CHAPTER 1

TYPICAL STATED CHOICE EXPERIMENTS

COPYRIGHTED MATERIAL

People make choices all the time; some of these decisions are of interest to governments and businesses. Governments might want to model demand for health services in the future or to assess the likely electoral impact of a decision to allow logging in a national park. Businesses want to predict the likely market for new goods and services.

Information about choices can be captured from sources like supermarket scanners. However, this "revealed preference" data tells you nothing about products that do not yet exist. Here an experimental approach can help. Such experiments are called "stated preference" or "stated choice" experiments. This book describes the best way to design generic stated preference choice experiments, from a mathematical perspective.

Stated choice experiments are widely used in business although often not published. According to the results of a survey sent out to businesses, there were about 1000 commercial applications in the United States in the 1970s and there were about 400 per year in the early 1980s (Cattin and Wittink (1982); Wittink and Cattin (1989)). Wittink et al. (1994) found less extensive use in Europe in the period 1986–1991, but Hartmann and Sattler (2002) have found that the number of companies using stated choice experiments and the number of experiments conducted had more than doubled in German-speaking Europe by 2001. The range of application areas has also increased and now includes transport (Hensher (1994)), health economics (Bryan and Dolan (2004)), and environmental evaluation (Hanley et al. (2001)), among others.

In the rest of this chapter we will define some concepts that we will use throughout the book. We will use published choice experiments from various application areas to illustrate these concepts. These examples will also illustrate the range of issues that needs to be

The Construction of Optimal Stated Choice Experiments. By D. J. Street and L. Burgess

Copyright © 2007 John Wiley & Sons, Inc.

addressed when designing such an experiment. The mathematical and statistical issues raised will be considered in the remainder of this book.

1.1 DEFINITIONS

Stated choice experiments are easy to describe. A *stated choice experiment* consists of a set of choice sets. Each choice set consists of two or more options (or alternatives). Each respondent (also called subject) is shown each choice set in turn and asked to choose one of the options presented in the choice set. The number of options in a choice set is called the *choice set size*. A stated choice or stated preference choice experiment is often called a *discrete choice experiment* and the abbreviations *SP experiment* and *DCE* are very common. We will look at the design of choice experiments for the simplest stated preference situation in this book — the so-called *generic* stated preference choice experiment. In such an experiment, all options in each choice set are described by the same set of attributes, and each of these attributes can take one level from a set of possible levels.

An example involving one choice set might ask members of a group of employees how they will travel to work tomorrow. The five options are {drive, catch a bus, walk, cycle, other} and these five options comprise the choice set. Each respondent will then choose one of these five methods of getting to work.

This simple example illustrates the fact that in many choice experiments people are forced to choose one of the options presented. We call such an experiment a *forced choice* experiment. In this case, being compelled to choose is easy since the respondents were employees (so would be going to work) and every possible way of getting to work was in the choice set since there was an option "other". Thus the list of options presented was *exhaustive*.

Sometimes a forced choice experiment is used even though the list of options presented is not exhaustive. This is done to try to find out how respondents "trade-off" the different characteristics of the options presented. A simple example is to offer a cheap flight with restrictive check-in times or a more expensive flight where there are fewer restrictions on check-in times. In reality, there might be intermediate choices, but these are not offered in the choice set.

However, there are certainly situations where it simply does not make sense to force people to choose. People may well spend several weeks deciding which car to buy and will defer choice on the first few cars that they consider. To allow for this situation, choice experiments include an option variously called "no choice" or "delay choice" or "none of these". We will just talk about having a *none option* to cover all of these situations.

A related situation arises when there is an option which needs to appear in every choice set. This can happen when respondents are being asked to compare a new treatment with an existing, standard treatment for a medical condition, for instance. We speak then of all choice sets having a *common base option*.

Sometimes just one option is described to respondents who are then asked whether or not they would be prepared to use that good or service. Usually several descriptions are shown to each respondent in turn. This is called a *binary response* experiment, and in many ways it is the simplest choice experiment of all. It does not allow for the investigation of trade-offs between levels of different attributes, but it gives an indication of combinations of levels that would be acceptable to respondents.

When constructing choice sets, it is often best to avoid choice sets where one option is going to be preferred by every respondent. In the flight example above, there will be respondents who prefer to save money and so put up with the restrictive check-in conditions, and there will be respondents for whom a more relaxed attitude to the check-in time will be very important. It seems obvious, though, that a cheap flight with the relaxed check-in conditions would be preferred by all respondents. An option which is preferred by all respondents is called a *dominating* or *dominant option*, and it is often important to be able to design choice experiments where it is less likely that there are choice sets in which any option dominates (or where there is an option that is dominated by all others in the choice set). There is a discussion about dominating options with some discussion of earlier work in Huber and Zwerina (1996).

When constructing options which are described by two or more attributes, it can be necessary to avoid unrealistic combinations of attribute levels. For example, when describing a health state and asking respondents whether they think they would want a hip replacement if they were in this health state, it would be unrealistic to describe a state in which the person had constant pain but could easily walk 5 kilometers.

Throughout this book, we will only consider situations where the options in the experiment can be described by several different *attributes*. Each attribute has two or more *levels*. For the flight example we have been describing, the options have two attributes, the cost and the check-in conditions. In general, attributes need to have levels that are plausible and that are varied over a relevant range. For example, health insurance plans can be described by maximum cost to the subscriber per hospital stay, whether or not visits to the dentist are covered, whether or not visits to the physiotherapist are covered, and so on. Although attributes like cost are continuous, in the choice experiment setting we choose a few different costs and use these as discrete levels for the attribute. Thus we do not consider continuous attributes in this book.

Finally, we stress that we will be talking about *generic* stated preference choice experiments throughout this book. We do not consider the construction of optimal designs when the options are labeled, perhaps by brand or perhaps by type of transport, say, and hence the attributes, and the levels, depend on the label. We only consider designs that are analyzed using the MNL model.

1.2 BINARY RESPONSE EXPERIMENTS

As we said above, in binary response experiments the respondents are shown a description of a good or service, and they are asked whether they would be interested in buying or using that good or service. For each option they are shown, they answer "yes" or "no".

One published example of a binary response experiment appears in Gerard et al. (2003). This study was carried out to develop strategies that were likely to increase the participation rates in breast screening programs. The goal of a breast screening program is to achieve a target participation rate across the relevant population since then there should be a reduction in breast cancer mortality across that population. To get this participation rate requires that women participate at the recommended screening rate. The aim of the study described in Gerard et al. (2003) was to "identify attributes of service delivery that eligible screenees value most and over which decision makers have control". The attributes and levels used in the study are given in Table 1.1.

Given these attributes and levels, what does a respondent actually see? The respondent sees a number of options, like the one in Table 1.2, and just has to answer the question. In this particular survey each respondent saw 16 options (invitations) and so answered the question about each of these 16 different possible invitations in turn.

Table 1.1 Attributes and Levels for the Survey to Enhance Breast Screening Participation

Attributes	Attribute Levels
Method of inviting women for screening	Personal reminder letter Personal reminder letter and recommendation by your GP Media campaign Recommendation from family/friends
Information included with invitation	No information sheet Sheet about the procedure, benefits and risks of breast screening
Time to wait for an appointment	1 week 4 weeks
Choice of appointment times	Usual office hours Usual office hours, one evening per week Saturday morning
Time spent traveling	Not more than 20 minutes Between 20 and 40 minutes Between 40 and 60 minutes Between 1 and 2 hours
How staff at the service relate to you	Welcoming manner Reserved manner
Attention paid to privacy	Private changing area Open changing area
Time spent waiting for mammogram	20 minutes 30 minutes 40 minutes 50 minutes
Time to notification of results	8 working days 10 working days 12 working days 14 working days
Level of accuracy of the screening test	70% 80% 90% 100%

Table 1.2 One Option from a Survey about Breast Screening Participation

	Screening Service
How are you informed	Personal reminder letter
Information provided with invitation	No information sheet
Wait for an appointment	4 weeks
Appointment choices	Usual office hours and one evening per week
Time spent traveling (one way)	Not more than 20 minutes
How staff relate to you	Reserved manner
Changing area	Private changing area
Time spent having screen	40 minutes
Time waiting for results	10 working days
Accuracy of the results	90%

Imagine that your next invitation to be screened is approaching.
Would you choose to attend the screening service described above?
(tick one only)
Yes ◯ No ◯

The statistical question here is: "which options (that is, combinations of attribute levels) should we be showing to respondents so that we can decide which of these attributes, if any, is important, and whether there are any pairs of attribute levels which jointly influence the decision to participate in the breast screening service?" The design of such informative binary response designs is described in Section 7.1.2.

1.3 FORCED CHOICE EXPERIMENTS

In a forced choice experiment, each respondent is shown a number of choice sets in turn and asked to choose the best option from each choice set. There is no opportunity to avoid making a choice in each choice set.

Severin (2000) investigated which attributes made take-out pizza outlets more attractive. In her first experiment, she used the six attributes in Table 1.3 with the levels indicated. A sample choice set for an experiment looking at these six attributes describing take-out pizza outlets is given in Table 1.4. There are three things to observe here.

The first is that all the attributes have two levels; an attribute with only two levels is called a *binary* attribute, and it is often easier to design small, but informative, experiments when all the attributes are binary. We focus on designs for binary attributes in Chapters 4 and 5.

The second is that the question has been phrased so that the respondents are asked to imagine that the two choices presented to them are the last two options that they are considering in their search for a take-out pizza outlet. This assumption means that the respondents are naturally in a setting where it does not make sense not to choose an option, and so they are forced to make a selection even though the options presented are not exhaustive.

Finally, observe that each respondent has been shown only two options and has been asked to state which one is preferred. While it is very common to present only two options in each choice set, it is not necessarily the best choice set size to use; see Section 7.2 for a discussion about the statistically optimal choice set size. Larger choice sets do place more cognitive demands on respondents, and this is discussed in Iyengar and Lepper (2000), Schwartz et al. (2002), and Iyengar et al. (2004).

Table 1.3 Six Attributes to be Used in an Experiment to Compare Pizza Outlets

Attributes	Attribute Levels
Pizza type	Traditional Gourmet
Type of Crust	Thick Thin
Ingredients	All fresh Some tinned
Size	Small only Three sizes
Price	\$17 \$13
Delivery time	30 minutes 45 minutes

Table 1.4 One Choice Set in an Experiment to Compare Pizza Outlets

	Outlet A	Outlet B
Pizza type	Traditional	Gourmet
Type of crust	Thick	Thin
Ingredients	All fresh	Some tinned
Size	Small only	Small only
Price	\$17	\$13
Delivery time	30 minutes	30 minutes

Suppose that you have already narrowed down your choice of take-out pizza outlet to the two alternatives above. Which of these two would you choose?
(tick one only)

Outlet A ◯ Outlet B ◯

Most forced choice experiments do not use only binary attributes. Chapter 6 deals with the construction of forced choice experiments for attributes with any number of levels. For example, Maddala et al. (2002) used 6 attributes with 3, 4, 5, 3, 5, and 2 levels, respectively, in a choice experiment examining preferences for HIV testing methods. The attributes, together with the attribute levels, are given in Table 1.5, and one choice set from the study is given in Table 1.6. Each respondent was presented with 11 choice sets and for each of these was asked to choose one of two options. As the respondents were all surveyed at HIV testing locations a forced choice experiment was appropriate.

Table 1.5 Attributes and Levels for the Study Examining Preferences for HIV Testing Methods

Attribute	Attribute Levels
Location	Doctor's office Public clinic Home
Price	$0 $10 $50 $100
Sample collection	Draw blood Swab mouth Urine sample Prick finger
Timeliness/accuracy	Results in 1-2 weeks; almost always accurate Immediate results; almost always accurate Immediate results; less accurate
Privacy/anonymity	Only you know; not linked Phones; not linked In person; not linked Phone; linked In person; linked
Counseling	Talk to a counselor Read brochure then talk to counselor

Both of the experiments discussed above used only six attributes. Hartmann and Sattler (2002) found that about 75% of commercially conducted stated choice experiments used 6 or fewer attributes and they speculated that this might be because commonly available software often used to generate choice experiments would not allow more than 6 attributes. However, there are choice experiments where many attributes are used; see Section 1.7.2.

1.4 THE "NONE" OPTION

As we said above, sometimes it does not make sense to compel people to choose one of the options in a choice set and so some choice experiments include an option variously called "no choice" or "delay choice" or "none of these" in each choice set.

Often an existing forced choice experiment can be easily modified to include an option not to choose. For instance, in the pizza outlet experiment described in the previous section,

Table 1.6 One Choice Set from the Study Examining Preferences for HIV Testing Methods

Attribute	Option A	Option B
Location	Doctor's office	Public clinic
Price	$100	$10
Sample collection	Swab mouth	Urine sample
Timeliness/accuracy	Results in 1–2 weeks; almost always accurate	Immediate results; less accurate
Privacy/anonymity	In person; not linked	Only you know; not linked
Counselling	Talk to a counselor	Read brochure then talk to counselor

Which of these two testing methods would you choose? *(tick one only)*
Option A ◯ Option B ◯

we could change the question to ask: "Suppose that you have decided to get a take-out meal. Which of these pizza outlets would you select, or would you go somewhere else?"

Dhar (1997) looks at the situations in which consumers find it hard to choose and so will opt to defer choice if they can. Haaijer et al. (2001) summarize his results by saying "respondents may choose the no-choice when none of the alternatives appears to be attractive, or when the decision-maker expects to find better alternatives by continuing to search. ...adding an attractive alternative to an already attractive choice set increases the preference of the no-choice option and adding an unattractive alternative to the choice set decreases the preference of the no-choice." In Section 7.1.1, we discuss how to design good designs when there is a "none of these" option in each choice set.

1.5 A COMMON BASE OPTION

Some choice experiments have a common (or base) option, sometimes called constant comparator, in each choice set, together with one or more other options. This is often done so that the current situation can be compared to other possibilities. A typical example arises in medicine when the standard treatment option can be compared to a number of possible alternative treatment options.

Ryan and Hughes (1997) questioned women about various possible alternatives to the surgical removal of the product of conception after a miscarriage (note that some such treatment is essential after a miscarriage). They identified the attributes and levels given in Table 1.7 as being appropriate.

In the choice experiment, the common base, which is the current treatment option of surgical management, was described as "having a low level of pain, requiring 1 day and 0 nights in hospital, taking 3–4 days to return to normal activities, costing $350, and there would be complications post-surgery". This raises the question of whether it is sensible to have an option in which it is known with certainty that there will be complications; it might have made more sense to talk of the probability of complications given a particular treatment. This interesting non-mathematical issue will not be addressed in this book.

Table 1.7 Five Attributes to be Used in an Experiment to Investigate Miscarriage Management Preferences

Attributes	Attribute Levels
Level of pain	Low Moderate Severe
Time in hospital	1 day and 0 nights 2 days and 1 night 3 days and 2 nights 4 days and 3 nights
Time taken to return to normal activities	1–2 days 3–4 days 5–6 days More than 7 days
Cost to you of treatment	\$100 \$200 \$350 \$500
Complications following treatment	Yes No

Having a common option in all choice sets is not as good as allowing all the options to be different from one choice set to another but when a common base is appropriate we show how to design as well as possible for this setting in Section 7.1.3.

1.6 AVOIDING PARTICULAR LEVEL COMBINATIONS

Sometimes a set of level combinations of at least two of the attributes is unrealistic and sometimes a set of level combinations is clearly the best for all respondents and so will always be chosen.

We discuss examples of each of these situations here. We give some ideas for how to design choice experiments when these circumstances pertain in Chapters 4, 5, and 8.

1.6.1 Unrealistic Treatment Combinations

To illustrate this idea consider the descriptions of 5 attributes describing health states devised by EuroQol; see EuroQoL (2006). These attributes and levels are given in Table 1.8 and are used to describe health states for various purposes. In the context of a stated preference choice experiment we might describe two health states and ask respondents which one they prefer.

But even a quick look at the levels shows that some combinations of attribute levels do not make sense. A health state in which a person is "Confined to bed" in the mobility attribute is not going to be able to be linked with "No problems with self-care" in the self-care attribute. Thus it is necessary to determine the level combinations that are *unrealistic* before using choice sets that are generated for these attributes using the techniques developed in later chapters.

Table 1.8 Five Attributes Used to Compare Aspects of Quality of Life

Attributes	**Attribute Levels**
Mobility	No problems in walking about Some problems in walking about Confined to bed
Self-Care	No problems with self-care Some problems with self-care Unable to wash or dress one's self
Usual Activities (e.g., work, study housework)	No problems with performing one's usual activities Some problems with performing one's usual activities Unable to perform one's usual activities
Pain/Discomfort	No pain or discomfort Moderate pain or discomfort Extreme pain or discomfort
Anxiety/Depression	Not anxious or depressed Moderately anxious or depressed Extremely anxious or depressed

1.6.2 Dominating Options

Many attributes can have ordered levels in the sense that all respondents agree on the same ordering of the levels for the attribute. In the levels presented in Table 1.8 it is clear that in every attribute the levels go from the best to the worst. Thus a choice set that asks people to choose between the health state (No problems in walking about, No problems with self-care, No problems with performing one's usual activities, No pain or discomfort, Not anxious or depressed) and any other health state is not going to give any information - all respondents will choose the first health state. We say that the first health state *dominates* the other possible health states.

It is possible to have a choice set in which one option dominates the other options in the choice set even though the option is not one that dominates all others in the complete set of level combinations. So if we use 0, 1, and 2 to represent the levels for each of the attributes in Table 1.8 then the best health state overall is 00000. But a choice set that contains (00111, 01222, 12211), for example, has a *dominating option* since 00111 has at least as good a level on every attribute as the other two options in the choice set. It is not clear, however, whether 01222 or 12211 would be preferred since it is not necessarily true that the same utility values apply to the same levels of different attributes.

We discuss ways of avoiding choice sets with dominating options in Chapters 4, 5, and 8.

1.7 OTHER ISSUES

In this chapter we have discussed a number of different types of choice experiments that have been published in the literature. We have indicated that, in later chapters, we will describe how to construct optimal designs for binary responses, for forced choice stated preference experiments, for stated choice experiments where a "none" option is included in every choice set, and for stated choice experiments where a common base option is included in every choice set. In all cases we consider only generic options analyzed using the MNL model. In this section we want to mention a couple of designs that we will not be constructing and to discuss briefly some non-mathematical issues that need to be considered when designing choice experiments.

1.7.1 Other Designs

In some experiments, options are described not only by attribute levels but also by a brand name or label; for example, the name of the airline that is providing the flight. Such experiments are said to have *branded* or *labeled* options (or alternatives), and these have *alternative specific* attributes.

In many situations, people choose from the options that are available at the time they make their choice rather than deferring choosing until some other option is available. Experimentally, we can model this by having a two-stage design process. We have a design which says what options are available and another design that determines the specific options to present given what is to be available. Such designs are termed *availability designs*.

We will not be discussing the construction of designs for branded alternatives or for availability experiments in this book. The interested reader is referred to Louviere et al. (2000).

1.7.2 Non-mathematical Issues for Stated Preference Choice Experiments

The first issue, and one that we have alluded to in Sections 1.3 and 1.4, is the question of task complexity and thus of respondent efficiency. If a task is too complicated (perhaps because there are too many attributes being used to describe the options in a choice set or because there are too many options in each choice set), then the results from a choice experiment are likely to be more variable than expected. Aspects of this problem have been investigated by various authors, and several relevant references are discussed in Iyengar and Lepper (2000). Louviere et al. (2007) found that completion rates are high even for what would be considered large choice experiments in terms of the number of attributes and the number of options.

A related issue concerns the number of choice sets that respondents can reasonably be expected to complete. If there are too many choice sets, then respondents may well become tired and give more variable results over the course of the experiment. This situation has been investigated by Brazell and Louviere (1995) and Holling et al. (1998).

The discrete choice experiment task needs to be thought about in context. Are the choices one time only or are they repeated? How important is the outcome of the decision? (In a medical setting you could be asking people to think about life-or-death decisions.) How familiar is the context? The choice of a health insurance provider may be familiar while other choices, such as for liver transplant services, may require the provision of detailed information in the experiment so that respondents can make an informed choice.

The attributes that are used to describe the options in a choice experiment need to be appropriate and plausible, and the combinations that are presented in the experiment need to be realistic; otherwise, respondents may take the task less seriously or be confused by it.

The mechanics of setting up a stated preference experiment are outlined in Chapter 9 of Louviere et al. (2000). Dillman and Bowker (2001) discuss many aspects of both mail and internet surveys.

There are various sources of bias that have been identified as sometimes occurring in choice experiments. One is what has been termed *affirmation bias* when respondents choose responses to be consistent with what they feel the study objectives are. This is why experimenters sometimes include attributes that are not of immediate interest to mask the main attributes that are under investigation.

A second possible source of bias is called *rationalization bias*, where responses are given that justify the actual behavior. This serves to reduce cognitive dissonance for the respondents. A third possible source of bias results from the fact that there are no transaction costs associated with choices in a stated preference study. Some respondents try to respond in a way that they believe will influence the chance of, or the magnitude of, changes in the real world. This is termed *strategic* or *policy response bias*. Finally, people may not be prepared to indicate preferences which they feel are socially unacceptable or politically incorrect. These terms are defined and used in Walker et al. (2002).

Other sources of bias can be related to the actual topic under investigation. Carlsson (2003) was investigating business passengers preferences, and some of his options were more environmentally friendly, but more expensive, than other options, and he spoke of respondents perhaps aiming to have a "warm glow" or "purchasing moral satisfaction" when making choices.

Severin (2000) has shown, in a paired comparison experiment, that if there are a large number of attributes describing each option, then respondents find the task more difficult and she has suggested that about 8 or 9 attributes seem to be effectively processed; see her thesis for more details. For a discussion of the role of cognitive complexity in the design of choice experiments, readers are referred to DeShazo and Fermo (2002) and Arentze et al. (2003) and references cited therein.

These and other psychological and cognitive issues are beyond the scope of this book, and we refer the reader interested in such issues elsewhere. The papers by Iyengar and Lepper (2000) and Schwartz et al. (2002) and the references cited therein provide a good starting point to find out more about these issues.

1.7.3 Published Studies

As we have said before, stated preference choice experiments are used in many areas. Here we give references to a few published studies, together with a very brief indication of the question being investigated and the type of design being used.

Chakraborty et al. (1994) describe a choice experiment to investigate how consumers choose health insurance. As well as the actual company offering the insurance, 23 other attributes were used to describe health insurance plans. Respondents were presented with choice sets with 4 options in each and asked to indicate their preferred plan.

Hanley et al. (2001) describe a stated preference study to investigate demand for climbing in Scotland. Each choice set contained two possible climbs and a "neither of these" option. They also give a table with details of about 10 other studies that used DCEs to investigate questions in environmental evaluation.

Kemperman et al. (2000) used an availability design to decide which of four types of theme parks would be available to respondents in each of spring and summer. Each choice set contained four theme parks and a "none of these" option.

McKenzie et al. (2001) describe a study in which five attributes that are commonly used to describe the symptoms of asthma were included and three levels for each of these chosen. Patients with moderate to severe asthma were shown "pairs of scenarios characterized by different combinations of asthma symptoms", and were asked which of the two scenarios they thought would be better to have or whether they felt there was no difference.

Ryan and Gerard (2003) and Bryan and Dolan (2004) give examples of the use of DCEs in the health economics context.

Scarpa et al. (2004) discuss experiments to "characterise the preference for fifteen different attributes related to water provision".

An example of a choice experiment that involved labeled options is given in Tayyaran et al. (2003). The authors were interested in investigating whether telecommuting and intelligent transportation systems had an impact on residential location. Each choice set contained three residential options as well as a "none of these" option. The residential options were "branded" as *central cities*, *first-tier satellite nodes*, and *second-tier satellite nodes*. Each residential option was described by 7 attributes, 4 of which were the same for all of the locations.

Walker et al. (2002) used two stated choice experiments to model tenants' choices in the public rental market.

1.8 CONCLUDING REMARKS

In this chapter we have seen that there are a number of areas where stated choice experiments have been applied and that there are a number of issues, both mathematical and non-mathematical, which need to be considered in the construction of the best possible experiments for a given situation. In the remainder of this book we look at the mathematical issues that need to be considered to design a good generic stated choice experiment to be analyzed using the MNL model. In the next chapter, we collect a number of results about factorial designs which are intimately connected with the representation of options by attributes. We follow this with a discussion about parameter estimators and their distribution. Over the following three chapters we show how to get the best designs for any number of attributes, each attribute being allowed to have any number of levels and with choice sets of any size. In the penultimate chapter we consider how to construct good designs for other situations, such as the inclusion of a "none" option in every choice set. The final chapter illustrates the application of the results in the book to the construction of a number of experiments that we have designed in the last five years.