

Problem Set #7 Econ 572

Introduction

Provided by the University of California, Irvine this data represents the results of a direct marketing campaign for a Portuguese banking institution. The marketing campaigns relied on phone calls, often requiring multiple contacts with the same client to determine whether they would subscribe ('yes') or not ('no') to the product (bank term deposit). The data and its accompanying information is available [here](#). The dataset involves 17 inputs, including the target variable, across 45,211 instances. The list of inputs includes: age, job (occupation description), marital (status), education, default (binary), balance (yearly average), housing (loan), loan (personal), contact, day_of_week, month, duration, campaign, pdays (number of days passed since last contact), previous (number of previous contacts), poutcome (outcome of previous campaign), and y (whether the client subscribed to the term deposit). See the link attached above for more information.

Methodology

Using Stata, I created a *probit* model that included all available variables. Various variables were considered categorical and required encoding before being incorporated into the model (see syntax), including converting the y variable into a binary variable indicating whether the client subscribed (1) or not (0). The *probit* model estimates the marginal effects of selected variables on the probability that a client subscribes to a term deposit, holding all other factors constant. The results are available via the supplementary PDF submitted with this report (*PS7_ProbitResults.pdf*).

Using a significance level of 0.05, several variables in the model are found to be statistically meaningful. The estimated marginal effects provide valuable insight into which strategies have the greatest impact on a client's likelihood to subscribe to a term deposit. This analysis presents an opportunity for the bank to identify and invest in the most effective marketing actions.

Through the probit model, I aim to answer the following three questions:

- 1) What is the probability of a client subscribing based on the duration (seconds) of the phone call?

The marginal effect of the duration of the call is highly significant and carries a positive. According to the model, each additional minute on the call increases the probability of subscription by approximately 1.7%. This suggests longer calls are associated with higher probability of success, indicating that the bank could train their team to keep the calls longer and give the client time to understand the offer rather than keeping the calls short.

- 2) How does the month in which a client is contacted influence the likelihood of a successful subscription?

Several months have significant effects on the probability of a client subscribing to a term deposit. According to the model, clients contacted in March and October are 18% and 16% more likely to subscribe, respectively. In contrast, those contacted in January and November are 7% and 5% less likely to subscribe. This suggests there is seasonal behavior the bank can leverage to create successful campaigns.

- 3) What is the probability of a client subscribing based on the outcome of the previous campaign?

Based on the probit model, the only statistically significant category within the previous campaign outcome variable is "success." This indicates that the categories "other" and "unknown" do not have a meaningful impact on the probability of a client subscribing. Holding all other variables constant, the probability of a client subscribing increases by approximately 32% when the outcome of the previous campaign was a success. This highlights the value of targeting clients with a positive engagement history.

Summary

Initially, I expected that the number of days since the last contact (*pdays*) or the number of previous contacts would have a significant impact on the likelihood of a client subscribing. However, the results showed otherwise, and I was able to focus on the variables that truly influence successful campaign outcomes. I believe a few limitations in the data could be addressed to make the investigation more robust. For example, the time of the call and perhaps the day of the week (e.g., Monday vs. Friday) could provide a more detailed analysis. Additionally, information about the terms of the deposit - such as the length of the deposit or the interest rate - could offer further insight. Overall, this has been a very insightful exercise, and I found it highly engaging. The syntax used to conduct this exercise is available below:

Syntax:

```
encode job, gen(job_n)
encode marital, gen(marital_n)
encode education, gen(education_n)
encode _default, gen(default_n)
encode housing, gen(housing_n)
encode loan, gen(loan_n)
encode contact, gen(contact_n)
encode month, gen(month_n)
encode poutcome, gen(poutcome_n)
gen y_binary = (y == "yes")
probit y_binary age balance day duration campaign pdays previous i.job_n i.marital_n
i.education_n i.default_n i.housing_n i.loan_n i.contact_n i.month_n i.poutcome_n
margins, dydx(*)
```