HUMAN SCIENCES STUDENTS' DIFFICULTIES IN PARAMETRIC TESTS: A CONTRIBUTION TO STATISTICS EDUCATION

Noëlle Zendrera
Institut de Psychologie et de Sociologie Appliquées (IPSA),
Université Catholique de l'Ouest (UCO), France
noelle.zendrera@uco.fr

The purpose of this article is to present part of the results of our research in statistical education. In particular, we have studied difficulties faced by undergraduate students in human sciences when solving a concrete and complete problem within a parametric hypothesis test. Two methods of collecting data have been used: the written proof and the individual clinical task interview. Results obtained by written proof among 90 students have been quantified in order to give a global vision of the seven variables studied. The written test has also facilitated the process permitting us to select 10 students for the task interviews. The results obtained in the interviews show that these students displayed original conceptions regarding many concepts implied in these kind of tests, different to certain misconceptions commonly encountered in the literature.

INTRODUCTION

Inferential statistics, and especially parametric tests, are very useful methods in experimental sciences (Batanero, 2001). Nevertheless, students and professional users of this kind of tests use to be strongly prone to many conceptual difficulties when learning and applying them (Bakan, 1966; Batanero et Díaz, 2006; Lecoutre et al., 2003; Poitevineau, 2004; Régnier, 2006). It is then essential to follow through the research regarding these difficulties, to identify them more accurately and to understand how learners conceptualise, in order to modify their way of teaching, to avoid misconceptions and to favour "conform conceptions".

Our research looks into the conceptions worked out by Human Sciences students–Psychology undergraduate students–towards the concepts and techniques involved in a parametric test. The experimentation has been held precisely regarding the "single mean parametric hypotheses test". This test allows us to compare the (unknown) value of a "sample population mean" to a (known a prior) standard value. This test has been chosen for this research because we consider it as the "easier" parametric test.

RESEARCH METHODOLOGY

Our research took place in two successive and complementary phases; in each phase, there has been a distinct method of collecting data. In the first phase, data have been collected by way of a "written task" (an "open question"); 90 students participated in it. In the second phase, data have been collected in "individual clinical task interviews"; this kind of interview gives a much more freedom to the interviewee. 10 individual clinical task interviews have been carried out with 10 students, selected among the 90 students on the first phase; the selection went by definite and clear criteria. The fact of using two different collecting data methods has allowed us to go into students' conceptions in depth and makes this research quite original because most of the educational studies concerning the hypotheses tests use rather "closed ended questionnaires" or "multiple choice questionnaires". These questionnaires are easier and quicker to analyse; nevertheless they prevent the researcher from reaching the student's real and personal thinking.

In both phases of this research, students have been asked to solve a given task: they had to solve "entirely" a single mean test problem. This is a second originality of this research because in most studies the learners are only asked to answer some closed questions relating to a circumscribed part of a test and not to a whole problem. We now include the problem presented:

> *A Psychology Research Laboratory has created a psychological scale to assess the psychomotor development in children and teenagers. This scale gives an index that measures the development degree for each age group. The higher this index is, the higher the psychomotor development degree is. The scale mean is, for the French children with eleven years old, situated on 200 points.*

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

A group of psychology researchers apply this scale to a group, composed by eleven years old French children born with low weight, and therefore likely to present an retarded psychomotor development. For the 31 children compounding this group, the researchers obtain an index mean of 190 points, with a sample deviation of 30 points.
Based on these results, can we consider that, for the age of eleven years, the index mean of children born with low weight is different from the general index mean? Use a 5% level of significance.

For analysing the found solutions afforded by the students, we have applied the "content analysis" method, established on "interpretation". Each solution found has been analysed with a common grid, based on expressly constructed variables. All the interviews have been recorded and transcribed (verbatim).

RESULTS AND DISCUSSION

We present here the results obtained regarding the "three sets" and the "three means" involved in the single mean test problem used in this experimentation. The concepts involving the problem analysed are: concepts of "sample" and "sample mean $\overline{x}$"; the concepts of "sample population" and "sample population mean μ"; and "reference population" and the "reference population mean μ_0" (the standard mean).

As shown in table 1, in the problem used in this experimentation, the sample is a subset of "31 children, eleven years old, born with low weight"; the value of the sample mean is here 190. The sample population is the set of "all the eleven years old children born with low weight"; the value of the sample population mean is always unknown. And finally the reference population is the set of "all the eleven years old children, all weights at birth taken together"; the value of the reference population mean "μ_0" (the standard) is here 200.

Table 1. The "three sets" and the "three means" involved in the single mean test problem submitted

The Reference Population	*The Sample Population*	*The Sample*
Population used as a reference	Large population from which the sample was taken	A representative subset of the "sample population"
All the children eleven years old, all weights at birth taken together	All the eleven years old children born with a low weight	31 children, eleven years old, born with a low weight
The Reference Mean, a standard	*The Sample Population Mean*	*The Sample Mean*
A specified value	The population parameter	The sample statistics
symbol: μ_0	symbol: μ	symbol: $\overline{x}$
value: known	value: unknown	value: known, calculated from sample data
$\mu_0 = 200$	μ: ?	$\overline{x} = 190$

We first present the results obtained by the "written task" among the 90 students (Figure 1). For the single mean test problem used in our experimentation, only 25,6% of our 90 students have produced a solution corresponding to what we expected ("conform" solution). Most of students, 66,7% have produced a solution not corresponding to what we expected ("erroneous", mistaken solutions). A little minority of students, 7,8%, have been unable to give a solution ("undetermined" solution). The identification of the "three sets" and the "three means" involved in the problem submitted appears to be difficult to the majority (74,5%) of our 90 students.

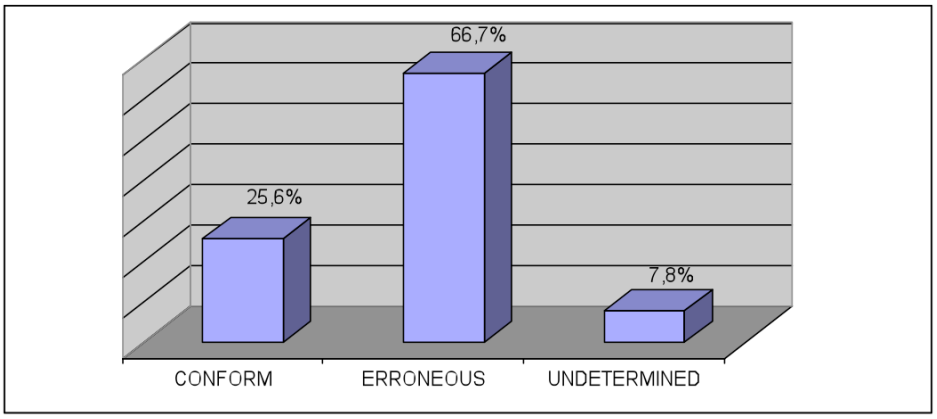

Figure 1. Results regarding the "three sets" and the "three means"
involved in the single mean test problem submitted by way of the written task (N = 90 students)

Let's see now the accurate results obtained by way of the 10 clinical task interviews. For the single mean test problem used in our experimentation, and as shown in table 2, the concepts of "sample" and "sample mean" are well understood by our 10 interviewed students: all of them have elaborated "*conform* conceptions" regarding these two concepts (conform: satisfying what we expect).

Table 2. Conceptions elaborated by the 10 interviewed regarding the "three sets" and the "three means" involved in the single mean test problem submitted

The Sample and the Sample Mean	*The Sample Population and the Sample Population Mean*	*The Reference Population and the Reference Mean (a standard)*
10 "conform" conceptions	1 "conform" conception	7 "conform" conceptions
any "misconceptions"	4 "misconceptions"	2 "misconceptions"
any "lack of conceptualization	5 "lacks of conceptualization"	1 "lack of conceptualization"

On the contrary, the concepts which provoke the most understanding difficulties among our 10 interviewed students are the concepts of "sample population" and "sample population mean" (see table 2). Indeed, only one student has elaborated "*conform conceptions"* regarding them; four students have clearly elaborated *misconceptions* ("wrong" conceptions) towards them. At last, five students have been completely unable to conceptualise them and have cleared them when solving the test problem; these five students demonstrate a real *"lack of conceptualisation"*. This latest result is very forward-looking because we have identified real "lacks" of conceptualisation and not only "confusions" "no discriminations", as reported in many statistics education research papers (Castro Sotos, 2007; Vallecillos, 1995).

The concepts of "reference population" and "reference population mean" (see table 2) provoke less understanding difficulties than the "sample population" and "sample population mean", but more than the "sample" and the "sample mean".

REFERENCES

Bakan, D. (1966). The tests of significance in psychological research. *Psychological Bulletin, 66*, 423-437.

Batanero, C. (2001). *Didáctica de la Estadística.* Granada: Universidad de Granada, Grupo de Investigación en Educación Estadística.

Batanero, C., & Díaz, C. (2006). Methodological and didacticial controversies around statistical inference. *InActes des 38èmes Journées de Statistique (JdS 38).* Clamart, France, 29 may-2 june 2006. Paris: Société Française de Statistique [CDROM].

Castro Sotos A.E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2,* 98–113.

Lecoutre, M.-P., Poitevineau, J., & Lecoutre, B. (2003). Even statisticians are not immune to misinterpretations of null hypothesis significance tests. *International Journal of Psychology, 38*(1), 37-45.

Poitevineau, J. (2004). L'usage des tests statistiques par les chercheurs en psychologie : aspects normatif, descriptif et prescriptif. *Mathématiques et Sciences Humaines (MSH-MSS), 42(*167), 5-25.

Régnier, J.-C. (2006). Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ? *In* C. Castela & C. Houdement (Eds.), *Actes du séminaire national de didactique des mathématiques 2005* (pp. 13-37). ARDM & IREM Paris-VII.

Vallecillos, A. (1995). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. *Recherches en Didactique des Mathématiques, 15*(3), 53-81.