CHAPTER 2
RESEARCH DESIGN AND METHODOLOGY

2.1. Sample

In this research, two types of research methods have been applied: (1) laboratory experiment, and (2) questionnaire survey. The essence of an experiment is control within the research project. That is, in studying the impact of the independent variable upon the dependent variable the experimenter attempts to hold all other variables constant (i.e. control them) while manipulating the independent variable (Ryan, Scapens and Thoebald, 1992). We conducted two laboratory experiments (discussed in chapters 3 and 5) to assume a real-world situation of a target cost-setting process where to determine the target cost in an environment of product design and development a department was created. All the subjects were undergraduate students in the College of Policy and Planning Sciences at the University of Tsukuba who were well taught about the target costing system beforehand.

The research questions dealt in the laboratory experiment were also investigated by data collected through questionnaire survey to see whether the impact of independent variable upon the dependent variable also exist in the uncontrolled real-world situation (discussed in chapters 4, 6 and 7). Four major industries in Japan satisfying two major characteristics are chosen for the study: first, manufacturing industries having the features of discrete manufacturing processes, regular model changes etc.; second, the companies that have listed their stocks at Tokyo stock exchange, Part I. These are machinery, electrical and electronics, transportation equipment and precision machinery. To seek the reason of practicality and to provide comparisons to prior research, the questionnaire has been validated through the questionnaire of the previous research conducted by Kobe University Management Accounting Research Group. Questionnaires were mailed to 518 companies on October 10, 1996. One hundred and forty six companies were responded by the deadline of November 10, 1996, among which three responses could not be used due to their incomplete answers. The response rate is 28.19% and the effective response rate is 27.61%. Industry-wise effective response rate is presented in Table 2.1. Invited companies were guaranteed confidentiality and, in return for their participation, were promised a feedback as a paper in some journal.

Table 2-1. Industry classification for 146 organization providing survey data in November, 1996

Questionnaires	Machinery	Electrical & Electronics	Transportation Equipment	Precision Machinery	Total
Sent	194	198	90	36	518
Returned	49	49	38	10	146
Return %	25.26	24.74	42.22	27.78	28.19
Effective return %	24.74	24.24	41.11	27.78	27.61

We can present the total structure of the target costing system as covered by the questionnaire survey in the following way (See next page):

*Complexities and uncertainties in the external environment

Ques. 42. Degrees of complexities and uncertainties in the business environment

Complexities' indicator: 1) Product market diversification
2) Production technology commonization
3) Sales promotion devices

Uncertainties' indicator: 4) Competitiveness of representative product market
5) Frequency of new product and new technology developments
6) Correctness of market demand prediction

*Business strategies

Ques. 43. Factors emphasized by the top management corresponding to environmental changes
1) Market strategies
2) Technology strategies

*How the company is coping with target costing

Ques. 1. Form of the company organization
Ques. 2. Products' model life
Ques. 3. Level of target costing implementation
Ques. 4. Starting time of target costing
Ques. 5. Objective of target costing
Ques. 6. Starting stage of target costing

*Target costing organization

Ques. 7. Location of target costing office
Ques. 8. Roles of target costing office
Ques. 9. Structure of the target costing project team
Ques. 10. Relationship between product development and design department
Ques. 11. Power or authority in deciding target cost allocation rate
Ques. 12. Performance evaluations method of product designers
Ques. 13. Degree of participation of each department in each development step of target costing

*Suppliers' relationship

Ques. 36. Degree of information possessions concerning borrowed blueprint manufacturers
Ques. 37. Frequency of visiting suppliers
Ques. 38. Sending design proposals by the suppliers
Ques. 39. Request to the approved blueprint manufacturers for VE proposals
Ques. 40. Degrees of sharing profit and risk relating to the design and production special parts
Ques. 41. Degrees of sharing profit and risk relating to the production of other parts

*Target sales price setting

Ques. 14. Methods of setting target sales price

*Target profit setting

Ques. 15. Extent/range of products for target profit setting
Ques. 16. Profit bases for target profit
Ques. 17. Methods of target profit determination

*Target cost setting methods

Ques. 18. Methods of determining target cost

*Distribution, attainability and follow-up of target cost

Ques. 19. Way of target cost decomposition
Ques. 20. Attainability of target cost
Ques. 21. Follow-up of target cost achievement levels
Ques. 22. Use of information on non-achievement of target cost

*VE activities in the detailed design stage

Ques. 23. Level of target costing conduction
Ques. 24. Extent of the duties of the VE team leaders
Ques. 25. Departments getting VE training
Ques. 26. Departments participating in VE activities
Ques. 27. Degrees of information used for conducting VE

*Use of cost tables

Ques. 28. Cost tables preparations
Ques. 29. Contributions of each department in cost table preparation
Ques. 30. Usefulness of cost tables
Ques. 31. Frequency of cost tables updating
Ques. 32. Introduction of computer system in cost table
Ques. 33. Combining CAD system to cost tables
Ques. 34. Information contained in the cost tables

*Production preparation stage

Ques. 35. Controllability in production preparation stage

Based on above model structure, the process of target costing system can be diagrammed as in Figure 2-1.

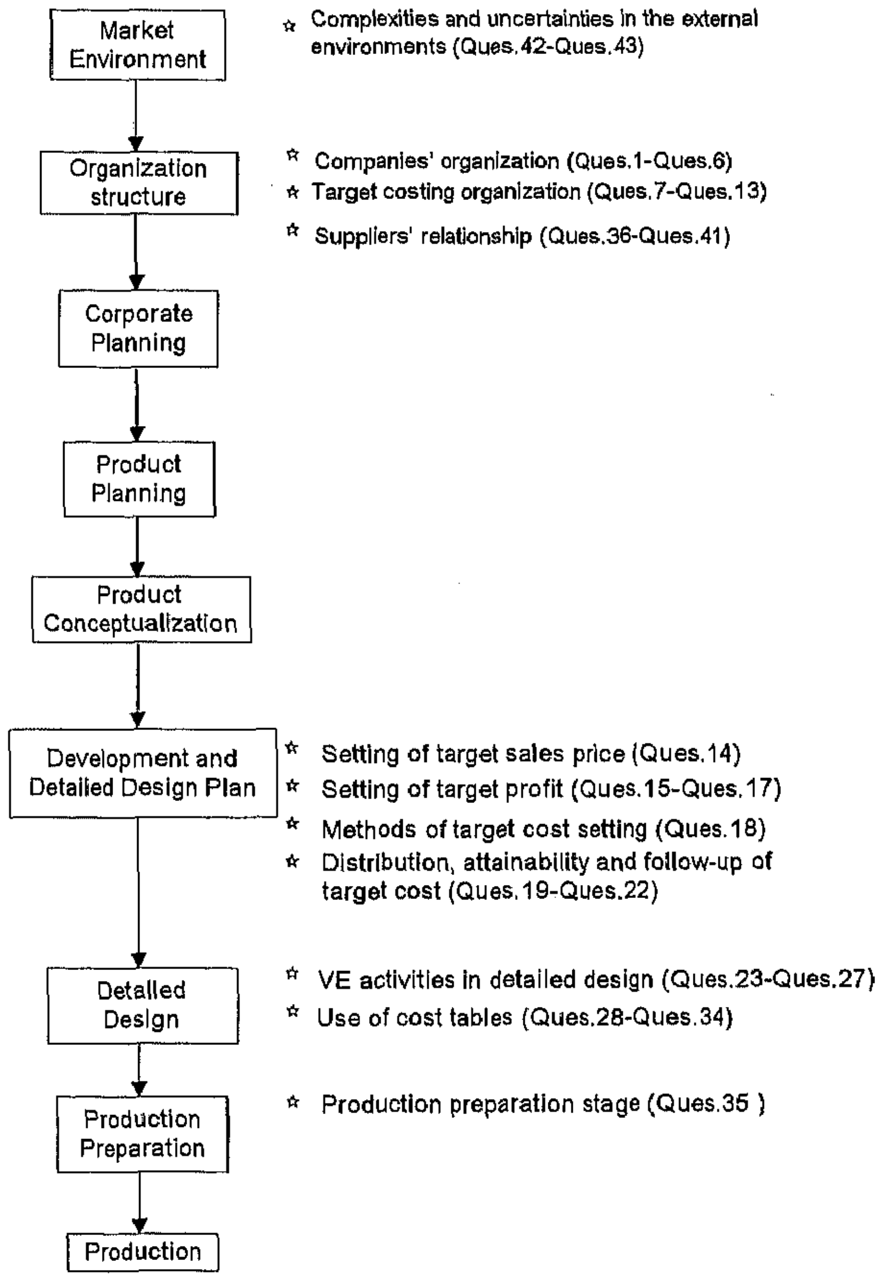

Figure 2-1. Process of target costing system

2.2. Variables and their Categories used in Survey research

Question	Name of the variables	Categories
Ques. 17-1	Target profit determined in the middle-range profit plan ⇒ *TP1*	• Lower utilization of *TP1* ⇒ $TP1_L$ • Medium utilization of *TP1* ⇒ $TP1_M$ • Higher utilization of *TP1* ⇒ $TP1_H$
Ques. 17-3	Target profit by cost reduction rate of the existing or similar products ⇒ *TP2*	• Lower utilization of *TP2* ⇒ $TP2_L$ • Medium utilization of *TP2* ⇒ $TP2_M$ • Higher utilization of *TP2* ⇒ $TP2_H$
Ques. 17-2	Target profit based on past actual performance of the concerned product ⇒ *TP3*	• Lower utilization of *TP3* ⇒ $TP3_L$ • Medium utilization of *TP3* ⇒ $TP3_M$ • Higher utilization of *TP3* ⇒ $TP3_H$
Ques. 18	Target cost methods ⇒ *TC*	• Subtractive ⇒ *SUB* • Combination ⇒ *COM* • Adding-up ⇒ *ADD*
Ques. 11	Participation ⇒ *PAR*	• Non-participation ⇒ *NP* • Joint participation ⇒ *JP* • Participation ⇒ *P*
Ques. 12	Performance evaluation ⇒ *PE*	• Uncontrollable ⇒ *UC* • Controllable ⇒ *C*
Ques. 20	Target cost achievement level ⇒ *TCAL*	• *TCAL* 70% ⇒ *TCAL-1* • *TCAL* 80% ⇒ *TCAL-2* • *TCAL* 90% ⇒ *TCAL-3* • *TCAL* 100% ⇒ *TCAL-4*

2.3. Measurement Framework for Survey Research

Chapter 4 covers the effects of tightness of target cost and profit methods on target cost achievement level while chapter six deals with the effects of participation and performance evaluation factors. In chapter 7, we verified the combination effects of target cost determination methods, participation and performance-evaluation factors on target cost achievement. The effects of the variables *TC*, *TP1*, *TP2* and *TP3* on *TCAL* are tested in analyzing the effects of tightness of target cost and profit methods on target cost achievement level, while the effects of the variables *PAR* and *PE* on *TCAL* are examined in analyzing the effects of behavioral factors in target cost allocation on target cost achievement level. For the purpose of chapter 7, we examined the effects of *TC*, *PAR* and *PE* on *TCAL*.

TCAL, that is, target cost achievement level has been considered as the response variable of the model to see how the various behavioral factors are influencing the cost-reduction performance. The justification for considering *TCAL* as the response variable is as follows: in a target costing environment, all the members (they may be different departments of a company or various companies in an industry) are striving to reduce cost considering the quality and functionality of the products. Since their sole objective is to reduce cost, all of them will try to increase their target cost achievement levels and there will be no difference among the companies in this effort. Among other performance criteria, ratio of cost of goods sold, development lead-time, or ROI may be good measures, however, these may vary from industry to industry. So to study the cost-reduction performance of all industries, the variable of target cost achievement level may be a better choice as the response variable in comparison to other performance measurement criteria. A short description of the all the variables used in survey research are given below:

Target profit (Ques. 17). Target profit is measured in three ways, each stating a particular method of determining target profit: (a) by target return on sales ratio determined in middle-range profit plan, (b) by target cost reduction rate of existing or similar products, (c) by target return on sales ratio based on past actual performance of the related products. It is found that most of the companies are using three types of profit setting methods simultaneously and at the same degree. It may be due to the use of different types of profit setting methods for different products. Therefore, the researcher decided to consider three profit-setting methods independently.

Target cost (Ques. 18). Target cost is coded into three categories ranging from subtractive method, combination method and adding-up method.

Participation (Ques. 11). Participation is measured mainly by five items asking the respondents to indicate to what degree the designers can participate in the allocation process of the target cost of product into its parts elements. The responses range from only product manager's participation (scored as 1) to participation of product designers only (scored as 5).

Performance evaluation (Ques. 12). Performance evaluation is coded into three categories, evaluation by group performance, evaluation by individual performance and others. It is found that 31 companies chose the third option and some of them mentioned that the company considers product's quality as the performance evaluation criterion. Due to the irrelevance of this type of response to the research objective, this category has been excluded from the analysis.

Target cost achievement (Ques. 20). Target cost achievement is measured by using five items asking the respondents to indicate the level to which they are achieving target cost. Since only five companies achieve target cost up to 60%, it seems better to merge this with 70%, thus converting the five levels into four.

2.4. Research Design of the Experiment relating to the Effect of Tightness and Target Information Types on Cost-reduction Performance

There were two independent variables:

(1) factors relating to types of target information, and

(2) factors relating to tightness of target.

With three levels of target information and two levels of tightness a 3 × 2 matrix will be formed. Table 2-2 shows this.

Table 2-2. Cost-reduction process in setting product's target cost

Tightness / Target Information Types	Tight	Loose
Total ideal target (*IT*) Previous actual (*PA*) Total attainable target (*TT*)	**Group A** Total Subjects =16 $IT=(PA)\times(1+.15)$ $TT=(PA)\times(1+.10)$	**Group D** Total Subjects =16 $IT=(PA)\times(1+.15)$ $TT=(PA)\times(1+.05)$
Total attainable target (*TT*)	**Group B** Total Subjects =16 $TT=(PA)\times(1+.10)$	**Group E** Total Subjects =16 $TT=(PA)\times(1+.05)$
Previous actual (*PA*) Incremental Target (*INT*)	**Group C** Total Subjects =16 $INT=(PA)\times(.10)$	**Group F** Total Subjects =16 $INT=(PA)\times(.05)$

2.5. Research Design of the Experiment relating to the Effect of Participation and Performance Evaluation Factors on Target Costing Performance

There were two independent variables:
(1) factors relating to participation, and
(2) factors relating performance-evaluation information.

With two levels of participation and two levels of performance-evaluation information, a 2 × 2 matrix will be formed. Table 2-3 presents this.

Table 2-3. Decision-making process in setting target cost for parts

	Groups in determining the allocation rate of parts target cost	
	Nonparticipative (Top-down)	Participative (Bottom-up)
Expost performance evaluation of groups engaged in setting target cost for products (evaluation by group performance)	Group1 Team A Team B (15 subjects in each team = 15 × 2 = 30)	Group2 Team A Team B (15 subjects in each team = 15 × 2 = 30)
Expost performance evaluation of groups engaged in setting target cost for parts (evaluation by individual performance)	Group3 Team A Team B (15 subjects in each team = 15 × 2 = 30)	Group4 Team A Team B (15 subjects in each team = 15 × 2 = 30)

2.6. Statistical Methods Applied

2.6.1. In case of Laboratory Experiments

For the statistical analyses, we used Analysis of Variance (ANOVA). Our experimental data fulfill all the required conditions for the use of ANOVA:

(1) each of the samples is drawn from a normal population,

(2) all of the population has the same variance, and

(3) the absence of many factors that might affect our conclusions concerning the factor(s) to be studied.

2.6.2. In case of Questionnaire Survey

Statistical methodology for categorical data has only recently reached the level of sophistication achieved early in this century by methodology for continuous data. The recent development of methods for categorical data was stimulated by the increasing methodological sophistication of the social and biomedical sciences. Statisticians developed regression-type models for categorical responses to meet the need for analyses of multivariate discrete data sets.

A categorical variable is one for which the measurement scale consists of a set of categories. For instance, smoking status might be measured as "never smoked," "former smoker," and "current smoker." There are many types of categorical variables. Many categorical scales have a natural ordering, which are called *ordinal* variables. Examples are the appraisal of the tightness of a target cost or profit method (loose, medium-tight, tight). Categorical variables for which levels do not have a natural ordering, that is, unordered scales are called *nominal* variables. One example of this variable is methods of target cost determination (subtractive, combination, adding-up). An *interval* variable is one that does have numerical distances between any two levels of the scale. For example, length of time of using target costing system is an interval variable. In the measurement hierarchy, interval variables are highest, ordinal variables are next, and nominal variables are lowest. Statistical methods designed for variables of one type can also be used with variables at higher levels, but not at lower levels. For instance, statistical methods for ordinal variables can also be used with interval variables (by using only the ordering of levels and not their distances); they cannot be used with nominal variables, since categories of such variables have no meaningful ordering. Normally, it is best to apply methods appropriate for the actual scale (Agresti, 1990, 1996).

Methods for analyzing associations in two-way and three-way contingency tables using Chi-square test help us investigate effects of explanatory variables on categorical response variables. However, modeling the effects helps us do this more efficiently than Chi-square test, as it can handle more complicated situations, such as analyzing simultaneously the effects of several explanatory variables. In addition, the model-building paradigm focuses on estimating parameters that describe the effects, which is more informative than mere significance testing (Agresti, 1996, 71). A good-fitting model evaluates effects of explanatory variables, describes association and interaction linkages, and produces improved estimates of response probabilities (Agresti, 1990, 79).

In Generalized Linear Model (GLM), mainly two types of models exist for categorical data: (1) Loglinear model, and (2) Logistic regression model. For the analysis of a set of nominal and ordinal variables, researchers may select either logistic regression or loglinear modeling (Agresti, 1990; Hosmer and Lemeshow, 1989). There is hardly any clear guidelines exist for helping researchers select one statistical tool over another as their primary data analysis method. This dilemma is further exacerbated by the tendency among advocates of various statistical techniques to overweigh the benefits of their favorite technique while minimizing its shortcomings. Researchers are confronted with both of these confusing conditions when making the decision to use loglinear modeling or logistic regression. In the absence of uniform rules for selecting one technique over the other, this choice will ultimately rest on the researcher's goals, statistical criteria, and the researcher's level of statistical training and willingness to learn new statistical techniques (Tansey, White, Long and Smith, 1996).

Given the categorical nature of much of the data in management research it is unfortunate that the multivariate techniques, such as loglinear modeling and logistic regression, are underutilized. Unfortunately, we are unaware of any presentation in the statistical theory literature that explicitly compares and contrasts loglinear modeling and logistic regression as tools in applied research (Tansey, White, Long and Smith, 1996).

For our research purpose, we used proportional-odds model for cumulative probabilities under the family of logistic regression. For clear understanding, we will discuss about logistic regression, its benefits, and the reasons to use it, multicategory logit model, and cumulative logit models for ordinal responses and finally, proportional-odds model.

2.6.2.1. Logistic regression model

Logistic regression models are special cases of GLMs. It is particularly appropriate in estimating a binary dependent variable using a maximum likelihood estimation procedure. There are two broad applications of logistic regression for applied research:

1. predicting group membership for the dependent variable;
2. measuring the "instantaneous rate of change in the probability of occurrence

of an event with change in a given predictor" (Demaris, 1990, 273).

Logistic regression gives the probability of accurately classifying the presence of an event, the absence of an event, and the overall pooled rate of all the sample cases across both those representing the presence or absence of a certain type of event. For a binary response, say, Y, there are two possible outcomes of 1 and 0, which are in a generic terminology "success" and "failure". A binary response is sometimes called a *Bernoulli variable*. Its distribution is specified by probabilities $P(Y = 1) = \pi$ of success and $P(Y = 0) = (1 - \pi)$ of failure. The value of π changes as the value x of X changes, and $\pi(x)$ reflects its dependence on x value.

The relationships between $\pi(x)$ and x are usually nonlinear rather than linear. A fixed change in X may have less impact when π is near 0 or 1 than when π is near the middle of its range. In practice, nonlinear relationships between $\pi(x)$ and x are often monotonic, with $\pi(x)$ increasing continuously as x increases, or $\pi(x)$ decreasing continuously as x increases. For a binary response Y and a quantitative explanatory variable X, let $\pi(x)$ denote the "success" probability when X takes value x. The logistic regression model has linear form for the logit of this probability,

$$\log it[\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x. \tag{2.1}$$

The formula implies that $\pi(x)$ increases or decreases as an S-shaped function of x. The link function is the *logit* transformation $\log[\pi/(1-\pi)]$ of π, symbolized by logit(π). Logistic regression models are often called *logit models*. The logit is the natural parameter of the binomial distribution, so the logit link is its canonical link.

An alternative formula for logistic regression refers directly to the success probability. This formula uses the exponential function $\exp(x) = e^x$, in the form

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}. \tag{2.2}$$

Linear Approximation Interpretations

The parameter β determines the rate of increase or decrease of the S-shaped curve. The sign of β indicates whether the curve ascends or descends, and the rate of change increases as $|\beta|$ increases. When the model holds with $\beta = 0$, the right-hand side of (2.2) simplifies to a constant. Then $\pi(x)$ is identical at all x, so the curve becomes a horizontal straight line. The binary response Y is then independent of X. Unlike the linear probability model, the logistic regression model permits the rate of change to vary as x varies. The term $\beta\pi(1-\pi)$ approximates the change in the probability per unit change in x.

Odds Ratio Interpretations

Another interpretation of the logistic regression model uses the odds and the odds ratio. For model (2.1), the odds of response 1 (i.e. the odds of a "success") are

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}(e^{\beta})^{x}. \tag{2.3}$$

This exponential relationship provides an interpretation for β: The odds increase multiplicatively by e^{β} for every one-unit increase in x. That is, the odds at level $x + 1$ equal the odds at x multiplied by e^{β}. When $\beta = 0$, $e^{\beta} = 1$, and the odds do not change as x changes. The logarithm of the odds, which is the logit transform of $\pi(x)$, has the linear relationship.

2.6.2.2. Comparative advantage of logistic regression

Methodological Accessibility: Logistic Regression's Similarity to OLS Regression

1. Since logistic regression designates a left-hand response variable and a set of right-hand predictors in the same manner as OLS regression, new users of this technique will intuitively understand the overall goal of this technique (i.e. for each predictor pattern how much variation in the response variable can be attributed to each predictor controlling for the remaining predictors). Logistic regression provides Wald Z and likelihood ratio test statistics to estimate the magnitude, standard error and sign for the impact of each predictor on the outcome variable.
2. Logistic regression provides researchers a superior opportunity to estimate each predictor's quantitative impact on the response variable.

Estimating the Conditional Odds for the Dependent Variable

Logistic regression can provide point estimates and confidence bound intervals for estimating the natural log of the ratio that an event occurs compared to its non-occurrence, given a certain set of conditions measured by the predictors. This information can only be extracted form loglinear modeling by experienced users capable of transforming selected loglinear model into an equivalent logit model (Aldrich and Nelson, 1984; Cramer, 1991; Demaris, 1993).

Logistic Regression's Superior Flexibility for Analyzing Mixed Data Sets

Logistic regression is a more flexible instrument than loglinear modeling for analyzing a mixed set of nominal/ ordinal and interval variables (Hosmer and Lemeshow, 1989). In contrast most uses of loglinear modeling are currently restricted to categorical data relying on nominal and ordinal variables (Agresti, 1990).

2.6.2.3. When to choose logistic regression

1. Loglinear models are most natural when at least two variables are response variables. When only one variable is a response, it is more sensible to use logit models directly (Agresti, 1996).
2. When the variables are any of the nominal, ordinal and interval.
3. The research goal is not to uncover all types of variable associations but to estimate the size, significance and sign of a predictors on binary response.
4. The research goal does not focus on discovering all significant higher order interactions but emphasizes on discovering main effects and 1st order (two-variable) interactions
5. There is no significant sampling or structural zero.

Ultimately, however, a researcher's substantive goals will have a major influence in determining which statistical tool is chosen. For instance, the use of loglinear modeling is more appropriate in situations where applied researchers are interested in the various pairwise and higher order associations among a set of independent variables. In contrast, logistic regression is a more powerful tool for applied research aimed at producing precise predictions regarding the log odds ratio of a binary dependent variable (Tansey, White, Long and Smith, 1996).

2.6.2.4. Multicategory logit models

This section presents generalizations of logistic regression models that handle multicategory responses. At each combination of levels of the explanatory variables, the models assume that the response counts for the categories of *Y* have a multinomial distribution. Like logistic regression models and unlike loglinear models, multicategory logit models treat one classification as a response and the other variables as explanatory.

Logit Models for Nominal Responses

Suppose Y is a nominal variable with J categories. The order of listing the categories is irrelevant. Let $\{\pi_1,\ldots,\pi_J\}$ denote the response probabilities, satisfying $\sum_j \pi_j = 1$. When one takes independent observations based on these probabilities, the probability distributions for the number of outcomes that occur for each of the J types is the multinomial. It specifies the probability for each possible way of allocating the n observations to the J categories. Multicategory logit models simultaneously refer to all pairs of categories, and describe the odds of response in one category instead of another. Once the model specifies logits for a certain $(J - 1)$ pairs of categories, the rest are redundant.

Baseline Category Logits

Logit models for nominal responses pair each response category with a baseline category, the choice of which is arbitrary. When the last category (J) is the baseline, the *baseline-category* logits are

$$\log\left(\frac{\pi_j}{\pi_J}\right), \qquad j = 1,\ldots,J-1$$

For $J = 3$, for instance, the logit model uses $\log(\pi_1/\pi_3)$ and $\log(\pi_2/\pi_3)$. The logit model using the baseline-category logits with a predictor x has form

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad j = 1,\ldots,J-1 \qquad (2.4)$$

The model consists of $J - 1$ logit equations, with separate parameters for each. That is, the effects vary according to the response category paired with the baseline. For an arbitrary pair of categories a and b,

$$\log\left(\frac{\pi_a}{\pi_b}\right) = \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right)$$

$$= (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)$$

$$= (\alpha_a - \alpha_b) + (\beta_a - \beta_b)x. \qquad (2.5)$$

Thus, the logit equation for categories a and b has intercept parameter $(\alpha_a - \alpha_b)$ and slope parameter $(\beta_a - \beta_b)$.

Estimating Response Probabilities

One can alternatively express the multicategory logit model directly in terms of the response probabilities, as

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_h \exp(\alpha_h + \beta_h x)}, \qquad j = 1, \ldots, J-1 \qquad (2.6)$$

2.6.2.5. Cumulative logit models for ordinal responses

When response categories are ordered, logits can directly incorporate the ordering. This results in models having simpler interpretations and potentially greater power than ordinary multicategory logit models.

The cumulative probabilities are the probabilities that the response Y falls in category j or below, for each possible j. The jth cumulative probability is

$$P(Y \le j) = \pi_1 + \ldots + \pi_j, \qquad j = 1, \ldots, J.$$

The cumulative probabilities reflect the ordering, with $P(Y \le 1) \le P(Y \le 2) \le \ldots \le P(Y \le J) = 1$. Models for cumulative probabilities do not use the final one, $P(Y \le J)$, since its necessarily equals 1. The logits of the first $J - 1$ cumulative probabilities are

$$\operatorname{logit}[P(Y \le j)] = \log\left(\frac{P(Y \le j)}{1 - P(Y \le j)}\right)$$

$$= \log\left(\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots \pi_J}\right), \qquad j = 1, \ldots, J-1.$$

These are called *cumulative logits.*

Each cumulative logit uses all J response categories. A model for the jth cumulative logit looks like an ordinary logit model for a binary response in which categories 1 to j combine to form a single category, and categories j + 1 to J form a second category. In other words, the response collapses into two categories. Ordinal models simultaneously provide a structure for all J − 1 cumulative logits. For J = 3, for instance, models refer both to $\log[\pi_1/(\pi_2+\pi_3)]$ and $\log[(\pi_2+\pi_3)/\pi_3]$.

2.6.2.6. Proportional-odds model

For a predictor X, the model

$$\text{logit}[P(Y \le j)] = \alpha_j + \beta x, \qquad j = 1, \dots, J-1, \qquad (2.7)$$

has parameter β describing the effect of X on the log odds of response in category j or below. In this formula β does not have a j subscript, so the model assumes an identical effect of X for all J − 1 collapsings of the response into binary outcomes. Interpretations for this model refer to odds ratios for the collapsed response scale, for any fixed j. For two values x_1 and x_2 of X, the odds ratio utilizes cumulative probabilities and their complements,

$$\frac{P(Y \le j \mid X = x_2)/P(Y > j \mid X = x_2)}{P(Y \le j \mid X = x_1)/P(Y > j \mid X = x_1)}$$

The log of this odds ratio is the difference between the cumulative logits at those two values of x. This equals $\beta(x_2 - x_1)$, proportional to the distance between the x values. The same proportionality constant (β) applies for each possible point j for the collapsing. Because of this property, model (2.7) is called a *proportional odds model.* Its interpretation is that the odds of making response $\le j$ are $\exp[\beta(x_2 - x_1)]$ times higher at $x = x_2$ than at $x = x_1$. In particular, for $x_2 - x_1 = 1$, the odds of response below any given category multiply by e^{β} for each unit increase in X. When the model holds with $\beta = 0$, X and Y are statistically independent.

When $\beta_i > 0$ in model (2.7), each cumulative logit increases as x_i increases, so each cumulative probability increases. This means that relatively more probability mass falls at the low end of the Y scale; that is, Y tends to be smaller at higher values of x_i. To make $\beta_i > 0$ have the more usual meaning of Y tending to be larger at higher values of x_i, we replace β in (2.7) by $-\beta$.

Explanatory variables in a cumulative logit model can be continuous or categorical. The ML fitting process uses an iterative algorithm simultaneously for all j. When the categories are reversed in order, one gets the same fit, but the sign of $\hat{\beta}$ reverses. Unlike other logit models, cumulative logit models making the proportional odds assumption are not equivalent to loglinear models.