GUY LORIES

CHAPTER 2

THE ROLE OF STATISTICAL AND FORMAL TECHNIQUES IN EXPERIMENTAL PSYCHOLOGY

1. INTRODUCTION

Statistical models play a large but sometimes unclear role in experimental psychology. The recommended practice for the analysis of experimental results has been made very explicit but objections to it have been made repeatedly over twenty or thirty years without bringing about significant changes in this standard strategy. We will describe what we think what the practice actually achieves.

To make our point we will first describe how statistical techniques are used in experimental psychology. We will examine various problems and objections that have repeatedly been raised against this practice. These objections appear sound in and by themselves and they have been raised by authorized voices but they have been repeatedly ignored. We will suggest this misunderstanding persists because statistical models are used to gain a measure of experimental control and are specifically not used for measurement or for theory building. In other words, our point is that statistical techniques are often used in the service of experimental control only, and not to model the phenomenon of interest. To conclude we will say a word of a few modelling approaches and the problems they face.

2. EXPERIMENTAL "PARADIGMS" AND AVERAGING ACROSS SUBJECTS

Experimental psychology is an empirical endeavour with all the paraphernalia of modern science. It is quantitative, it has produced laws (e.g. the law of practice, Fitts law etc.) and it has a number of received (consensual) experimental and methodological strategies. Statistics play an important role in these strategies, in the training of professionals, in scientific practice and in the publication process. This role, however has come under attacks and it has been suggested repeatedly that the way we analyze our data is responsible for a lack of cumulative progress (Cohen 1994; Schmidt, 1996, 1992; Bakan 1966; Morrison and Henkel 1970; Rozeboom 1960; Hogben1957; Lykken 1968; Meehl 1967, 1978, 1986, 1990a, 1990b; Oakes 1986; Gigerenzer 1993; Pollard 1993; Tukey 1969, 1991; Thompson 1992). We only cite these works to demonstrate that considerable opposition has taken place.

Cohen (1994) provides a reasonably short review and points to most of these references. Papers published along these lines are usually countered by

R. Franck (ed.), The Explanatory Power of Models, 31–49.

reaffirmations of the standard position (e.g. Hagen 1997, Wainer 1999) with unclear conclusions to the argument, but a large number of unsolved "technical" problems continue to plague research (see Judd, McClelland and Culhane 1995 for a recent review of these). It has also been suggested that this state of affairs could only follow from deep incomprehension of several statistical concepts among practitioners and textbook authors (Dawes 1988; Gigerenzer 1993; Pollard 1993; Greenwald 1993). Some authors provide empirical evidence that misunderstandings are indeed common among academic practitioners (Oakes 1986, Kahneman and Tversky 1993, Pollard 1993, Zuckerman, Hodgins, Zuckerman and Rosenthal 1993). Although a debate is taking place within the American Psychological Association (Wainer 1999), research practice has not been clearly modified yet[1]. For instance repeated calls to give more attention to the problem of power in significance testing have apparently not been followed (Sedlmeier and Gigerenzer 1989). We cannot, of course, ambition to bring such a debate to its end. We will only try to pinpoint what we believe is the reason why the standard methodology keeps being used in experimental psychology, what it does achieve for the discipline. This "standard approach" constitutes a specific mode of relating theory to data that we feel is of interest in the context of this volume.

The most important difficulty with psychology experiments is probably that some factors can actually never be controlled experimentally (the subject personal history, for instance). More generally, one cannot make the subjects identical at the beginning of the experiment[2].

The solution chosen to solve these problems was to introduce the "standard" statistical methodology. Data are aggregated across subjects and the differences between subjects are treated as error variance[3]. The object of interest is the expected value of the dependent variable(s) in the various experimental conditions. A well-controlled experimental situation ensures that the conditions required by this statistical approach are met.

This control problem is in turn taken care of by a methodology designed to tighten experimental control. Experimental psychology has developed the notion of an experimental paradigm: a narrowly defined experimental situation used to investigate the effect of varying the level of one or a few independent variables on one or a few dependent variables. This same situation is used repeatedly in several similar experiments. It must ensure by design that the response strategies of the subjects remain comparable. This may seem a heavy requirement but the introduction of such a methodology is necessary in any case because many experimental procedures are fragile in the sense that subtle changes to the experimental situation can wipe out an effect completely either because the changes deeply alter the meaning of the experimental situation, either because heterogeneity is introduced when some subjects are sensitive to this change and some are not[4].

Although each experimental situation or "paradigm" is different, all experiments share a number of common features. Most of the time some material is provided and a task is given to several groups of subjects under the various experimental conditions. For instance, the subjects are given test problems to solve under various environmental constraints. Usually a number of college students participate in order to receive credit for a class. They "volunteer" in the sense that they have been given

the choice to participate in one experiment instead of another. They are assigned randomly to the experimental conditions and one—or a few—aspect of their behaviour is recorded to provide a single—or a few—dependent variable. The experiment assesses the effect of the manipulated variable—called independent—on the dependent variable. The various experimental conditions are the levels of the independent variable. Because it is manipulated, the independent variable is attributed a causal influence on the dependent variable. The difference between experimental conditions is demonstrated by aggregating across subjects and comparing group (i.e. condition) means using, for instance, a simple t statistic or an appropriate ANOVA in which the dependent variable is the analysis variable while the experimental conditions provide for one or several factors. Within group (between subjects) variance provides the error term. The main published result is usually this statistical test and the rejection of the corresponding null hypothesis (H0: no difference between conditions).

To make things clear we will briefly summarize a study by Loftus and Palmer (1974). Although 25 years old, this example is a piece of excellent research and its methodology cannot really be considered questionable by current standards. The authors showed their subjects a short movie of a car accident and asked them to answer a few questions. According to the experimental condition, the form of the question changed. The subjects were asked "at what speed did car A hit car B" but the verb actually used varied and was smash, collide, bump, hit or contact according to the experimental condition. The subjects answered by a speed estimate in miles per hour. This was the dependent variable. One way to analyze these data is to use a one factor, 4-level, ANOVA with subjects as a random factor nested under group. The analysis that was actually run by Loftus and Palmer is slightly more complex but it shows that the average estimated speed does change with the wording of the question, thus demonstrating the potential of leading questions in analog (i.e. court room) situations.

Again, this is a piece of excellent research and these characteristics are not unusual. The procedure implements the standard approach in experimental psychology. It is usually referred to as Null Hypothesis Testing (NHT from now on) because it relies almost exclusively on the testing and rejection of a null hypothesis. We will now consider this strategy and what it can be expected to achieve.

3. THE CONTROVERSIES AROUND NHT

3.1. The effect size controversy

A first puzzling aspect of the procedure in our example is that although the study contains information regarding the difference in speed estimates, confidence intervals around this difference are not discussed. Obviously, the main point of interest in the study is only that such intervals do not include zero. We are also not told under which circumstances the effect would be maximal or what variables would allow us to predict its size. The theory presented by the authors actually does not deal with these questions. The dependent variable also has no formalized

http://www.springer.com/978-1-4020-0867-2

The Explanatory Power of Models
Bridging the Gap between Empirical and Theoretical Research in the Social Sciences
Franck, R. (Ed.)
2002, X, 310 p., Hardcover
ISBN: 978-1-4020-0867-2