® AN APPROACH TO REPORT WRITING IN STATISTICS COURSES

FRANCIS, Glenda
Swinburne University of Technology
Australia

Historically, teaching statistics meant teaching students to do a few complex calculations, with very little time left over for interpretation or report writing. Even with the advent of modern computer packages that have removed some of the previous time constraints, report writing is still often overlooked, or considered unimportant. It is assumed that students will somehow work out how to write reports for themselves or that 'someone else' will teach them how to do it. In practice, most students find report writing extremely difficult. They are unclear about what should be included in the report, they tend to include a lot of 'technical terms' which they don't fully understand and they don't really expect 'statistical' reports to make sense. Coupling these problems with an often poor level of English expression leads to reports which can be totally incomprehensible. Report writing needs to be taught explicitly, and in the context of understanding what you are trying to convey to your audience.

This paper presents an approach to report writing which has been developed in second and third year statistics subjects designed for psychology students. The approach involves giving students a process to follow, clear instructions on the sort of language which is appropriate and some model reports to use as a guide. But most importantly, the approach focuses on first understanding the meaning of the results.

INTRODUCTION

Historically the production of statistics and graphs was such a time consuming and challenging process that almost all of the effort in teaching statistics went into these mechanical skills. Report writing was generally ignored completely. However, changes are underway. With the development of graphical calculators and accessible computer software packages such as Minitab and SPSS, it is now possible to spend much more time on concepts and on communication of results. Ritter, Starbuck and Hogg (2001) comment that

> Excellent oral and written communication skill development should be a core part of a statistics program. (page 17)

Garfield, Hogg, Schau and Whittinghill (2002), in reviewing recent changes in statistics courses, note that

> there was also a common theme among many instructors who stated that they focus more on concepts and big ideas and on data analysis and interpretation and less on computation, formulas, and theory (para 33).

Unfortunately, at least in regards to the reporting of statistics, text books are lagging somewhat behind. A casual look at even some of the more recent textbooks (Argyrous, 1996; Heiman, 1998; Sanders & Smidt, 2000; Croucher, 2002) revealed that in most the emphasis was on the theory and production of the statistics, rather than on reporting. This is particularly evident in the exercises given at the end of each chapter. You might find some exercises which request a 'conclusion from the significance test' but nothing that asks for a report on the study. What impression does this make on the students? Clearly, since the bulk of the exercises are related to producing a test statistic, finding a p value and reporting whether the test is significant or not, these must be the really important skills in statistics.

The implicit assumption here is that report writing will come naturally, picked up by a process of osmosis, or that someone else will teach them how to do that – after all, what do mathematicians know about teaching writing skills! Unfortunately, report writing does not come

easily to students; it needs to be taught explicitly, and as an integral part of the process of performing a statistical analysis.

WHAT MAKES REPORT WRITING SO DIFFICULT?

It has been widely documented that students often enter statistics courses with a very negative attitude towards statistics (Gordon 1995; Francis 2002). Gal and Ginsburg (1994) suggest that such negative attitudes towards statistics contribute to students' difficulty in learning statistical concepts. Associated with the anxiety many students feel towards statistics, is a view of statistics as divorced from the real world, rather than a source of information about the real world. Martin (2003) comments that:

> Students often come to their first statistics class with the preconception that statistics is confusing and dull (abstract para 1).

Often students don't really expect statistical reports to make sense. Consider, for example, the following response on a second year statistics exam question, based on data collected from a sample of Australian women. The question involved a bivariate regression analysis, which showed that women with more pre-school aged children tended to do more housework, but one of the students concluded that:

> 'For each additional hour of housework, on average women have an additional 17 pre-school aged children'.

In any other context, the student would realise that this statement was nonsense – but this was a 'statistics' question, so somehow rational thought had been turned off.

Convincing students that 'statistics' are meant to make sense and that reports should convey something to people with no statistical training is a good first step in any statistics training. The next step is to help students to understand what the analysis is about and what the statistics are actually telling you. Consider the question from an examination in a second year statistics service subject for psychology students, shown in Table 1:

Table 1 *Question On Examination Paper for Second Year Statistics Students*

A researcher was interested in examining how people conform to norms concerning littering. In particular, the researcher wanted to determine whether the tendency of secondary school students to litter depended on the amount of litter already in the area. Students were handed a pamphlet as they entered a room that already had either 0, 4 or 16 pamphlets lying on the ground. Students were then observed to determine whether or not they dropped their pamphlet on the ground. Using the results below, write a report on whether the researcher's hypothesis was supported.

Littering * Amount of litter Crosstabulation

			Amount of litter			
			none	some (4 pieces)	A lot (16 pieces)	Total
Littering	did not drop	Count	102	102	69	273
		% within Amount of litter	85.7%	85.0%	57.5%	76.0%
	dropped pamphlet	Count	17	18	51	86
		% within Amount of litter	14.3%	15.0%	42.5%	24.0%
Total		Count	119	120	120	359
		% within Amount of litter	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	34.045[a]	2	.000
Likelihood Ratio	32.603	2	.000
Linear-by-Linear Association	26.106	1	.000
N of Valid Cases	359		

The following student responses were obtained:

Student 1

> What research has found was that overall students were less likely to drop the pamphlet, with 76% not dropping at all under the three circumstances. (Chi-square Equation $\chi^2=(2) = 34$, $p < .001$).

Clearly this student has no idea what the analysis is about. They are not looking at the relationship between amount of litter present and littering behaviour, and are just blindly reporting the chi-square statistic without any notion of what it conveys.

Student 2

> Generally there was very little difference between the adults who littered and the amount of litter that was around.

Not only does this response show some basic problems in the student's understanding, it is also another example of a student writing something which they would not write in a non-statistical setting.

Even when students understand a particular analysis, they can still find it difficult to write a cohesive report. The response from Student 3 seems to show some understanding of the analysis, but an inability to express it clearly.

Student 3

> The results produced in the crosstabulation only support the hypothesis slightly that the tendency of secondary students to litter depended on the amount of litter already in the area. There was little distribution between the litter dropped in a room with no pamphlets already on the ground and the room with 4 pamphlets already on the ground, reporting a column percentage

> difference of .07% (one piece of litter). However, whilst 14.3% dropped litter in the room with no pamphlets, 42.5% dropped litter in the room with 16 pamphlets. Using the chi squared test, we can see there is a significant difference ($\chi^2(2) = 34.05$, $p < .01$) between the amount of litter dropped depending on the amount of litter already in the area.

Part of the difficulty in writing reports relates to poor English expression skills, as the response from student 4 amply demonstrates (and yes, English is their first language)

Student 4

> The researcher's hypothesis was supported. Dependant upon the amount of litter was around the school children indicated their lack to place their pamphlets in the bin. This is indicated by the crosstabulation table. Correlation is significant as 0.000 (2-tailed). Pearson's Chi-Square value at 34.045.

Added to problems with basic English expression, students often don't understand some of the 'subtleties' of the English language. For example, they find it difficult to see how the following two statements differ: 'There is a slight tendency for older people to watch less television' (correlation is weak) and 'Older people tend to watch slightly less television' (regression coefficient is small).

Even if the student has good basic English skills, there is still the added difficulty of understanding and using statistical terms correctly. One common error is to talk about the 'sample population', and another is to refer to a relationship as 'insignificant' – when they really mean it was not statistically significant. Also, there is often confusion over the term 'variable'. In one example in our second year statistics subject for psychology students, we hypothesise that males are more likely to be satisfied with their cars than females. When we ask the students what are the two variables involved here, it's not uncommon to be told 'male' and 'female'.

One of the most common errors though is to use the terms 'difference', 'relationship' and 'interaction' interchangeably. This leads to statements such as:

> there was no significant difference between age and hours spent watching TV

and when combined with lack of clarity on what a variable is:

> 'Generally there was very little difference between the adults who littered and the amount of litter around'.

The language used in statistics is not self-evident to students; it needs to be taught explicitly.

Having conquered the use of English and of statistical term, there is yet another hurdle for students to overcome. They are often unaware of what belongs in a report and what does not. A common mistake is to try to give the reader a statistics lesson, often at the expense of providing the relevant information, as in the following example. Here students were meant to compare the average hours of work for males and females, and comment on whether the difference is significant:

> Of the responses to this question, the average number of hours worked was 44.09. The mean working week for male respondents was just over 48 hrs, so we would conduct a hypothesis test to see if female employees worked less hours. The fact that some of the 624 factors were marked a -1 may not give a true indication of hours worked, these people could have worked 50 hours each last week but due to the fact no information was given they cannot be taken into account. In this case we would reject the null hypothesis. We can see that, on average, female employees worked less hours then male employees within the company.

HOW DO YOU TEACH STUDENTS TO WRITE COMPREHENSIBLE REPORTS?

While there are many recent articles which emphasise the importance of written communication skills, for example Love (1998), Ludlow (2002) and Roback (2003), they say little about how to develop such skills. Those who do comment on how to develop writing skills

emphasise the need for students to practice, to receive feedback on their initial attempts and to see examples of 'good' reports (for example Smith, 1998; Spurrier, 2001).

The following observations come from many years of teaching statistics to business and psychology students:

- Many students find it difficult to identify the variables in a study, often confusing categories with variables. Time needs to be spent on developing an understanding of what a variable is, and students need practice in identifying the variables involved in specific scenarios.
- There is a tendency for students to rush into an analysis of the results without first reflecting on the aims of the study and what they expect the outcomes to be. Asking students to verbalise their expectations before they look at the data forces them to think in non-statistical terms. It is much easier to say whether the results are consistent with some prior expectation than it is to look at the results of an analysis and ask 'What are these results telling me?' This step also helps students to understand the results they have produced – an essential step in writing a meaningful report.
- The sorts of phrases typically used in describing relationships need to be explicitly pointed out. For example, when looking at the relationship between gender and satisfaction with life, the way the relationship is discussed will depend on whether the satisfaction is measured on a metric scale or a categorical one (satisfied/not satisfied). If satisfaction is measured on a metric scale, then you might report that 'males were on average less satisfied than females', or that 'males tended to be more satisfied than females', but you wouldn't say that 'males were more likely to be satisfied than females'.
- Students benefit from a consistent style of report writing across all examples and solutions to exercises. When reports were written in a variety of styles to demonstrate a range of possibilities, students tended to complain that the solutions were inconsistent. They did not have the experience or skills necessary to recognise that two different phrases might be conveying the same information, and so they could not construct a picture of what information was being included in every report.

Keeping these observations in mind, the following approach to report writing was developed. The sorts of phrases that are used for different types of variables are given explicitly, the features which each report must contain are listed and an example report is given. All of the examples and all of the solutions to exercises follow this same model. Students are strongly encouraged to use the examples as models for their own reports, until they become more experienced and confident with their report writing. This leads to a lot of very similar answers on examination, which is rather boring to mark, but at least the reports (generally) make sense. The students appreciate having a model they can follow which gives them at least one way in which they can write a sensible conclusion or report.

Students are also given a process to follow which leads them through the steps of selecting an appropriate analysis, reflecting on the results and writing a report. Table 2 shows how this process is put into practice for the relationship between two categorical variables. The example used is the question on littering shown in Table 1.

Table 2 *Process for Exploring the Relationship between Two Categorical Variables*

Process in General Terms	*Process in Practice for Littering Example*
What are the variables involved in this study?	'amount of litter present' (none, 4 pieces or 16 pieces) and 'littering behaviour' (drop pamphlet or do not drop pamphlet).
How are they measured (categorical or metric)?	Both variables are categorical.
Is there an explicit hypothesis here – based on previous research? If so, what is it – for two categorical variables, phrase the hypothesis in terms of which group is more (or less) likely to - - -	There is no explicit hypothesis in the question.
If there is no explicit hypothesis – say what you	We might guess that people are more likely to drop the pamphlet

expect to find. This could be a wild guess – use the words more (or less) likely to - - -	if there is a lot of litter present than if there is no litter present.
Who are you comparing? This is the IV, place in columns and request column %	We are comparing participants in the 'no litter' environment to participants in the 'lots of litter' environment, so amount of litter is the independent variable. This is placed in the columns of the crosstabulation and column percentages calculated.
In your initial sentence of expectations which level of the dependent variable were you focussing on? Look at all of the percentages in this row – which group is actually more or less likely to?	In our hypothesis we focussed on how likely people were to drop the pamphlet, so this is the row of the table we'll focus on. Looking at the percentages in this row, we can see that the percentages look pretty much as we expected – people were more likely to drop the pamphlet when there were 16 pieces of litter present than when there was no litter present.
Give an overview statement that comments on this (don't use any percentages, just say more or less likely)	"In our sample of 359 school students, people were more likely to drop their pamphlet when the room had a lot of litter on the floor than when there was no litter on the floor."
Give the percentages to back this up.	"While 42.5% of students dropped their pamphlet when there was a lot of litter on the floor, only 14.3% of students dropped their pamphlet when there was no litter on the floor and 15.0% of students dropped their pamphlet when their was some litter on the floor.
Is there a consistent pattern in other rows? No – then describe other rows as well	Because there are only two categories of the dependent variable here, a consistent pattern emerges for both rows of the table, so there is no need to mention the percentages from the 'non-littering' row – they add nothing to the story.
Use chi-square to comment on the significance of the relationship.	"The relationship between amount of litter and littering behaviour is significant ($\chi^2(2) = 34.05$, $p < .001$)."

Note that this process gives the students a framework around which to write their report, gives the specific language appropriate to discussing the relationship between two categorical variables (more/less likely) and asks the students to focus on what they expect to find – that is, exactly what are they hypothesising. Having a clear picture of the question/hypothesis allows them to write a more meaningful report.

Students are provided with many exercises on which to practice their report writing, and solutions are provided for all of these exercises. They are also given the opportunity to submit answers for correction on an informal basis as well as on regular assignments. The examination questions focus on report writing and the assessment schedules always include some marks for style. To get full marks a report must do more than just contain all of the relevant pieces of information; it must be well written and easy to read. What we choose to include in assessments clearly conveys to students what we value. If we want students to see communication of results as an integral part of statistics, we must include this in our assessments.

CONCLUSION

While there have been many improvements in the teaching of statistics over recent years, and an acknowledgement of the importance of being able to communicate results, little has been said on how to train students to write reports. If we want students to write effective reports on statistical analyses we have to include specific training on report writing as part of our subjects. This cannot be left to some capstone subject at the end of statistical offerings, it must be integrated at all levels.

Students must be given clear guidelines on what reports should contain, what sort of language should be used and a framework around which to construct their reports. Report writing needs to be given a prominent place in exercises throughout each statistics subject, as well as in the assessment. With a greater emphasis on communicating the findings from statistical analyses, perhaps students will be more able to see statistics as an aid to understanding the world we live in, rather than a hindrance.

REFERENCES

Argyrous, G. (1996). *Statistics for social research*. Melbourne: Macmillan

Croucher, J. (2002). *Statistics: Making business decisions.* Roseville: McGraw-Hill Australia.

Francis, G. (2002). Choosing to study independently – when is it a good idea. In L. Pereira-Mendoza (Ed.), *Proceedings of the Fifth International Conference on Teaching of Statistics*, Voorburg, The Netherlands: International Statistical Institute.

Gal, I. & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education,* 2.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: the status of educational reform efforts. *Journal of Statistics Education*, 10.

Gordon, S. (1995). A theoretical approach to understanding learners of statistics. *Journal of Statistics Education,* 3

Heiman, G. (1998). *Understanding research methods and statistics.* Boston:Houghton Mifflin.

Love, T. (1998). A project-driven second course. *Journal of Statistics Education* Vol 6

Ludlow, L. (2002). Rethinking practice: using faculty evaluations to teach statistics. *Journal of Statistics Education,* 10

Martin, M. (2003). "It's like... you know": The use of analogies and heuristics in teaching introductory statistical methods. *Journal of Statistics Education,* 11.

Ritter, M., Starbuck, R., & Hogg, R. (2001). Advice from prospective employers on training BS statisticians," *The American Statistician*, 55, 14.

Roback, P. (2003). Teaching an advanced methods course to a mixed audience. *Journal of Statistics Education,* 11.

Sanders, D., & Smidt, R. (2000). *Statistics, A first course (*6th edn.). McGraw Hill

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education,* 6.

Spurrier, J. (2001). A capstone course for undergraduate statistics majors. *Journal of Statistics Education,* 9.