

Predictive Analytics

Exercise 4: Specification and Tests

Univ.-Prof. Dr. Sven Müller

Business Administration, esp. Operations Managements
Otto-von-Guericke-University Magdeburg



Fashion Industry Consulting “Choice Lab”

- ▶ Choice Lab Inc.: collecting and processing data on companies operating in the fashion industry
- ▶ Business-to-business market covering producers and distributors (these are the customers of Choice Lab) of clothes, shoes and accessories
- ▶ Task: factors characterizing customer departure
- ▶ Data includes an unbalanced panel from 2000 until 2002
- ▶ Observations: firms (customers of Choice Lab)
- ▶ Choice: =1, if customer defects, 0 if customer stays with Choice Lab
- ▶ Attributes: no
- ▶ Characteristics: customer ID, year of observation, some financial and economic indicators and the type of products that the customer has purchased

“Choice Lab” Products

- 1 *Fashion Industry Analysis Report*: key figures in the past 5 years on the clothing, shoes and accessories sectors
- 2 *Fashion Credit Info*: contains an overview of the company's management, history, and accounting figures
- 3 *Individual Accounts Database*: copies of past 5 annual accounts on companies
- 4 *Customized Individual Business Monitoring*: product launches, financial information, ownership change
- 5 *Web Access Real-Time Fashion Data*: real-time fashion data
- 6 *CD-Fashion*: most complete database of companies
- 7 *CRM-F Integrated*: Internet based customer relations management system
- 8 *Internet-Credit*: access to credit information of companies
- 9 *Open Fashion Data Base*: real-time new product launches, mergers, acquisitions and bankruptcies
- 10 *Other Customized Solutions*

“Choice Lab” Data

Variable	Description
Choice	Equals 1 if customer drops next year; 0 otherwise
ID	Company ID
IndAnalysis	Equals 1 if product 1 has been purchased; 0 otherwise
CreditInfo	Equals 1 if product 2 has been purchased; 0 otherwise
Accounts	Equals 1 if product 3 has been purchased; 0 otherwise
Monitor	Equals 1 if product 4 has been purchased; 0 otherwise
Web	Equals 1 if product 5 has been purchased; 0 otherwise
CD	Equals 1 if product 6 has been purchased; 0 otherwise
CRM	Equals 1 if product 7 has been purchased; 0 otherwise
Internet	Equals 1 if product 8 has been purchased; 0 otherwise
OpenDB	Equals 1 if product 9 has been purchased; 0 otherwise
Other	Equals 1 if product 10 has been purchased; 0 otherwise
Age	Number of years the client has existed
Rating	Client credit rating: 100 represents the best and 0 the worst (this is a proxy for the current financial condition of the client)
Year	Year of observation
NegProfit	Equals 1 if profit < 0; 0 otherwise
NegEquity	Equals 1 if equity < 0; 0 otherwise
LRSC	Equals 1 if company is a limited responsibility stock owned company; 0 otherwise
LRC	Equals 1 if a company is a limited responsibility company; 0 otherwise
NbEmpl	Total number of employees
LnNbEmpl	Natural log of the number of employees
LnAge	Natural log of the age of the company

Get Your Data in Shape

```
1 # Data -----  
2 database      <- read.delim("marketing.dat")  
3 database      <- database[order(database$ID), ]
```

Challenge Questions

Analyze the Choice Lab data along the following questions

- Q1 Start with a simple model that includes only the *number of employees*, the *age of the customer company*, and whether the companies has a *negative profit*. Come up with a specification that has the opt out alternative (leave) as the reference alternative. Perform a likelihood ratio test with respect to the Null-Model! Compare this value to the one reported by R.
- Q2 Now consider a log-transformation of the *age* and *number of employees* variables. How do the coefficients change?
- Q3 Compare the log-transformation of age with a piecewise linearization of the age variable. Which specification would you prefer? Why?

Challenge Questions Continued

- Q4 Introduce the variables *rating* and $Year = 2002$ to your model. Consider an interaction of these two new variables. How should this interaction be interpreted in terms of utility? What does the result tell us?
- Q5 Now add the Choice Lab products *Internet*, *CRM*, and *Web* to you model. What do the coefficients tell us?
- Q6 Compare all your models in terms of model fit with an appropriate statistical test. Which model do you prefer and why?