Variable description

Erlend Dancke Sandorf

15 12 2021

Data

Description

The data set contains data from a novel stated preference survey conducted in the UK in January of 2020.

The data is in the "wide" format with one row per choice situation.

Treatments

Treatment 1: A standard stated choice experiment with 3 experimentally designed alternatives and a "buy none" option.

Treatment 2: A standard stated choice experiment with 6 experimentally designed alternatives and a "buy none" option.

Treatment 3: A standard stated choice experiment with 9 experimentally designed alternatives and a "buy none" option.

Treatment 4: A sequential stated choice experiment with up to 9 experimentally designed alternatives and a "buy none" option. The "buy none" option was always visible and to see another alternative a respondent had to actively choose to reveal one by clicking a button under the choice task with the text "Reveal another bottle of wine". Up to 9 alternatives could be revealed.

Treatment 5: Same as T4, but each time a respondent clicked the button to reveal another alternative, they also had the option of indicating with a tick box which alternative was their current most preferred.

Treatment 6: Same as T5, but instead of only selecting one, respondents could select up to 3 alternatives that they were actively considering.

Treatment 7: Same as T6, but instead of only selecting a maximum of 3 alternatives, a respondent could indicate as many alternatives as they wanted, up to 9, that they were considering at any given point in the sequence.

Treatment 8: Same as T4, but search was costly. Cost was imposed by adding a random time delay from when an alternative was revealed until a respondent could choose to reveal another alternative. In this treatment the time delay was the same across alternatives and choice tasks.

Treatment 9: Same as T8, but the time delay varies across choice tasks within a respondent.

Treatment 10: Same as T9, but the time delay now also varies across alternatives within each choice task.

Data gathering and implementation

The data was gathered at the end of January 2020. The survey was programmed in Shiny. A live version of the survey instrument including how each variable was coded can be found here: https://survey.inspire-project.info.

Source code for the survey can be found here: https://github.com/edsandorf/inspire-project-survey

Sample

The sample was drawn from a non-probability based internet panel. The sample consisted of people aged 18 and above living in the UK at the time of the survey. No quotas were implemented in sampling. We do not have data on response rates, but they are probably not great.

Example of the implementation

To see a live version of the choice tasks: https://choice-tasks.inspire-project.info

Attribute descriptions

The attributes were determined based on what information was commonly displayed on supermarkets' shelves, websites, attributes that commonly feature in wine reviews, and feedback from informal focus groups.

To obtain the best levels for the attributes, we scraped the websites of three large supermarket chains in the UK to get the attribute information for all wines they sell. We only considered wines that were sold in 75cl bottles.

Country of origin

The attribute gives the country of origin for the wine. The levels of the attribute was limited to only include the 8 most common countries: Australia, Chile, France, Italy, New Zealand, South Africa, Spain and USA.

Grape variety

The attribute gives the grape variety used in the wine. If the wine contained more than one grape it was labelled as a blended. Based on the web scraping, we only included the 10 most common grape varieties: Blend, Cabernet Sauvignon, Chardonnay, Chenin Blanc, Malbec, Merlot, Pinot Grigio, Pinot Noir, Sauvignon Blanc and Tempranillo.

Color

Whether the wine was a red, white or rosé. This variable is (almost) perfectly correlated with grape variety, that is, you can infer the color of the wine based on the grape. Furthermore, it is closely linked to characteristic of the wine where red wines were given on the scale from light to full bodied and white wine and rosé were on a scale from sweet to dry. Takes on the levels: Red, white and rosé

Alcohol by volume

Alcohol by volume measures the alcoholic content of the wine. The attribute could take on the following levels: 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5, 15.0.

Characteristic

The characteristic of the wine was given on a five-point Likert scale. For white and rosé wines the scale ranged from dry to sweet, and for red wines from light bodied to full bodied.

Organic

Organic indicated whether the wine was organic.

Price

The price attribute was limited to include 35 levels in varying increments with smaller increments around the most common values and larger increments for very cheap or very expensive wines. We included no wine bottles less than £4 and not bottles costing more than £20. Bottles exceeding £20 are more often found in specialty shops.

Design

The full factorial design included 1,848,000 possible alternatives, i.e., attribute combinations. We excluded all unfeasible combinations such as grape variety and color, and grape variety and wine characteristic. This left us with 381,920 feasible alternatives (hereafter bottles of wine).

Using the data we scraped from the supermarkets, we calculated the frequency of each attribute level and under the assumption that the levels of one attribute were independent of another, calculated the probability that a given attribute combination would exist in reality, i.e., the likelihood of a particular wine to exist.

We used these probability weights to sample individual profiles from the restricted factorial each time a respondent entered the survey. The idea was to enhance the realism of our random design by making it more likely that the alternatives shown to an individual could exist in reality. Said another way, more likely to be familiar to the respondent.

Thre random experimental design ensures plenty of variation in the data and that we eliminate order and path dependency in choices between respondents. Furthermore, the random design also means that non-revealed alternatives allocated to an individual in treatments 4-10 are unlikely to systematically bias our estimates or affect our ability to make inference.

Table with variable names and descriptions

In the table the '*' is a placeholder for the numbers 1 through 10 and is the alternative indicator in the data set in "wide" format.

Variable name	Description
resp_id	A unique respondent id
ct	A choice task indicator from 1 to 10

Variable name	Description
treatment	An integer specifying which treatment a respondent was in
no_choice	An indicator equal to 1 if no choice was made in a given choice task
choice	Choice variable that takes on the value of the chosen alternative
impossible_choice	An indicator equal to 1 if a respondent chose an alternative that had not been revealed. This is likely an issue with data registration.
revealed_*	Indicator taking on the value of 1 if the alternative was revealed.
search_in_period_*	Indicator taking on the value of 1 if the respondent chose to reveal another alternative in choice task *
never_searched	Indicator taking on the value of 1 if a respondent never revealed an additional alternative
country_*	The country of origin in alternative *
color_*	The color of the wine in alternative *
alcohol_*	Alcohol my volume in alternative *
grape_*	Grape variety in alternative *
characteristic_*	Characteristic of the wine in alternative *
organic_*	Whether the wine was organic in alternative *
price_*	The price of the wine in alternative *
australia_*	An indicator for whether the wine was Australian in alternative *. Derived from the country of origin variable
chile_*	An indicator for whether the wine was Chilean in alternative *. Derived from the country of origin variable
france_*	An indicator for whether the wine was French in alternative *. Derived from the country of origin variable
italy_*	An indicator for whether the wine was Italian in alternative *. Derived from the country of origin variable
new_zealand_*	An indicator for whether the wine was New Zealandic in alternative *. Derived from the country of origin variable
south_africa_*	An indicator for whether the wine was South African in alternative *. Derived from the country of origin variable
spain_*	An indicator for whether the wine was Spanish in alternative *. Derived from the country of origin variable
usa_*	An indicator for whether the wine was American in alternative *. Derived from the country of origin variable
red_*	An indicator for whether the wine was red. Derived from the color variable.
rosé_*	An indicator for whether the wine was rosé. Derived from the color variable.
white_*	An indicator for whether the wine was white. Derived from the color variable.
blend_*	An indicator for whether the wine was a blended. Derived from the grape variety variable.
cabernet_sauvignon_*	An indicator for whether the wine was a Cabernet Sauvignon. Derived from the grape variety variable.
chardonnay_*	An indicator for whether the wine was a Chardonnay. Derived from the grape variety variable.
chenin_blanc_*	An indicator for whether the wine was a Chenin Blanc. Derived from the grape variety variable.
malbec_*	An indicator for whether the wine was a Malbec. Derived from the grape variety variable.
merlot_*	An indicator for whether the wine was a Merlot. Derived from the grape variety variable.
pinot_grigio_*	An indicator for whether the wine was a Pinot Grigio. Derived from the grape variety variable.
pinot_noir_*	An indicator for whether the wine was a Pinot Noir. Derived from the grape variety variable.

Variable name	Description
sauvignon_blanc_*	An indicator for whether the wine was a Sauvignon Blanc. Derived from the grape variety variable.
tempranillo_*	An indicator for whether the wine was a Tempranillo. Derived from the grape variety variable.
dry_*	An indicator for whether the wine was dry. Derived from the wine charactersitic variable. This variable is coded zero for all red wines.
semi_dry_*	An indicator for whether the wine was semi dry. Derived from the wine charactersitic variable. This variable is coded zero for all red wines.
medium_dry_*	An indicator for whether the wine was medium dry. Derived from the wine charactersitic variable. This variable is coded zero for all red wines.
semi_sweet_*	An indicator for whether the wine was semi sweet. Derived from the wine charactersitic variable. This variable is coded zero for all red wines.
sweet_*	An indicator for whether the wine was sweet. Derived from the wine charactersitic variable. This variable is coded zero for all red wines.
light_*	An indicator for whether the wine was light bodied. Derived from the wine charactersitic variable. This variable is coded 0 for all white and rosé wines.
light_medium_*	An indicator for whether the wine was light to medium bodied. Derived from the wine charactersitic variable. This variable is coded 0 for all white and rosé wines.
medium_*	An indicator for whether the wine was medium bodied. Derived from the wine charactersitic variable. This variable is coded 0 for all white and rosé wines.
medium_full_*	An indicator for whether the wine was medium to full bodied. Derived from the wine charactersitic variable. This variable is coded 0 for all white and rosé wines.
full_*	An indicator for whether the wine was full bodied. Derived from the wine charactersitic variable. This variable is coded 0 for all white and rosé wines.
organicno_*	An indicator for whether the wine was not organic. Perfectly correlated with organicyes
organicyes_*	An indicator for whether the wine was organic. Perfectly correlated with organicno
red_character_*	A five point likert scale from light- to full bodied giving the characteristic of red wines
white_character_*	A five point likert scale from dry to sweet giving the characteristic of white and rosé wines
time_delay_alt_*	The time delay before another alternative could be revealed in milliseconds. This is only available for treatments 8, 9 and 10.
consideration_set_+_alt_*	A boolean equal to true if a respondent indicated that alternative * was part of the consideration set in search period +

Important notes

1. Because of a bug when collecting the data, we do not have data on the length of search for the 10th choice task. Depending on the analysis, this task may have to be dropped from the data. However, because the design is random, it is unlikely to bias parameter estimates.

References