SOME EMPIRICAL EVIDENCES ON LEARNING DIFFICULTIES ABOUT TESTING HYPOTHESES

Angustias Vallecillos
Universidad de Granada, Departamento de Didáctica de la Matemática
Facultad de Ciencias de la Educación, Campus de Cartuja
18071 Granada, Spain
e-mail: avalleci@platon.ugr.es

SUMMARY

In spite of the numerous references to the problems derived from incorrect uses of statistical tests or from misinterpretations of their results by experimental researchers in different areas, the educational world has remained detached from them until now. In this paper we present a survey of educational experimental research on this topic, as well as a summary of the results in our own comprehensive assessment of undergraduates' learning difficulties concerning statistical tests. We point to some difficulties and errors that underlie the problems described and should be taken into account to improve the teaching and learning of the topics. Students' conceptions about key concepts in statistical tests are also described.

1. INTRODUCTION

There are numerous references to incorrect uses of significance tests and to erroneous interpretations of their results in experimental research, particularly in Psychology. As a summary, we quote the book by Morrison & Henkel (1970). New proposals to reform the uses of significance tests and the editorial politics of AERA journals are also described in Thompson (1996). More recently, Poitevineau (1998) describes a study on the methodology of analysing experimental data in psychology, where he still returns to the topic of uses and abuses of hypothesis tests. A large number of related references can be found in the bibliography prepared by Thompson at the following web address: http://acs.tamu.edu/~bbt6147 /.

Although, only recently educational research is paying some attention to the teaching of statistical inference as possible source of the problems described, which are related to its applications to real problems. More rarely experimental research to show incorrect planning and other teaching problems, learning difficulties or errors in inference is being carried out.

In this work we synthesise some published results on students' difficulties about statistical tests. Our aim is to make teachers and lecturers involved in the teaching of these topics conscious of the empirical evidence of problems in our students' learning, and of the need to pay more attention to them as well as to integrate research results in our teaching. Without intending to be exhaustive, we have selected published results of research centred on the learning of statistical inference. Our intention is to show the diversity of problems described, even in a relatively scarce specific educational research on this question.

We have classified these results according to key concepts on which our own experimental investigation on the learning of hypothesis tests by undergraduates was based (Vallecillos, 1994).

2. THE LOGIC OF HYPOTHESIS TESTING

Falk & Greenbaum (1995) present a criticism of the logical structure of statistical tests analysing the possible causes of the persistence in using significance tests, in spite of the misconceptions described. In particular, they suggest the existence of profound psychological reasons leading people to believe that they cope with chance and minimise their uncertainty when obtaining a significant result.

In our own research, we found many students who were not able to capture the logic of a process that lead to choose one of two statistical hypotheses, not just by a strictly subjective process, but on the basis of the data obtained from a random sample from the reference population.

For example, many students believed that when a test was correctly carried out this proved the truth of the null hypothesis, as in the case of a deductive procedure. Four different conceptions regarding the type of proof that hypotheses tests provide were differentiated: **a**) *Conception of the test as a decision-making rule*; **b**) *Conception of the test as a procedure for obtaining empirical support for the hypothesis being researched*; **c**) *Conception of the test as a probabilistic proof of the hypotheses and* **d**) *Conception of the test as a mathematical proof of the truth of the hypothesis* (Vallecillos, 1995; Vallecillos, 1996a).

3. LEVEL OF SIGNIFICANCE

Falk (1986) suggested that the verbal ambiguity in the term "Type I error" used for the significance level concept, which is a well defined *conditional* probability whereas "Type I error" *is not conditionally* phrased, may provoke confusion among the students, who may believe that they are dealing with the probability of a single event. Other authors, such as Oakes (1986) and Pollard & Richardson (1987) also described errors in interpreting statistical tests in researchers and statistically knowledgeable people. Williams (1998) carried out an experimental research with 18 university students from an introductory statistics course to analyse their understanding of the significance level. She proposed conceptual task and two classical problems to them in order to analyse their procedural knowledge. She concludes that most student's knowledge of the significance level was limited and highlighted several problems associated with students' knowledge of the concept.

We also have found a great variety of interpretations of significance level, most of them incorrect. These are not limited to exchanging the two events that intervene in the definition of α, but they rather include a variety and are associated to diverse incorrect interpretations of the conditional probabilities in its definition or to the test results (Vallecillos, 1996b). Not all the explanation given by Falk & Greenbaum (1995) to the mechanism producing the misconceptions concerning the significance level could be applied in our research, where we found students who correctly understood conditional probabilities, who, nevertheless, misunderstood the significance level. These students' misunderstandings were related to conceptual errors linked to others concepts in statistical tests. Some students did not realise the random nature of the statistics sample and did not relate the sampling distribution to the critical and acceptance regions and to the level of significance. Some other students' incorrect interpretation of the significance level, was due to the fact that they exchanged the two events intervening in the definition of the same one, that is, they confused the probability that the null hypothesis is true given that it had been rejected with the level of significance. Indeed, this erroneous interpretation of the significance level described previously and related to misunderstanding conditional probability (Falk & Greenbaum, 1995), but, in our opinion, a better explanation would be that this is linked to the incorrect interpretation of the test results, believing that they demonstrate the truth of the null hypothesis as in case of a deductive procedure. A strong interrelationship between misconceptions on the significance level and those regarding the global test logic is therefore evident, and until now this has not be taken into account in teaching. In addition to those students who show a correct conception of the significance level, we determined three main incorrect conceptions among the students who took part in our research: **a**) *Significance level as conditional probability referred to one of the hypotheses*; **b**) *Significance level as simple probability of the null hypothesis* and **c**) *Significance level as error probability*, Vallecillos, 1998b).

4. STATISTICS, PARAMETERS, SAMPLING DISTRIBUTIONS

Schuyten (1991) informs of the difficulty that the simultaneous use of concepts at different concretion levels suppose for students. For example, the mean of the sample, that of the population and that of the sampling distribution of means is used simultaneously in inference, where they have different levels of abstraction that confuse and make the students' work more difficult. She also informs of other additional difficulties, such as the use of formal language (incorrect notation in correctly understood concepts and appropriate notation linked to an incorrect understanding of the concept); interpretation of tables and graphical representations.

In our research, a frequent error in the students' answers was confusing the sampling statistics with the population parameter (in our case the sample mean $\bar{x}$ with the population mean μ). With the written information provided by the students in their answers to the questionnaire, it is not possible to discover the possible causes of this very widespread error. In later interview carried out to a selected group of students they were asked specifically for the interpretation given and the notation used to refer to these concepts. We were able to verify that, at least in the cases analysed, there was not a problem of incorrect use of the notation appropriate for each concept, but that the students did not take into account the different means and the distributions implied, such as the sampling distribution (Vallecillos, 1998a).

5. UNDERSTANDING THE ROLE OF THE HYPOTHESES

The establishment of hypotheses appropriate to the situation, that answer the problems we are trying to solve, is obviously, crucial and basic. However, this first step in solving a problem of hypothesis test and in applying the test process presents great comprehension problems for students, who are unable to identify the hypotheses adapted in each case. They do not understand the role played by them in the process or they confuse the null and alternative hypotheses (Vallecillos, 1996b). In particular, we have determined three main conceptions concerning the null hypothesis in a test: ***a**) Null hypothesis as hypothesis to prove;* **b**) *Null hypotheses referred either to the population or the sample and* **c**) *Null hypothesis referring to only one population or only one parameter*.

6. CONCLUSIONS

We have experimentally found errors and learning problems affecting many concepts, aspects and applications in statistical tests. Some of these errors have their origin in the previous levels of teaching (confusion among populations and samples, insufficient understanding of the sampling variability and sampling process, etc.) which outlines the need to work these concepts in the pre-University levels. Experimental research serves to detect possible teaching problems at different teaching levels. Their results are a solid base to plan and scientifically assess the teaching with the purpose of reaching its higher efficiency. At the moment, the empirical evidences obtained in our field are scarce although we hope that more researchers would be encouraged to continue making an effort to obtain a more precise knowledge about what our students really learn. A final recommendation to these researchers is involving University lecturers and secondary school teachers in the research teams, because their experience and collaboration can be very valuable to improve educational research.

Acknowledgments: To the Research Projects PB96-1411 and PB97-0826 supported by DGES, MEC, Madrid, Spain.

7. REFERENCES

Falk, R. (1986). Misconceptions of statistical significance. Journal of structural learning, 9, 83-96.

Falk & Greenbaum (1995). Significance test die hard. Theory & Psichology, 5(1), 75-98.

Morrison, D.E. & Henkel, R.E. (Eds.) (1970). The significance test controversy. Chicago: Aldine.

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioural sciences. Chichester: J. Wiley & Sons, Inc.

Poitevineau, J. (1998). Méthodologie de l'analyse des données experimentales: Étude de la practique des tests statistiques chez les chercheurs en psychologie, approaches normative, prescriptive et descriptive. Ph. D. Université de Rouen.

Pollard, P. & Richardson, J.T.E. (1987). On the probability of making Type I errors. Psychological Bulletin, 10, 159-163.

Thompson, B. (1996). AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reform. Educational Researcher, 25(2), 26-30.

Vallecillos, A. (1994). Estudio teórico-experimental de errores y concepciones sobre el contraste estadístico de hipótesis en estudiantes universitarios. Tesis Doctoral. (Theoretical and experimental study on errors and conceptions about hypothesis testing in university students. Ph. D.). Universidad de Granada.

Vallecillos, A. (1995). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. (University students' understanding of the logic of hypothesis testing). Recherches en Didactique des Mathématiques, 15(3), 53-81.

Vallecillos, A. (1996a). Students' conceptions of the logic of hypothesis testing. Hiroshima Journal of Mathematics Education, 4, 43-61.

Vallecillos, A. (1996b). Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas. (Statistical inference and teaching: didactical analysis of the tests of hypotheses). Granada: Comares.

Vallecillos, A. (1998a). Research and Teaching of Statistical Inference. Proceeding of the International Conference on the Teaching of Mathematics, 296-298. Boston: J. Wiley & Sons, Inc.

Vallecillos, A. (1998b). Experimental Study on the Learning of the significance level concept. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.): Proceedings of the ICOTS 5, 1475-1476. Nanyang Technological University, Singapore.

Williams, A. M. (1998). Students' understanding of the significance level concept. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.): Proceedings of the ICOTS 5, 743-749. Nanyang Technological University, Singapore.

RÉSUMÉ

Malgré les nombreuses références aux problèmes qui se dérivent d'un usage incorrect des tests statistiques ou de l'interprétation erronnée de leurs résultats par les chercheurs expérimentaux dans différents domaines, l'univers de l'éducation ne s'est pas senti concerné jusqu'à maintenant par ce problème. Dans cet article, nous présentons une révision d'autres articles où il est question d'une recherche expérimentale sur ce sujet. Nous présentons de même un résumé des résultats d'une importante recherche centrée sur l'évaluation compréhensive de l'apprentissage des tests d'hypothèse par des étudiants universitaires. Nous signalons certaines difficultés et erreurs qui pourraient être à l'origine des problèmes auxquels nous faisions allusion précédemment et qui devront être tenus en compte dans l'enseignement afin d'améliorer l'apprentissage de ces sujets. Nous décrivons également des conceptions d'étudiants à propos de concepts clés des tests statistiques.

What are hypothesis tests? Covers null and alternative hypotheses, decision rules, Type I and II errors, power, one- and two-tailed tests, region of rejection.Â A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true. Hypothesis testing refers to the formal procedures used by statisticians to accept or reject statistical hypotheses. Statistical Hypotheses. The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Surprisingly, empirical evidence for the hypothesis has remained inconsistent. In the present article, we reinterpret the original uncanny valley hypothesis and review empirical evidence for different theoretically motivated uncanny valley hypotheses.Â Some of these hypotheses could be derived from established psychological constructs and theories. In some cases, minor adjustments to the original uncanny valley hypothesis could be justified.Â Fourth, we will review empirical evidence for the formulated hypotheses based on the adopted evaluation criteria (Section Review of Empirical Evidence). Finally, we will discuss the implications and limitations of our findings and consider open questions in uncanny valley research (Section Discussion). If you really understand the hypothesis, the fact that some people are able to consistently or almost consistently beat market is itself a contradiction to the hypothesis. Ideally, nobody should be able to beat a buy and hold strategy in the long run. Then there is strong empirical proof that market anomalies do exist. Read http://www.investopedia.com/arti for some examples. 1k views Â· View 1 Upvoter.