Jurnal Pendidikan Malaysia 43(1)(2018): 25-33
DOI: http://dx.doi.org/10.17576/JPEN-2018-43.01-04

Pedagogical Issues in Hypothesis Testing

(Isu-isu Pedagogi dalam Pengujian Hipotesis)

MIHIR DASH*

ABSTRACT

Hypothesis testing is a statistical technique which is used to evaluate assumptions about a population on the basis of sample data, to determine the extent to which they are tenable. Hypothesis testing is the most widely-applied statistical technique, particularly because of the emphasis on hypothesis development and testing in the scientific method. Unfortunately, students and researchers are quite prone to making mistakes and misinterpreting inferences in hypothesis testing. These mistakes and misinterpretations tend to arise from insufficient understanding of the probability and sampling theory underlying the logic of hypothesis testing. The present study attempts to identify the causes of different types of mistakes made in hypothesis testing, in order to suggest pedagogical strategies to avoid these mistakes. The data for the study was collected from a sample of postgraduate management students in Bangalore, India, using specially-designed business decision-making case lets based on hypothesis testing. The analysis focuses on the incidence of different types of mistakes that the respondents committed, particularly with respect to the type of tests, and uses multiple linear discriminant analysis to identify the factors impacting the overall inference, i.e. the correct taking of the decision and the correct drawing of the conclusion. The key finding of the study is that both the formulation and computation factors play a significant role in taking the overall inference. Further, in each panel, the critical discriminator was found to be the aspect for which the incidence of mistakes was highest. With increasing complexity of the hypothesis test, the computation factor was found to become more important. In panels A and B (tests for a single population mean and proportion, respectively), formulation aspects were found to be the most significant discriminators, and in panel C (test for equality of means), both formulation and computation aspects were significant; on the other hand, for the remaining panels (test for independence, one-way ANOVA, and two-way ANOVA), only computation aspects were significant. The study contributes to the literature by proposing some pedagogical strategies for teaching of different types of hypothesis tests based on the findings.

Keywords: Hypothesis testing; scientific method; mistakes; misinterpretations; pedagogical strategiess

ABSTRAK

Pengujian hipotesis adalah teknik statistik yang digunakan untuk menilai andaian mengenai populasi berdasarkan data sampel, untuk mengenal pasti sejauh mana ia dapat dipertanggungjawabkan. Pengujian hipotesis adalah teknik statistik yang paling banyak digunakan, terutamanya teknik ini memberi penekanan kepada perkembangan hipotesis dan ujian dalam kaedah saintifik. Malangnya, pelajar dan penyelidik agak cenderung melakukan kesilapan dan menyalahtafsirkan kesimpulan dalam ujian hipotesis. Kesalahan dan salah tafsiran ini cenderung timbul daripada pemahaman yang tidak mencukupi tentang teori kebarangkalian dan pensampelan yang mendasari logik pengujian hipotesis. Kajian ini cuba untuk mengenal pasti punca pelbagai jenis kesilapan yang dibuat dalam pengujian hipotesis serta untuk mencadangkan strategi pedagogi bagi mengelakkan kesilapan-kesilapan ini terus berlaku. Data untuk kajian ini diambil dariapada sampel pelajar pengurusan pasca siswazah di Bangalore, India dengan menggunakan reka bentuk pembuat keputusan perniagaan yang direka khusus berdasarkan ujian hipotesis. Analisis ini memberi tumpuan kepada kejadian pelbagai jenis kesilapan yang dilakukan oleh responden, terutamanya berkenaan dengan jenis ujian, dan menggunakan analisis diskriminasi berganda untuk mengenal pasti faktor-faktor yang mempengaruhi kesimpulan keseluruhan, iaitu pengambilan keputusan yang betul dan yang betul lukisan kesimpulan. Temuan utama kajian ini adalah bahawa kedua-dua faktor perumusan dan perhitungan memainkan peranan penting dalam mengambil kesimpulan keseluruhan. Selanjutnya, dalam setiap panel, diskriminator kritikal didapati merupakan aspek kesilapan kesilapan tertinggi yang dilakukan. Dengan meningkatkannya kesukaran pengujian hipotesis, maka faktor perhitungan dijumpai menjadi lebih penting dalam pengujian hipotesis. Dalam panel A dan B (ujian bagi min dan populasi masing-masing), aspek perumusan didapati sebagai diskriminator yang paling penting. Dalam panel C (ujian untuk kesamaan sarana), aspek perumusan dan perhitungan adalah penting; Di sisi lain, untuk panel yang tinggal (ujian untuk kebebasan, ANOVA sehala, dan ANOVA dua hala), hanya aspek pengiraan yang signifikan. Kajian ini menyumbang kepada ulasna kepustakaan dengan mencadangkan beberapa strategi pedagogi untuk mengajar pelbagai jenis pengujian hipotesis berdasarkan hasil kajian.

Keywords: Pengujian hipotesis; kaedah sainstifik; kesilapan; salah interpretasi; strategi pedagogi

Introduction

Hypothesis testing is the most widely-applied statistical technique in applied empirical research, particularly because of the emphasis on hypothesis development and testing in the scientific method. The classical empirical paradigm involves the identification and definition of problems, the definition of response variables or measures, the formulation of hypotheses, the collection of sample data, and the application of hypothesis tests to reject hypotheses that are not supported by the sample data. The importance of hypothesis testing is especially underlined by its inclusion in two core courses in any typical postgraduate program, statistics and research methodology. It also finds echoes in other courses such as marketing research, and psychometric research. Unfortunately, students and even researchers are quite prone to make mistakes in setting up and executing hypothesis tests. The present study attempts to identify the causes of different types of mistakes made in hypothesis testing, in order to suggest pedagogical strategies to avoid these mistakes.

There is an extensive literature addressing pedagogical issues in hypothesis testing. The broad areas of concern are that of the understanding of the probabilistic and statistical concepts underlying hypothesis testing, both among students and among teachers (refer to Liu (2005) and Aquilonius (2005) for an extensive review of the literature), and the transmission of these concepts in classroom settings, especially the role of pedagogy in developing students' understanding of statistical reasoning (Garfield 2002; Garfield & Ben-Zvi 2003).

Two key concepts in hypothesis testing are randomness and variability. Psychological research has found that most people had poor intuition regarding these key concepts (Kahneman & Tversky 1973; Shaughnessy 1992). Falk (1986) found that many students and researchers did not have a clear understanding about the conditional probabilities involved in hypothesis testing, resulting in an unclear understanding of p-values. Quilici and Mayer (1996) studied how students used worked examples to learn how to categorize statistics word problems, including hypothesis testing problems. Their study examined the interplay of surface characteristics and structure characteristics in statistics word problems, and suggested that examples emphasizing structure characteristics were more effective in helping students to determine which statistical test to use when solving statistics word problems.

Garfield (2002) suggested a model for statistical reasoning, analyzing the learning of statistical concepts into five hierarchical levels: idiosyncratic reasoning, wherein students use statistical terms superficially, without fully understanding them; verbal reasoning, wherein students have verbal understanding of some of the concepts, but are unable to apply their understanding in practical situations; transitional reasoning, wherein students are able to recognize some aspects of a statistical process, but are unable to properly integrate these aspects; procedural reasoning, wherein students are able to correctly recognize the aspects of a statistical process, but do not fully integrate them or do not fully understand the process; and lastly, integrated process reasoning, wherein students have complete understanding of a statistical process, coordinating the rules and behavior. She suggested that statistical teaching should lay more emphasis on the development of students' statistical reasoning and not just the computational procedure, using graphical simulation of sampling distributions, with varying sample sizes and population parameter values, to develop students' reasoning about sampling distributions.

Link (2002) analyzed the mistakes made by students in setting up of statistical hypotheses. He considered four categories of errors: mistakes in identifying the correct population parameter; mistakes of using the sample statistic rather than the population parameter; mistakes in specifying the hypothesized value, sign, or direction of inequality; and meaningless statements. He stressed the importance of a proper formulation of null and alternative hypotheses, and found that, in particular, misspecification of the alternative hypothesis increased the chance of making mistakes in subsequent steps in the hypothesis testing process. He also found that students were generally comfortable in using the test statistic formulae, but faced difficulty in interpreting the p-value. He suggested that this was due to their lack of understanding of the difference between the probability statement about the observed value of the sample statistic and the probability statement about the appropriate test statistic.

Liu (2005) explored teachers' understanding of probability and statistical inference, to understand the conceptual and pedagogical issues that teachers face in order to teach probability and statistics effectively in the classroom. He found that a majority of teachers did not understand the logic of hypothesis testing, as their understanding of "unusualness" of the observed value of the sample statistic was not based on the concept of sampling distribution. This suggested that this was due to their lack of understanding of hypothesis testing as a tool and of the types of questions for which hypothesis testing can be applied. He suggested that teachers would be able to teach hypothesis testing more effectively by orienting them in thinking through the sampling distribution foundations of statistical inference.

Aquilonius (2005) investigated students' reasoning process in hypothesis testing, particularly the concepts of population and sample, the concept of p-value, and the drawing of conclusions about the population. He found that students were able to understand the importance of random samples, but they did not properly understand the mathematical nature of random sampling, confusing randomness with representativeness, and as a result, they did not fully understand the concept of sampling distribution. He also found that the students had a mechanical approach to p-values, in that they were readily able to arrive at statistical decisions (i.e. rejecting or not rejecting the null hypothesis), but often had difficulty in translating their decisions to the context of the original problem. He found that the increased use of examples in the teaching of hypothesis testing helped the students overcome this difficulty over time.

Smith (2008) studied the development of understanding of statistical hypothesis testing among undergraduate students in introductory statistical courses using quantitative and qualitative methods. She found that introductory statistics students do not develop strong, connected understandings of statistical hypothesis testing overall. Even though they are able to perform the procedures, students generally did not have strong understanding of the concepts, logic, and uses of the method. She also found that students did not understand the role of indirect reasoning and inference in implementing and interpreting the results of a statistical hypothesis test. She suggested that statistical instruction should focus more on the development of student understanding of overall statistical hypothesis testing, particularly on the development of understanding of the logic of indirect reasoning and its role in statistical hypothesis testing, along with uncertainty and variability. Krishnan and Idris (2015) provided an overview of the problems faced by students in the learning of hypothesis testing. They suggested a classification of twenty-one distinct types of mistakes in hypothesis testing under six broad groups, viz. formulation of the hypotheses, calculating the test statistic, determining the critical region, determining the critical value and the p-value, making a decision, and communicating the decision in the context of the problem.

The literature reviewed above has highlighted the importance of the problem of students' and teachers' understanding of hypothesis testing, and the need for appropriate pedagogical approaches to foster a clear understanding of the probability and sampling theoretic concepts underlying hypothesis testing. The present study examines students' mistakes in applying hypothesis testing, as in Link (2002) and Krishnan and Idris (2015), in order to identify the causes of different types of mistakes made in hypothesis testing, in order to suggest pedagogical strategies to avoid these mistakes.

Methods

The data for the study was collected from a sample of 116 postgraduate management students in Bangalore, India. The data was collected using specially-designed caselets (i.e. mini-cases), each of which related to a business problem which required decision-making based on hypothesis testing using the sample data provided to them. The respondents represented a good cross-section of postgraduate management students: 72.4% of the respondents were male and 27.6% female; all of the respondents were in the age group 21-25 years; 37.1% of the respondents were from engineering background, 29.3% science, 28.4% commerce, and 5.2% arts. In terms of the tests applicable in the caselets, 18.1% were based on the test for a single population mean, 15.5% on the test for a single population proportion, 19.0% on the test for equality of means, 22.4% on the χ^2 test for independence, 12.1% on the one-way ANOVA test, and 12.9% on the two-way ANOVA test.

The analysis focuses on the incidence of different types of mistakes that the respondents committed in the application of hypothesis testing in the caselets provided to them. The incidence of mistakes was analyzed with respect to the type of tests. Further, factor analysis was applied to identify patterns in the coincidence of these mistakes. Finally, multiple linear discriminant analysis was used to identify the factors impacting the overall inference, i.e. the correct taking of the decision and the correct drawing of the conclusion. The first discriminant model used both of the factor scores as discriminating variables; the second model used the original variables as discriminating variables, via stepwise discriminant analysis, thereby identifying the significant discriminating variables, overall, as well as for each type of test. For detailed explanations of factor analysis and discriminant analysis, the reader may please refer to Hair et al (2006).

Results and Discussion

The incidence of the different types of mistakes, overall and by type of test, is presented below in Table 1.

Overall, as shown in Table 1, the most frequent mistakes were found to be the incorrect drawing of the conclusion, the incorrect taking of decision, incorrect identification of one/two tailed test, and incorrect identification of the critical value. Incorrect identification of the test statistic formula and incorrect computation of the calculated value of the test statistic were moderately-frequent mistakes. These results are similar to the typology identified by Link (2002) and Krishnan and Idris (2015). The most frequent mistakes in testing for a single population mean were found to be the incorrect taking of the decision and drawing of the conclusion, followed by incorrect formulation of H_0 and H_1, incorrect identification of one/two tailed test, and incorrect identification of the critical value.

The most frequent mistakes in testing for a single population proportion were found to be the incorrect formulation of H_0 and H_1, incorrect identification of one/two tailed test, incorrect drawing of the conclusion, and incorrect identification of the critical value. In this case, there was no incidence of mistakes in identification of the test statistic and its sampling distribution, and relatively low incidence of mistakes in the test statistic formula, the computation of the "calculated value of the test statistic" and taking the decision.

The most frequent mistakes in testing for equality of two population means were found to be the incorrect test statistic formula and the computation of the "calculated

TABLE 1. Incidence of mistakes in hypothesis tests overall and by type of test

	Overall	A	B	C	D	E	F
Formulation of H_0	37.93%	71.43%	94.44%	4.55%	15.38%	0.00%	46.67%
Formulation of H_1	36.21%	66.67%	88.89%	4.55%	15.38%	0.00%	46.67%
Identification of one/two tailed test	59.02%	66.67%	83.33%	31.82%	-	-	-
Juxtaposition of H_0, H_1	7.76%	23.81%	0.00%	4.55%	7.69%	0.00%	6.67%
Identification of test statistic	17.24%	14.29%	0.00%	59.09%	11.54%	0.00%	6.67%
Test statistic formula	41.38%	28.57%	11.11%	81.82%	15.38%	85.71%	40.00%
Computation of the calculated value of test statistic	48.28%	28.57%	16.67%	81.82%	15.38%	92.86%	80.00%
Identification of the test statistic: distribution	23.28%	14.29%	0.00%	63.64%	11.54%	7.14%	40.00%
Identification of the test statistic: degrees of freedom	37.76%	19.05%	0.00%	72.73%	16.00%	14.29%	73.33%
Identification of critical value	53.45%	66.67%	77.78%	77.27%	15.38%	14.29%	73.33%
Taking of the decision	59.48%	76.19%	11.11%	77.27%	30.77%	92.86%	86.67%
Drawing of the conclusion	71.55%	76.19%	83.33%	77.27%	34.62%	92.86%	86.67%

Panel: A: single population mean; B: single population proportion; C: equality of means; D: independence; E: one-way ANOVA; F: two-way ANOVA

value of the test statistic," followed by incorrect identification of the critical value, incorrect taking the of the decision, incorrect drawing of the conclusion, incorrect identification of the degrees of freedom, and incorrect identification of the test statistic. In this case, there was relatively low incidence of mistakes in formulation of H_0 and H_1, and juxtaposition of H_0 and H_1, and moderate incidence of mistakes in identification of one/two tailed test. For the χ^2-test for independence, there was relatively lower incidence of mistakes in all aspects, and moderate incidence of mistakes in taking of the decision and drawing of the conclusion.

The most frequent mistakes in the case of one-way ANOVA were found to be the incorrect taking of the decision and drawing of the conclusion, and computation of the "calculated value of the test statistic," followed by incorrect test statistic formula. In this case, there was no incidence of mistakes in formulation of H_0 and H_1, and juxtaposition of H_0 and H_1, and moderate incidence of mistakes in the identification of the sampling distribution of the test statistic and its degrees of freedom, and identification of the critical value.

The most frequent mistakes in the case of two-way ANOVA were found to be the incorrect taking of the decision and drawing of the conclusion, and computation of the "calculated value of the test statistic," followed by incorrect identification of the degrees of freedom of the sampling distribution of the test statistic. In this case, there was low incidence of mistakes in juxtaposition of H_0 and H_1 and in the identification of the test statistic, and moderate incidence of mistakes in the test statistic formula and identification of the sampling distribution of the test statistic, and in the formulation of H_0 and H_1. The results of the factor analysis are presented in Table 2 below.

TABLE 2. Factor analysis of the different types of mistakes

	Rotated Component Matrix			Component Score Coefficient Matrix	
	Component 1	Component 2	Communalities	Component 1	Component 2
Formulation of H_0		0.9234	0.8917	-0.0589	0.3812
Formulation of H_1		0.9181	0.8743	-0.0523	0.3791
Juxtaposition of H_0, H_1		-0.6155	0.4101	-0.0598	-0.2556
Identification of test statistic	0.8289		0.7025	0.2668	0.0556
Test statistic formula	0.7954		0.6667	0.2545	-0.0725
Calculated value of test statistic	0.8129		0.6766	0.2604	-0.0481
Distribution of test statistic	0.8634		0.7513	0.2777	0.0360
Identification of critical value	0.5356		0.5578	0.1746	0.2181
Cronbach alpha	0.8243	0.8177			
Percentage of variance explained	38.93%	30.21%			

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization
K.M.O. measure of sampling adequacy = 0.577
Bartlett's test for sphericity chi-square = 746.62**

The types of mistakes were categorized into two groups, as shown in Table 2, those related to formulation of the hypothesis test, i.e. the formulation of the null and alternative hypotheses, and those related with the computation of the hypothesis test, i.e. the identification of the test statistic and its formula, the computation of the calculated value of the test statistic, the identification of the distribution of the test statistic and the critical value. Both of these identified constructs were found to be highly reliable, together explaining 69.14% of the total variance. The overall discriminant analysis results are presented in Table 3, while the discriminant analysis results for each of the panels are presented in Tables 4-9.

The overall discriminant analysis results shown in Table 3 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the computation factor playing a marginally higher role (by approximately 25%). Among the individual aspects, the significant discriminators were

TABLE 3. Overall discriminant analysis results

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-1.6682		-2.4821	
Formulation factor		1.7859	0.7529		
Formulation of H_0					
Formulation of H_1				1.9519	0.9264
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		2.3786	0.9288		
Identification of test statistic					
Test statistic formula				1.9720	0.9014
Computation of the calculated value of test statistic					
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value				1.1629	0.4947
Group centroids	incorrectly performed	-0.5343		-0.4580	
	correctly performed	1.4649		1.7811	
Wilks' lambda		0.5567		0.5392	
p-value		0.0000		0.0000	
Percentage Correctly Classified		79.31%		95.69%	

TABLE 4. Discriminant analysis results for panel A

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-2.7782		-0.7772	
Formulation factor		2.2521	0.9001		
Formulation of H_0				3.8258	0.5887
Formulation of H_1				5.4402	0.8014
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		1.2569	0.4973		
Identification of test statistic					
Test statistic formula					
Computation of the calculated value of test statistic					
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value					
Group centroids	incorrectly performed	-0.2700		-0.3742	
	correctly performed	1.6198		2.2452	
Wilks' lambda		0.6742		0.5185	
p-value		0.0288		0.0027	
Percentage Correctly Classified		85.71%		80.95%	

TABLE 5. Discriminant analysis results for panel B

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-7.5574		-0.2722	
Formulation factor		4.8455	1.0075		
Formulation of H_0				4.8990	1.0000
Formulation of H_1					
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		6.7135	0.9136		
Identification of test statistic					
Test statistic formula					
Computation of the calculated value of test statistic					
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value					
Group centroids	incorrectly performed	-0.5245		-0.2722	
	correctly performed	2.6227		1.3608	
Wilks' lambda		0.3925		0.7059	
p-value		0.0009		0.0202	
Percentage correctly classified		94.44%		88.89%	

the formulation of the alternative hypothesis and the test statistic formula, with roughly equal discriminating power. Overall, these two individual factors were able to classify 95.69% of the sample cases correctly.

The discriminant analysis results for the test for a single population mean shown in Table 4 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the formulation factor playing a higher role by approximately 81%. Among the individual aspects, the significant discriminators were the formulation of the null hypothesis and the formulation of the alternative hypothesis, with the formulation of the alternative hypothesis playing a higher role by approximately 36%. Overall, these two individual factors were able to classify 80.95% of the sample cases correctly.

The discriminant analysis results for the test for a single population proportion shown in Table 5 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis

TABLE 6. Discriminant analysis results for panel C

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-0.7111		-1.5405	
Formulation factor		0.7510	0.1789		
Formulation of H_0				4.8990	1.0000
Formulation of H_1					
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		4.8823	0.8459		
Identification of test statistic					
Test statistic formula				3.3892	0.6778
Computation of the calculated value of test statistic					
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom				3.3892	0.7352
Identification of critical value					
Group centroids	incorrectly performed	-1.1936		-1.3412	
	correctly performed	4.0581		4.5600	
Wilks' lambda		0.1580		0.1294	
p-value		0.0000		0.0000	
Percentage correctly classified		95.45%		95.45%	

test, with roughly equal discriminating power. Among the individual aspects, the only significant discriminator was the formulation of the null hypothesis, which was able to classify 88.89% of the sample cases correctly.

The discriminant analysis results for the test for a equality of two population means shown in Table 6 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the computation factor playing a considerably higher role by approximately 373%. Among the individual aspects, the significant discriminators were the test statistic formula and the identification of the degree of freedom, with the identification of the degree of freedom playing a marginally higher role by approximately 8%. Overall, these two individual factors were able to classify 95.45% of the sample cases correctly.

The discriminant analysis results for the test for independence shown in Table 7 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the computation factor playing a considerably higher role by approximately 200%. Among the individual aspects, the only significant discriminator was the test statistic formula, which was able to classify 82.77% of the sample cases correctly.

TABLE 7. Discriminant analysis results for panel D

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-3.1509		-2.7024	
Formulation factor		0.7852	0.2841		
Formulation of H_0				4.8990	1.0000
Formulation of H_1					
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		2.6207	0.8517		
Identification of test statistic					
Test statistic formula				3.2171	1.0000
Computation of the calculated value of test statistic					
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value					
Group centroids	incorrectly performed	-0.9809		-0.9151	
	correctly performed	0.5193		0.5147	
Wilks' lambda		0.6444		0.6614	
p-value		0.0064		0.0023	
Percentage correctly classified		84.62%		82.77%	

The discriminant analysis results for the one-way ANOVA test shown in Table 8 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the computation factor playing a marginally higher role by approximately 23%. Among the individual aspects, the only significant discriminator was the computation of the calculated value of the test statistic, which was able to classify 92.86% of the sample cases correctly.

The discriminant analysis results for the two-way ANOVA test shown in Table 9 indicate that both the formulation factor and the computation factor play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test, with the computation factor playing a marginally higher role by approximately 17%. Among the individual aspects, the only significant discriminator was the computation of the calculated value of the test statistic, which was able to classify 93.33% of the sample cases correctly.

CONCLUSION

The key finding of the study is that both the formulation and computation factors play a significant role in taking the correct decision and drawing the correct conclusion from the hypothesis test. In particular, in each panel, the critical discriminator was found to be the aspect for which the incidence of mistakes was highest. Also, with increasing complexity of the hypothesis test, the computation factor was found to become more important. In panels A and B (tests for a single population mean and proportion, respectively), formulation aspects were found to be the most significant discriminators, and in panel C (test for equality of means), both formulation and

TABLE 8. Discriminant analysis results for panel E

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-0.8213		-0.5151	
Formulation factor		3.1949	0.6713		
Formulation of H_0					
Formulation of H_1					
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		4.1731	0.8239		
Identification of test statistic					
Test statistic formula					
Computation of the calculated value of test statistic				3.6056	1.0000
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value					
Group centroids	incorrectly performed	-0.4303		-0.2377	
	correctly performed	5.5945		3.0905	
Wilks' lambda		0.2625		0.5385	
p-value		0.0006		0.0076	
Percentage correctly classified		100.00%		92.86%	

TABLE 9. Discriminant analysis results for panel F

		Model I		Model II	
		Coeff.	Std. Coeff.	Coeff.	Std. Coeff.
[Constant]		-2.9422		-0.7506	
Formulation factor		2.1017	0.8173		
Formulation of H_0					
Formulation of H_1					
Identification of one/two tailed test					
Juxtaposition of H_0, H_1					
Computation factor		3.2478	0.9576		
Identification of test statistic					
Test statistic formula					
Computation of the calculated value of test statistic				3.7528	1.0000
Identification of the test statistic: distribution					
Identification of the test statistic: degree of freedom					
Identification of critical value					
Group centroids	incorrectly performed	-0.4154		-0.4619	
	correctly performed	2.7000		3.0022	
Wilks' lambda		0.4359		0.3846	
p-value		0.0069		0.0005	
Percentage correctly classified		93.33%		93.33%	

computation aspects were significant; on the other hand, for the remaining panels (test for independence, one-way ANOVA, and two-way ANOVA), only computation aspects were significant. The study contributes to the literature by proposing the following pedagogical strategies for teaching of different types of hypothesis tests based on the findings.

For the tests for a single population mean and proportion, emphasis should be placed on appropriate formulation of the hypotheses. For these tests, the test statistics are relatively simple and easy to compute, and the sampling distributions are straightforward. The formulation aspects, particularly the issue of whether to use a one-tailed or two-tailed test, are usually sources of confusion for students. Students also often get confused whether the test involves a mean or a proportion. Also, as these tests build the foundation for statistical reasoning for the student, the instructor should take care to make them understand how the formulation of the hypotheses affects the outcome of the test (i.e. the decision and the conclusion). Simulation could be a useful tool in these contexts to help students understand the role of formulation in hypothesis testing.

For the test for equality of two population means, emphasis should be placed both on appropriate formulation of hypotheses and on computation aspects. In this case, as above, students often get confused about whether to formulate the test as a one-tailed or two-tailed test; but they also can get confused about which test statistic formula to use and what are the appropriate degrees of freedom. Yet another source of confusion for students is the use of the paired-samples test as against the independent-samples test. All these potential sources of confusion must be addressed by the instructor, and illustrated with appropriate examples.

The chi-squared test for independence was the test for which there was least incidence of mistakes at different steps. Nevertheless, there was approximately a 30% to 35% incidence of mistake in taking the decision and drawing the conclusion. The chi-squared test is a very popular test as it is relatively straightforward and can be applied in several different contexts, but confusion can arise with the test statistic formula. This difficulty can be readily addressed by the instructor by making students to write the formula explicitly when applying the test, and to break down the calculation into steps.

For one-way and two-way ANOVA, emphasis should be given to computation of the calculated value of the test statistic, which may partly arise due to confusion with the test statistic formula. ANOVA is a more advanced test than the other tests, and can be applied with many variations. As with the chi-squared test for independence, the instructor can address possible confusions by making students to write the formula explicitly when applying the test, and to break down the calculation into steps.

There are some limitations inherent in the study. The sample size for the study was relatively small. Also, the respondents were all postgraduate management students, and many of them were engineers and science graduates, so that the results may not be widely generalisable. The methodology for the study could be extended by additionally using observational data to understand how the respondents approached the problems logically and to identify the root causes of their mistakes.

REFERENCES

Aquilonius, B.C. 2005. How do college students reason about hypothesis testing in introductory statistics courses? PhD Dissertation, University of Santa Barbara, California.

Falk, R. 1986. Misconceptions of statistical significance. *Journal of Structural Learning* 9: 83-96.

Fischbein, E. & Schnarch, D. 1997. The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education* 28 (1): 96-105.

Garfield, J. 2002. The challenge of developing statistical reasoning. *Journal of Statistics Education* 10 (3).

Garfield, J. & Ben-Zvi, D. 2005. *The Challenges of Developing Statistical Literacy, Reasoning and Thinking*. Netherlands, Kluwer.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. 2006. *Multivariate Data Analysis*. 6th Ed., Harlow, Essex, U.K.: Pearson Education Ltd.

Hubbard, R. & Armstrong, J.S. 2006. Why we don't really know what "statistical significance" means: a major educational failure. *Journal of Marketing Education* 28(2): 114-120.

Kahneman, D. & Tversky, A. 1982. Subjective probability: A judgment of representativeness. in *Judgment Under Uncertainty: Heuristics and Biases*, edited by Kahneman, D., Slovic, P. & Tversky, A. New York: Cambridge University Press.

Konold, C. 1989. Informal conceptions of probability. *Cognition and Instruction* 6(1): 59-98.

Konold, C. 1991. Understanding students' beliefs about probability. In. *Radical Constructivism in Mathematics Education*, edited by von Glasersfeld, E. Dordrecht: Kluwer Academic Publishers.

Konold, C., Pollatsek, A., Well, A.D., Lohmeier, J. & Lipson, A. 1993. Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education* 24(5): 393-414.

Krishnan, S. & Idris, N. 2015. An overview of students' learning problems in hypothesis testing. *Jurnal Pendidikan Malaysia* 40(2): 193-196.

Liu, Y. 2005. Teachers' understandings of probability and statistical inference and their implications for professional development. PhD Dissertation, Graduate School of Vanderbilt University, Nashville, Tennessee.

Link, W.C. 2002. An examination of student mistakes in setting up hypothesis testing Problems. Paper presented at Louisiana-Mississippi Section of the Mathematical Association of America held in Spring 2002.

Mendez, H. 1991. Understanding the central limit theorem. Doctoral Dissertation, University of California, Santa Barbara. *Dissertation Abstracts International,* 53, 2722.

Nisbett, R.E., Krantz, D.H., Jepson, D. & Kunda, H. 1983. The use of statistical heuristics in everyday inductive reasoning. *Psychological Review* 90: 339-363.

Quilici, J.L. & Mayer, R.E. 1996. Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology* 88: 144-161.

Shaughnessy, J.M. 1992. Research in probability and statistics: Reflections and directions. in Grouws, D.A. (ed.). *Handbook of Research on Mathematics Teaching and Learning*. New York: MacMillan.

Smith, T.M. 2008. *An Investigation into Student Understanding of Statistical Hypothesis Testing*. Doctoral Dissertation, Digital Repository at the University of Maryland.

Well, A.D., Pollatsek, A., & Boyce, S. 1990. Understanding the effects of sample size on the mean. *Organizational Behavior and Human Decision Processes* 47: 289-312.

Mihir Dash
Alliance University
Bangalore, India.
mihirda@rediffmail.com

Address for correspondence e-mail: mihirda@rediffmail.com

Received : 17 April 2017
Accepted : 15 Februari 2018