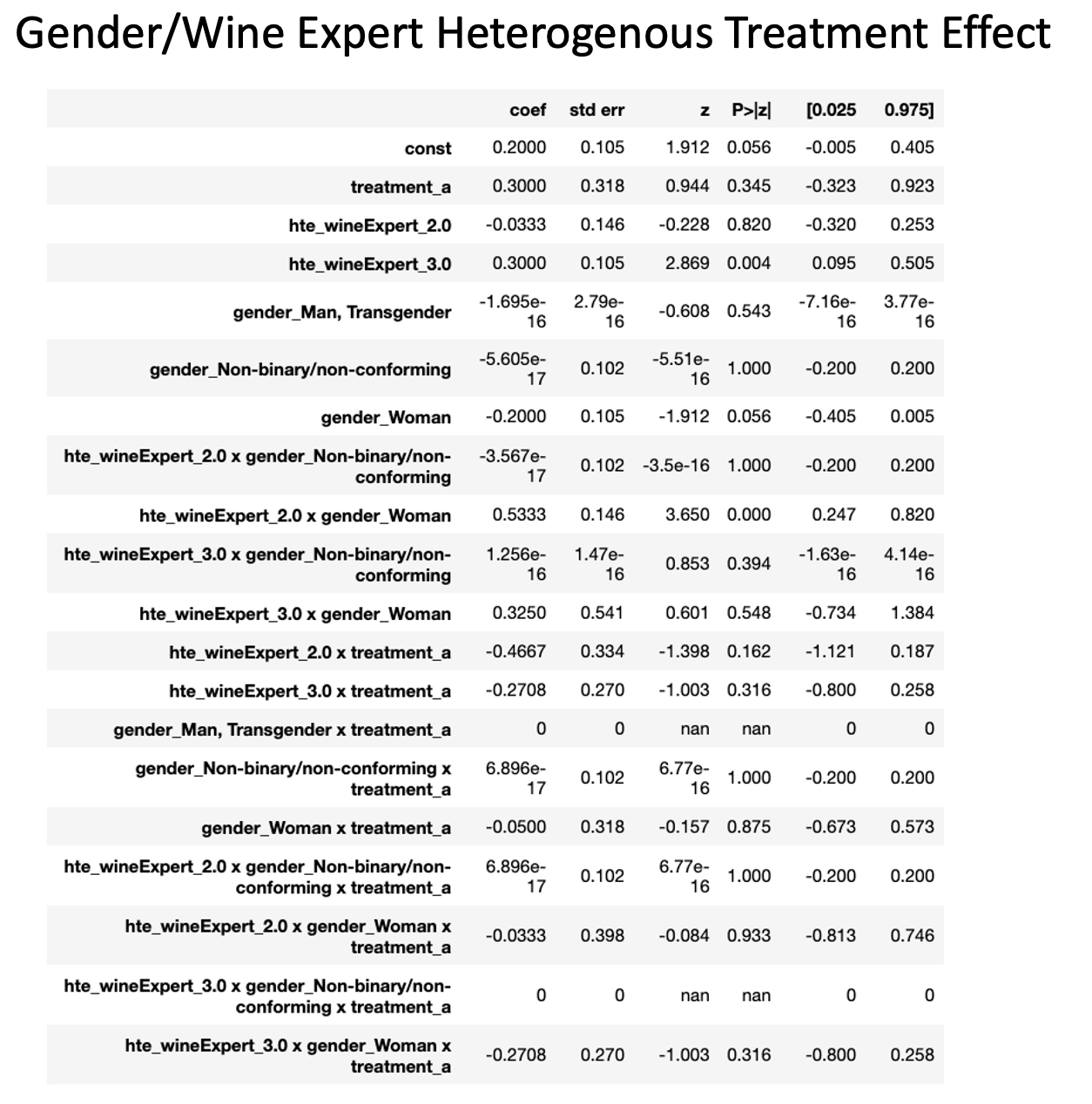


Figure 9: Gender/Wine Expert HTE regression

In the analysis above, we regressed the outcome variable onto the treatment assignment, the self-reported wine expertise category, and gender category, and the interaction terms between all the variables. We see that none of the interactions between the covariates and the treatment are significant, implying that we do not have any heterogeneous treatment effects. In the chart above, we show the HTE analysis for just treatment A. An identical analysis was run for treatment B, showing that there were not heterogeneous treatment effects for treatment B either. The “nan” found in some of the rows correspond to no samples in the interaction group.

One other interesting take away from the analysis was that the standalone “hte\_wineExpert\_3.0” variable, the binary classification for someone reporting themselves as an “Intermediate-Level Wine Expert”, did have a statistically significant p-value of 0.004. Based on this regression, we expect people who self-identify as “Intermediate-Level” wine experts would on average score 30% higher than those who identify as “No Expertise”. Given the number of experiments run & the combination of covariate slicing, we were concerned that this significant p-value may be the result of the multiple testing problem and actually be a false positive. To gain more confidence in this finding, we ran two more regressions in which we predicted the outcome variable against the self-reported wine expertise as a likert style ordinal variable and categorical variable.



Figure 10: ATE with controls for self reported wine expertise score

As shown above, when treating the “Intermediate-Level” wine expertise as a category, we see that the coefficient is significant at the 0.1 level and very close to being significant at the 0.05 level. Further, when we treat the ranking of self-reported wine expertise as a Likert scale continuous variable, we find that the “WineExpertise\_Int” variable is significant at the 0.05 level. While this test does not conclusively support the significance of intermediate-level expertise in predicting wine accuracy, we believe that given the different ways we tested the hypothesis & the logical nature of the claim, it is reasonable to believe that this is not a false positive due to multiple testing.

Another potential issue with the implementation of the experiment is spillover. As previously mentioned, we tested subjects one-by-one. While we asked participants to not discuss the experiment with others who have not taken the experiment, we had no formal way to enforce this request. As such, we wanted to test whether participants who we tested later had a statistically significantly different outcome than those who we tested earlier. To check this, we split the participants into two groups based on if they were tested in the first or second half. Then we regressed the outcome variable onto the participants' temporal group assignment.

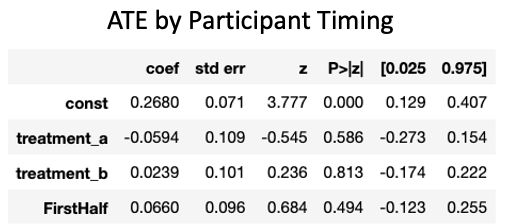


Figure 11: Spillover Check

In the results above, we see that the p-value for the “FirstHalf'' is not statistically significant. We can interpret this to mean that those in the first half did not perform significantly worse or better than those in the second half. Based on this result, we can assume that spillover effects did not significantly impact our study.

To estimate the ATE, which is functionally the difference in treatment group means relative to the control group, we ran 6 linear regression models and computed the heteroskedasticity robust standard errors using HC3 covariance. We had one control, and two treatment groups which were used in all of these regressions. The regressions differed in the way the outcome variable was defined. In the first regression, we used the overall percentage accuracy as the outcome variable being measured. We also ran 4 separate regressions where only the percentage accuracy associated with a single bottle of wine was used as the outcome variable in each regression. Finally, we create an outcome variable by splitting the wines into binary categories of cheap vs expensive and measuring the percentage accuracy of guessing a wine as cheap or expensive even if the specific guess within each group for the individual bottle was not accurate.

The chart below shows the results of the first model where we regress the subject’s price picking accuracy against whether they were assigned to treatment A or treatment B. We find that treatment A and treatment B have ATEs of -6.5% and 1.6% respectively. However, the p-values associated with both these ATEs are not significant at the 5% level of significance. This could suggest that our treatment of showing the labels and the bottles to participants did not affect their perception of price. It is worth noting that we could also have a power issue given our sample size of only 50 people. The direction of the ATE which is negative for treatment group A is opposite of what we had in our pre-experimental hypothesis.

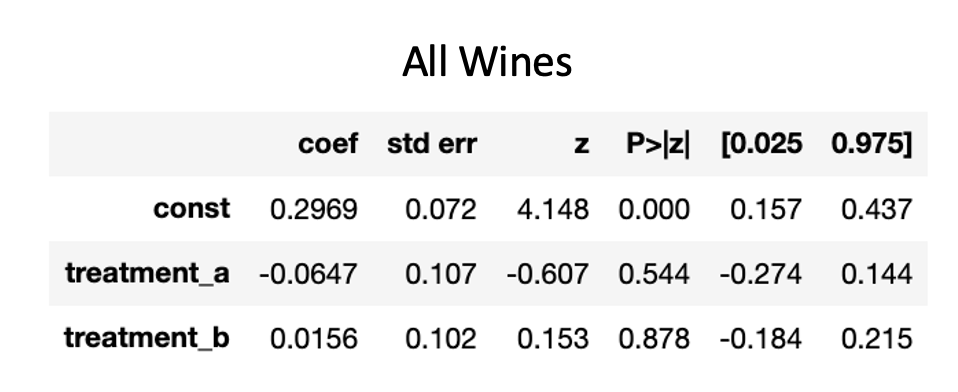
****

Figure 12: Experimental ATE for both treatment groups

The following tables show the results from the four separate regressions where only the percentage accuracy associated with a single bottle of wine was used as the outcome variable in each regression. Similar to the regression for full wines, we find that the coefficients of treatment A and treatment B in each of these regressions are not statistically significant. Given the findings, we cannot conclude that either treatment A or treatment B improved the subjects ability to predict the price of a wine.

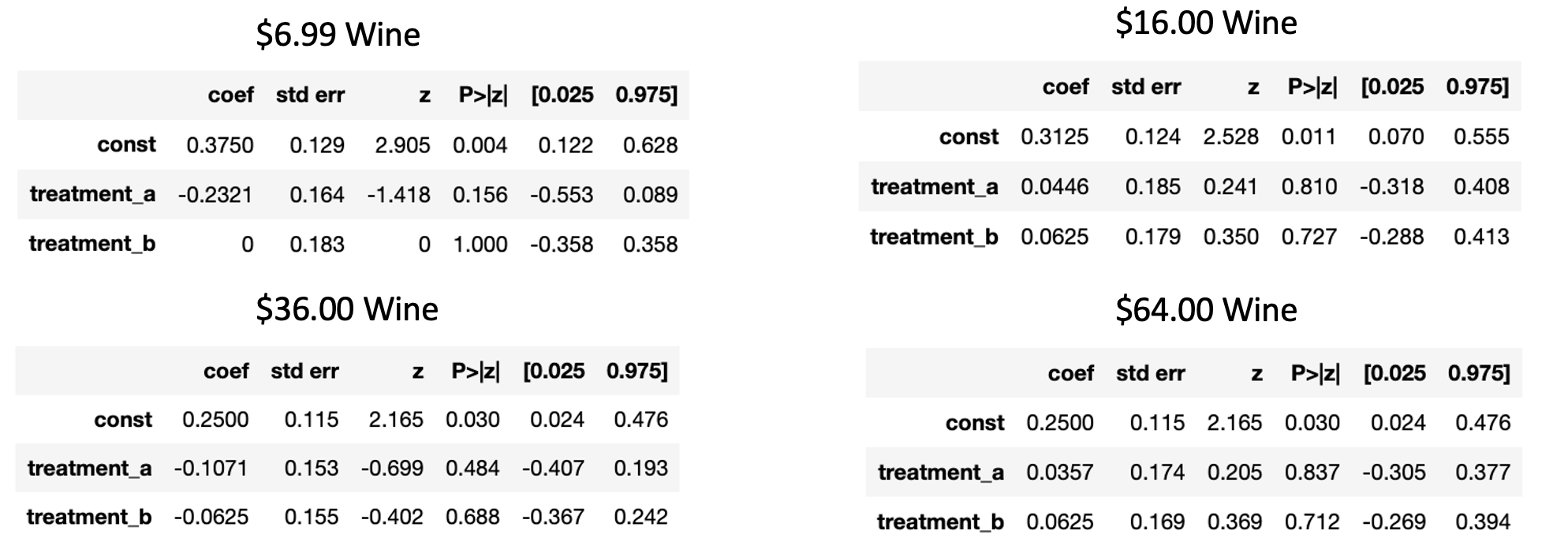
****

Figure 13: Experimental ATE for percentage accuracy computed for each wine

As previously shown, we know that the subjects who identified as a “Intermediate-Level Expertise” were statistically significantly better at predicting wine prices than those who identified as “No Expertise” or “Beginner-Level Expertise”. We therefore also, regressed people’s total wine accuracy on treatment A, treatment B, and a dummy variable corresponding to if the subject identified as “Intermediate-Level Expertise”

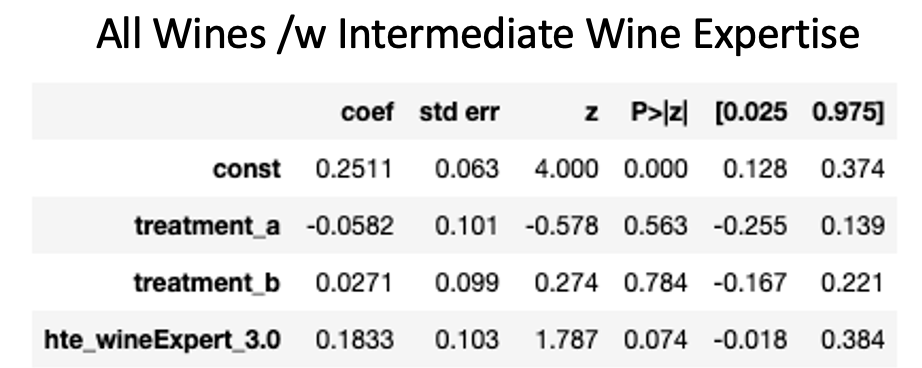


Figure 14: Experimental ATE controlled for the significant covariate (intermediate wine expertise)

We again find that treatment a and treatment b are not significant and whether you are an “Intermediate Wine Expert” is significant only at the 0.1 threshold level.

Additionally, we ran a fully saturated model including all of the collected covariates. Again, we find that none of the coefficients are statistically significant. One interesting difference in this model compared to the prior models is that in the fully saturated version of the experiment, the constant is also non-significant.

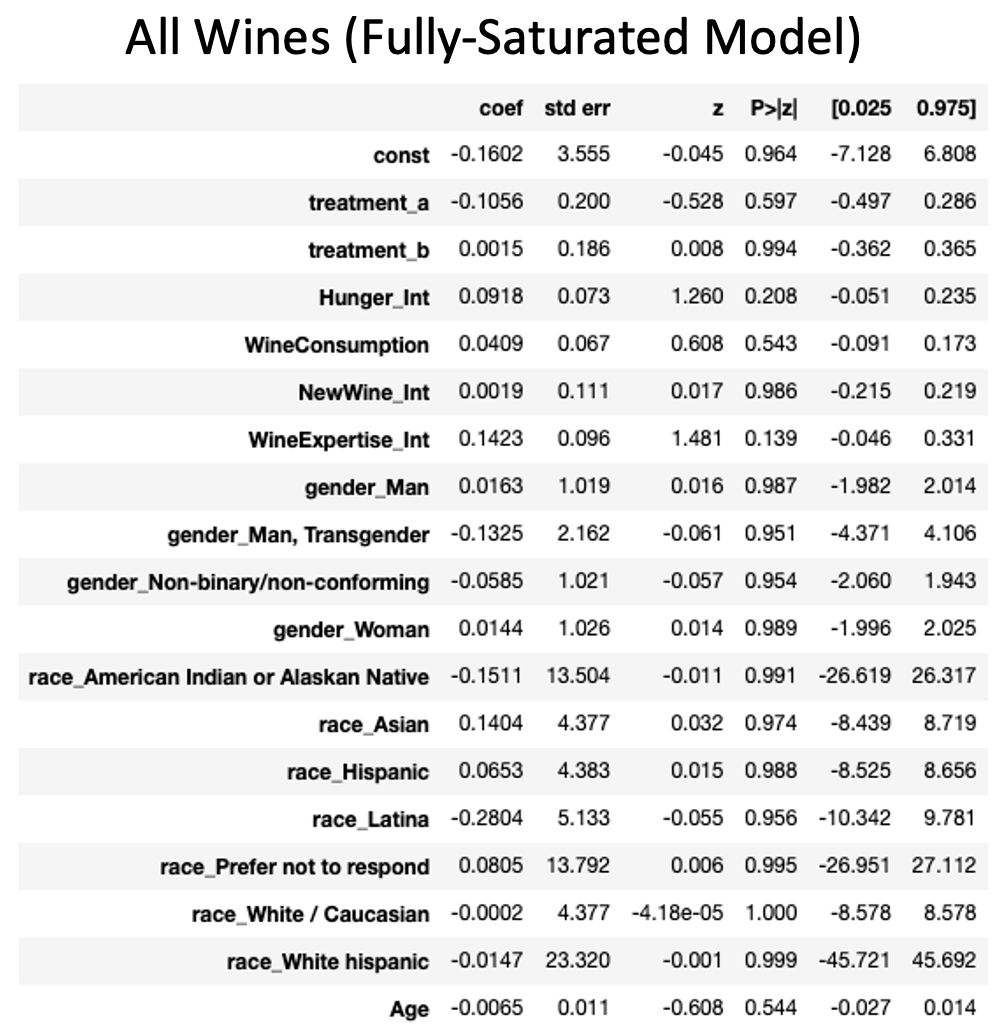


Figure 15: Experimental ATE for the fully saturated model

Finally, based on our results, we felt that it was relatively clear that the additional information provided by treatment A and treatment B was not statistically significantly aiding people in matching specific wines. However, we wanted to see if people were unable to differentiate due to the granular nature of the wine. In other words, with four wines within $6.99 and $64.00, there was not enough range for people to be able to accurately determine the price of a given wine. To test this hypothesis, we split the wines into the cheaper two and more expensive two. We then calculated the percentage accuracy of guessing whether a wine was in the cheap or expensive category. For example, if a subject assigned either the $6.99 or $16.00 wine to the $6.99 category, they would be considered correct.

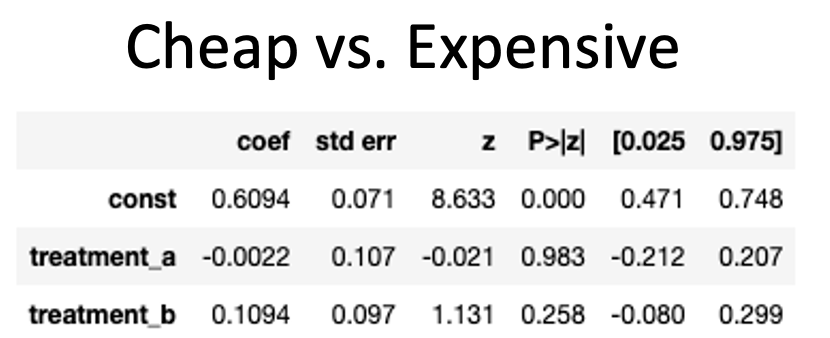


Figure 16: Experimental ATE for percentage accuracy computed within cheap and expensive categories

Again, we see that the p-value is not significant for either of the treatment effects. This result does further support the statement made above that people were relatively successful in predicting whether a wine was in the cheap or expensive category. Here a constant coefficient of 0.6094 shows that subjects correctly assigned wines to expensive or cheap ~60% of the time.

# 4. Learnings

Looking back on this project, there are a number of decisions that we felt could have improved the implementation of the experiment.

First, we should have attempted to get a greater sample size and further diversify our subject population. Our primary method of recruiting was posting on the Berkeley iSchool Slack channels & reaching out to individual friends and family. One option that could have potentially increased our sample size would be to promote the event to non-iSchool through email recruiting or other more general methods. Similarly, we found that no subjects stated that they had a wine expertise greater than “Intermediate-Level” expertise. By diversifying our sampling population, perhaps to include people who are more experienced with wine, we could have better generalized our results.

After the experiment we had a number of questions that could have been answered by gathering more covariate data. For example, we could have asked participants what their grape varietal of choice is to see if those who liked Pinot Noir performed better as well as checking for covariate balance across the groups.

Finally, we noticed that many of the subjects who were assigned to the treatment groups did not really look at the information or the bottle itself. Given that we didn’t not explicitly state that they should look at the provided info, we did not view this as non-compliance. However, looking back, if we wanted to truly understand the impact of seeing the bottle and label information, we could have presented those to the subjects prior to providing the actual wine itself. That way we could have ensured that participants focused on the content of the provided information.

# 5. Future Work

This experiment also gave a number of ideas for future research. First, our study used exclusively Pinot Noir from France. It would be interesting to try a similar experiment with different grapes across different regions. Another experiment could include different types of grapes within the same experiment to see how the variety impacts the outcome. Additionally, there are other versions of the same experiment that we considered as part of our design that could be informative. For example, we could give the same wine in front of two different bottles to see how just the aesthetic of the label, agnostic of the taste of the wine, impacts people's perception of the prices.

As mentioned in the conclusion, we are exclusively looking at how the information from the wine bottle impacts people’s perception during the tasting phase. There is a reasonable chance that actually tasting the wine distracts people, who pay less attention to the other information provided. However, the customer interacts with the bottle and wine label information when purchasing the wine as well. It may be interesting to run the experiment without having the customer actually taste the wine, but make the price determination exclusively based on the wine information and bottle.

# 6. Conclusion

Based on the results shown above, we are unable to reject the null hypothesis that presenting information found on a wine label and the bottle itself has no impact on the ability of subjects in our sampled population to perceive the price of wine at the time of testing.

At this time, we are unable to determine whether we failed to reject the null hypothesis as a result of not having enough power (too small of a sample size) to detect an effect, or whether there truly is no effect.

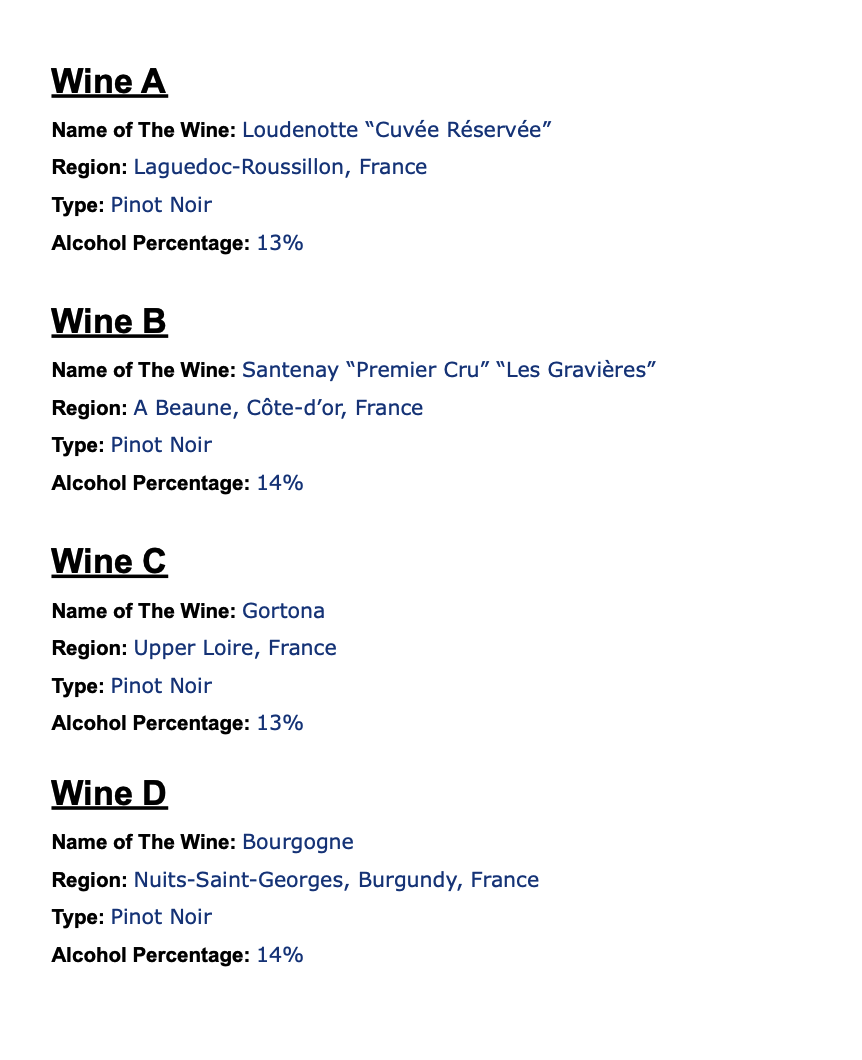
# 7. References

1. Morrot, Gil, et al. “The Color of Odors.” *Brain and Language*, vol. 79, no. 2, Nov. 2001, pp. 309–320, 10.1006/brln.2001.2493.
2. Goldstein, Robin. “DO MORE EXPENSIVE WINES TASTE BETTER? EVIDENCE from a LARGE SAMPLE of BLIND TASTINGS.” *AgEcon Search*, 2008, ageconsearch.umn.edu/record/37328. Accessed 12 Dec. 2022.
3. Hodgson, Robert T. “An Examination of Judge Reliability at a Major U.S. Wine Competition.” *Journal of Wine Economics*, vol. 3, no. 2, 2008, pp. 105–113, 10.1017/s1931436100001152. Accessed 27 June 2019.

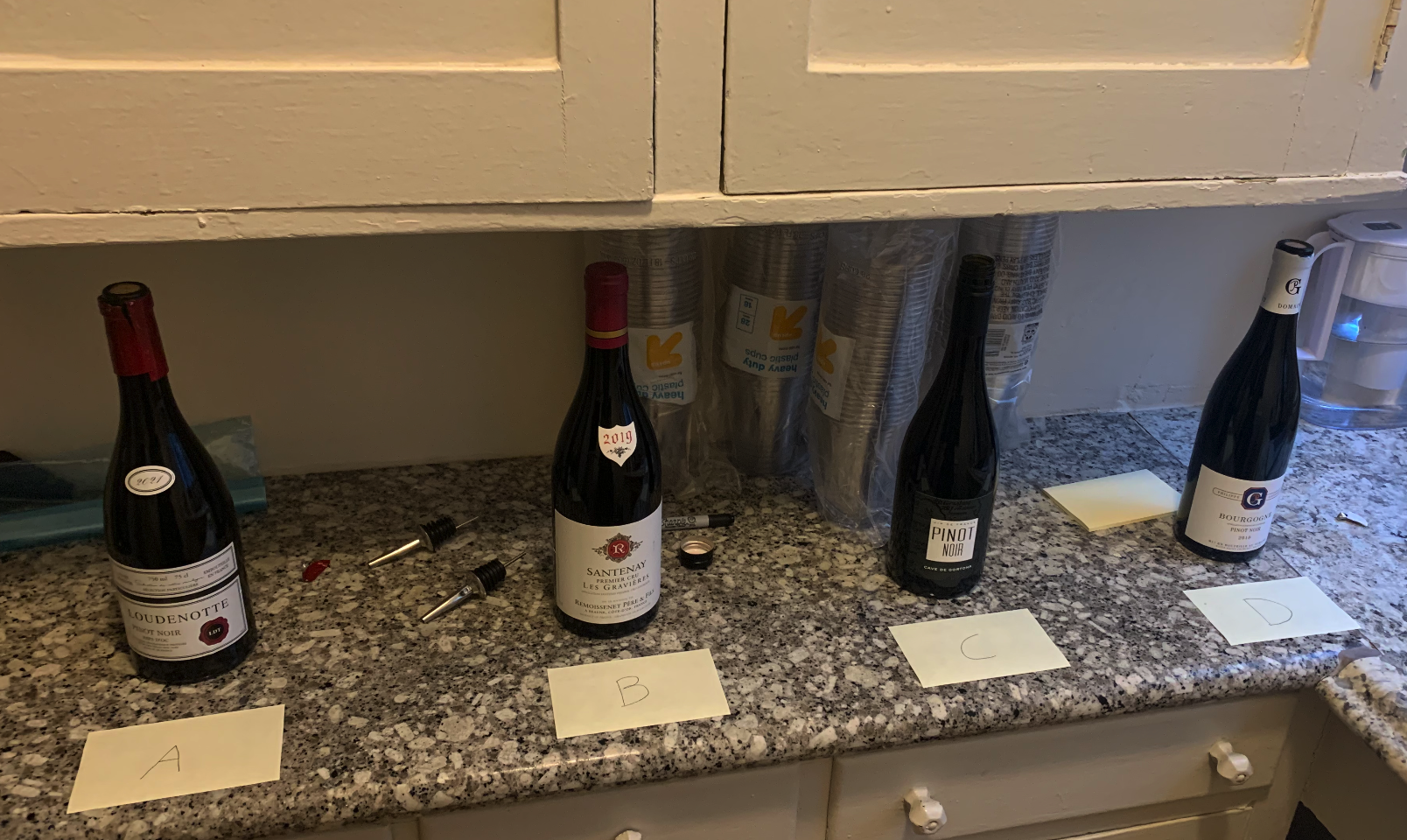
# 

# 8. Appendix

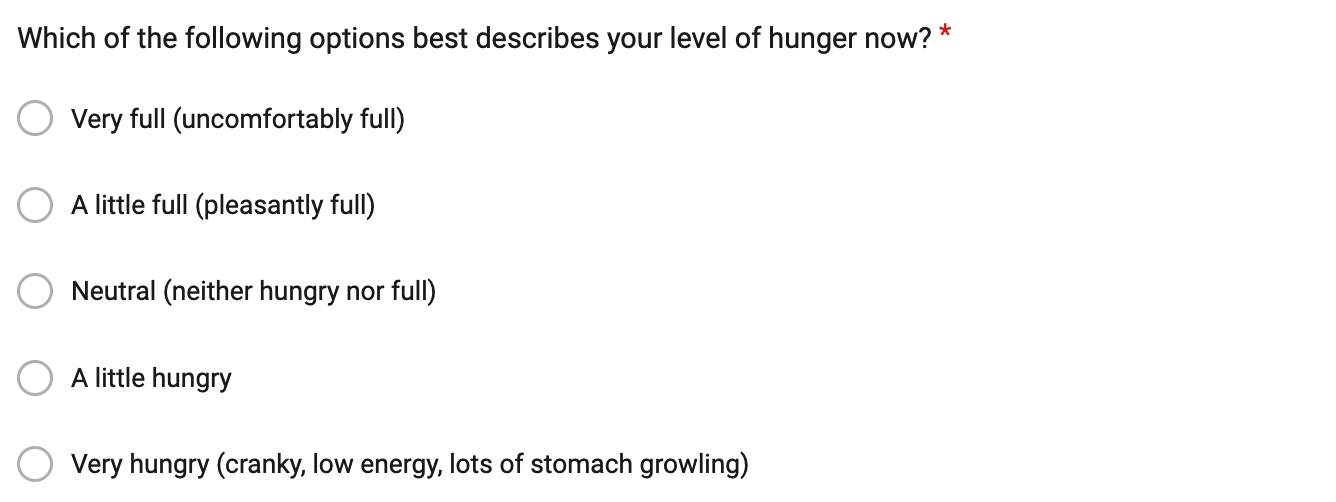
### **Treatment A: Wine labels containing limited information about the wine in a printed version.**

****

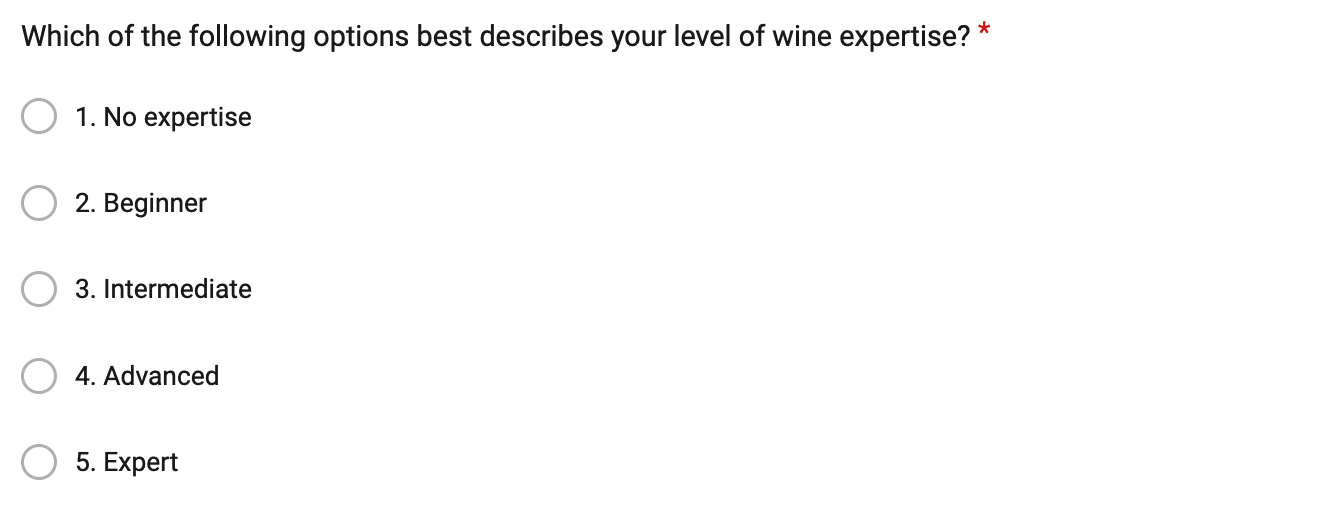
### **Treatment B: Bottles were presented to subjects**

****

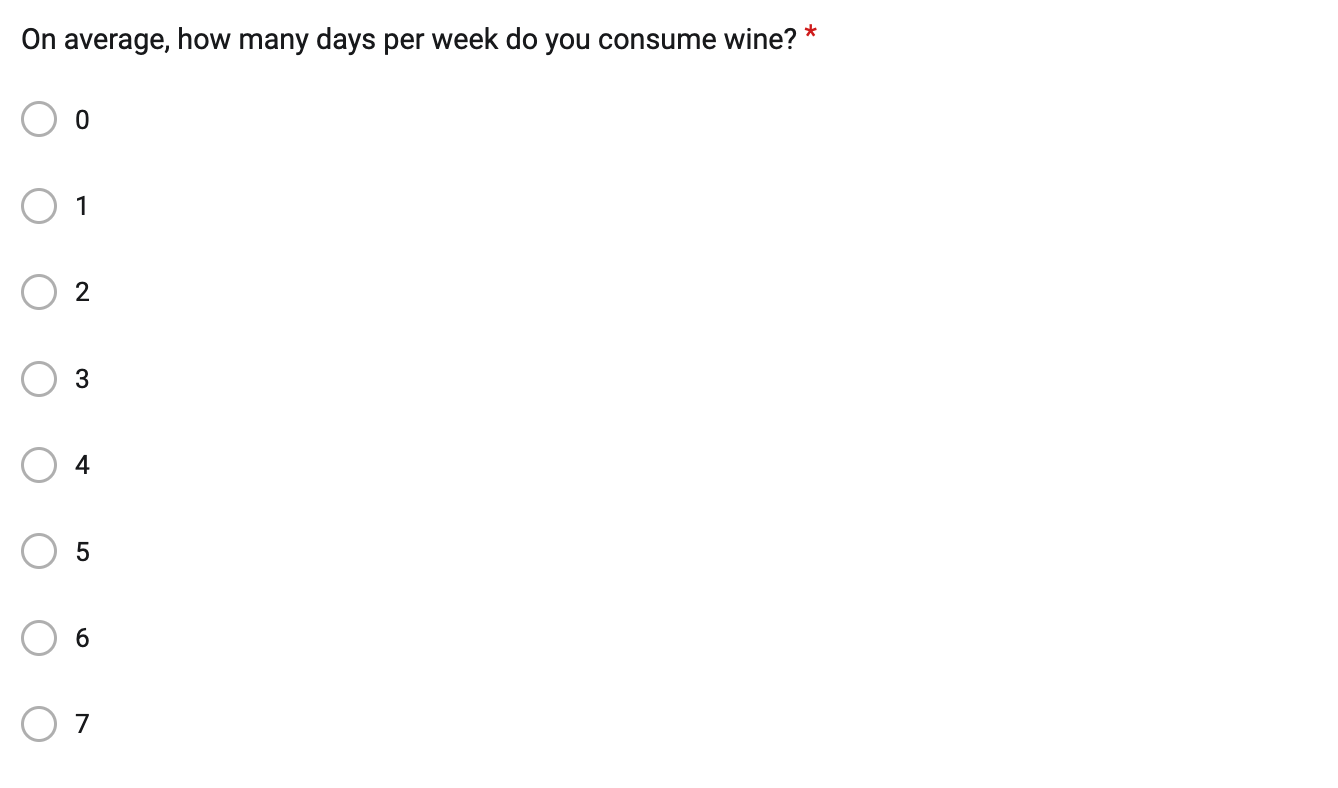
### **Survey: Hunger level question**



### **Survey: Wine expertise question**



### **Survey: Wine consumption question**



### **Survey: Wine adventurousness question**

