SIMPLIFYING REGRESSION MODELS USING DIMENSIONAL ANALYSIS

G A Vignaux
Institute of Statistics and Operations Research,
Victoria University of Wellington,
PO Box 600, Wellington, New Zealand

J L Scott
Department of Management Systems, University of Waikato,
Private Bag, Hamilton, New Zealand

September 12, 2001

SUMMARY

Dimensional Analysis can make a contribution to model formation when some of the measurements in the problem are of physical factors. The analysis constructs a set of independent dimensionless factors that should be used as the variables of the regression in place of the original measurements. There are fewer of these than the originals and they often have a more appropriate interpretation. The technique is described briefly and its proposed role in regression discussed and illustrated with examples. We conclude that Dimensional Analysis can be effective in the preliminary stages of regression analysis when developing formulations involving continuous variables with several dimensions.

1 Introduction

Henderson and Velleman (1981) argued that, instead of automatic stepwise regression, one should use an interactive approach. They point out that "the data analyst knows more than the computer" and "can use knowledge of the subject matter" to guide the analysis. In support of that contention we will show that the physical dimensions of variables involved in the problem also contain information which can be used to formulate models. Such models will satisfy the first and essential requirement of any model: that it be dimensionally homogeneous. The method used is Dimensional Analysis.

Dimensional Analysis, often shortened to the conventional acronym, DA, has been a working tool of physicists and engineers for many years (for example, Rayleigh (1915), Bridgman (1963), Bender (1978), and Yalin (1971)) but its use has also been advocated in other disciplines. Stahl (1961, 1962) discussed the use of DA in mathematical biology in two major papers and Rashevsky (1960) also uses the technique in his book on mathematical biophysics. Naddor (1966) and Vignaux (1986) argued for its use in Operations Research; Finney (1977) considers the dimensions of statistics, showing how DA can provide rapid, useful checks for statistical algebra; while Rutherford (1975) shows there is a place for DA in data analysis and in planning experiments to reduce the number of factors to be considered.

These works focus on DA in model building. Our paper takes a different slant, discussing DA and its potential application in data analysis and regression. We demonstrate its use, showing both the reduction in effort and more effective modelling that can result.

We begin with a summary of DA and the dimensional issues that need to be considered when building a regression model. We then demonstrate its use on the Black Cherry Tree data published by Ryan et al. (1976) and show that the model form that they and Atkinson (1982) found is generated almost automatically from a preliminary analysis of dimensions. We then present a second illustration using automotive data previously published in Biometrics. We conclude with comments on the process, the benefits, and the limitations of the approach.

2 Dimensional Analysis

In essence, DA restructures the variables of a model into a smaller number of independent dimensionless products using the constraint that all terms of the model must have the same dimensions (the requirement for dimensional homogeneity). For the reader unfamiliar with the process, the method of finding these dimensionless products is described in most applied mathematics and physical modelling textbooks, such as Langhaar (1951). There are also a number of books written about DA which explain this simple process (e.g., Huntley (1967)). In this section, therefore, we only outline the method while briefly discussing the issues that need to be considered when using DA to build a regression model. Reading this section in conjunction with the later illustrations should give a good feel for the suggested process.

Each physical variable, x_i, is measured in units with dimensions composed of a few fundamental dimensions such as length, mass, and time, conventionally represented as $[M], [L]$ and $[T]$. For example, a wind speed has dimensions of a length divided by a time, conventially represented as $[LT^{-1}]$, a mass has dimensions $[M]$, a density dimensions of $[ML^{-3}]$, rainfall, measured in volume per square metre, has dimensions $[L^3L^{-2}]$ or $[L]$. In DA this is a fundamental feature. We are not interested in whether we measure mass in kilograms or pounds, length in metres or feet, or time in minutes or seconds. We can even measure some lengths in metres and some in feet. The only result of this is that an additional purely numerical factor appears in the model. In cases where heat and temperature are important we may have a fundamental dimension of temperature $[\theta]$.

In problems in economics or operations research, we often add a fundamental dimension of cost with dimensions [\$] represented in objective functions. The cost of material might have dimensions $[\$M^{-1}]$. For an inventory theory example where [\$] is used see Naddor (1966); Vignaux (1986); Vignaux and Jain (1988).

In addition to the variables set out in the problem, there may also be some dimensional constants such as viscosity, rate constants, density of fluids such as water or air, and the acceleration of gravity. These constants are those which occur naturally in the area of discourse. Often they correspond to the recognition of a physical relationship known to be involved such as Stokes' law for the drag experienced by small bodies in moving fluids. Clearly, the analyst's background knowledge of the phenomena involved in the problem will influence the choice of dimensional constants.

In addition to the dimensional variables and constants there may also be some appropriate non-dimensional values such as ratios or scale factors. An aspect ratio of height to width of a shape or a ratio of two velocities are typical examples. Give these the symbols R_k. They will take no part in the Dimensional Analysis itself but will re-enter the procedure when we come to describe and fit the final model. For the moment, we leave out the R_k from the analysis.

Every term in *any* valid physical relationship or equation between the variables must have the same dimensions. Apples can only be added to apples not to oranges. From this observation, Buckingham (1914) showed, in his famous Pi Theorem, that the relationship can be reformulated as a function of a set of *dimensionless* products of the variables. The original relationship of n dimensioned variables is written as the equation

$$f(x_1, x_2, \ldots, x_n) = 0. \tag{1}$$

If this is dimensionally homogeneous (that is that the dimensions of all terms are the same) the Pi theorem states that we can express it as a new function of a set of dimensionless parameters, written conventionally as πs,

$$\phi(\pi_1, \pi_2, \ldots \pi_{n-m}) = 0. \tag{2}$$

There are only $n-m$ dimensionless products π_j of the original n x_i. Here m is the number of fundamental dimensions (e.g., $[M], [L], \ldots$) in the relationship.

Sets of π_js can be found in a relatively mechanical manner[1], described in many books on DA, as mentioned earlier. We then have a new function that is equivalent to the old one but with fewer variables. However, while the number of π_j is fixed, there can be many sets of these products. We can use this to highlight those original variables or those effects we consider may be important by manipulating the π_j through raising them to powers, multiplying them or dividing one by another. Again, knowledge of the subject is helpful. Some well-known dimensionless products such as Reynold's number appear again and again in some subject areas. They are recognised as representing the relative power of two important effects (for Reynold's number, the inertial and viscous forces in fluids and hence the tendency for turbulence). In summary, we choose a set that is meaningful to the case.

To complete the model to be fitted we complete our new function of π_j by including the original R_k, the dimensionless values of the original problem set aside previously, giving the dimensionally homogeneous equation

$$h(\pi_1, \pi_2, \ldots, \pi_{n-m}, R_1, \ldots, R_k) = 0. \tag{3}$$

We now have a new data modelling or regression problem with up to m fewer variables than the original. This in itself is a considerable advantage. Indeed

[1] Briefly, to satisfy the requirement that the relationship be homogeneous in each of the m fundamental dimensions we choose a basis set of m (dimensionally) independent x_i and express the remaining x_i in terms of this basis

it may even do away with the need for a regression at all. For example, the standard physics demonstration of DA includes the derivation of the formula for the period of a pendulum which collapses from a 4-variable problem to one involving only the determination of a single (dimensionless) constant. A similar collapse is seen in the derivation of the optimum lot-size in deterministic inventory problems and is shown by Naddor (1966) and Vignaux (1986); Vignaux and Jain (1988). One problem analysed in this paper collapses in a similar way.

3 DA and Regression

From the viewpoint of dimensions, ordinary regression is a simplistic way of inventing new pseudo-laws because it will take variables of any dimensions and invent dimensioned constants – the regression coefficients – to form a correct dimensional relationship, one that is homogeneous. These constants will change in value with changes in measurement units (kg to pounds, for example) and thus they are not true constants of nature. Any transformation of the data before analysis is usually dimensionally incorrect. For instance, taking logs of height variables is equivalent to a series of power terms of height, each with different dimensions. This destroys dimensional homogeneity.

In contrast, using the dimensionless variables constructed by a DA, any coefficients in a fitted model are also dimensionless and will not change if the units of measurement are changed. Further, any transformations are legitimate and the model remains dimensionally homogeneous. For example, taking logs of a dimensionless variable is quite legitimate. What is more, the π_j will include fundamental nonlinear relationships between the variables as our illustrations will show.

Having carried out a DA we can now use all the power of traditional data analysis and regression to find and fit the unknown relationship using the new π_j variables. DA cannot disclose the final form of the function nor any of the numerical constants Typically this analysis involves trying several simple expressions using the chosen π_j. We can now confidently carry out our data analysis in the knowledge that any function we find is guaranteed to be dimensionally homogeneous with the advantages described above.

To illustrate the use of DA in regression we look at a simple example with only one fundamental dimension which illustrates both the simplicity and speed with which it develops a suitable solution to the fitting problem. We then go on to a more complicated problem involving car petrol consumption.

4 The Black Cherry Tree Data

The Black Cherry Tree data gives the volume, v (cubic feet), height, h (feet) and diameter, d (inches) of a sample of 31 trees from the Allegheny National Park in the USA. It was published in the Minitab Handbook as a student example of regression by Ryan et al. (1976) and is reprinted in Hand et al. (1994) as data set 210.

There have been several published analyses of the data (Atkinson (1982); Atkinson (1994); Cook and Weisberg (1983); Everitt (1994)) starting with a simple linear regression of v against d and h, and through trial and error evolving into a regression including a d^2 term which Everitt found significant. Atkinson (1982) concluded that transformation to $v^{1/3}$ gave a better fit. This resulted in a regression model of the form

$$v^{1/3} = \beta_0 + \beta_1 d + \beta_2 h + \beta_3 d^2. \tag{4}$$

Atkinson (1994) examined both a model in $\log v$ and using the "conical" model, discussed below.

In the next section we show how DA can be applied to the same problem. It reaches a similar equation directly with an ultimate form reflecting nature and an improved fit over (4).

4.1 A Dimensional Analysis

We are given that the volume, v, is connected by some unknown function of height, h, and diameter, d. Write this in the form

$$f(v, d, h) = 0. \tag{5}$$

Statistical error is ignored at this point and will be brought in at the regression stage. We now force the requirement that this function be dimensionally homogeneous. On examining the dimensions of the variables v, d, and h we note that they all involve a single fundamental dimension, L. Thus d and h have dimensions L and v has dimension L^3. Buckingham's Pi theorem shows that the function of 3 dimensioned variables in Eq. 5 can now be rewritten in terms of only $3 - 1 = 2$ dimensionless products.

For our model, out of the possible sets of dimensionless products that the DA could generate, we arbitrarily chose

$$\pi_1 = \frac{v}{h^3}, \quad \pi_2 = \frac{d}{h}. \tag{6}$$

π_1 looks like a shape factor: it is the ratio of the actual volume of a tree to that of a cube with the dimensions of the tree's height. π_2 is a shape factor indicating the ratio of diameter to height – pictured as the relative "thickness" of the tree.

Another useful pair of dimensionless products might have been v/d^3 and d/h and this, it turned out after the analysis, would actually have been somewhat simpler to fit[2]. However we retain the original choice as our example.

We can rewrite Eq. 5 using the new variables as

$$\pi_1 = \phi(\pi_2). \tag{7}$$

This is equivalent to Eq. 2. $\phi()$ is the unknown function to be determined by regression or from other considerations.

By substituting for the π_js, Eq. 7 is equivalent to the nonlinear equation

$$\frac{v}{h^3} = \phi(\frac{d}{h}). \tag{8}$$

We could start our fitting by assuming a few simple models for $\phi()$ such as a constant (dimensionless, of course) or expressions like $a + b\pi_2$, $b\pi_2$, $a + b\pi_2 + c\pi_2^2$ or even $a + b/\pi_2$. In these expressions, the parameters a, b, c are also all dimensionless.

4.2 Consequent regression studies

We carry out a regression analysis using π_1 and π_2 instead of v, h, and d. Fitting the models in turn from the list above, we obtain a satisfactory fit at the third attempt. Neither the intercept term nor the linear term are significant.

Substituting for the πs, we find the fitted model is

$$\frac{v}{h^3} = 0.302(\frac{d}{h})^2. \tag{9}$$

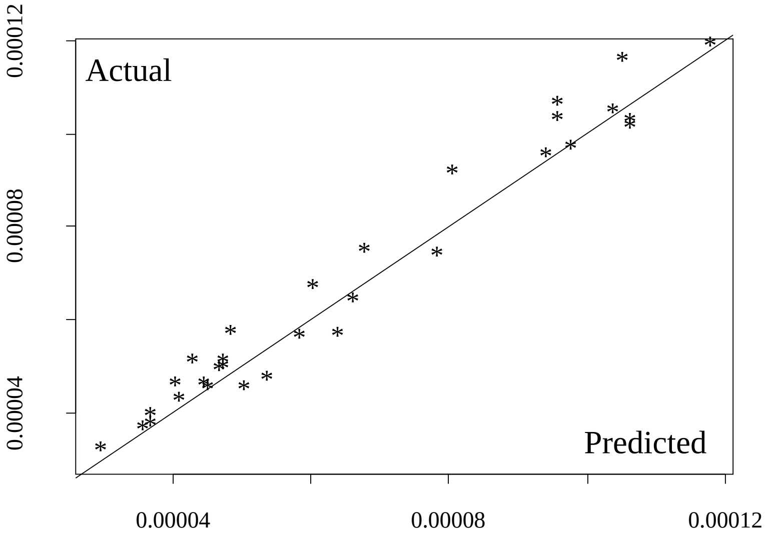
[2] Since the final relationship would then have been linear between the pair

This can be re-expressed as

$$v = 0.302hd^2 \tag{10}$$

The standard error of the estimated constant is 0.0039. In the discussion to Atkinson (1982) Professor Sprent suggested that a “forester or biologist will certainly be attracted to the model of a tree as something very like a cone”. Our solution has exactly this form though the constant is different from that for the cone ($\pi/12 = 0.2618$). A plot of π_1 against predicted values is shown in Figure 1.

Figure 1: A plot of actual against predicted values of π_1 for the dimensionless quadratic regression.



With a good fit evident, the analysis has stopped at this point. If this was not the case, other simple models from the list would have been tried. (No dimensionless ratios or scale factors were introduced as they appeared unnecessary)

5 A Further Dimensional Analysis

A more controversial form of DA, called “vectorised” or “directional” DA allows one to ascribe different fundamental dimensions to different directions of length (labelled x, y, and, vertically, z). Vertical lengths, such as heights, would have dimension L_z and this is different from that along a horizontal axis, L_x. Thus h would have dimensions, L_z. Diameter, d, being measured horizontally, has an overall dimension of a single length, L but as it shares dimensions along the x and y axes equally, it is given a dimensional formula $\mathrm{L}_x^{0.5}\mathrm{L}_y^{0.5}$. Volume, v would have dimensions $\mathrm{L}_z\mathrm{L}_x\mathrm{L}_y$ rather than L^3. Massey (1978, 1986) argues strongly against this technique but, if it is valid, and many analysts find it so, it has the considerable advantage of increasing the number of fundamental dimensions with a corresponding decrease in the number of final dimensionless products.

In our problem, using vectorised dimensions results in a *single* dimensionless variable, π_1. The resulting equation can be re-expressed as:

$$\pi_1 = \frac{v}{hd^2} = a, \tag{11}$$

where a is an unknown constant. This equation is the same as Eq. 10 which we arrived at earlier by a more roundabout route. The "vectorised" DA gave this result instantly. Given Eq.11, we would only have to determine the value of the constant.

6 Example - Hocking's Automobile Data

This example illustrates the use of DA as a preliminary analysis in a more complicated setting. This data for a number of different automobiles from 1974 *Motor Trend* magazine was used in *Biometrics* as a test case by Hocking (1976) to investigate automatic regression methods. The objective was to find an expression that best predicts miles per gallon or gas consumption from the other known variables. Henderson and Velleman (1981) also examined this data. They argued for a philosophy of computer assisted data analysis - a collaboration between the analyst and the computer in an interactive mode. This is in accord with the philosophy of 'data analysis' rather than blind regression.

6.1 Examining the variables

The data (reproduced in Henderson and Velleman (1981)) contains 11 variables, of which 6 are already dimensionless, being either ratios (such as the

Table 1: The Dimensional variables and their dimensions

symbol	*variable*	*Description*	*Dimensions*
m	MPG	miles per gallon	$[LG^{-1}]$
h	HP	horsepower	$[ML^2T^{-3}]$
w	WT	weight	$[M]$
T	$QSEC$	Quarter mile time	$[T]$

Final drive ratio, with the symbol DRAT in the original dataset) or numbers (such as the Number of cylinders, symbol CYL). Four are dimensional in nature: Miles per gallon, MPG, Horsepower, HP, vehicle weight, WT, and Quarter mile time, QSEC. The remaining variable, Cylinder displacement, symbol DISP is intermediate in nature in that it has dimensions (cubic inches in this case) but is really a surrogate for the more fundamental variable, engine size. A measure of engine size would be preferable and is included in some similar sets of consumption data. It would therefore have been included had it been available. For this reason, DISP should really be included. On the other hand, since it is likely to be strongly correlated with HP, we have dropped it from the analysis. Table 1 lists the four dimensioned variables, the symbols we will use, and their dimensions.

DA can only work on the dimensional variables, attempting to link them together in a way that is physically sensible. The remaining dimensionless variables and ratios (the Rs of Section 2 and the surrogate variables) are, of course, appropriate candidates for inclusion in a data analysis or a regression; they are of the correct dimensionless product form that we hope to construct from the other variables and may carry important information about the behaviour of the system. They are included later.

If one considers how these four variables might be collected together into the form of dimensionless products, one immediately is struck by the fact that m [LG^{-1}] cannot be combined with the others sensibly because of the [G] (gallons) term which does not appear in any of the other three. In many DAs we would immediately remove m from consideration and try and combine the rest into a dimensionless group. Unfortunately, in this case that is the dependent variable of the problem and it is hardly wise to remove it right at the start.

A gallon is usually a volume measure [L^3]. As used here, it measures the volume of fuel (as in MPG) - a combination of volume and energy and therefore a different measure of volume to the length based one; one therefore deserving of the separate dimension [G]. We introduce a new dimensional variable, e [ML^{2}T^{-2}G^{-1}], the energy content of a gallon of fuel. In this problem e is constant but could be a real variable if a more extensive data set contained examples with different fuels. m [LG^{-1}] e always occur together in the analysis.

A dimensional analysis will connect these variables together to produce dimensionless products but it can do better if further physical information is supplied. Such information include the obvious relationship between horsepower, acceleration and mass, the possible effect of rolling or frictional resistance, and the effect of air resistance. All these will have "laws" that will connect the variables to each other or to new dimensional constants.

6.2 A Dimensional Analysis of the car data

Not only should we include the energy coefficient of petrol, e, but we should also involve the acceleration of gravity, $g[LT^{-2}]$, and a variable, $d[L]$, representing the "quarter-mile" distance over which QSEC is measured. We are then left with 7 dimensional variables and constants. Buckingham's theorem tells us that if they are to occur in a physical relationship they must occur in the form of dimensionless products. We have 7 variables and 4 fundamental dimensions (M, L, T, G) giving us $7 - 4 = 3$ dimensionless products to connect them.

A dimensional analysis reveals the following dimensionless products. To ensure functional independence, we have chosen a particular set of π_j so that the major variables m, h and T are constrained to have power 1 in respective π_j.

1. the first represents the effect of energy use and weight, perhaps due to rolling or frictional resistance:

$$\pi_1 = \frac{mwg}{e} \tag{12}$$

2. the second represents the effect of power and weight on acceleration:

$$\pi_2 = \frac{h}{wd^{0.5}g^{1.5}} \tag{13}$$

3. the third is a little difficult to understand because T has been forced to have power 1. $\pi_3^{-2} = d/gT^2$ would be more physically understandable as representing a dimensionless measure of acceleration.

$$\pi_3 = \frac{Tg^{0.5}}{d^{0.5}} \tag{14}$$

6.3 Regression with the new dimensionless products

DA can get us no further. In the absence of further physical or engineering input, we must leave the final determination of the relationship between these factors and the other dimensionless factors of the original problem to the usual methods of regression.

The problem is to find a prediction of m from values of the other factors. Expanding π_1 and rearranging, we have:

$$\frac{mwg}{e} = g(\pi_2, \pi_3, R_1, R_2), \tag{15}$$

Where $g()$ is some unknown function, R_1 is DRAT, the Final Drive ratio, and R_2 is CYL, the Number of cylinders, and there could be other similar dimensionless terms. Re-expressing m, this is:

$$m = \frac{e/g}{w} g(\pi_2, \pi_3, R_1, R_2.) \tag{16}$$

From this point DA has nothing more to contribute. We would therefore analyse and fit Equation 15 by adding an error term and using normal data analysis and regression methods. These can include fitting a nonlinear model of the set of π_j and the R_k. This is legitimate because it does not break dimensionless homogeneity.

After examining the automobile data, Henderson and Velleman suggested that it might be better to use m^{-1} (gallons-per-mile) as the dependent variable and indicated a “loose theoretical argument” from the physics of the problem suggesting that this should be proportional to w. Their argument is, we believe, stronger than they claim, if we consider the dimensions of the variables. The DA shows that the form mw is indeed an appropriate combination and appears as the main component of our π_1. e/g is, of course, a constant for this set of data.

Further analysis by Henderson and Velleman showed that the next most important combination was the ratio h/w (using our symbols). This is suggested by the DA which gives this ratio as the variable part of π_2. In fact they showed a strong relationship between m^{-1} and w and h/w (their Table 2) which is suggested in equation (15).

Aitkin and Francis (1982), in response to the paper, showed a regression model based on the logs of the same variables and indicated that DRAT, the Final Drive Ratio, (R_1 in our model) was also important. In their response, Henderson and Velleman relate that they had consulted an automotive engineer who suggested the theoretical relationship $m^{-1} \propto w$. We derived this easily from the dimensional analysis.

7 Conclusions

The DA method reveals how an equation describing a physical system can be reduced to a function of a set of dimensionless products which are usually fewer than the prescribed set of explanatory variables. This may even reduce the problem to simply determining a constant in exceptional cases.

Used in regression, DA focuses on fewer terms than conventional regression and forces the analysis to reflect nature, rather than look for chance relationships. The models generated are guaranteed to be dimensionally homogeneous. Further, the constants determined in fitting the models are, unlike in standard regression analysis, true dimensionless constants, unaffected by changes in units of measurement.

Further benefits indicated by our illustrations are that DA can suggest other variables that might be useful for a more complete model (e.g., fuels with a range of energy content in the automobile example); inherent nonlinearities are implicit within the dimensionless terms and do not have to be discovered by data analysis (e.g., the d^2 component in the Cherry Tree example).

"Vectorised" DA, while its theoretical base is contestable, gives a smaller set of π_j which were very effective, at least in the Black Cherry Tree case. Conventional DA will include this smaller set but, if quick results are sought, then "vectorised" DA is available.

Potentially, therefore, DA is a powerful preliminary stage to conventional regression analysis. The method is particularly powerful in situations with continuous variables and several dimensions. No single recipe of sequential steps is available but there are certain ingredients:

Firstly, the process needs a thorough understanding of the problem environment. From this we decide on the key variables to be used in the model. The dimensional units for each variable are then established. Potential dimensional constants are then judiciously added.

A set of dimensionless products (the π_j) are next found using the standard techniques. The approach to choosing the π_j relates to choosing variables in regression studies. Certain of the original x_i variables are likely to be the ones with important effects. Choose a set of π_j where those originals are separated, each π_j carrying one of the important variables.

However, the set of π_j is not unique; by changing the basis, other, equally valid, sets of products can be found and there is some uncertainty about the best set to use in a particular problem. Further research is needed here.

We then add potential ratios or scale factors R_k.

Simple forms of equations using the π_j and R_k are then generated. These are tested by regression analysis and the results compared in the usual process of fitting different models. When a good fit is not obtained then the stages offering analyst choice need to be revisited. While no natural order is (clearly) evident, we suggest working back up:

- look for other potential ratios or scale factors,
- choose a different set of π_j,
- look for other candidate dimensional variables and repeat the DA.

The DA approach has its limitations which we now discuss briefly:

- It can be applied only to systems involving physical quantities, several dimensions, and with positive, continuous variables.
- As with any method of analysis, it depends on being furnished at the outset with all the physical explanatory variables affecting the situation. If one fails to include an important effect, the model will attempt to compensate but cannot replace real knowledge. We will always need some understanding of the effects that are in operation.
- It will seldom reveal whether or not any variables have been omitted. If inconsistencies occur, such as one dimension associated with only one variable, then DA reveals this as the automobile example shows.
- It cannot handle made-up, or surrogate factors easily. These should be replaced by their physical representations.
- It does not provide numerical values as does a complete mathematical analysis. These have to be determined externally by regression or data analysis, once DA has been applied.

- One factor that must be considered is the error structure of the resulting models. For example in the Car Data, minimising the errors in the dependent variable m will not necessarily result in the same model as that if we minimised the errors in π_1 which combines m with other variables. DA offers no particular insight into the quality of the error structure but this is really no different from any other transformations that statisticians traditionally undertake.

- There is uncertainty about the set of π_j to use. This is unsatisfactory though it may be inevitable. Some choices force themselves on us by experience. For example, Reynold's number will almost always be a component of physical systems involving its effects and we may as well choose it as one of the π_j right from the start. We need further investigation into methods for choosing the best sets of the π_j.

These limitations are not severe and, as our examples show, DA can provide an almost automatic model form which is an ideal preliminary to regression or data analysis particularly in situations involving continuous physical variables in several dimensions.

ACKNOWLEDGEMENTS

We thank Harold Henderson for the valuable comments he made on an earlier draft of this paper and to the referees for useful suggestions.

References

Aitkin, M. and Francis, B. (1982). Interactive Regression Modelling. *Biometrics*, 38:511–516.

Atkinson, A. C. (1982). Regression Diagnostics, transformations and constructed variables. *J R Statist Soc B*, 44(1):1–36.

Atkinson, A. C. (1994). Transforming both sides of a tree. *The American Statistician*, 48(4):307–313.

Bender, E. A. (1978). *An introduction to mathematical modeling.* John Wiley and Sons, NY.

Bridgman, P. W. (1963). *Dimensional Analysis.* Yale University Press, revised edition.

Buckingham, E. (1914). On physically similar systems; Illustrations of the use of dimensional equations. *Phys.Rev.*, 4:345–76.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1):1–10.

Everitt, B. S. (1994). *A Handbook of Statistical analyses using S-Plus.* Chapman and Hall.

Finney, D. J. (1977). Dimensions of Statistics. *Applied Statistics*, 26(3):285–289.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., and Ostrowski, E. (1994). *A Handbook of Small Data Sets.* Chapman and Hall, London.

Henderson, H. V. and Velleman, P. E. (1981). Building Multiple Regression Models interactively. *Biometrics*, 37:391–411. (discussion 38, 511-516, June 1982).

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32:1–49.

Huntley, H. E. (1967). *Dimensional Analysis.* Dover Publications, New York.

Langhaar, H. L. (1951). *Dimensional Analysis and the theory of models.* Wiley.

Naddor, E. (1966). Dimensions in Operations Research. *Operations Research*, 14:508–514.

Rashevsky, N. (1960). *Mathematical biophysics : physico-mathematical foundations of biology.* Dover Publications, New York, 3rd edition.

Rayleigh, J. W. S. (1915). The principle of similitude. *Nature*, 95(66):591 and 644.

Rutherford, J. R. (1975). Design of experiments and data analysis: a scientific approach. In Gupta, R. P., editor, *Applied Statistics*, pages 271–287. North Holland.

Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1976). *Minitab Student Handbook*. Duxbury Press.

Vignaux, G. A. (1986). Dimensional Analysis in Operations Research. *NZ Operations Research*, 14(1):81–92.

Vignaux, G. A. and Jain, S. (1988). An approximate inventory model based on Dimensional Analysis. *Asia-Pacific Journal of Operational Research*, 5(2):117–123.

Yalin, M. S. (1971). *Theory of Hydraulic Models*. MacMillan.