DOCUMENT RESUME

ED 382 680 TM 023 221

AUTHOR Hummel, Thomas J.
TITLE Qualitative Dependent Variables in Counseling Research.
PUB DATE Apr 95
NOTE 43p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classification; *Counseling; *Educational Research; Estimation (Mathematics); *Qualitative Research; *School Choice; Simulation; *Statistical Significance
IDENTIFIERS Choice Behavior; *Dependent Variables; Order Analysis

ABSTRACT

Two types of qualitative dependent variables are presented for use in counseling research: choices from an unordered set of categorical alternatives and ordered, categorical counseling outcomes. To investigate school choice behavior, the conditional logit model and analysis are introduced. The conditional logit model can include the attributes of the people who make particular choices and also the attributes of the choices themselves that make them more or less attractive. To investigate categorical counseling outcomes, a model and analytical procedure for studying ordered categories is introduced. For both choice behavior and outcomes, parameter estimates and significance tests are presented with examples based on simulated data. These parameter estimates and significance tests are viewed as preliminary to presenting the probabilities of category membership. These probabilities are presented graphically using a spreadsheet model. It is argued that these probabilities should be the primary focus of investigation, for they are the results that can most directly affect actions to be taken by counselors and clients. Six graphs are included. (Contains 10 references.) (Author)

# Qualitative Dependent Variables in Counseling Research

By

Thomas J. Hummel
University of Minnesota

A Paper Presented at the Annual Meeting of the
American Educational Research Association
San Francisco, 1995

## Abstract

Two types of qualitative dependent variables are presented for use in counseling research: choices from an unordered set of categorical alternatives and ordered, categorical counseling outcomes. To investigate choice behavior, the conditional logit model and analysis are introduced. The conditional logit model can include the attributes of the people who make particular choices and also the attributes of the choices themselves that make them more or less attractive. To investigate categorical counseling outcomes, a model and analytical procedure for studying ordered categories is introduced. For both choice behavior and outcomes, parameter estimates and significance tests are presented with examples based on simulated data. These parameter estimates and significance tests are viewed as preliminary to presenting the probabilities of category membership. These probabilities are presented graphically using a spreadsheet model. It is argued that these probabilities should be the primary focus of investigation, for they are the results that can most directly affect actions to be taken by counselors and clients.

## Acknowledgments

A number of sources were consulted, integrated, and eventually became foundational to this paper. To ensure that they receive adequate credit, they are listed here in alphabetical order: Bock (1975, Chapter 8), Greene(1993, Chapter 21), Greene (1992, Part VIII), Johnson and Kotz (1970a, Chapter 21), Johnson and Kotz (1970b, Chapter 22), Johnson and Kotz (1972, Chapter 42), Luce (1959), Maddala (1983, Chapters 2 & 3), and McFadden (1973).

# Qualitative Dependent Variables in Counseling Research

Traditionally, counseling research has focused on providing quantitative data that can sometimes be difficult to translate into actionable information. Counselors and their clients are often concerned with questions that are immediate and personal. What is the probability that a particular counseling intervention will result in a desirable outcome? How likely is it that a client will fall into the category of those who will benefit from a particular treatment? What is the probability that an individual will choose a certain action or option under certain conditions? Typically, counselors who turn to research for guidance must attempt to interpret regression weights and effect sizes associated with various independent variables and try to extrapolate information that will help them and their clients in making important individual decisions.

It is the thesis of this paper that there could be great value in a type of research that focuses on how independent variables interact to affect the probability that a particular client would have a particular outcome or be in a particular outcome category, or would choose a certain action or option. Such an approach could yield research results that might be more relevant and useful for clients and their counselors who must make important decisions about treatment interventions.

In the first section of the paper, we will be focusing on unordered categories, giving attention to the study of choice behavior. This is of interest because it is a key human behavior and because it allows us to introduce a comprehensive statistical model, the conditional logit model, and an analytical technique, conditional logit analysis, that looks at independent variables related to the choices themselves as well as independent variables that describe choosers. These techniques, not often used in counseling research, offer a new way to add a qualitative element to the traditional quantitative tools.

In the second section, we will deal with ordered categories, those that can be rank ordered with respect to some underlying continuum, such as the commonly used five point Likert scale. The ordered categories model can be applied to treatment outcome. This is an important area, for it holds the promise of letting us give clients the probabilities of various outcomes that they need to manage their own care effectively. Despite its importance, treatment outcome will receive somewhat less attention in what follows. This is primarily due to the lack of widely agreed upon, operationally defined outcome continua. Therefore outcome categories are an area more for future development than for current research.

For both choice and outcome categories we will present methods for non-linear estimation that can provide the independent variable weights required to predict the probability of a client being in one choice or outcome category or another, or that a client will choose a certain alternative. We will also provide simulation models to show how independent variables and weights combine to demonstrate dynamically changes in probability as a result of changes in the values of the independent variables.

## The Choice Model

Why do people make the choices they make? Different models have been put forward, but the one described here has been called the *utility-maximizing* model (McFadden, 1973). According to this model, an individual is presented with set of alternatives from which one will be chosen, and each alternative has a certain value, or utility, or that individual. The individual selects the alternative with the highest associated utility. For example, a person in psychological need is presented with four alternative sources of help, say, "Friends/Family," "Social Worker," "Psychiatrist," and "Psychologist." The four corresponding utilities for this person might be [1.645, -2.564, 1.769, 0.457]. In this case, the person would chose the third alternative, "Psychiatrist," because it has the highest utility, namely, 1.769. If another person were to choose, the vector of utilities might be [2.307, 0.571, 1.654, .003], in which case the first alternative, "Friends/Family," would be chosen.

Considering a population of individuals, the possible utilities associated with each of these four alternatives can be symbolized as $u_1$, $u_2$, $u_3$, and $u_4$, or in vector notation as $\mathbf{u} = [u_1, u_2, u_3, u_4]$. The highest value in $\mathbf{u}$ determines the choice, and in this model, there are never any ties.

To code the actual choice that has been made we can use another vector. For the above examples, choosing alternative three would be coded as $[0, 0, 1, 0]$ and choosing alternative one as $[1, 0, 0, 0]$. To represent a choice, we can define the array $\mathbf{y} = [y_1, y_2, y_3, y_4]$. For the first of the preceding two examples, $y_3 = 1$ and rest of the values in $\mathbf{y}$ would be zero.

With this setup we can begin to consider the probability of making a particular choice. Let us suppose that the individuals in a certain population are presented with J alternatives. (In the psychotherapy example above, J = 4.) If the various alternatives are indexed by j, with j = 1 to J, then, for a randomly selected individual, the probability that a particular choice will be made is $\Pr[y_j = 1] = \Pr[u_j > u_{j'}]$ for all $j \neq j'$. Because we are now considering more than one person's choice, we could include a subscript for each person, e.g., $u_{ij}$, but for simplicity, this second subscript is omitted and we will assume that $\mathbf{y}$ and $\mathbf{u}$ vary across individuals.

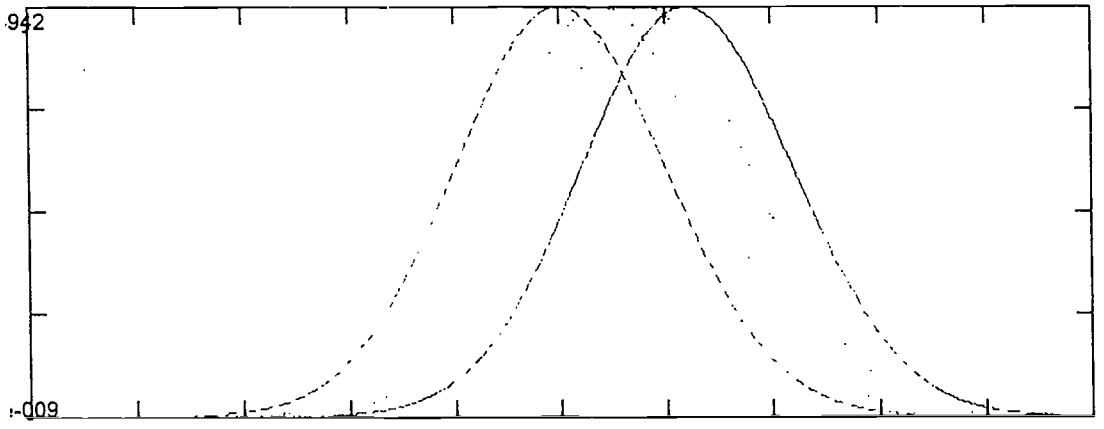
To make this last equation more concrete, we can imagine a simple experiment in which we repeatedly draw individuals from a population, present them with four alternatives and record their choices. In the long run, the proportion of people selecting alternative two, i.e., $y_2 = 1$, will converge to $\Pr[y_2 = 1]$. This proportion will be determined by the

number of times the utility for alternative two exceeds the utilities of the other alternatives, i.e., the frequency of $u_2 > u_1$, $u_2 > u_3$ and $u_2 > u_4$.

The probability $\Pr[u_j > u_{j'}]$, for all $j \neq j'$, can be determined if one knows the distribution of **u**. Perhaps the easiest way to introduce the determination of probabilities is to assume that each $u_j$ has a normal distribution. Further, we will assume that the $u_j$ are independently distributed of one another, each with a variance equal to one. The means of the utilities, $\mu_j$, may or may not be equal.

If the $\mu_j$ were all equal, then the choice probabilities for each alternative would be equal, and if they were different, the choice probabilities would be unequal. For example, if $\mu_1 = 0.0$ and $\mu_4 = 1.2$, in sampling individual choices we would expect $u_4 > u_1$ more often than $u_1 > u_4$, and therefore, we would expect alternative four to be chosen more frequently than one.

If we focus for the moment on the distributions of the individual utilities, we can represent the situation using the four overlapping distributions depicted below.




The distribution to the far right, with the solid line, represents the distribution of $u_4$, while the leftmost distribution, the one with a dashed line, represents the distribution of $u_1$. A choice is modeled by randomly drawing a vector, **u**, containing one value from each distribution, and taking the chosen alternative as the one whose utility is highest. Focusing on just alternatives one and four, it is clear that the utility for distribution four is more likely to be higher than that for distribution one.

Continuing with the above example, if we knew the values of all the $\mu_j$, we could determine the probability that alternative four would be chosen in the following manner:

$$\Pr[y_4 = 1] = \Pr[u_4 > u_1 \;\&\; u_4 > u_2 \;\&\; u_4 > u_3]$$

$$= \Pr([(u_4 - u_1 > 0) \;\&\; (u_4 - u_2 > 0) \;\&\; (u_4 - u_3 > 0)].$$

To actually compute the probability, we must consider the joint distribution of the three random variables, $u_4 - u_1$, $u_4 - u_2$, and $u_4 - u_3$, or $\mathbf{d}_4 = \begin{bmatrix} d_{41} \\ d_{42} \\ d_{43} \end{bmatrix} = \begin{bmatrix} u_4 - u_1 \\ u_4 - u_2 \\ u_4 - u_3 \end{bmatrix}$. These three random variables are still normally distributed, but they are no longer independently distributed with unit variance. The three now have an intercorrelation of 0.5 and each has a variance equal to two. The means of distributions one through four depicted above were 0.0, 0.4, 0.8, and 1.2, respectively, and, therefore, the means of the three difference variables are 1.2, 0.8, and 0.4. To compute the probability of choice four, we must integrate a multivariate normal distribution with the following parameters:

$$\mu_4 = \begin{bmatrix} \mu_4 - \mu_1 \\ \mu_4 - \mu_2 \\ \mu_4 - \mu_3 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 0.8 \\ 0.4 \end{bmatrix} \text{ and } \Sigma_4 = \begin{bmatrix} \sigma_4^2 + \sigma_1^2 & \sigma_4^2 & \sigma_4^2 \\ \sigma_4^2 & \sigma_4^2 + \sigma_2^2 & \sigma_4^2 \\ \sigma_4^2 & \sigma_4^2 & \sigma_4^2 + \sigma_3^2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

In this case, the probability of choice four is obtained by integrating the following trivariate normal distribution: $\int_S f(\mathbf{d}_4)d\mathbf{d}_4$, where

$S = [d_{41}, d_{42}, d_{43} \mid 0 \le d_{41} \le \infty,\ 0 \le d_{42} \le \infty,\ 0 \le d_{43} \le \infty]$ and

$f(\mathbf{d}_4) = \dfrac{1}{\sqrt{(2\pi)^3 |\Sigma_4|}} \exp\left[-\dfrac{1}{2}(\mathbf{d}_4 - \mu_4)'\Sigma_4(\mathbf{d}_4 - \mu_4)\right]$. In an analogous fashion, we could compute the probabilities for alternatives one, two, and three. Given the assumptions made above and the means of the distributions, the probabilities for choices one through four, respectively, are 0.086, 0.162, 0.284, and 0.468. These can only be obtained by numerically integrating multivariate normal distributions, and as one might expect, this is a computationally intensive approach, one that quickly becomes impractical as the number of alternatives increases. Also, as conditional probabilities are computed to reflect the influence of attributes of the choices and the choosers in the model for the average utilities, a topic dealt with below, the computation required would increase as a function of the complexity of the experimental design employed.

A way around the computational difficulties would be to use a different distribution to compute the probabilities, one that has a "closed form" that allows the probabilities to be computed without numerical integration. A distribution that is very useful in this case is the multivariate logistic distribution. There are two distinct reasons for using this distribution. One reason is that it provides a good approximation to the multivariate

normal distribution. The other is that in certain cases the multivariate logistic distribution is the correct distribution to use, rather than the multivariate normal distribution.

Most discussions of how closely the logistic distribution approximates the normal distribution focus on the univariate counterparts of the multivariate distributions discussed here. An exception would be Bock (1975) (See Gupta, 1963). He gives an example in which a multivariate logistic distribution approximation of a multivariate normal distribution probability is off by only 0.001, certainly a trivial difference. One can construct examples where the approximation is not this good, but generally speaking, if one takes into account the difference in the variances between the usual multivariate normal distribution assumed and the usual multivariate logistic distribution assumed, then the approximation is good over a large class of locations for the distributions.

A different perspective from the one just presented, one that is preferred here, begins with the focus on the probabilities. Using the above example, suppose 0.086 proportion of a population would choose to see Friends/Family for a personal problem, 0.162 proportion a Social Worker, 0.284 a Psychiatrist, and 0.468 a Psychologist. Now these proportions might have come about because of the normally distributed utilities described above, where the means were 0.0, 0.4, 0.8, and 1.2. Alternatively, these same proportions might have arisen because the utilities were independently distributed according to the *Type I extreme value distribution* with "location parameters" equal to $\xi_1 = 0.000$, $\xi_2 = 0.633$, $\xi_3 = 1.194$, and $\xi_4 = 1.694$. Paralleling the approach taken above, we can again define the probability that a person would choose to see a psychologist as:

$$\Pr[y_4 = 1] = \Pr[u_4 > u_1 \;\&\; u_4 > u_2 \;\&\; u_4 > u_3]$$

$$= \Pr([(u_4 - u_1 > 0) \;\&\; (u_4 - u_2 > 0) \;\&\; (u_4 - u_3 > 0)].$$

As above, we must consider the joint distribution of the three random variables, $u_4 - u_1$, $u_4 - u_2$, and $u_4 - u_3$. Differences of this nature between random variables distributed according the Type I extreme value distribution are themselves jointly distributed according to a multivariate logistic distribution. As with the normal distribution, the fact that the utilities are independently distributed results in the differences, $u_4 - u_1$, $u_4 - u_2$, and $u_4 - u_3$ having a correlation of 0.5.

Earlier, it was stated that the multivariate logistic distribution allows for a straightforward way to compute probabilities. For example, the probability that a member of this population would choose to see a psychologist is equal to:

$$\Pr[y_4 = 1] = \frac{e^{\xi_4}}{e^{\xi_1} + e^{\xi_2} + e^{\xi_3} + e^{\xi_4}} = \frac{e^{1.694}}{e^{0} + e^{0.633} + e^{1.194} + e^{1.694}} = 0.468.$$ In general, if there are J alternatives to choose from, and j = 1 to J indexes the alternatives, the probability that the $j^{th}$ alternative will be chosen is:

$\Pr[y_j = 1] = \frac{e^{\xi_j}}{\sum_{j=1}^{J} e^{\xi_j}}$. Using this approach to calculate the probabilities for alternatives one, two, and three (Friends/Family, Social Worker, and Psychiatrist), we obtain the same probabilities as we did with the multivariate normal distribution.

The point here is this: when we are comparing the multivariate normal and logistic distributions, we can fix the choice probabilities and compare the utilities, or we can fix the utilities and compare the probabilities. The perspective taken here is that we should focus more on the probabilities and take whatever utility estimates our probability model gives us. After all, social workers trying to increase their "market share" will judge their success by the increase in the percentage of clients who seek them out, not by an increase in an average utility.

Assuming the choice probabilities are the same for both models (normal and logistic), if we wanted to compare the utilities for the normal and logistic distributions, we should not simply compare the means of the normal distributions $(\mu_1, \mu_2, \mu_3, \mu_4)$ with the location parameters for the Type I extreme value distributions $(\xi_1, \xi_2, \xi_3, \xi_4)$. The reason is that the linear model for normally distributed utilities can be written as $u_j = \mu_j + \varepsilon$. The individual's part, $\varepsilon$, would frequently be assumed to be normally distributed with a mean of zero and standard deviation of one. Therefore, $\varepsilon$ follows a *standard* normal distribution. This is a *standardized* distribution, because it has a mean of zero and a standard deviation of one. The linear model for utilities distributed according the Type I extreme value distributions can be written as $u_j = \xi_j + \varepsilon$. The individual part, $\varepsilon$, is assumed to follow the *standard form* of the Type I extreme value distribution. *Standard form* means that this distribution's location parameter, $\xi$, is equal to zero, and its scale parameter, $\theta$, is equal to one. The standard form of the Type I extreme value distribution is *not a standardized* distribution because it has a mean of 0.577 and a standard deviation of $\sqrt{\pi^2/6} = 1.283$. The fact that the normal and the extreme value distributions have different means, 0.000 and 0.577, has no effect, because it is the difference between the means of the utility distributions that is important. Adding or subtracting a constant to all the means does not affect the mean differences. The difference in standard deviations between the two distributions does have an effect. We need to judge the mean differences on the proper scale. If we subtract the lowest average utility from the highest and then divide by the appropriate standard deviation, we get for the normal distribution $\frac{0.4 - 0}{1} = 0.4$, $\frac{0.8 - 0}{1} = 0.8$, $\frac{1.2 - 0}{1} = 1.2$, and for the Type I extreme value distribution, we get

$\frac{0.633 - 0.0}{1.283} = 0.493$, $\frac{1.194 - 0.0}{1.283} = 0.931$, $\frac{1.694 - 0.0}{1.283} = 1.320$. Clearly, the values are more similar after being transformed to a common scale. This suggests that if you collect

data from a population where the utilities are normally distributed and then fit a model using the multivariate logistic distribution, you will have no problem getting a fit for the probabilities. The estimates of the differences between the means will be off somewhat, but this discrepancy that can be reduced by proper scaling.

The preceding paragraph shows how to compare the utilities from the two different probability models. In what follows, some additional comparisons will be made, but the primary focus will be on the probabilities, as stated earlier,.

Before comparing the multivariate logistic distribution and the multivariate normal distribution, we stated that the similarities between the two provide some justification for using the multivariate logistic distribution as an approximation of the multivariate normal distribution. This assumes that the normal distribution is the correct distribution. The normal distribution, however, has no special justification other than that many individual difference variables are approximately normally distributed.

One can develop an argument for assuming the multivariate logistic distribution to be the correct distribution. In this case, there would be little interest in considering the multivariate normal as an approximation for the multivariate logistic distribution, since the former comes with considerable computational baggage.

To justify the use of the multivariate logistic distribution we again assume the "utility-maximizing" model which conceptualizes the chooser as "utility maximizing," i.e., the chooser will select the alternative with the highest perceived utility, or value. Next, we need to determine if a characteristic of the multivariate logistic model agrees with how a population of choosers is expected to behave.

Suppose a population of choosers selects Psychiatrist over Social Worker on a two-to-one basis, there being only these two alternatives available to them. In this case, the probability of selecting a Psychiatrist would be 0.667 and for Social Worker it would be 0.333, representing the two-to-one ratio just described. Now suppose that a new alternative becomes available to the choosers, namely Psychologist, and that this alternative is chosen 0.400 proportion of the time. That would leave 0.600 to be shared between the Psychiatrist and Social Worker alternatives. If it is reasonable to assume that after the introduction of Psychologist, this population would still prefer Psychiatrist over Social Worker on a two to one basis, then we can deduce the probabilities for Psychiatrist and Social Worker. They would be 0.400 and 0.200, respectively. What this says, then, is that introducing a new alternative changes the proportions of the population seeking the various helpers, but it does not change the odds ratios of the original alternatives, for Psychiatrist is still preferred two to one over Social Worker. If this is the expected behavior of members of the population under consideration, then the multivariate logistic distribution is the proper one to use.

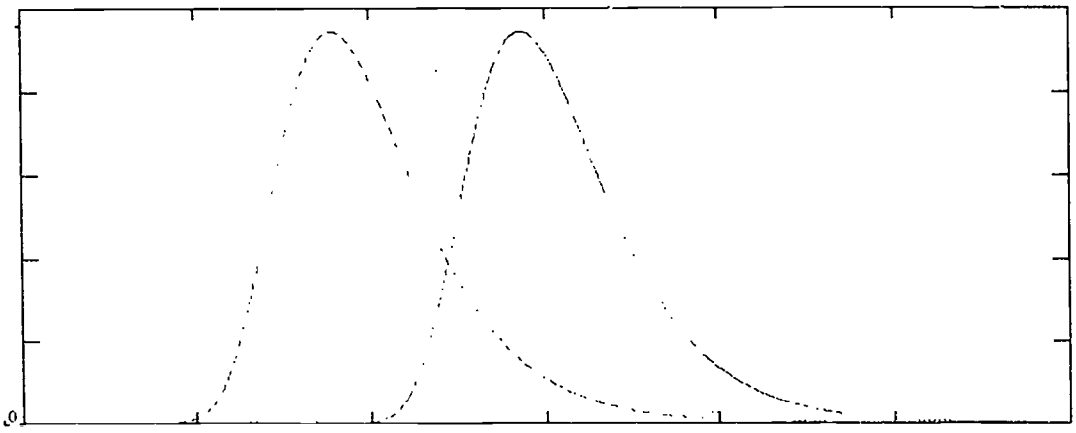
The appropriateness of the multivariate logistic distribution can be demonstrated as follows: if the initial probabilities for Social Worker and Psychiatrist are defined as

$\frac{e^{\xi_1}}{e^{\xi_1}+e^{\xi_2}}$ and $\frac{e^{\xi_2}}{e^{\xi_1}+e^{\xi_2}}$, respectively, then the odds ratio is $\frac{\frac{e^{\xi_2}}{e^{\xi_1}+e^{\xi_2}}}{\frac{e^{\xi_1}}{e^{\xi_1}+e^{\xi_2}}} = \frac{e^{\xi_2}}{e^{\xi_1}}$. When Psychologist is introduced to the alternative set, the probabilities for Social Worker and Psychiatrist are redefined as $\frac{e^{\xi_1}}{e^{\xi_1}+e^{\xi_2}+e^{\xi_4}}$ and $\frac{e^{\xi_2}}{e^{\xi_1}+e^{\xi_2}+e^{\xi_4}}$, respectively, but the odds ratio remains constant, $\frac{\frac{e^{\xi_2}}{e^{\xi_1}+e^{\xi_2}+e^{\xi_4}}}{\frac{e^{\xi_1}}{e^{\xi_1}+e^{\xi_2}+e^{\xi_3}}} = \frac{e^{\xi_2}}{e^{\xi_1}}$. It is clear that the multivariate logistic distribution provides the desired property of a consistent odds ratio for selecting a Psychiatrist over a Social Worker regardless of whether or not Psychologist is in the set. If this agrees with the expected behavior of the population of choosers, then the multivariate logistic distribution provides the correct model.

This property of keeping odds ratios constant as alternatives are included or excluded from the choice set is called *independence of irrelevant alternatives* (IIA). Greene (1993) refers to it more clearly as independence of *other* alternatives.

If the utility-maximizing model holds, then it can be shown that a necessary and sufficient condition for the IIA property (and the multivariate logistic distribution) is that $\varepsilon$, in the model $u_j = \xi_j + \varepsilon$, is independently and identically distributed according to the standard form of the Type I extreme value distribution. In this case, the probability density of $\varepsilon$ is $f(\varepsilon) = e^{-\varepsilon - e^{-\varepsilon}}$.

Returning to the example introduced above for Friends/Family, Social Worker, Psychiatrist, and Psychologist and using the values $\xi_1 = 0.000$, $\xi_2 = 0.633$, $\xi_3 = 1.194$, and $\xi_4 = 1.694$, we can plot the distributions for the individual utilities as we did for the normal distribution model.



The above distributions, which represent, from left to right, the utility distributions for Friends/Family, Social Worker, Psychiatrist, and Psychologist, are positively skewed and more leptokurtic than their normal counterparts above. Their shape has no particular intuitive appeal, but given the preceding location parameters, these distributions result in the probabilities of choosing Friends/Family, Social Worker, Psychiatrist, and Psychologist being 0.086, 0.162, 0.284, and 0.468, respectively.

Before we end this section, we will consider a situation where the IIA assumption would not be justified. Suppose that in a particular geographical location, 60% of the people see psychiatrists and 40% see psychologists with Ph.D.'s. The odds ratio is 1.5 to 1. If a training program in the area started graduating psychologists with Doctor of Psychology degrees, a new alternative would be available to potential clients. It is very likely, however, that the type of graduate degree would be unimportant to clients, and therefore, some of those seeing Ph.D.'s would begin to see PsyD's. If the clients preferring a psychologist were equally divided between the two graduate degrees, the effect of this would be to reduce the proportion seeing Ph.D.'s to 0.20.

In this situation, there is no reason to believe that those seeing psychiatrists would change to psychologists simply because they have a PsyD. Psychiatrists would continue with 60% of the clients, and this would give a psychiatrist-to-Ph.D. psychologist odds ratio of 3-to-1, a change from the previous ratio of 1.5 to 1. In this example, this change would violate the IIA assumption, for the odds ratio has changed with the introduction of a new alternative. In attempting to avoid this situation, a good rule of thumb is to include only alternatives that are discernibly different to choosers. This does not mean that discernibly different categories must have different utilities. It just means that two or more categories should not be effectively treated as one category by choosers due to the trivial (from the chooser's perspective) nature of the difference between categories.

The above section presented a model of choice for a population of individuals. For each alternative there was a distribution of utilities presented, with the location of the distribution determined by the utility characteristic of the population as a whole. Variability in the distribution was presented as a function of individuals' unique tastes and

decision rules. This model of aggregate choice behavior has been developed in the field of econometrics by McFadden (1973).

There is no reason why the choice model just described could not be applied to a single individual. That is exactly how the model using the multivariate logistic distribution was first developed in psychology by Luce (1959). For an individual making a choice, each alternative would have a fixed utility. Situational factors representing random influences both inside and outside the individual make this a probabilistic process. In this setup, an individual repeatedly offered a set of alternatives would tend to select some more often than others. For example, an individual might eat more oranges than bananas, and more bananas than apples, but at any given choice point it would not be completely certain which fruit would be selected. Perhaps it would be helpful to view this model as analogous to true score/error score model in psychometric theory. An individual has a true score, with independent error scores distributed around the true score. It is the error scores that can cause one's scores to vary from testing to testing.

Research Design and Analysis

If one were only going to ask a sample of people to choose from among a set of alternatives, there would be little motivation to learn about a choice model and a new type of analysis. For example, if the research called for a sample to choose from Friends/Family, Social Worker, Psychiatrist, and Psychologist for help with a personal problem, we could simply assume a multinomial probability model and analyze the data using a $\chi^2$ goodness-of-fit test. In this section, however, we will consider much more complex research designs than the one in this paragraph.

Generally, choice research gets more complex in two ways: by considering additional attributes of the choices and by including attributes of the choosers. While the primary choice attribute of interest in our ongoing example is type of helper, there are other attributes of the choices that could be considered. For example, three of the categories have fees associated with their service. We might vary cost per session to see if lowering or raising fees would affect people's choices. Other attributes of the choices might be convenience (how much time does it take to travel to the helper), availability (what is the time between calling for an appointment and actually seeing the professional), helper's gender, helper's theoretical orientation, and so on. Different combinations of attributes might make a type of helper more or less attractive in comparison to the other alternatives.

Attributes of the choosers might be gender, race, the type of insurance coverage they have (if any), the types of professionals they have seen in the past, the specific nature of their problem, how eager they are to take medication, and so on. Combinations of chooser attributes like these could define groups with very different choice behaviors. Also, a chooser attribute could be the chooser's membership in various treatment groups. For example, people may have been randomly assigned to groups that viewed different video tapes about the mental health professions, with the experimenter's interest being in how the tapes affected the subjects' choice behavior.

Some readers may reflect on the preceding two paragraphs and associate to research designs where analysis of variance or regression analysis is used to investigate the influence of similar attributes on a dependent variable (or dependent variables) of interest. This is a useful insight; for in what follows, we will use the design matrix concept, a concept fundamental to the general linear model approach to analysis of variance and regression analysis. Also, we will use many of the design principles used in research with quantitative dependent variables.

The link between more common research methods and research on choice models is through the utility distributions. The model, $u_j = \xi_j + \varepsilon$, becomes $u_{ij} = \xi_{ij} + \varepsilon_{ijk}$. The subscript, j, still refers to the $j^{th}$ alternative, while i refers to the $i^{th}$ combination of attributes (I = 1 to I) and k refers to the $k^{th}$ person presented with the $i^{th}$ combination of attributes (k = 1 to $K_i$). ($K_i$ allows the number of subjects for each attribute combination to vary. Likewise, the number of alternatives could vary from choice set to choice set by using $J_i$ instead of J.). In this setup, the $\xi_{ij}$ are defined by the following linear model: $\xi_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_m x_{ijm} + \cdots + \beta_M x_{ijM}$, where m = 1 to M, the number of parameters in the model. If we define $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \\ \vdots \\ \beta_M \end{bmatrix}$ and $\mathbf{x}_{ij} = \begin{bmatrix} x_{ij1} \\ \vdots \\ x_{ijm} \\ \vdots \\ x_{ijM} \end{bmatrix}$, then the probability for the $k^{th}$ person receiving the $i^{th}$ combinations of attributes with respect to the $j^{th}$ alternative is

$$P_{ijk} = P_{ij} = \frac{e^{\beta' x_{ij}}}{\sum_{j=1}^{J} e^{\beta' x_{ij}}}$$

. The reason for dropping the subscript, k, is because all of the $K_i$ people receiving the $i^{th}$ combination are predicted to have the same probability for the $j^{th}$ alternative. This is analogous to regression analysis were individuals with the same values on the independent variables are predicted to have the same score. The difference is that each person is predicted to have J probabilities, where $\sum_{j=1}^{J} P_{ijk} = \sum_{j=1}^{J} P_{ij} = 1$.

The predicted probabilities are conditional in the sense that they depend upon the values of the independent variables for the $i^{th}$ combination of attributes. Also, the probability model assumes the errors are independently distributed according the Type I extreme value distribution, because the multivariate logistic distribution is used to compute the probabilities. This model is referred to as the *conditional logit model* and is attributed to McFadden (1973).

In actual practice, one would have to estimate $\beta$, with, say, $\hat{\beta}$. Using $\hat{\beta}$, one would have a different (less accurate) set of predicted probabilities, namely, $\hat{P}_{ijk} = \hat{P}_{ij} = \frac{e^{\hat{\beta}' x_{ij}}}{\sum_{j=1}^{J} e^{\hat{\beta}' x_{ij}}}$, with

$$\sum_{j=1}^{J} \hat{P}_{ijk} = \sum_{j=1}^{J} \hat{P}_{ij} = 1.$$

To estimate the parameter vector we will use a statistical package, LIMDEP, Version 6.0, available from Econometric Software, Inc. The conditional logit analysis is carried out by the program's procedure for estimating the "Discrete Choice Model," which is described Chapter 41 of the User's Manual.

To estimate $\hat{\beta}$, LIMDEP uses the maximum likelihood method. Maximum likelihood estimation is not covered very well (if at all) in some applied statistics courses. More emphasis is placed on least squares estimation, especially in the context of regression analysis. For this reason, we will introduce here the basic concepts underlying maximum likelihood (MLE) estimation in the context of conditional logit analysis.

To simply the comments about MLE, we will suppose that we have drawn a sample of K subjects for a single combination of attributes (i.e., I = 1), and we ask each person to make a choice. As each subject chooses, he or she generates a vector, $\mathbf{y}_k$, with J elements. For the elements of this vector, $y_{jk} = 1$ if the $j^{th}$ alternative was selected, otherwise $y_{jk} = 0$. Given this setup, the likelihood (probability) of the vector for the $k^{th}$ subject can be written as $L_k = P_1^{y_{1k}} P_2^{y_{2k}} \cdots P_j^{y_{jk}} \cdots P_J^{y_{Jk}} = P_{j'k}$, where $j'$ is the value of j corresponding to the choice that was made. (One way to motivate this definition of the likelihood of a single vector is to think of $L_k$ as a multinomial probability based on J events and only a single repetition or trial (i.e., n = 1).) Since the subjects' response vectors, $\mathbf{y}_k$, are independently distributed, the likelihood of the sample is equal to the product of the likelihoods of the vectors, or $L = \prod_{k=1}^{K} P_{j'k}$, where, again, $j'$ is understood to equal the number of the choice that was made by the $k^{th}$ subject. Of the K terms in the preceding product, there are only J unique values. If we let $n_j$ equal the number of subjects who chose the $j^{th}$ alternative, then $K = \sum_{j=1}^{J} n_j$ and $L = \prod_{k=1}^{K} P_{j'k} = \prod_{j=1}^{J} P_j^{n_j}$. If there were four choices (J = 4), then $L = P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}$.

If we knew the values of the $P_j$, we could compute the likelihood of the particular sample we drew, but, of course, these values are unknown. Therefore, we must estimate the unknown values. Suppose we had two set of estimates, $\tilde{P}_j$ and $\bar{P}_j$, and that we estimated

the likelihood of the sample using each estimate, i.e., $\tilde{L} = \prod_{j=1}^{J} \tilde{P}_j^{n_j}$ and $\check{L} = \prod_{j=1}^{J} \check{P}_j^{n_j}$. If $\tilde{L}$ were much larger than $\check{L}$, we would know that it is much more likely that our data came from a population with parameters $\tilde{P}_j$ than from a population with parameters $\check{P}_j$. We would therefore take $\tilde{P}_j$ as our estimate over $\check{P}_j$ because our data are more likely to have come from a population with the $\tilde{P}_j$ as parameters. This seems reasonable, for we would not want the estimates that are less consistent with the data.

If we agree that we should always take the estimate that gives the highest likelihood for the sample we have drawn, then we are led to accept as our estimates those values that maximize the likelihood function. These maximum likelihood estimates are symbolized as $\hat{P}_j$, and they have the property that $\hat{L} = \prod_{j=1}^{J} \hat{P}_j^{n_j}$ is greater than for any other set of estimates. We therefore take as our estimates those values that maximize the likelihood of the sample we have drawn. We do this because it seems reasonable think that these values would be more "similar" to the true values, the $P_j$, than other values that would make our data less likely. (Maximum likelihood estimates have been shown to have many desirable properties that recommend them, properties such as consistency, efficiency and sufficiency. These properties make them good estimators and are preferred to the vague notion of "similarity" to a parameter in all but the most casual discussions.)

With this build up, it seems somewhat anticlimactic to note that the maximum likelihood estimators are the sample proportions, i.e., $\hat{P}_j = \frac{n_j}{K}$, for these are certainly easy to compute. Our ultimate interest, though, is to estimate $\beta$, and our estimate, $\hat{\beta}$, must yield the appropriate probability values, i.e., the $\hat{P}_j$. This means that $\hat{\beta}$ must satisfy the equation $\frac{e^{\hat{\beta}'x_j}}{\sum_{j=1}^{J} e^{\hat{\beta}'x_j}} = \hat{P}_j$ for all j = 1 to J.

Examples

Following in this section are examples of five relatively simple research designs that introduce the basics of the design, analysis, and interpretation of choice experiments. All experiments are based on simulated data; therefore the populations that generated the data are known. The underlying strategy in this section is analogous to that used in learning the use of radar in nautical environments. To learn radar interpretation, one picks a perfectly clear day and continually alternates between observing what is on the radar screen and what is clearly visible as one looks outside. By alternately looking at radar "targets" and the actual objects being depicted, one learns what boats, buoys, islands, and so on look like on the screen. In what follows, the "radar screen" is the output from the LIMDEP program and "what's really out there" are the parameters of the populations from which the data were sampled.

To make the interpretation more demanding, it was decided to generate the choice data using normally distributed utilities and then make the correspondence between LIMDEP's parameter estimates based on Type I extreme value distributions and the actual parameters of the normal distributions. It became clear in the initial draft of this paper that constantly working back and forth between the two distributions made the presentation too convoluted. To remedy this, the following examples are interpreted with respect to extreme value distribution parameters that would result in the same population probabilities as their normal distribution counterparts. A discussion of the relationship of the Type I extreme value distribution parameters to actual normal parameters has been moved to an appendix. The only consequence for the following examples is that the parameters discussed are not found to be the "nice, neat numbers" usually found in simulations. Those will be found in the appendix, where the normal distribution parameters are discussed.

*Example 1*: As an example we have generated a sample of K = 50 observations where I = 1 and J = 4. This data could have come from a study where subjects were asked to choose from among Friends/Family, Social Worker, Psychiatrist, and Psychologist. In this sample one person chose alternative one ($n_1 = 1$), five chose alternative two ($n_2 = 5$), 14 chose alternative three ($n_3 = 14$), and 30 chose alternative ($n_4 = 30$). The maximum likelihood estimates are the sample proportions:
$\hat{P}_1 = 0.02$, $\hat{P}_2 = 0.10$, $\hat{P}_3 = 0.28$, and $\hat{P}_4 = 0.60$. The independent variables are "choice-specific" dummy variables: $\mathbf{x}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$, and $\mathbf{x}_4 = \begin{bmatrix}0\\0\\0\end{bmatrix}$.
Since proportions add to one, only three proportions are free to vary, and accordingly, only three elements are required in $\hat{\beta}$:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}.$$

The equations $\frac{e^{\hat{\beta}'x_j}}{\sum_{j=1}^{J} e^{\hat{\beta}'x_j}} = \hat{P}_j$ for j = 1 to 4, due to the simplicity of the dummy variables, can be written as:

$$\frac{e^{\hat{\beta}_1}}{e^{\hat{\beta}_1} + e^{\hat{\beta}_2} + e^{\hat{\beta}_3} + e^{0}} = 0.02$$

$$\frac{e^{\hat{\beta}_2}}{e^{\hat{\beta}_1} + e^{\hat{\beta}_2} + e^{\hat{\beta}_3} + e^{0}} = 0.10$$

$$\frac{e^{\hat{\beta}_3}}{e^{\hat{\beta}_1} + e^{\hat{\beta}_2} + e^{\hat{\beta}_3} + e^{0}} = 0.28$$

$$\frac{e^{0}}{e^{\hat{\beta}_1} + e^{\hat{\beta}_2} + e^{\hat{\beta}_3} + e^{0}} = 0.60$$

Note that $\hat{\beta}_4$ is set to zero, therefore, as stated, only three estimates are free to vary and they are interpreted relative to zero. We must solve the preceding system of nonlinear equations for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$. The solution is obtained iteratively using the Newton method, a numerical analysis technique that will not be covered here. The solution for the above equations is $\hat{\beta}_1 = -3.4012$, $\hat{\beta}_2 = -1.7918$, and $\hat{\beta}_3 = -0.7621$.

To test the hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, LIMDEP uses a likelihood ratio test. Under the hypothesis, the $\beta$'s are restricted to zero and, therefore, $P_1 = P_2 = P_3 = P_4 = P = 0.25$. The likelihood of the sample under this restriction is $P^{n_1}P^{n_2}P^{n_3}P^{n_4} = 0.25^{1} \cdot 0.25^{5} \cdot 0.25^{14} \cdot 0.25^{30}$. Without the restriction on the $\beta$'s, the values $\hat{\beta}_1 = -3.4012$, $\hat{\beta}_2 = -1.7918$, $\hat{\beta}_3 = -0.7621$, and $\hat{\beta}_4 = 0.0$ are used to compute the probabilities, and this leads to the sample proportions $\hat{P}_1 = 0.02$, $\hat{P}_2 = 0.10$, $\hat{P}_3 = 0.28$, and $\hat{P}_4 = 0.60$, i.e., the maximum likelihood estimates. Using these values, the likelihood of the sample is $P_1^{n_1}P_2^{n_2}P_3^{n_3}P_4^{n_4} = 0.02^{1} \cdot 0.10^{5} \cdot 0.28^{14} \cdot 0.60^{30}$. The ratio of the two likelihoods,

$\lambda = \frac{P^{n_1} P^{n_2} P^{n_3} P^{n_4}}{P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}}$, which varies between 0 and 1, is an index of the adequacy of the null hypothesis. The larger the ratio, the stronger the support for the hypothesis, the smaller the ratio, the weaker the support. At some point, as the ratio decreases in size, we would make a decision to reject the null hypothesis. To test the hypothesis, the statistic $-2\ln(\lambda)$ is used. It is compared to a $\chi^2$ distribution, with the degrees of freedom equal to the number of independent variables (df = 3 for the current example). LIMDEP prints out the log likelihood for the denominator $\left(\ln\left(P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}\right) = -48.57124 \text{ for this example}\right)$, the log likelihood of the numerator $\left(\ln\left(P^{n_1} P^{n_2} P^{n_3} P^{n_4}\right) = -69.31472\right)$ and the value of the test statistic $\left(-2\left(\ln\left(P^{n_1} P^{n_2} P^{n_3} P^{n_4}\right) - \ln\left(P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}\right)\right) = 41.48696\right)$. The p-value for the $\chi^2$ in this example is 0.0000000044, which would lead to rejecting the hypothesis.

The sample just analyzed was drawn from a population were the true probabilities are 0.086, 0.162, 0.284, and 0.468. As described in the previous section, the location parameters for the Type I extreme value distributions were $\xi_1 = 0.000$, $\xi_2 = 0.633$, $\xi_3 = 1.195$, and $\xi_4 = 1.694$. The first step in judging the adequacy of the estimates is to give them a common reference point. Given the independent variables used, $\hat{\beta}_4$ is "estimated" to be zero. Since the probabilities are unaffected by adding or subtracting a constant from the location parameters, we subtract 1.694 from each location parameter so that we have a more appropriate comparison to the estimates just computed, i.e., $\hat{\beta}_4 = \xi_4 = 0.000$. After this adjustment, we have $\xi_1 = -1.694$, $\xi_2 = -1.061$, $\xi_3 = -0.501$, and $\xi_4 = 0.000$. $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ now estimate the adjusted values $\xi_1$, $\xi_2$, and $\xi_3$, respectively (e.g., $\hat{\beta}_1 = -3.4012$ is estimating $\xi_1 = -1.694$). The standard errors for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ as given by LIMDEP are $\hat{\sigma}_{\hat{\beta}_1} = 1.017$, $\hat{\sigma}_{\hat{\beta}_2} = 0.4830$, and $\hat{\sigma}_{\hat{\beta}_3} = 0.3237$. All the estimates are within two standard errors of the parameters they estimate. The p-values for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are 0.00082, 0.00021, and 0.01854, respectively, and therefore, all the hypotheses for the individual weights $\left(\text{e.g., } H_0: \beta_1 = 0\right)$ would be rejected. The p-values for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are computed by LIMDEP by assuming that under the hypothesis that $\hat{\beta}_j / \hat{\sigma}_{\beta_j}$ is distributed according to a standard normal distribution.

*Example 2*: As a second example, we continue with the above setup, but our sample has an additional 950 observations drawn from the same population, for a total of K = 1000. The maximum likelihood estimates are the sample proportions, and in this sample, they are equal to $\hat{P}_1 = 0.083$, $\hat{P}_2 = 0.162$, $\hat{P}_3 = 0.295$, and $\hat{P}_4 = 0.460$. These are much better estimates of the population values, as one would expect. The solution for the above equations is $\hat{\beta}_1 = -1.7124$, $\hat{\beta}_2 = -1.0436$, and $\hat{\beta}_3 = -0.4443$. These values, not surprisingly, agree very closely with the adjusted location parameters, $\xi_1 = -1.694$, $\xi_2 = -1.061$, and $\xi_3 = -0.501$. The standard errors for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are $\hat{\sigma}_{\hat{\beta}_1} = 0.1193$, $\hat{\sigma}_{\hat{\beta}_2} = 0.09136$, and $\hat{\sigma}_{\hat{\beta}_3} = 0.07459$. As would be expected, the standard errors have decreased a great deal, and all the estimates continue to be within two standard errors of the parameters they estimate. Due to the increased sample size, the p-values for the various hypothesis tests have decreased, and therefore, the hypotheses tested would all be rejected.

With this sample, we will begin to present the data the way LIMDEP accepts it. For this example, LIMDEP expects K = 1000 subjects with J = 4 rows for each subject (I = 1). Since everyone was presented with same set of choices and attributes, the values of X1, X2, and X3 are the same for everyone. The four sets of rows in the following table depict the choice of alternative one, two, three, and four, respectively. The first set appears in the data $n_1 = 83$ times, the second set $n_2 = 162$ times, the third $n_3 = 295$ times, and the fourth $n_4 = 460$ times, for a total of 4000 rows.

| Choice | X1 | X2 | X3 |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 |

Alternatively, if we had tabulated the data prior to submitting it to LIMDEP and knew the sample proportions, we could submit just four rows, with the proportions in place of the

choice data. In this case, LIMDEP would require an additional column of weights to tell it the number of people the proportions are based on.

Given the simplicity of *Example 1* and *Example 2*, one may question the effort required to compute $\hat{\beta}$, for it is hard to see its value when one has the sample proportions. That may be so for these examples, but for more complex examples, the above approach allows us to specify a probability model, to see which independent variables have a significant impact on the probabilities, and to develop simulation models that allow us to estimate probabilities for combinations of attributes not included in our data collection.

*Example 3*: In this example, with one change, we are using the same four-choice setup we have been using involving Friends/Family, Social Worker, Psychiatrist, and Psychologist. The change is that we are simulating an experiment that includes an additional attribute, travel time to the helper. The travel times are 10, 20, 30, 40, 50, 60, 70, 80, and 90 minutes. The subjects are asked, for example, whether they would choose a psychiatrist they had to travel 70 minutes to see, or a social worker they had to travel 50 minutes to see, and so on. For each alternative for each subject a time is randomly selected with replacement. In this experiment, there are J = 4 rows in each block, K = 1 one subjects per block, and I = 1000 blocks. It is possible that time might be constant across the four alternatives presented to the subject, although we would only expect this to happen about once in our sample of I = 1000. The experimenter believes that time influences the utilities in a linear fashion and fits a model that reflects that. The true state of affairs is, however, that time has no effect.

The following table contains the first 16 rows (4 blocks) of the data.

| Choice | X1 | X2 | X3 | Time |
|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 80 |
| 0 | 0 | 1 | 0 | 30 |
| 0 | 0 | 0 | 1 | 50 |
| 1 | 0 | 0 | 0 | 40 |
| 0 | 1 | 0 | 0 | 50 |
| 0 | 0 | 1 | 0 | 80 |
| 0 | 0 | 0 | 1 | 60 |
| 1 | 0 | 0 | 0 | 40 |
| 0 | 1 | 0 | 0 | 40 |
| 1 | 0 | 1 | 0 | 40 |
| 0 | 0 | 0 | 1 | 20 |
| 0 | 0 | 0 | 0 | 60 |
| 0 | 1 | 0 | 0 | 60 |
| 0 | 0 | 1 | 0 | 80 |
| 0 | 0 | 0 | 1 | 90 |
| 1 | 0 | 0 | 0 | 30 |

Since for this data, time has no impact on the utilities, the sample proportions are the same as the previous example. When we fit the model with four independent variables, the weights are $\beta_1 = -1.7125$, $\beta_2 = -1.0444$, $\beta_3 = -0.4447$, and $\beta_4 = -0.0004$. The first three weights are almost identical to their counterparts in the previous example. The tests for these three weights, as well as the overall test on the full model, are significant, with p-values less than 0.00001. The p-value of the fourth weight, for Time, is $p = .8161$, which would lead us to a correct decision about the influence of Time.

*Example 4*: With one exception, the setup for this example is the same as for *Example 3*. The difference is that Time has an influence in this data. The utility for each helper alternative had $t = -0.006721 \cdot (\text{Time} - 50)$ added to it.[1] This is an additive effect that causes the utility to be reduced for travel times greater than 50 minutes and increased for travel times of less than 50 minutes. The first four observations are the same as for the previous example, so no table is included. The sample proportions are only slightly changed, 0.082, 0.160, 0.287, 0.471. The weights for this data set are $\hat{\beta}_1 = -1.7568$, $\hat{\beta}_2 = -1.0744$, $\hat{\beta}_3 = -0.4921$, and $\hat{\beta}_4 = -0.0065$. The p-values for the full model and the first three weights are less than 0.00001. For $\beta_4$, the p-value is 0.00002. While $\beta_4$ is small compared to the other weights, it is 16 times larger than it was in the previous example and can exert a noticeable influence on the *conditional* probabilities. For example, if the travel times for the four alternatives were, 50, 50, 10, and 90 minutes,respectively, then given these travel times, the choice probabilities are predicted to be 0.083, 0.164, 0.382, and 0.371. Taking time into account, we see that about 10% of the participants would shift to a psychiatrist if one were "in the neighborhood" and the nearest psychologist was 90 minutes away.

[1]The equation $t = -0.006721 \cdot (\text{Time} - 50)$ is a close approximation to the definition of the time effect that would have been required had the probabilities been defined using Type I extreme value distributions. See the appendix for an explanation of how we arrived at the value, -0.006721.

The manner in which time was included in *Example 3* and *Example 4* represents only one of a number of possibilities, and reflects the researcher's belief that a simple linear relationship exists between time and choice. If the researcher believed that for each alternative a different linear relationship held, then the first four blocks would have looked like the following:

| Choice | X1 | X2 | X3 | Time1 | Time2 | Time3 | Time4 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 80 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 30 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 50 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 40 |
| 0 | 1 | 0 | 0 | 50 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 80 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 60 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 40 |
| 0 | 1 | 0 | 0 | 40 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 40 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 20 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 60 |
| 0 | 1 | 0 | 0 | 60 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 80 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 90 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 30 |

Handling time as it is in the preceding table allows a different slope for each helper alternative. This would allow, for example, time to have less of an effect on the more popular alternatives and a stronger effect on the less popular ones.

Another alternative for dealing with time is to "dummy code" it. The following table gives an example of this for the first four blocks (observations). Only the columns for time are included.

| T10 | T20 | T30 | T40 | T50 | T60 | T70 | T80 |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

The coding pattern in the preceding table is straightforward. A "1" is placed in the column that corresponds to the Time level presented, except for 90 minutes, which is represented by a row of all zeroes. This type of coding could handle a variety of very complex relationships between time and choice, but whatever the relationship was, it would be assumed not to interact with type of professional. To account for complex interactions, additional columns would be required. One could use the 24 columns that would result if "product variables" were constructed by multiplying the three choice-specific variables times the eight time variable.

In the three examples employing time, the same sets of time levels were used with all helper categories. There is nothing in conditional logit analysis that requires this, for a different set of times could have been used with, say, Friends/Family.

To this point, we have dealt only with attributes of the choices. In the next example, we introduce an attribute of the choosers.

*Example 5*: This example continues with the same setup as *Example 4*. Time is included as a single column in the set of independent variables (see the first table in the previous example) and has the simple additive effect described above. To the design we add Gender, with two levels, Females and Males. The first 500 observations in the data are simulated responses from females, and the remaining 500 are from males. In an ordinary regression approach, the design matrix would be augmented with a single column "dummy coded" to reflect gender. For example, we could code Females as "1" and Males as "0."

If we did this, the set of J rows for a given subject would have one of its columns containing J ones or J zeroes. For example, if the first observation were augmented with a column for Gender, for a female it would look like this:

| Choice | X1 | X2 | X3 | Time | Gender |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 80 | 1 |
| 0 | 0 | 1 | 0 | 30 | 1 |
| 0 | 0 | 0 | 1 | 50 | 1 |
| 1 | 0 | 0 | 0 | 40 | 1 |

For a male, an example is:

| Choice | X1 | X2 | X3 | Time | Gender |
|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 20 | 0 |
| 0 | 0 | 1 | 0 | 70 | 0 |
| 1 | 0 | 0 | 1 | 50 | 0 |
| 0 | 0 | 0 | 0 | 80 | 0 |

We cannot use this approach in the conditional logit model. Since for both females and males, the gender variable is constant across all alternatives, it cannot affect the choice. This is because in assessing the impact of an independent variable on choice, the conditional logit analysis uses the deviation of the observations in a column around the column mean *within that block*. This is in contrast to the ordinary regression approach, where deviations are taken around the means *of the entire column*. To make gender vary across the choices, we use the product variables that result from multiplying the choice-specific variables and the variable for gender. We are assessing, therefore, the interaction of choice and gender. The following two tables give an example of an observation for a female and a male. An example of a female's observation is:

| Choice | X1 | X2 | X3 | Time | X1G | X2G | X3G |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 80 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 30 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 | 50 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 40 | 0 | 0 | 0 |

For a male, an example of an observation is:

| Choice | X1 | X2 | X3 | Time | X1G | X2G | X3G |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 20 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 70 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 50 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 80 | 0 | 0 | 0 |

To study the effect of Gender, we will assume that Time = 50 for the four alternatives so that its effect is removed and we can concentrate our attention on the helper alternatives and Gender. In order for Gender to have an effect, there must be differential increments/decrements of the four helper utilities associated with females and males. For females, the population model was modified so that the helper utilities were changed by 0.3738, 0.8648, 0.3698, and 0.0 for Family/Friends through Psychologist, respectively. For males, the corresponding weights were changed by -0.3702, -0.9142, -0.3529, and 0.0. The sum of each helper utility and its corresponding change for females results in utilities of -1.3203, -0.1961, -0.1297, and 0.0. The sum of each helper utility and its corresponding change for males results in utilities of -2.0643, -1.9751, -0.8524, and 0.0. The largest effect is clearly for Social Worker, the second alternative, where the difference between females and males is -0.1961 - (-1.9751) = 1.7790. This magnitude of effect can substantially shift the utility distributions and result, as we will see shortly, in major shifts in the choice probabilities. The effects for alternatives one and three, are 0.7441 and 0.7227, respectively.

As should be clear from the preceding two tables, the choice model now has seven independent variables, three for choice-specific variables, one for Time, and three for the interaction of choice and Gender. The weights estimated by LIMDEP for the data are: $\hat{\beta}_1 = -2.1778$, $\hat{\beta}_2 = -1.9457$, $\hat{\beta}_3 = -0.82827$, $\hat{\beta}_4 = -0.0069723$, $\hat{\beta}_5 = 0.76716$, $\hat{\beta}_6 = 1.8304$, and $\hat{\beta}_7 = 0.67602$. The last three estimates, $\hat{\beta}_5$, $\hat{\beta}_6$, and $\hat{\beta}_7$, are reasonable estimates of the population Gender effects reported above, which we will designate as $\beta_5 = 0.7441$, $\beta_6 = 1.7790$, and $\beta_7 = 0.7227$. On the other hand, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ bear much less resemblance to the original location parameters, $\xi_1 = -1.694$, $\xi_2 = -1.061$, and $\xi_3 = -0.501$. This discrepancy is due to the coding scheme used for gender, which caused the choice-specific variables and the product variables (choice multiplied times gender) to be nonorthogonal. Had gender been coded 0.5, -0.5, instead of 1, 0, the product variables would have been orthogonal to the choice-specific variables and the values of the first three estimates would have been -1.7942, -1.0305, and -0.49026, respectively. These values are in much closer agreement with actual location parameters. Regardless of the

[1] The fact that deviations in the "corrected" design matrix are taken about the within-block column means has an effect on the rank of this matrix. If J, and K, are constant in all blocks, then the matrix will have IJK rows and M columns, and the maximum the rank can be is the minimum of IJ - J and M. This is the reason why the number of choice-specific variables is one less than the number of choices.

coding scheme, the sums of the products of the estimates and the values of the independent variables lead to the same probabilities. Therefore, while the estimates may look different as a function of the coding scheme used, in combination with the values of the independent variables they lead to the same probabilities for the four choice alternatives.

The weight for Time, $\hat{\beta}_4 = -0.0069723$, remains close to the population value -0.006721, given in *Example 4*. Since the values of the time variable were selected at random, they would not be expected to correlate with the other variables.

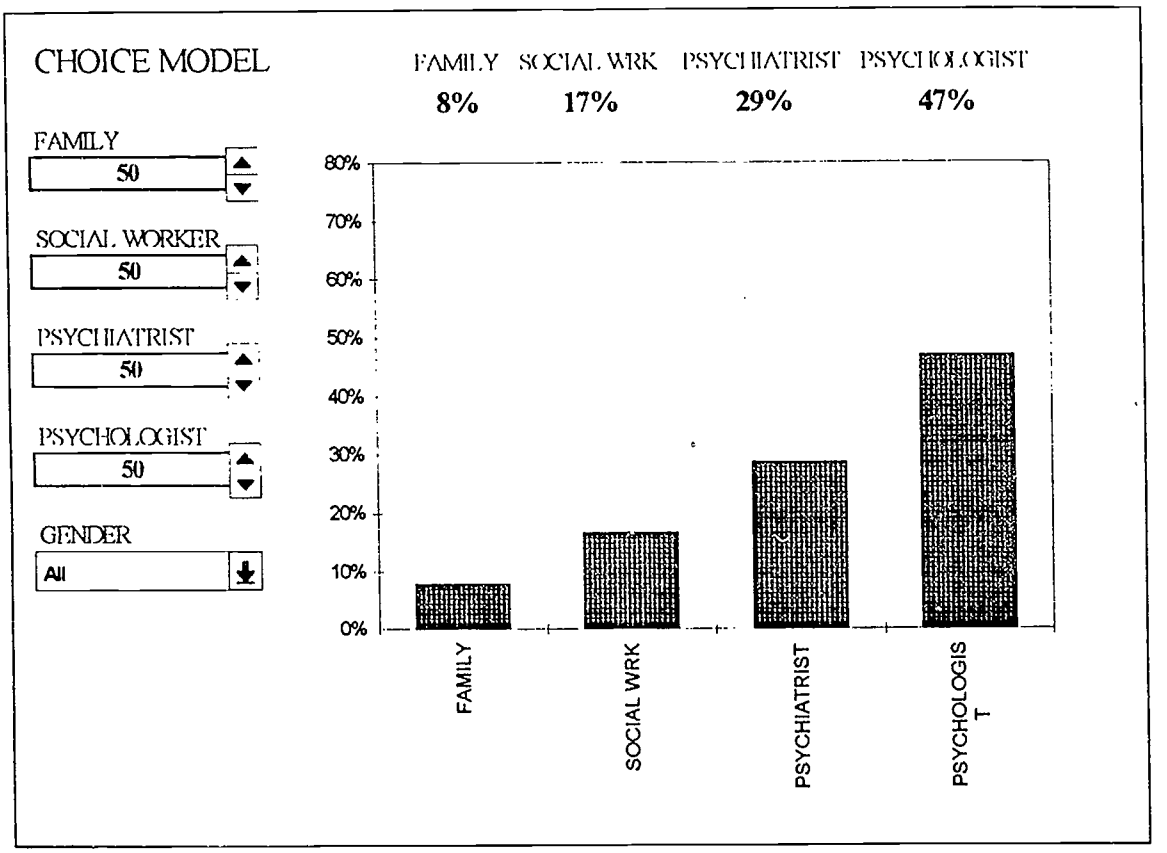
The preceding five examples demonstrate designs and associated analyses employing attributes of the choices and of the choosers. They are, of course, very simple research designs and serve only to introduce the conditional logit model and analysis. In practice, experiments designed to study choice could involve many attributes and the resulting experimental designs could be far too complex to allow all combinations of the independent variables to be included. In this case, researchers can turn to confounded designs, namely fractional factorial and incomplete block designs. While these designs are not typically found in counseling research, choice experiments could cause their greater use.

*Relationship to the multinomial logit model*: What has been presented here as the conditional logit model has sometimes been called the multinomial logit model. In this paper we have followed what little consensus there seems to be by using the term *conditional logit model* whenever the model involves attributes of the choices *or* attributes of the choices and attributes of the choosers. We reserve the term *multinomial logit model* for designs that do not have attributes of the choices. For example, if one were studying career choices and used as independent variables such predictors as gender, ethnic group, parents' educational level, parents' income, and so on, then we would call it the multinomial logit model because attributes of the careers were not included in the model.[1] With respect to the above examples, if the choice-by-gender product variables were the only independent variables in the model, it would be a multinomial logit model and analysis. For this case, one could use LIMDEP's LOGIT command, although the data would be structured differently than in *Example 5*.[2]

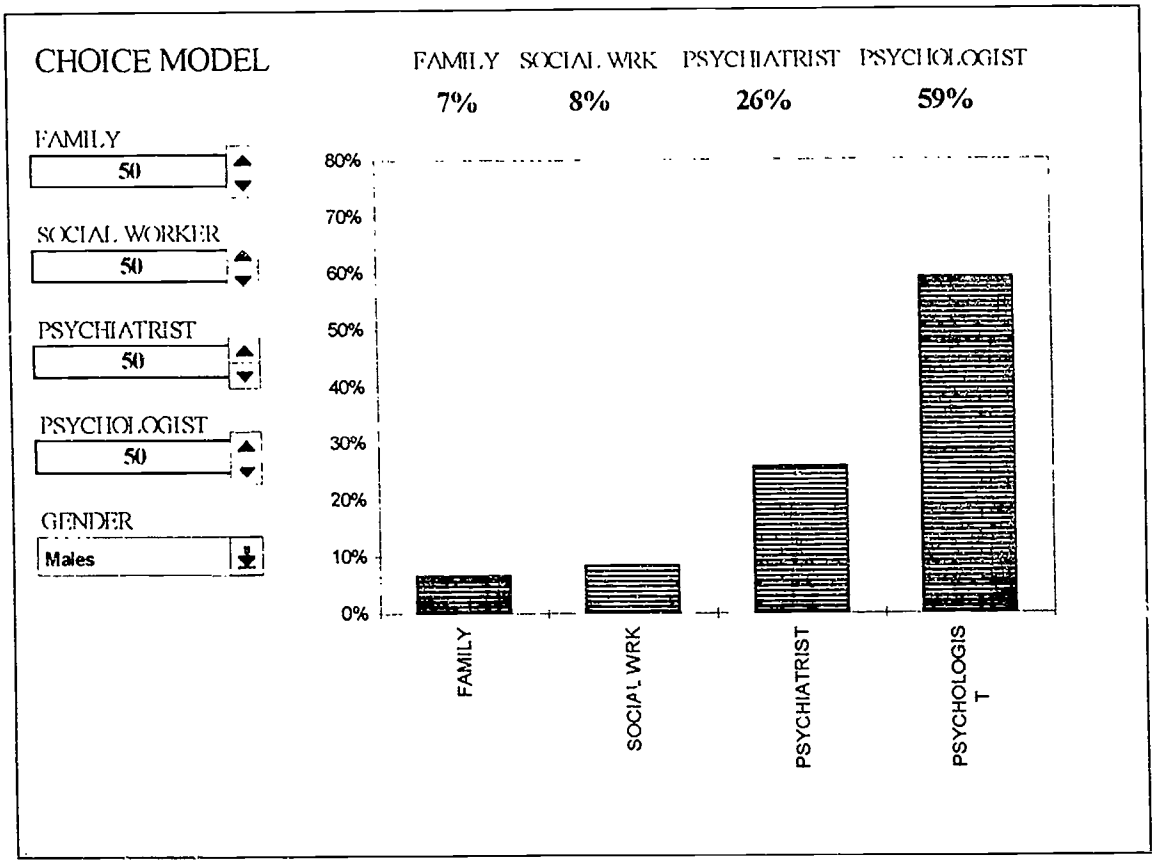
Model Simulation: Earlier in this paper, it was suggested that primary attention might well be placed on the probabilities of the choice alternatives, rather than the estimators used in computing them. The impact of the independent variables can best be seen through their effect on the conditional probabilities associated with the various levels of the independent variables. The following is an Excel spreadsheet for *Example 5*. Since

[1] See Maddala, 1983, p. 42, for the formal equivalence of the two models.
[2] See Chapter 40 of the Version 6.0 User's Manual.

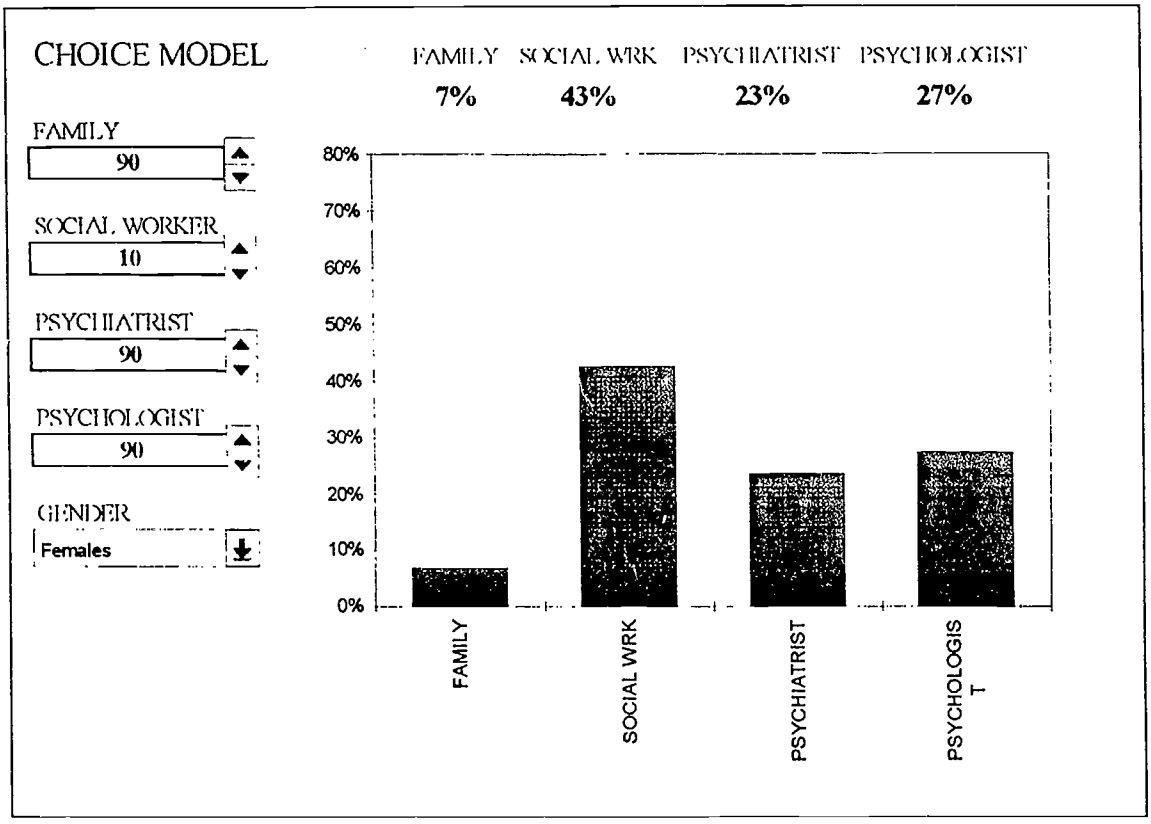
Time is constant across the four alternatives and Gender is set to "All," the percentages reported reflect the overall sample proportions for the four categories.




Leaving Time constant, but setting Gender to Males, we can observe the conditional probabilities for men unaffected by time.




As a last example, Gender is set to Females, and Social Worker is made the most convenient alternative, with the others all having a travel time of 90 minutes. Under these conditions, Social Worker is clearly the preferred alternative.




With only the three preceding figures, it is hard to capture the dynamic nature of the above spreadsheet model. Using "spinners" to change the Time values and a "drop down menu" to vary Gender, the spreadsheet's instant recalculation of values and redrawing the graph produces an animated presentation that goes far in revealing the variables' effects, both singly and in combination, on the choice probabilities. In our experience, this has been more revealing than any attempt to ferret out meaning by focusing exclusively on the estimators.

## Ordered Outcome Categories

Ordered categories are those that can be rank ordered with respect to some underlying continuum. All that is initially postulated is that the second category has more of something than the first, the third has more than the second, and so on. A common example would be a five point Likert scale from "Strongly Disagree" to "Strongly Agree." While it seems reasonable to assume that, in general, those choosing "Strongly Agree" have more of a particular attitude than those who merely "Agree," it would usually not be reasonable to assume that everyone who chose "Agree" had exactly the same strength of attitude. Likewise, the amount of difference between, say, "Strongly Disagree" and "Disagree" would not necessarily be equal to the difference between, say, "Agree" and "Strongly Agree."

The focus in this section will be on a set of ordered categories that could represent treatment outcomes in counseling. For this discussion, the categories to be considered are "Significantly Deteriorated," "Deteriorated," "No Discernible Change," "Improved," and "Significantly Improved."[1] Using a set of categories such as these could eventually lead to giving potential clients the probabilities of these five counseling outcomes. Clients could then weigh their chances for success in counseling in much the same way as patients do when they face surgery or some other medical procedure.

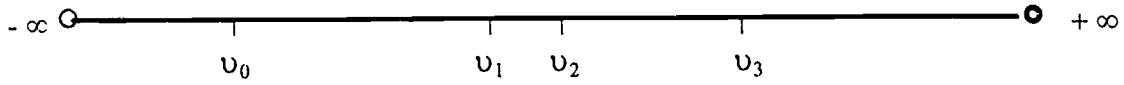
Using categorical outcomes could lead to a different type of quantitative integration of research findings. The meta analysis of the future is envisioned as being based on categorical outcomes that are agreed upon by researchers and practitioners, and a set of independent variables that also have some wide acceptance. This assumes agreement on diagnostic categories, outcome criteria, treatment specifications, and client and counselor characteristics of importance. This agreement does not currently exist, and may never exist. Accordingly, this section is somewhat abbreviated in comparison to the section on "unordered categories," because choice research and similar studies can be carried out immediately by individual researchers. Lacking the kind of cooperation required for useful categorical outcome research, the remainder of this section will briefly introduce the underlying statistical model and an available method of analysis in the hope that it will stimulate discussion and lead to increased interest in and understanding of the benefits of using categorical outcomes in counseling.

## The Ordered Probit Model

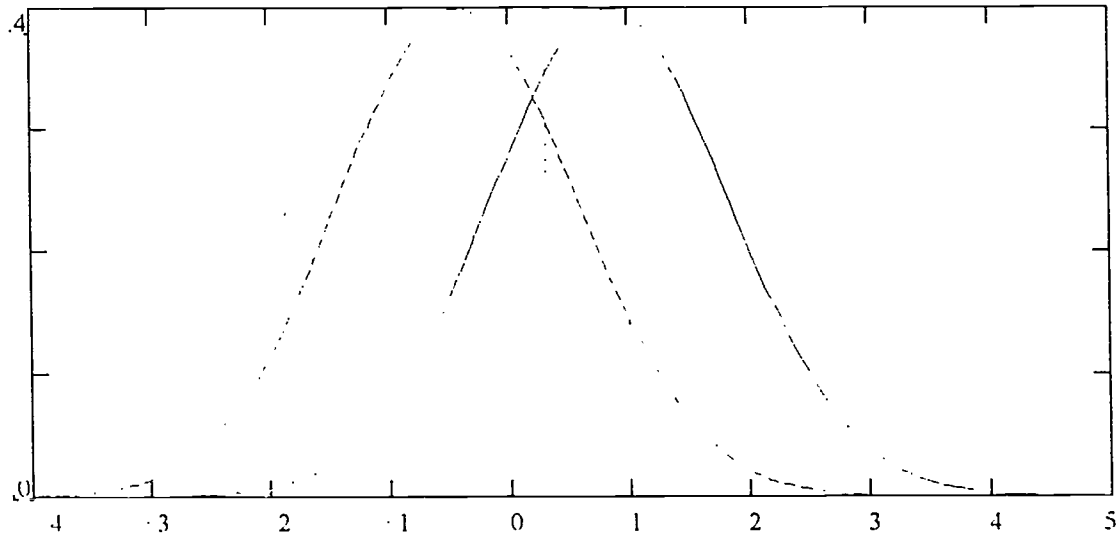
Let us suppose that for a particular type of problem or diagnostic category that raters have been trained to rate individuals' improvement or deterioration and to assign them to the following five categories: "Significantly Deteriorated," "Deteriorated," "No Discernible Change," "Improved," and "Significantly Improved." These categories are numbered 0, 1, 2, 3, and 4, respectively, and they correspond to segments on a continuum defined by the points $\upsilon_0$, $\upsilon_1$, $\upsilon_2$, and $\upsilon_3$. If the amount a person changes falls below $\upsilon_0$, he or she is placed in category "0." When people fall between $\upsilon_0$ and $\upsilon_1$, they are in category "1." People between $\upsilon_1$ and $\upsilon_2$ are in "2," and those between $\upsilon_2$ and $\upsilon_3$ are in "3." Those above $\upsilon_3$ are in category "4." The values of upsilon in this example are $\upsilon_0 = -1.65$, $\upsilon_1 = -0.30$, $\upsilon_2 = 0.30$, and $\upsilon_3 = 1.00$. Category 2, "No Discernible Change," is defined by $\upsilon_1 = -0.30$ and $\upsilon_2 = 0.30$, an interval that covers zero and represents that part of the continuum where raters cannot tell whether the change was positive or negative. The two categories to the left of this interval represent successively more negative changes, while the two to the right represent successively more positive

[1] The word "Significantly" is meant to mean "Marked" and has no statistical meaning as it would in the context of hypothesis testing in the phrase "significantly different."

changes. The following number line graphically depicts the outcome continuum.




For the particular diagnostic category related to this continuum, we will assume that the independent variables that determine the amount of change are socio-economic level (upper, middle, lower), ethnicity (majority, minority), gender (female, male), and treatment (yes, no). The 24 = 3 x 2 x 2 x 2 combinations of these variables define 24 populations. Determining the rater's assignment, there is for each population an underlying random variable, $u_i = \beta' \mathbf{x}_i + \varepsilon$, where $\beta$ is a vector of parameters, $\mathbf{x}_i$ is the vector of independent variable values for the $i^{th}$ population and $\varepsilon$ is independently distributed according to the standard normal distribution. Given this setup, each of the 24 populations can have a different location on the outcome continuum, and the location determines the proportion of its distribution that falls in each of the five categories. The following figure contains two such distributions, one located at -0.433 and the other at 0.833.




The vertical dotted lines in this figure are located at $\upsilon_0$, $\upsilon_1$, $\upsilon_2$, and $\upsilon_3$, defining the area in each category for each distribution.

The dependent variable in this model is, of course, category assignment, and it is defined as y = 0, 1, 2, 3, or 4, depending on the category assigned being category 0, category 1, and so on. The following equations define the probability of being assigned to each category:

$$\Pr(y = 0) = \Pr(u_i < \upsilon_0) = \Phi(\upsilon_0 - \beta' \mathbf{x}_i)$$

$$\Pr(y = 1) = \Pr(\upsilon_0 < u_i < \upsilon_1) = \Phi(\upsilon_1 - \beta' \mathbf{x}_i) - \Phi(\upsilon_0 - \beta' \mathbf{x}_i)$$

$$\Pr(y = 2) = \Pr(\upsilon_1 < u_i < \upsilon_2) = \Phi(\upsilon_2 - \beta' \mathbf{x}_i) - \Phi(\upsilon_1 - \beta' \mathbf{x}_i)$$

$$\Pr(y = 3) = \Pr(\upsilon_2 < u_i < \upsilon_3) = \Phi(\upsilon_3 - \beta' \mathbf{x}_i) - \Phi(\upsilon_2 - \beta' \mathbf{x}_i)$$

$$\Pr(y = 4) = \Pr(u_i > \upsilon_3) = 1 - \Phi(\upsilon_3),$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. Again, these equations make it clear that different populations can have different probabilities associated with the categories.

LIMDEP, the statistical package discussed previously, has a procedure for estimating the parameters of the ordered probit model. To demonstrate the use of this program, 100 simulated observations were generated for each of the 24 populations, for a total of 2400 observations. A main effects model was used for the data, with the effects for the four independent variables being the following: 0.3333, 0.0, and -0.3333 for social-economic status; 0.5 and -0.5 for gender; 0.3, and -0.3 for ethnicity; and for treatment, 0.7 and -0.7. The sum of the effects for each of the 24 combinations resulted in the means for the 24 populations varying from -1.833 to 1.833.

The data matrix submitted to LIMDEP had 2400 rows and five columns. The first column was for y, with 0, 1, 2, 3, and 4 representing the categories. The independent variables were dummy coded using contrast coefficients. Socio-economic level had a single column coded for a linear effect.

With respect to model estimation, LIMDEP's output included 8 maximum likelihood estimators: a constant, four coefficients for the independent variables, and three points on the outcome continuum. The sample size and effects used guaranteed statistical significance for estimators.

For purposes of interpretation, we symbolize the estimators in the following manner: $\hat{\beta}_0$ for the constant, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ for the four independent variables, and $\hat{\upsilon}_1$, $\hat{\upsilon}_2$, and $\hat{\upsilon}_3$ for the three points on the continuum. The values for these estimators are $\hat{\beta}_0 = 1.6948$, $\hat{\beta}_1 = -0.35349$, $\hat{\beta}_2 = -1.0033$, $\hat{\beta}_3 = -0.64269$, $\hat{\beta}_4 = -1.3699$, $\hat{\upsilon}_1 = 1.3906$, $\hat{\upsilon}_2 = 1.9799$, and $\hat{\upsilon}_3 = 2.7130$. Corresponding to $\beta$, defined above, is $\hat{\beta}' = \left[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4\right]$.

Relating the estimators to the population values is straightforward for $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$, but less so for the others. The first independent variable was coded as -1, 0, 1 and multiplying these values with $\hat{\beta}_1$ results in 0.35349, 0.0, -0.35349, which are close to the socio-economic effects, 0.3333, 0.0, -0.3333. The other three independent variables were all coded -0.5, 0.5 and carrying out the multiplications with $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ we have 0.50165, -0.50165 for gender, 0.321345, - 0.321345 for ethnicity, and 0.68495, - 0.68495 for treatment, which compare favorably with the population effects 0.5, -0.5 and 0.3, -0.3 and 0.7, -0.7, respectively.

To make the correspondence between the $\upsilon_0$, $\upsilon_1$, $\upsilon_2$, and $\upsilon_3$ and $\hat{\nu}_1$, $\hat{\nu}_2$, and $\hat{\nu}_3$ we must note that LIMDEP always sets $\hat{\nu}_0 = 0$ and then finds $0 < \hat{\nu}_1 < \hat{\nu}_2 < \hat{\nu}_3$. Our population model and LIMDEP both assume a normal distribution with unit variance, but the distributions are assumed to be centered at different locations. The "center" of our population model is "0" (for zero change), because given the usual constraint that effects sum to zero, the mean of the 24 population means is zero. For LIMDEP, the "center" is estimated to be $\hat{\beta}'\bar{\mathbf{x}}$, where $\bar{\mathbf{x}}$ contains the column means of the independent variables. Therefore, the distance, $\hat{\beta}'\bar{\mathbf{x}} - \hat{\nu}_0$, must estimate $0 - \upsilon_0 = 0 - (-1.65) = 1.65$. Due to the contrast coding employed, the means of the independent variables are all zero. As a result, $\bar{\mathbf{x}} = [1,0,0,0,0]$, and consequently, $\hat{\beta}'\bar{\mathbf{x}} = \hat{\beta}_0$. This means that $\hat{\beta}'\bar{\mathbf{x}} - \hat{\nu}_0 = \hat{\beta}_0 - \hat{\nu}_0 = 1.6948 - 0 = 1.6948$, and therefore, the negative of the constant, $-\hat{\beta}_0 = -1.6948$, is the estimate of $\upsilon_0 = -1.65$, i.e., $\hat{\upsilon}_0 = -1.6948$. To adjust the other three points for the difference in location, we must compute:

$\hat{\upsilon}_1 = -\left(\hat{\beta}_0 - \bar{\nu}_1\right) = -(1.6948 - 1.3906) = -0.3042;$

$\hat{\upsilon}_2 = -\left(\hat{\beta}_0 - \bar{\nu}_2\right) = -(1.6948 - 1.9799) = 0.2851;$

$\hat{\upsilon}_3 = -\left(\hat{\beta}_0 - \bar{\nu}_3\right) = -(1.6948 - 2.7130) = 1.0182.$

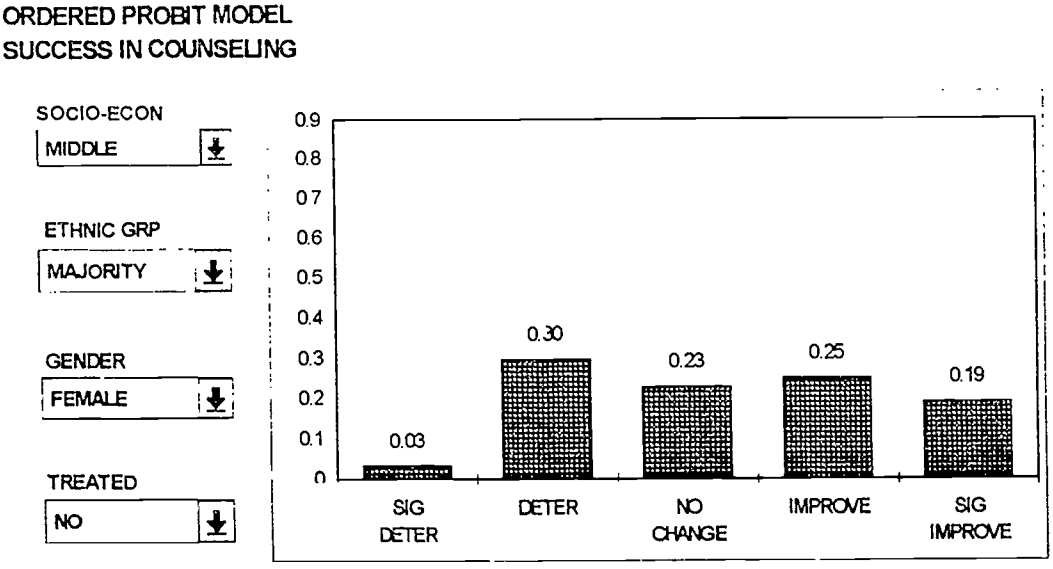
These last three values compare fairly well with the parameters, $\upsilon_1 = -0.30$, $\upsilon_2 = 0.30$, and $\upsilon_3 = 1.00$.

Had a different coding scheme been used for the independent variables, the results would have been different. For example, had the coding scheme been 0,1,2 for the first independent variable and 0,1 for the others, the constant would have equaled 3.5565, not 1.6948. The other estimates would have been virtually the same. With this alternative coding, $\hat{\beta}'\bar{\mathbf{x}} \neq \hat{\beta}_0$ and $\hat{\beta}'\bar{\mathbf{x}}$ would have to be used in finding $\hat{\upsilon}_0$, $\hat{\upsilon}_1$, $\hat{\upsilon}_2$, and $\hat{\upsilon}_3$.

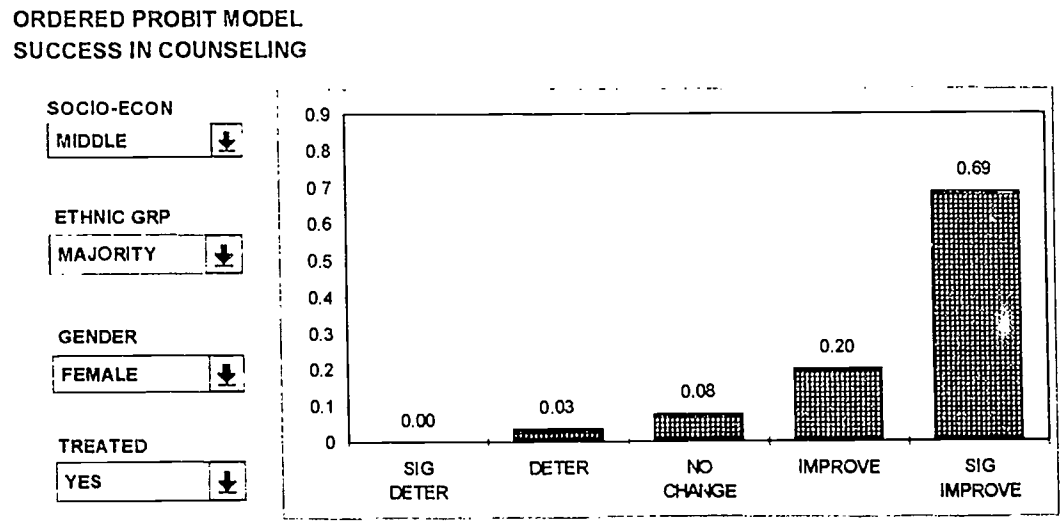
As with choice modeling above, the ordered probit model requires some effort to understand the estimates, even in a situation were the data has a very simple structure. An alternative is to focus more on the probabilities of category membership associated with

various combinations of the independent variables. To accomplish this, we turn again to a simulation model.

Model Simulation: The following is a simulation of the model analyzed in the previous section. The first graphic shows the simulated probabilities for each category for a middle class, white female who would not receive treatment for her problem.




In the next graphic, the probabilities for the same woman are displayed, assuming she has been treated.



## Concluding Remarks

This paper focused on two types of categorical dependent variables, choices and treatment outcomes. Choice behavior is an important class of human behavior. Choices about sexual behavior (abstinence, protected, and unprotected), drug usage (abstinence, legal, and illegal), birth control, type of post high school education and so on can have lasting consequences in one's life. Later in life, some choices may be viewed as "life defining," in that the choice of another alternative may have led to a substantially different life. To investigate choice behavior, the conditional logit model and analysis were introduced. The conditional logit model allows us not only to study the attributes of the people who make particular choices, but also the attributes of the choices themselves that make them more or less attractive.

The second area of focus, treatment outcome, discussed the need to consider categories of treatment outcome. As consumers, potential clients want to know how likely it is that they will get better, or better than ever, or worse than they were. Knowing the probability of various outcomes is essential information in making an informed decision to seek help. To investigate categorical outcomes, a model and analytical procedure for studying ordered categories was introduced.

For both choice behavior and outcomes, parameter estimates and significance tests were presented for examples based on simulated data. These estimates and tests are analogous to the estimates and analysis of variance common to the linear models approach that dominates counseling research today. With categorical data, however, parameter estimates and significance tests are only preliminaries to presenting the probabilities of category membership. Probabilities of category membership are best presented graphically using a spreadsheet model. In this way, changes in independent variables can be related directly to changes in probabilities. It is argued that these probabilities should be the primary focus of the investigation, for they are the results that can most directly affect actions to be taken by clients and counselors.

# References


Bock, R. Darrell. (1975). Multivariate statistical methods in behavioral research. New York: McGraw Hill.

Greene, William H. (1992). LIMDEP user's manual and reference guide. New York: Econometrics Software, Inc.

Greene, William H. (1993). Econometric analysis, 2nd ed. New York: Macmillan Publishing Co.

Johnson, N.L., & Kotz, S. (1970a). Continuous univariate distributions –1. New York: John Wiley & Sons, Inc.

Johnson, N.L., & Kotz, S. (1970b). Continuous univariate distributions –2. New York: John Wiley & Sons, Inc.

Johnson, N.L., & Kotz, S. (1972) Distributions in statistics: continuous multivariate distributions. New York: John Wiley & Sons, Inc.

Luce, R. Duncan (1959). Individual choice behavior: a theoretical analysis. New York: John Wiley & Sons, Inc.

Maddala, G.S. (1983). Limited dependent and qualitative variables in econometrics. New York: Cambridge University Press.

McFadden, Daniel (1973) Conditional logit analysis of qualitative choice behavior. In P. Zarembka (ed.), Frontiers in Econometrics. New York: Academic Press.

MathSoft, Inc. (1994). Mathcad 5.0 [computer program] Cambridge, MA.

# Appendix

As was stated in the body of this paper, the simulated data used in the choice modeling examples were obtained by generating random numbers that were distributed $N(\mu_j,1)$, $j = 1$ to 4. This was done so that the examples could be used to demonstrate the effect of violating the assumption that the utility distributions for the choice alternatives were distributed according to Type I extreme value distributions. It became clear that introducing this complexity led to making the examples more difficult to read than was thought appropriate given the expository nature of this paper. This appendix explains the relationship of the location parameters of the Type I extreme value distributions described in this paper to the means of the normal distributions that were used in generating the data for the examples.

Examples 1, 2, & 3[1]

The location parameters for the four choices in *Examples 1*, *2*, and *3* were $\xi_1 = 0.0$, $\xi_2 = 0.633$, $\xi_3 = 1.194$, and $\xi_4 = 1.694$. The relationship of these values to the normal distribution means is described in the section entitled " The Choice Model." This appendix complements that section, but focuses more on how the data were generated.

To obtain the simulated data, four random numbers were generated for each observation, each number from a different normal distribution. The means for these distributions were $\mu_1 = 0.0$, $\mu_2 = 0.4$, $\mu_3 = 0.8$, and $\mu_4 = 1.2$, and each distribution had unit variance. The four random numbers were compared, and since they were simulated utilities, the choice for that observation corresponded to the highest number, or utility.

To determine the population probabilities, three difference variables were defined for each alternative as a function of the four utilities, $u_j$, for $j = 1$ to 4. For example, for alternative one, these difference variables were defined as $\mathbf{d}_1 = \begin{bmatrix} d_{12} \\ d_{13} \\ d_{14} \end{bmatrix} = \begin{bmatrix} u_1 - u_2 \\ u_1 - u_3 \\ u_1 - u_4 \end{bmatrix}$, with density $f(\mathbf{d}_1) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_1|}} \exp\left[ -\frac{1}{2} (\mathbf{d}_1 - \mu_1)' \Sigma_1 (\mathbf{d}_1 - \mu_1) \right]$, where $\mu_1 = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \mu_1 - \mu_4 \end{bmatrix}$ and $\Sigma_1 = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_3^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma_4^2 \end{bmatrix}$. With similar sets of difference variables defined for

[1] Mathcad Plus 5.0 was used to obtain the results in this and the following sections.

the other three alternatives, the choice probabilities were found by integrating the following four trivariate normal integrals:

$$P_1 = \int_S f(\mathbf{d}_1)d\mathbf{d}_1 \text{, with } S = [d_{12}, d_{13}, d_{14} \mid 0 \le d_{12} \le \infty,\ 0 \le d_{13} \le \infty,\ 0 \le d_{14} \le \infty]$$

$$P_2 = \int_S f(\mathbf{d}_2)d\mathbf{d}_2 \text{, with } S = [d_{21}, d_{23}, d_{24} \mid 0 \le d_{21} \le \infty,\ 0 \le d_{23} \le \infty,\ 0 \le d_{24} \le \infty]$$

$$P_3 = \int_S f(\mathbf{d}_3)d\mathbf{d}_3 \text{, with } S = [d_{31}, d_{32}, d_{34} \mid 0 \le d_{31} \le \infty,\ 0 \le d_{32} \le \infty,\ 0 \le d_{34} \le \infty]$$

$$P_4 = \int_S f(\mathbf{d}_4)d\mathbf{d}_4 \text{, with } S = [d_{41}, d_{42}, d_{43} \mid 0 \le d_{41} \le \infty,\ 0 \le d_{42} \le \infty,\ 0 \le d_{43} \le \infty].$$

Given $\mu_1 = 0.0$, $\mu_2 = 0.4$, $\mu_3 = 0.8$, and $\mu_4 = 1.2$, $P_1 = 0.086$, $P_2 = 0.162$, $P_3 = 0.284$, and $P_4 = 0.468$. To find the Type I extreme value distribution location parameters that would yield these probabilities, the following set of nonlinear equations were solved for $\xi_2$, $\xi_3$, and $\xi_4$, with $\xi_1 = \mu_1 = 0$:

$$\frac{e^0}{e^0 + e^{\xi_2} + e^{\xi_3} + e^{\xi_4}} = 0.086$$

$$\frac{e^{\xi_2}}{e^0 + e^{\xi_2} + e^{\xi_3} + e^{\xi_4}} = 0.162$$

$$\frac{e^{\xi_3}}{e^0 + e^{\xi_2} + e^{\xi_3} + e^{\xi_4}} = 0.284$$

$$\frac{e^{\xi_4}}{e^0 + e^{\xi_2} + e^{\xi_3} + e^{\xi_4}} = 0.468$$

The solution to these equations is $\xi_2 = 0.633$, $\xi_3 = 1.194$, and $\xi_4 = 1.694$. Since the normal distributions used had unit variance and the Type I extreme value distribution location parameters are for distributions with a variance of $\pi^2/6$, when comparing the parameters of the two types of distributions, the above location parameters should be divided by $\sqrt{\pi^2/6} = 1.283$. This results in values of 0.0, 0.494, 0.931, and 1.321.

Example 4

Using the same set of random numbers used in *Examples 2* and *3*, each number (utility) had the random variable, $-0.005(t_q - 50)$, added to it to simulate a time effect. Values of $t_q$, q = 1 to 9, were randomly drawn from the set: [10, 20, 30, 40, 50, 60, 70 ,80, 90].

The first step in finding a corresponding weight for the Type I extreme value distributions was to compute multivariate normal probabilities that reflected the time effect. Accordingly, four population means were defined as follows: $\mu_1 - 0.005(90-50)$, $\mu_2 - 0.005(60-50)$, $\mu_3 - 0.005(30-50)$, and $\mu_4 - 0.005(10-50)$, with the $\mu_j$ set to the values used above. Using these four means, trivariate integrals were evaluated to find the conditional probabilities for these times. When the means of the utility distributions for the four alternatives were changed to reflect the effect of 90, 60, 30, and 10 minutes, respectively, the resulting probabilities for the four alternatives were 0.052, 0.131, 0.287, and 0.529. In general, for *a given set of probabilities*, there is not a linear relationship between the means of the normal distributions and the location parameters for the Type I extreme value distributions. Therefore, even though the time effect is a linear function of time with normally distributed utilities, it would, in general, not be a linear function of time with utilities distributed according to Type I extreme value distributions. This latter statement assumes that the probabilities are the same for the two probability models. That being the case, a linear time effect for utilities distributed according to the Type I extreme value distribution would only approximate the slope parameter for normally distributed utilities. For this reason Mathcad's "minerr()" function was used to find the best fitting value as an approximate solution for the following equations:

$$\frac{e^{40\beta}}{e^{40\beta} + e^{\xi_2+10\beta} + e^{\xi_3-20\beta} + e^{\xi_4-40\beta}} = 0.052$$

$$\frac{e^{\xi_2+10\beta}}{e^{40\beta} + e^{\xi_2+10\beta} + e^{\xi_3-20\beta} + e^{\xi_4-40\beta}} = 0.131$$

$$\frac{e^{\xi_3-20\beta}}{e^{40\beta} + e^{\xi_2+10\beta} + e^{\xi_3-20\beta} + e^{\xi_4-40\beta}} = 0.287$$

$$\frac{e^{\xi_4-40\beta}}{e^{40\beta} + e^{\xi_2+10\beta} + e^{\xi_3-20\beta} + e^{\xi_4-40\beta}} = 0.529$$

With $\xi_2 = 0.633$, $\xi_3 = 1.194$, and $\xi_4 = 1.694$, the approximate value found was $\beta = -0.00672079$. Rescaling this value by dividing by $\sqrt{\pi^2/6}$ results in -0.00524018, which is close to -0.005, the weight actually used.

Example 5

Starting with the same data as used in *Example 4*, a gender effect was added in *Example 5*. For females, which were taken to be the first 500 observations in the data set, the values 0.15, 0.50, 0.15, and -0.15 were added, respectively, to the four utilities of each observation. For males, which were taken to be the remaining 500 observations in the data set, the values -0.15, -0.50, -0.15, and 0.15 were added to the utilities. With the time effect held constant at zero by setting Time equal to 50 for all alternatives, the means of the conditional utility distributions for females became 0.15, 0.90, 0.95, and 1.05. With these means, the probabilities for the four choices were 0.090, 0.277, 0.296, and 0.337. Using these probabilities and setting the fourth location parameter to zero, the following location parameters result, -1.32027, -0.19607, -0.12972, and 0.0. To remove the effect of the helper categories, the location parameters for those categories must be subtracted. Since the values reported above had $\xi_1 = 0$, however, those values must be relocated before subtraction, so that the last parameter is zero. This is easily accomplished by subtracting 1.69412 from $\xi_1$, $\xi_2$, $\xi_3$, and $\xi_4$. These relocated parameters are then subtracted from the females' location parameters and the differences are the changes in location for females reported in Example 5, namely, 0.3738, 0.8648, 0.3698, and 0.0. To compare these values to their counterparts for normal distributions. they must rescaled by dividing by $\sqrt{\pi^2/6}$ and then relocated so that the first location parameter is equal to 0.15. After these operations, the values are 0.15, 0.533, 0.147, and -0.141. These are reasonably close to the values used in generating the data, i.e., 0.15, 0.50, 0.15, and -0.15.

For males, when the values -0.15, -0.50, -0.15, and 0.15 were added to the utilities, again with the time effect held constant, the means of the conditional utility distributions for males became -0.15, -0.10, 0.65, and 1.35. With these means, the probabilities for the four choices were 0.075, 0.082, 0.252, and 0.591. Using these probabilities and setting the fourth location parameter to zero, the following location parameters result, -2.06433, -1.9751, -0.85239, and 0.0. Adjusting these parameters in the same way as for females results in the changes in location for males reported in Example 5, namely, -0.3702, -0.9142, -0.3529, and 0.0. To compare these values to their counterparts for normal distributions, they must be rescaled by dividing by $\sqrt{\pi^2/6}$ and then relocated so that the first location parameter is equal to -0.15. After these operations, the values are -0.15, -0.574, -0.136, and 0.139, and these are fairly close to the values used in generating the data, i.e., -0.15, -0.50, -0.15, and 0.15.

For the above examples, it is clear that even though one assumes the wrong distribution for the utilities, there is a set of "wrong" parameters for that "wrong" distribution that lead to the same ( or virtually the same) probabilities as would be obtained using the "right" set of parameters with the "right" distribution. While the "wrong" and "right" sets of parameters must necessarily be different, their difference is not great. We would most often be led to the same conclusions regarding the effects of the independent variables no matter which set of parameters we used.