

Focus on the statistical education of prospective engineers in South Africa

Temesgen Zewotir and Delia North

University of KwaZulu-Natal

Email: northd@ukzn.ac.za

The paper deals with the teaching of statistics to engineering students at tertiary level in South Africa. A number of suggestions are made in order to improve the statistical education of engineering students, thus potentially enabling future prospective engineers to optimise the power of statistics in their profession. Though the focus here is on suggesting ways to improve the statistical education of engineers at tertiary level, current changes in the school curriculum are eluded to, as this adds another dimension to early statistical education of future engineers.

Introduction

Statistics has been described as the science of learning from data. It includes everything from planning for the collection of data and data management, to end-product activities such as the drawing of conclusions from data and the presentation of results.

As is the case in many scientific professions, the engineering profession relies on numerical measurements to make decisions in the face of uncertainty. Whenever there is uncertainty or prediction involved, then statistics, with probability theory as major building block, plays a significant role. This has led to a great demand for familiarity with basic statistical techniques and inference procedures in the workplace. In particular, with the advances in technology and the associated increased ability to produce and process large masses of numeric readings, data handling and statistical techniques play an ever increasing role. It is thus very pleasing to note that the new curriculum currently being phased in at schools in South Africa (Department of Education, 2003), includes data-handling throughout the various levels of schooling. This is in direct contrast to what had been the case prior to the adoption of this new school curriculum.

As is the case all over the world, statistics courses forms an essential part of all engineering programmes at tertiary institutions in South Africa. These courses typically deal with descriptive data handling procedures, probability theory, common univariate distributions, bivariate distributions, estimation of parameters, tests of hypothesis and regression analysis. The course tends to dwell more on theory and less on applications of statistics, thus fostering an inwardly focused approach where theory plays the dominant role,

followed by a few techniques, with the hope that the value of the subject will speak for itself. It is argued here that the underlying purpose is implicit rather than explicit (McLean, 2000). This is the traditional approach to teaching statistics (see, for example, Moore, 1993; Bazargan, 2002; North & Zewotir, 2006). This setting creates a common criticism that undergraduate Engineering Statistics courses are too academic in focus, excessively theoretical, and divorced from real problems that appear in the engineering industry.

Engineering students generally attend statistics modules separately from non-engineering students studying statistics, thus giving the perfect opportunity for engineering-specific examples and applications of statistics to be used in such a module, yet it generally does not happen. Bearing in mind that South African scholars will in future be leaving school with basic statistical literacy skills, it is essential that the teaching of statistics to engineers needs to be modernised so that full use can be made of the higher level of statistical proficiency that will be present upon entering the tertiary institution.

The problem outlined above and need for change in statistical education of engineering students however, is not exclusive to South Africa. Several educators have described the need for specific changes in statistics education for engineers (see, for example, Box, 1990; Bisgaard, 1991; Hogg, 1994; Higgins, 1999; Disney, Bendell & McCollin, 1999; Acosta, 2000; Vardeman, 2002; and references therein). What is surprising, however, is that the discussion forum and research in this regard is limited in South Africa. Moreover, the general unpleasant perception about statistics amongst engineering students in South Africa is labelled and attributed to poor mathematical

background of the formerly disadvantaged racial groups (see, for example, Blinnaut & Venter, 2002; De Wet, 2002; Steffens, 1998). Though the disparate schooling system of apartheid, and its legacy, has its own impact in the South African learners' overall performance, it would be naïve to associate everything which is deficient with apartheid schooling. Students are only able to enrol in the engineering faculty if they meet the basic entry requirements as set by the faculty; in fact, only highly competent matriculants (regardless of their race) will qualify to enter the faculty and thus become prospective engineers.

An overview of the current engineering statistics in South Africa

Bear in mind that at most South African universities and technikons, engineering students' first encounter with statistics courses are at *third year level*. The initial stages of any engineering programme, by necessity, include a vast amount of calculus and numerical analysis. As mathematics is an essential key tool for statistics teaching and learning, it would seem reasonable to assume that teaching statistics to such a group of engineering students should be free of problems caused by poor mathematical preparation and one should thus find such a course to have a high pass rate. Unfortunately, this expectation is far from reality as engineering students find statistics courses difficult with a resulting poor pass rate in such courses.

Despite the effort that instructors of Engineering Statistics devote to the course, many students experience anxiety when they are required to take statistics courses, as these courses are rumoured to be difficult to pass. Cruise, Cash and Bolton (1985) argued that anxious students' image of statistics is generally not a very positive one, with the resulting failure rate of such students being an indicator of the negative effect that the anxiety has on their chances of passing. With this in mind, it is important to examine the failure rate of a typical statistics module to third year engineers at a South African university. As an illustration, we used the failure rate of the same course over a number of years. Note that this module is an essential part of the engineering programme and thus has to be passed prior to them graduating with a degree in engineering.

The number of passes and failures from 1997-2005 academic years is reflected in Table 1. Using the Cochran-Armitage trend test (Margolin, 1988; Agresti, 2002) we analysed the pattern of the failure rate of this

module over the last nine years to see if any significant trend developed over this period. The Cochran-Armitage statistic ($Z=11.2100$ and $p<0.0001$) provides strong evidence of a positive trend. This shows an increasing failure rate for students taking Engineering Statistics courses for the period 1997 to 2005. This is not just by chance, in fact, it is a statistically significant trend.

As is usual when a monotonic effect is observed, the linear logit model (Margolin 1988; Agresti 2002) was fitted with

$$\text{logit}(\pi_t) = \alpha + \beta t$$

where π_t is the failure rate at time $t=1,2,\dots,9$; $t=1$ indicates the academic year 1997, and $t=9$ is the academic year 2005. The results are reflected in Table 2. The estimated multiplicative effect of a unit increase in academic year on the odds of fails is $\exp(0.2137) = 1.238$.

Deviance and Pearson Chi-square divided by the degrees of freedom are used to detect over-dispersion or under-dispersion in the logistic regression. Values greater than 1 indicate over-dispersion, that is, the true variance of the failure rate is greater than what it should be under the given model. If this happens the resulting estimates are consistent, however, estimates of the variance are not. It can result in spuriously small standard errors of the estimates (Barron, 1992). This inconsistent variance estimate invalidates any hypothesis testing. The most common and most widely implemented approach to remedy this is the use of "quasi-likelihood" through the introduction of a scale term into the variance equation. This approach has the advantage that it inflates the variance of each of the observations by a like amount, so that the estimated values will be the same – just the associated standard errors will be

Year	Passed	Failed	Failure rate
1997	174	37	0.17536
1998	234	22	0.08594
1999	149	34	0.18579
2000	191	31	0.13964
2001	119	89	0.42788
2002	216	40	0.15625
2003	163	64	0.28194
2004	138	161	0.53846
2005	228	131	0.3649

Table 1. The number of passes and fails in Engineering Statistics at the UKZN

Parameter	Estimate	Standard error	Wald Chi-square	P value
Intercept	-2.1969	0.1282	293.4256	< 0.0001
Year	0.2137	0.0196	119.4811	< 0.0001
Testing Global Null Hypothesis: $\beta=0$				
Test		Chi-Square	DF	P-value
Likelihood Ratio		130.5489	1	< 0.0001
Score		125.6644	1	< 0.0001
Wald		119.4811	1	< 0.0001
Goodness of Fit				
Criteria	Value	df	Value/df	P-value
Deviance	117.6822	7	16.8117	< 0.0001
Pearson	119.2952	7	17.04217	< 0.0001

Table 2. Logistic regression analysis result

inflated. Logistic regression with quasi-likelihood over-dispersion is implemented in a wide variety of statistical packages, including SAS. Statistical hypothesis tests or confidence intervals using this adjusted fit provide valid inference (Allison, 1999).

The values of Pearson Chi-square and deviance divided by the degrees of freedom are significantly larger than 1. This evidence of over-dispersion indicates inadequate fit of the logit model. Nevertheless, limited inference can be made from the fit. This limited inference is only about the estimates of the parameters as they are consistent. Accordingly, the estimate of the logistic regression coefficient shows an increasing failure rate pattern.

We refitted the model by adjusting for over-dispersion. The result is presented in Table 3. As noted earlier, the adjustment does not change the parameter estimates. The values of Pearson Chi-square and deviance divided by the number of degrees of freedom are close to 1. All the statistical

tests, namely, the likelihood ratio, the score and Wald tests show that the failure rate increases over the academic years. On the average, the failure rate in year $(t+1)$ is $\exp(0.2137) = 1.238$ times year t failure rate. In other words, on the average, failure rate increases 23.8% a year.

Figure 1 displays the observed and logit model fitted values. The plots show the increasing pattern of failure rate. The results from two logistic regression model fits assure the existence of positive trend of failure rate. In the first model there is no allowance for over-dispersion, in the second the quasi-likelihood approach to over-dispersion is employed. All the analyses support our call for revisiting the current offering of Engineering Statistics at tertiary level in South Africa. We know that poor mathematical preparation cannot be the problem, as discussed above, yet there is strong evidence of increased failure rates in Engineering Statistics courses.

Parameter	Estimate	Standard error	Wald Chi-square	P value
Intercept	-2.1969	0.5258	17.46	< 0.0001
Year	0.2137	0.0802	7.11	0.0077
Testing Global Null Hypothesis: $\beta=0$				
Test		Chi-Square	DF	P-value
Likelihood Ratio		7.7654	1	0.0053
Score		7.4548	1	0.0068
Wald		7.1070	1	0.0077
Goodness of Fit				
Criteria	Value	df	Value/df	P-value
Deviance	6.9989	7	0.9998	0.428994
Pearson	7.0948	7	1.0135	0.419077

Table 3. The quasi-likelihood logistic regression analysis result

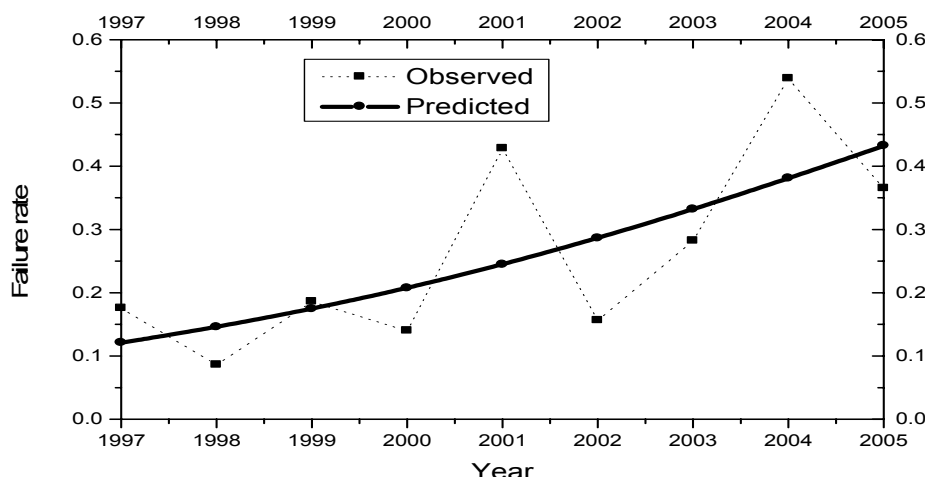


Figure 1. Observed and logistic regression predicted failure rates

According to the Engineering Council of SA records, between 1998 and 2004 50,570 people enrolled at South African universities for engineering courses and 8,900 graduated. This a graduation rate of 17.5 percent across all engineering disciplines. The graduation rate for engineers is even lower at universities of technology. Between 1998 and 2004 there were 139,820 enrolments and 14,250 graduates – a rate of 10 percent across all disciplines (South African Migration Project, 2007). This is further echoed by Boroughs (2007) who states that the work environment in South Africa is continually improving for black engineers, as affirmative action opens up more opportunities, but engineering educators note that the supply of engineering graduates is shrinking.

The reasons for an inability to succeed can be discussed by considering the curriculum, including what happens in individual courses. Steffens (1998) remarked that statistics syllabi in South Africa have traditionally been very theoretical and have deliberately shied away from official (“birth-and-death”) statistics. He also noted that a more balanced attitude has lately become popular internationally. We thus argue that the key to solving the problem of increasing failure rates amongst students in the Engineering Statistics courses may lie in examining the nature of the material in such a course. The overall goal must be to deliver a product which is relevant to the needs of future engineers and to structure the course in such a way as to maximise the possibility of motivating students about the need for statistics in their profession. This will go a long way towards replacing anxiety and negativity with recognition

of the relevance of statistics to their future careers. Clearly the professionals are not interested in the logic of statistical analysis, but will get motivated when learning statistical methods through hands-on experience related to solving problems in their discipline.

Problem-solving approach for engineering students

There is a growing body of literature providing suggestions and discussions strongly favouring the teaching of statistical concepts through a practical approach (Cobb, 1993; Forte, 1995; Rossman, 1995; Moore, 1997; Schaeffer, 1998; Moore, 2000, Gelman & Nolan, 2002; North & Zewotir, 2006), rather than the traditional mathematical approach. The focus of this approach is to promote general classroom activities and discussions on substantive application issues relevant to the students’ field of study, so that the student may *discover* statistical principles and the relevance thereof, rather than being able to prove mathematically why the principles hold. It is thus directed around a problem-solving approach, i.e. data to be collected as the result of a problem/question/statement to be analysed.

It is very pleasing to note that this is the approach that has been outlined in the new National Curriculum Statement (Department of Education, 2003), where a problem-solving approach has been taken throughout the data-handling sections. The added advantage of taking the problem-solving approach to curriculum development at tertiary level as opposed to only at school level, however, is that it lends itself to being discipline-specific and can thus be far more

effective in motivating future engineers as to the power of statistics in their future careers.

We believe that the content of an introductory statistics course for engineers should be determined by the types of problems that engineers are most likely to encounter. Further, we believe that the topics defined by solving such problems should be introduced in a manner that is similar to how these problems would be encountered in practice rather than being presented in a fashion that is determined for mathematical convenience. We are thus in favour of introductory Engineering Statistics courses being driven by problems rather than by techniques, with applied problems, rather than mathematical derivations, forming the basis for such a course. In decades gone by, large data sets were avoided in class as the computational power was a serious time constraint; however, with recent technological advances there is now no need to teach in the classic way.

In order to most effectively modernise a statistics course to engineers, one must start by initiating discussions between the school of statistics and the engineering faculty as a 'buy-in' from both these parties will be necessary prior to achieving the outcomes mentioned above.

The next step would be to consult with the customer to ascertain the exact nature of the desired product in order to be certain of relevance. When deciding on the appropriate material for an introductory statistics course to engineers, one might obtain information from several companies that employ large numbers of engineers as their input is vital when redesigning such a course. Also, the information obtained when performing numeric readings in engineering experiments in other courses is a valuable source of appropriate data capturing opportunities for an introductory statistics course. Above all, the data used must be seen to be collected in order to solve a problem and the student ideally needs to be part of the data capturing process in order for the statistical process to achieve maximum appreciation by the student (Moore, 2000).

Above all, we believe that an introductory course in statistics for engineers must be considered in conjunction with their entire curriculum. No matter how good an introductory course in statistics might be, if students are not asked to use this material in any subsequent courses, they will soon forget it and most probably question why they were required to take the course in the first place. Thus, we propose to enlarge our area of concern from just an introductory course in statistics to how the concepts from this course can be utilised,

reinforced, and enhanced in subsequent engineering courses. The statistical concepts obtained should be an integral part of all laboratory experiences in subsequent courses. All of this necessitates a true collaborative effort between the engineering faculty and the statistics lecturers as they will need to work together in order to determine where statistical techniques can be used and which techniques are most appropriate in other modules not lectured by the statisticians.

Conclusion

The type of introductory statistics course we are proposing will evolve as we gain information from industry and the engineering labs. The effectiveness of the course will increase as statistical content is added to the engineering labs and students are required to use statistical methods in their subsequent engineering curricula. There is no doubt that engineering students will become more motivated about learning statistics if they see the relevance thereof in subsequent modules in their programme, and ultimately, the power of statistics in the field of engineering will be appreciated by them.

The core of the effort would be to develop a revised laboratory program for engineers in order to highlight the benefits to be gained by appropriate utilisation of statistical techniques. In the spirit of continuous improvement and of designing quality into a product rather than trying to address problems after the manufacturing stage, one can emphasise not being satisfied with simply dealing with variability in existing experiments, but rather designing new experiments that better emphasise the engineering issues. As a start, however, it is essential to ensure that statisticians fully understand the engineering experiments as they are currently being conducted.

References

- Acosta, F.M. (2000). Hints for the improvement of quality teaching in introductory engineering statistics courses. *European Journal of Engineering Education*, 25, 276-289.
- Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.
- Allison, P.D. (1999). *Logistic Regression, Using the SAS System-Theory and Application*. Cary, NC: SAS Institute.
- Barron, D. (1992). The Analysis of Count Data: Overdispersion and Autocorrelation. *Sociological Methodology*, 22, 179-220.
- Bazargan, A. (2002). Teaching Statistics to Medical Doctors through Research Methods: A

- Case of Medical Education Research in Iran. *Proceedings of the Sixth International Conference on the Teaching of Statistics*. Retrieved July 20, 2005, from http://www.stat.auckland.ac.nz/~iase/publications/1/4f1_baza.pdf
- Bisgaard, S. (1991). Teaching statistics to engineers. *American Statistician*, 45, 274-284.
- Blinnaut, R.J. & Venter, I.M. (2002). Statistics Teaching Enhanced by Teamwork – a Multicultural Experience in South Africa. *Proceedings of the Sixth International Conference on Teaching of Statistics*. Retrieved July 20, 2006, from http://www.stat.auckland.ac.nz/~iase/publications/1/8g1_blig.pdf
- Boroughs, D. (2007). New Opportunities for South Africa. *American Society for Engineering Education*, 16(9). Retrieved June 7, 2007, from <http://www.prism-magazine.org/mayjune/html/global.html>
- Box, G. (1990). Commentary on Communications Between Statisticians and Engineers / Physical Scientists. *Technometrics*, 32, 251-252.
- Cobb, G.W. (1993). Reconsidering Statistics Education: a National Science Foundation Conference. *Journal of Statistics Education*, 1(1). Retrieved July 22, 2005, from <http://www.amstat.org/publications/jse/secure/v1n1/cobb.cfm>
- Cruise, J.R., Cash, R.W. & Bolton, L.D. (1985). Development and validation of an instrument to measure statistical anxiety. *Proceedings of Statistical Education section, American Statistical Association* (pp 92-98). Las Vegas: Nevada.
- Department of Education. (2003). *The National Curriculum Statement (General): Mathematics*. Retrieved June 07, 2007, from <http://www.education.gov.za/curriculum/SUBSTATEMENTS/Mathematics.pdf>
- De Wet, J.I. (1998). Teaching of Statistics to Historically Disadvantaged Students: the South African Experience. *Proceedings of the Fifth International Conference on Teaching of Statistics, Vol. II* (pp 573-577). Singapore: NTU University.
- Disney, J., Bendell, A. & McCollin, C. (1999). The future role of statistics in quality engineering and management. *The Statistician*, 48, 299-326.
- Forte, J.A. (1995). Teaching statistics without sadistics. *Journal of Social Work Education*, 31, 204-218.
- Gelman, A. & Nolan, D. (2002). *Teaching Statistics: A Bag of Tricks*. London: Oxford University Press.
- Higgins, J.J. (1999). Nonmathematical Statistics: a New Direction for the Undergraduate Discipline. *The American Statistician*, 53, 1-6.
- Hogg, R.V. (1994). A core in statistics for engineering students. *The American Statistician*, 48, 285-287.
- Margolin, B.H. (1988). Test for Trend in Proportions. In S. Klotz & N.L. Johnson (Eds.), *Encyclopaedia of Statistical Sciences, Volume 9* (pp 334-336). New York: John Wiley & Sons.
- McLean, A. (2000). The Predictive Approach to Statistics. *Journal of Statistics Education*, 8(3). Retrieved July 15, 2005, from <http://www.amstat.org/publications/jse/secure/v8n3/mclean.cfm>
- Moore, D. (1993). The Place of Video in New Styles of Teaching and Learning Statistics. *The American Statistician*, 47, 172-176.
- Moore, D. (1997). New Pedagogy and New Content: The Case of Statistics. *International Statistics Reviews*, 65, 123-165.
- Moore, D. (2000). *The Basic Practice of Statistics* (2nd Ed.). New York: WH Freeman & Co.
- North, D. & Zewotir, T. (2006). Teaching Statistics to Social Science Students: Making it Valuable. *The South African Journal of Higher Education*, 20, 503-514.
- Rossman, A.J. (1995). *Workshop Statistics: Discovery with Data*. New York: Springer-Verlag.
- Schaeffer, R.L (1998). Statistics Education – Bridging the Gaps among School, College and the Workplace. *Proceedings of the Fifth International Conference on Teaching of Statistics, Vol. I* (pp 19-26). Singapore: NTU University.
- Steffens, F.E. (1998). Statistical Education in the African Region: Private Experiences in South Africa and Namibia. *Proceedings of the Fifth International Conference on Teaching of Statistics, Vol. II* (pp 571-572). Singapore: NTU University.
- South African Migration Project. (2007). *Demand booms but skills lack*. Retrieved June 08, 2007, from <http://www.queensu.ca/samp/migration/news/article.php>
- Vardeman, S.B. (2002). Providing ‘Real’ Context in Statistical Quality Control Courses for Engineers. *Proceedings of the Sixth International Conference on Teaching of Statistics*. Retrieved July 18, 2005, from http://www.stat.auckland.ac.nz/~iase/publications/1/5e2_vard.pdf