MISUSE OF STATISTICAL METHODS IN BUSINESS APPLIED RESEARCH PROJECTS

WU YUAN and LOI SOH LOI

Division of Actuarial Science and Insurance

School of Accountancy and Business

Nanyang Technological University

MISUSE OF STATISTICAL METHODS IN BUSINESS APPLIED RESEARCH PROJECTS

ABSTRACT

A survey of statistical techniques used in the Applied Research Projects (ARP) was carried out amongst 196 ARP done by Business undergraduates for the academic year 1993/94 and a random sample of 50 ARP for the academic year 1994/95. The study shows that more than 60% of the ARP used statistical methods. It shows that statistics is widely used by our students. However, the reliability and validity of the research finding and the quality of the statistical techniques used in many ARP can be improved. This paper first gives a general picture of the survey and the common statistical methods used in the ARP, and then points out the common and unaware errors and misuses in these ARP. The paper ends with some recommendations and guidelines as references for future students embarking on their ARP.

Introduction

The use of statistics is becoming substantial in all fields of study. It is used in marketing research conducted by companies to establish the feasibility of their products and services and by the Department of Statistics to establish economic indicators. Since statistics is widely used by people to establish conclusions, it is crucial that statistical techniques be employed to the data concerned. Any misunderstanding of statistics might lead to the misuse of statistical methods resulting in wrong conclusions. If this happens, it will be disastrous for the user in terms of the validity of his/her analytical reports.

A survey of statistical techniques used in the Applied Research Projects (ARP) by our students in Nanyang Business School for the academic year 1992-93 was carried out by Loi & Lian (1993). The study shows that more than 70% of the ARP involved statistical analysis. It also shows an enormous diversity of statistical methods used. However, the reliability and validity of the research findings and the quality of the statistical techniques used in many ARP can be improved. Incidentally, even in good international journals, many of the papers subsequently examined by researchers have doubtful statistical results. For example, Schor & Karten (1966) reviewed 295 papers published in ten medical journals and concluded that only 28% of the papers were statistically acceptable, 68% were deficient, and 5% were "un-salvageable".

During the first semester of the academic year 1995-1996, one of the ARP teams did similar survey for the academic years 1993-1994 and 1994-1995. The results are quite consistent with that for the academic years 1992-1993. Loi & Lian (1993) focused on the presentation of data, especially graphic techniques. In this paper, we shall devote our attention to analysing the data. Firstly, we will proceed to identify the common statistical techniques employed by students in the academic years 1993-1994 and 1994-1995. Then, from these common statistical techniques we will highlight some of the common misuses and errors. We hope that by highlighting these misuses and errors, future students would avoid making the same mistakes.

Common Statistical Methods Used In The ARP

In our survey, we examined only these projects done by final year business students. Projects done by accountancy students will be discussed in a coming paper. There are a total of 196 ARP done by business students in the academic year 1993-1994 and 225 in the academic year 1994-1995. All ARP in the academic year 1993-1994 were examined and a random sample of 50 ARP for the academic year 1994-1995 was reviewed. See Tables 1 to 4 for a summary of our findings.

Table 1: No. of Statistical Techniques Used in 1993/1994

	No. of ARP	%
Projects with some statistical techniques	125	64
Projects without statistical techniques	81	36
Total	196	100

Table 2: Statistical Techniques Use in 1993/1994

	No. of ARP	%
Summary table	70	56
Graphs: Pie	34	27
Bar	36	29
Line	22	18
Statistical Inference		
Regression: Simple	7	6
Multiple	14	11
ANOVA	12	10
Student t-test	28	22
Chi-square test	10	8
Pearson Correlation	13	10
Spearman	4	3
Factor Analysis	5	4
Non-parametric	5	4
Others	14	11

Note: The total exceed 100% as multiple selections were allowed.

Table 3: No. of Statistical Techniques Used in 1994/1995

	No. of ARP	%
Projects with some statistical techniques	30	60
Projects without statistical techniques	20	40
Total	50	100

Table 4: Statistical Techniques Used in 1994/1995

		No. of ARP	%
Summary Table		27	90
Graph:	Bar	14	47
	Line	11	37
	Pie	13	43
Statistical Inference			
Regression:	Simple	4	13
	Multiple	7	23
ANOVA		6	20
Student t-test		9	30
Chi-square test		2	7
Pearson Correlation		6	20
Spearman Rank		1	3
Non-parametric		3	10
Factor Analysis		1	3
Others		4	13

Table 1 shows that 64% of ARP in the academic year 1993-1994 used statistical techniques. Among these projects which used some statistical techniques, 56% employed simple descriptive statistics (see Table 2). It can be seen from Table 2 that t-test, regression analysis, Pearson correlation, ANOVA and Chi-square test are commonly used techniques and the t-test is the most popular (22%). Regression analysis, without distinguishing between simple and multiple regression technique, is the second most popular technique (17%). The similarities of the ARP in the academic year 1994-1995 has been shown in Table 3 and 4. In the academic year 1994/1995, the regression analysis (36%) is the most common statistical technique employed, with t-test second one (30%). Therefore, we can conclude that t-test and regression analysis are the most frequently used statistical techniques by students in their ARP.

Errors In Common Statistical Techniques

t-tests

The t-test is used widely to compare two groups of measurements, but often incorrectly. The problems usually relate to the data not complying with the underlying statistical assumption that the two sets of data come from populations that are normal and independent. For example, in some ARP, students collected sets of data of ordinal/nominal scale from the response on certain qualitative factors and the size of data were small, then they used t-test to compare two groups of data. It clearly indicates that the students were not aware of the appropriate situation to use the t-test.

Besides the assumption, whether the two populations have the same variance is often ignored. In fact, there are two different t-tests for comparing two sets of data. One is T_1 which is used when the two populations have the same variance. T_1 is defined as follows:

$$T_1 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}, \qquad \text{where} \qquad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

$\overline{X}_1$ and S_1^2, and $\overline{X}_2$ and S_2^2 are the means and variances of independent samples of sizes n_1 and n_2, respectively. The other is T_2 which is used when the two populations have different variances. T_2 is defined as follows:

$$T_2 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})}}, \quad \text{with} \quad \nu = \frac{(S_1^2 / n_1 + S_2^2 / n_2)^2}{[(S_1^2 / n_1)^2 / (n_1 - 1)] + [(S_2^2 / n_2)^2 / (n_2 - 1)]}$$

degrees of freedom, and $\overline{X}_1$ and S_1^2, and $\overline{X}_2$ and S_2^2 are the means and variances of independent samples of sizes n_1 and n_2, respectively. If we do not have information about the variances of populations, we should use the F-test instead to test (i.e. to test the ratio of $F = s_1^2/s_2^2$) if the two populations have the same variance before we use t-test to test the means.

Another serious error is to ignore the fact that the two groups of measurements obtained from the two populations that are related. In this case, the paired t-test is needed, that is, $T = \frac{\overline{D} - D_0}{S_D / \sqrt{n}}$, where $\overline{D}$ is the sample mean of the paired differences, S_D is the sample standard deviation of the paired differences, and D_0 is the mean of the paired differences for the population(s).

Regression

The rationale for regression analysis is very different. In regression, we are interested in describing mathematically the dependence of one variable on one or more other variables. In simple linear regression we calculate the equation of the best straight line relating to the dependent variable y to the independent (explanatory) variable x. The appropriateness of a linear relationship can be verified easily by means of a scatter plot. The most important underlying assumption in regression is that the dependent variable y is normally distributed with the same variance for each value of x. Regression is used to estimate a dependence relationship. The resulting equation can be used to predict y from x for an individual.

The following are some examples of common misuse of the regression analysis:

(1) Predicting the Y variable for values of the X variable outside the range of the original data set.

(2) The fitting of a straight line where the data show curvature.

(3) The use of a Y on X regression equation to predict X from Y.

(4) The use of simple regression where there are heterogeneous subgroups.

(5) The use of R^2 as the measure of goodness of fit for multiple regression model rather than adjusted R^2. In other words, students are not used to the adjusted R^2 to select a better model with fewer independent variables.

(6) The use of a regression model where R^2 is not significantly different from 0.

(7) Not using dummy variables for nominal variables in the multiple regression model.

(8) Misinterpreting the results from the printout, say the estimated coefficients and beta coefficients.

Pearson Correlation

The differences in statistical analyses of regression and correlation have become greatly confused. This is probably because of the similarity of their mathematical, calculations rather than logic. The correlation coefficient is a measure of the degree of linear association between two continuous variables. If the relationship between the two variables is curved, the correlation may be an artificially low measure of association. Alternatively, the correlation may be artificially high if a few observations are very different from the rest. For these reasons, it is unwise to place any importance on the magnitude of the correlation without looking at a scatter plot of the data.

The main problem is that the test of significance of a correlation coefficient is based on the assumption of joint normality of the two variables. This is characterized by the data points having a roughly elliptical shape in the scatter diagram. If this is not so, the correlation will be misleading and the test of significance becomes invalid. The distributional assumption may be overcome either by transforming the data, or calculating the rank correlation which makes no important assumptions. In our survey, we found that many ARP teams used correlation analysis to ordinal/nominal data sets in their projects.

Correlation is much over-used technique, with the significant correlation coefficient often wrongly interpreted as important and , even worse, as necessarily indicating a causal relationship. Correlation is merely an investigative analysis, suggesting areas for further research and formulation of hypotheses rather than for testing them.

Chi-square test

The chi-square test is normally used for the following tests:

(1) The goodness of fit of a set of data to a specified probability distribution

(2) The independence of two variables

(3) Homogeneity for two or more populations

Generally, students showed a clear understanding of the usage of this test, but some did not know how to state their hypotheses or how to state them clearly. Consequently they may make wrong decisions. Another minor mistake is that the expected frequencies or the observations were too few.

ANOVA (Analysis of Variance)

ANOVA testing is an extension of the t-test for two populations that are normally distributed. It involves testing the means of two or more groups and is used only when all the variances are assumed to be equal. A common test for testing the equality of variances is **Hartley's test** (Ott, 1993).

The common error in ANOVA is forgetting to check the assumptions when using it. The three assumptions are: (1) normally distributed values in each group; (2) homogeneity of variances. (i.e. all populations have the same variance.); and (3) independence of errors. All students who used the ANOVA in their ARP failed to state and check these assumptions before proceeding.

Recommendations

In this section, we will attempt to give some guidelines and recommendations as references for future students to refer to before embarking on using statistics for their ARP.

Population and Sample

Because this is avast topic, we will merely highlight a few important points. In our survey, we found that a large proportion of students neither stated the target population for their study nor described how their samples were selected randomly. For instance, the target population in many ARP was NTU students. Our impression is that students distributed and collected their questionnaire to and from their friends, their classes, or their friends' classes, etc. It is natural that people will doubt if these "convenience" samples could represent the population of NTU. Subsequently people will distrust the conclusions made by the ARP because the samples could be seriously biased.

Rarely can we collect data on all the subjects of interest in a particular study. Samples provide a practical and efficient means to collect data. The sample serves as a model of the population. However, for us to extend our findings to the population, the model must be an accurate representation of the population, that is, the sample should be a random sample.

How large a sample size in a study should be is usually the first question addressed by a study team. There are many ARP teams who give little thought to sample size, choosing the most convenient number (say, 20, 50, 100, etc) for their study. Sample size is the most potent method of achieving estimates that are sufficiently precise and reliable for scientific inquiry. Small sample size may contribute to a conservative bias (Type II error) in the application of a statistical test. Increasing sample size obviously has a cost. Larger samples require more expenditure for collecting data, especially when interviews are used; following up on nonresponse; and coding and analyzing data. So choice of a sample size cannot be considered in a vacuum. A trade-off in cost, total error, and other design choices must be considered.

To begin the process of choosing a sample size, a number of factors must be examined sequentially. Prior to the process, one must determine the tolerable error of the estimates or power of the analysis. The computation of an efficient sample size for descriptive studies begins with the tolerable error: the standard error times the t-value for the selected confidence level. The variance or standard deviation of the variable must be estimated. Usually the standard deviation can be estimated from a previous study or the use of a small pilot study. For analytical studies, efficient sample size calculation is based on the power tests. A power test is use to indicate whether a particular sample size is sufficiently sensitive to detect the expected effect (Lipsey, 1989). The larger the standard error or the smaller the sample size, the more difficult it is to reject the null hypothesis. Failure to reject the null hypothesis when it is in fact false is described as a Type II error. The idea behind using the concept of power to calculate sample size is to maximise the chances of finding a real and important effect if it is there, and to enable us to be reasonably sure that a negative finding is strong grounds for believing that there is no important difference.

Analysis

The basic principle to be adhered to with respect to the statistical aspects of the ARP is that the methods should be described in sufficient detail to be fully understood, and so that anyone else with access to the raw data could, if desired, reproduce the results. The use of unusual forms of analysis should be justified, preferably with a reference, but all analyses done should be clearly described. One should describe clearly and exactly what was done. It may be necessary to demonstrate the validity of the assumptions for some analyses, say t-test, ANOVA and regression analysis.

Statistical Methods

Here are a few points to note when using any statistical technique. They can be applied to all methods and students should clearly state these points in their research. They are :

- Assumptions must be clearly stated if there are any.
- Both the null and alternate hypotheses should be stated and explained if possible. For example, if the null hypothesis is "employee growth will not change", then the term "employee growth" should be defined and explained.

- The level of significance must be specified. This is important as this would determine whether the null hypothesis will be rejected or accepted at the given level.
- The critical value should be included in the report. This value separates the rejection and non-rejection regions. The purpose of doing this is to help the reader understand why the author arrived at the particular conclusion for the statistical test.
- The nature of data determines the different types of tests. In general, for interval/ratio data, parametric tests should be used. For ordinal/nominal data, non-parametric tests (Siegel and Castellan, 1988) should be used. Table 5 lists the types of statistical techniques and their application in relation to the nature of data.

Table 5

	Dependent Variables are Nominal / Ordinal	Dependent Variables are Interval / Ratio
Independent Variables are Nominal / Ordinal	**Contingency Tables**	**Analysis of Variance**
Independent Variables are Interval / Ratio	**Discriminant Analysis**	**Regression Analysis**

Appendix and Bibliography

Raw data or any other supporting documents should be included in the appendix to validate the study. Reference materials and other correspondences related to the study should be included too.

Guidelines on the Use of Statistical Inferences

When using t-test, the following questions need to be asked:

- Is the data in of interval / ratio or nominal / ordinal form?
- Is one-sample or two-samples t-test required?
- If two-samples t-test is needed, are the populations related or independent?
- If the two samples are independent, are the population variances equal?
- If variances are assumed to be equal, is there any justification for it?

For Chi-square-test, the data can be in any form as it is primarily a non-parametric test. For test of independence, the data for the contingency table should be in frequencies instead of means. Moreover, the factors examined should be subdivided.

When using regression, students should plot the scatter diagram to supplement the regression result. This is to clarify the relationship between the independent variable(s) and the dependent variable. The data used, as stated in Table 5, should be in interval/ratio form.

ANOVA can only be used if the data is in interval/ratio form. The two common types of ANOVA must be stated to help the reader know the purpose of the investigation. There is an essential difference between the completely randomised design and randomised block design. A completely randomised design or one-way ANOVA is a test for the difference in means of several groups based on only one factor. The groups may be types of tire, brand of drug, etc.

A randomised block design involves repeated measurements or matched samples in order to evaluate the differences between two treatment conditions. The purpose of blocking is to remove as much variability as possible so that we may focus on the differences among the treatment groups. In comparison to the completely randomised design, the randomised block design is a more efficient analysis that reduces experimental error so as to produce more precise results. The randomised block design is used to test whether there are any treatment effects and block effects.

Conclusion

After reviewing the entire applied research projects (business) done for the academic year 1993/1994 and the sample of 50 applied research projects for the academic year 1994/1995, we found that more than 60% of ARP used statistical techniques. A large proportion of these projects had more or less statistical errors or misuse. Some of them had serious errors in which the students applied inappropriate statistical techniques to the data collected. Some of them did not present their findings and conclusions in an easily understandable form for the reader. The incorrect analysis of data is probably the worst misuse of statistical methods. The mishandling of statistical analysis is as bad as the misuse of any laboratory technique. It is of no value using good statistical techniques to analyze "poor" data (serious biased), as is analysing "good" data with inadequate or invalid statistical techniques.

There needs to be a greater appreciation of the importance of correct statistical thinking, and an improvement in the standard of ARP so that the errors discussed can be eradicated. We hope our findings and discussion will help future undergraduates who are embarking on their ARP to improve the quality of their ARP.

References

Lipsey, M. W. (1989). *Design Sensitivity: Statistical Power for Experimental Research.* Newbury Park, CA: Sage.

Loi, S. L. and Lian, R. (1993). *Statistical Techniques Used in Final Year Project.* Working Paper Series No. 33-93, SAB, NTU, Singapore.

Ott, R.L. (1993). *An Introduction to Statistical Methods and Data Analysis, 4th edition*. Duxbury Press: An Imprint of Wadsworth Publishing Company.

Schor, S. and Karten, I. (1966). *Statistical Evaluation of Medical Journal Manuscripts.* J. Am. Med. Ass., 195, 1123-8.

Siegel, S. and Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences, 2nd edition*. New York: McGraw-Hill Book Company.