Some Examples of Dimensional Analysis in Operations Research and Statistics

G A Vignaux*

Abstract

Operations Research is an unusual area for the application of Dimensional Analysis but one that has some potential. It has its own set of fundamental dimensions and a rich zoo of dimensional variables, particularly in inventory theory. So DA should prove valuable in developing solutions and preventing errors. It is certainly possible to improve the efficiency of one part of of OR, queues and discrete-event simulation, by using simple DA ideas which might be obvious to an engineer or physicist but appears not to be for those of other disciplines. On the other hand we have found Statistics to be less rich in applications though when it can be applied in particular areas it can prove useful. I illustrate the talk with a few examples.

1 Introduction

The early exponents of Operations Research (referred to as OR, here) were, in the main, applied mathematicians and physicists and had been educated in dimensional ideas. It is not surprising, then, that early OR workers often used, or attempted to use, Dimensional Analysis (DA) in their work as the idea of constructing models of decision situations grew. More recently, however, the discipline has been taken over by pure mathematicians, economists, and management experts and the technique is, unfortunately, rarely used. In this paper, I will give some examples of its use in OR and to widen the scope by describing our attempts to use DA in the more general statistical area.

*School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand

In an important paper, (Naddor, 1966a) encouraged those in OR to use dimensional methods and gave examples in queueing, inventory, and linear programming. He listed several ways the technique could aid the OR worker but noted the usual restrictions. In his book on inventory theory, (Naddor, 1966b), he emphasized dimensions but did not use DA to develop any formulae. He pointed out, in particular, the existence of many different costs, all with different dimensions.

Physicists and engineers, aware of the sharp admonition given by (Rayleigh, 1915) have long been careful to apply DA before setting out to design experiments. Statisticians, on the other hand, complain that they infrequently have an opportunity to involve themselves in the original design of experiment. When they do they are more concerned with setting up 'balanced designs' than explicitly eliminating variables. However, there have been attempts to use DA in statistics: (Rutherford, 1975) shows there is a place for DA in in planning experiments to reduce the number of factors to be considered and (Finney, 1977) showed that it can provide rapid useful checks for statistical algebra. However it is unusual for statisticians to use DA in their normal approach to data analysis. Searching for examples in the literature, (Vignaux and Scott, 1999) discovered that there are very few published datasets that offer opportunities for an attack by DA.

In this paper I give a few examples using standard (old-fashioned) DA in Operations Research and Statistics. I look first at some models in inventory and test an inventory meta-model approximation by (Ehrhardt, 1979) and subsequent papers, and find it dimensionally incorrect. An alternative model, based on DA, reported in (Vignaux, 1986) is presented. I look at some applications in queue theory and then some models found in the statistical literature.

2 DA in OR models

Operations Research involves constructing models of human organizational systems in order to help in making the best decisions possible. In this, the combination of good models of the system's behaviour and powerful optimization methods is important.

In the earliest classical textbooks in OR, (Sasieni et al., 1959) and (Flagle et al., 1967) use dimensional methods to check inventory formulae derived using more complicated analyses. (Naddor, 1966b) uses the concept of dimensions extensively in his book on inventory theory. The ideas are particularly powerful in this topic because several types of costs, differing only in their dimensional structure, can be defined. For example, Naddor uses three types of stockout (or penalty) cost with dimensions [$], [$$T^{-1}$], and [$$Q^{-1}T^{-1}$], which lead to different solutions for the inventory problem. Below we add another cost to this list.

(Sivazlian, 1971) used dimensionless numbers to reduce the number of variables he had to plot to present the results of determining $(s,\ S)$ inventory policies with gamma distributed demand in each period. (Silver, 1983) similarly used dimensionless numbers to plot indifference curves in a continuous review stochastic $(s,\ S)$ system.

In economics, an allied discipline, (de Jong, 1967) uses the same fundamental dimensions as we do in OR and discusses at some length the dimensional problems associated with interest rates and discounting.

2.1 Fundamental dimensions on OR

The fundamental dimensions involved in Operations Research are not as familiar as the classical three $[MLT]$ dimensions of mechanics and physics. OR problems will often have a component of time $[T]$ in common with mechanics but will usually also have some criterion of optimization, perhaps cost or profit, for which we will use the symbol [$] although it is not essential for the objective function in OR to be measured in money terms. Thus, the fixed cost of an order in an inventory system, the reorder cost, will have dimension of [$] and average cost per unit time will have dimensions of [$/$T$].

In many problems there are also components of material $[Q]$, such as lot sizes or production quantities. Demand, for example, has dimensions $[Q/T]$ and the cost of an item, [$/$Q$]. It is sometimes necessary to distinguish between different types of such materials and each can be given a different dimensional symbol. Naddor points out that linear programming models typically contain very many different types of quantities and constraints, perhaps running into thousands of fundamental dimensions. Some queue models will also include people, $[P]$ or customers, $[C]$. Probability is, of course, dimensionless in nature but random variables often have a dimension of their own, depending on the particular model.

2.2 Inventory Models

Inventory systems provide a particularly rich source of problems for the application of DA. In these problems we have a stock or inventory of items which are subject to deterministic or stochastic demand. There are costs of holding the inventory, higher costs if we run out of stock, and a costs for re-stocking. We must decide how much and when to order to replenish the stock. In practice there may be many thousands of item types (the technical term is stock-keeping-units, or SKUs).

An analysis of the simplest deterministic continuous demand inventory model provides an example of the use of DA that is almost as impressive as the traditional pendulum problem. In this simple inventory model, we must determine how often to order (or equivalently, how much to order at a time - the optimal order quantity, denoted Y, below). We are not allowed to go out of stock (i.e., to allow inventory to become negative). Lead-time is zero (or equivalently, since demand is deterministic, fixed).

We will consider the following quantities:

Table 1: Dimensions of inventory variables

Y	order quantity (lot size)	$[Q]$
M	demand rate	$[QT^{-1}]$
h	unit holding cost rate	$[\$Q^{-1}T^{-1}]$
K	order cost	$[\$]$

The time between orders in this model is just Y/M. We are seeking an expression for the optimum (minimum cost) order quantity, Y. A simple DA gives the following solution:

$$\frac{Y^2 h}{KM} = \text{const} \tag{1}$$

The traditional analysis, based on differentiating the total cost rate, gives what is now known as the Wilson Lot-size formula, or the "square-root formula" introduced by Wilson in 1934 (but, it was recently discovered, actually first published by (Harris, uary)

$$Y = \sqrt{\frac{2KM}{h}} \tag{2}$$

An interesting point about this formula is that order quantity is not, as a non-technical manager might suppose, proportional to demand rate. I have previously suggested that the ratio Y^2h/KM be named the Wilson Number, Wi. Perhaps that should now be changed to the Harris Number. It reflects a balance between the total holding cost rate and the reordering cost rate.

We can add other dimensional parameters to the model while still retaining its deterministic nature. For example, when the model is used in a production situation, we we add the maximum production rate, A, $[QT^{-1}]$, which limits the rate at which stock can be replenished. We can also add a number of penalty costs, each with different dimensions.

Table 2: More inventory dimensions

Y	order quantity (lot size)	$[Q]$
A	maximum supply rate	$[QT^{-1}]$
p	unit stockout (penalty) cost rate	$[\$Q^{-1}T^{-1}]$
C_4	a fixed penalty cost per stockout	$[\$]$
C_5	penalty cost on average size of a stockout	$[\$]$

A dimensional analysis of this system, still excluding leadtime effects as the model is deterministic, gives the following function

$$f\left(\underset{(1)}{\frac{hY^2}{MK}}, \quad \underset{(2)}{\frac{Yh}{MC_5}}, \quad \underset{(3)}{\frac{M}{A}}, \quad \underset{(4)}{\frac{h}{p}}, \quad \underset{(5)}{\frac{K}{C_4}}\right) = \text{const} \tag{3}$$

Describing the dimensionless numbers in order, we can see that:

1. is the Wilson Number, Wi;

2. balances the size-of-backlog penalty cost, C_5, and holding cost, h. This group appears in the continuous review stochastic models as the probability of no stockout per leadtime;

3. is a scale factor which becomes active for cases of limited production rate, A;

4. is a scale factor giving the ratio of holding to penalty costs and is active where the model allows stockouts;

5. is a scale factor balancing reorder cost and fixed penalty costs.

Ignoring one of these groups corresponds to making constraining assumptions about the model. For example, ignoring group (3) is equivalent to assuming an infinite replenishment rate. Ignoring all but (1) gives the Wilson lot-size formula as shown above.

Allowing groups (1) and (4) to be active gives us the formula for the lot size, Y, with a finite stockout cost, p.

$$Y^2 = k\frac{KM}{h} f\left(\frac{h}{p}\right) \tag{4}$$

We can get a closer approximation to the solution by considering reasonable forms for the function of h/p. We know that as the penalty cost for stockouts, p, becomes larger, the function tends to a constant, since Y still has a value even when stockouts are forbidden. Without loss of generality we can assume that this is 1.0 for the limiting case. It is reasonably easy to convince ourselves that a simple first model for this component is $(1 + ah/p)$, where a is a dimensionless constant.

The textbook solution for this problem gives $a = 1$ and the functional form as $(1 + h/p)$. Similar logic gives us a first attempt at the form of the production rate function using group (3) as $(1 - M/A)$ which is a good first approximation to $(1 + M/A)^{-1}$, the analytical solution.

In an interesting special development of this model, there are apparently some cases in practice where the holding cost is not proportional to the average level of stock but to its square. We can define another holding cost rate, h_2 [$\$Q^{-2}T^{-1}$]. A standard DA analysis gives a cube-root solution much more quickly than the standard analytic solution:

$$Y^3 = a\frac{MK}{h_2} \tag{5}$$

In the next section we will extend these deterministic models stochastic periodic-review inventory systems.

2.3 Ehrhardt's Power Approximation

A periodic-review inventory system observes the level of inventory only at regular intervals rather than continuously. Combining this with stochastic demand and a lead-time, the time lag between requesting a delivery and receiving it, produces models that are much more complicated than the deterministic systems.

Computing optimal policies for a periodic-review, single-item, back-logged inventory system with a (fixed) leadtime and independent stochastic demand proves to be difficult though there is a theoretical solution. This optimum policy is known as an (s, S) policy, that is one that orders if at a review time the inventory level is observed to fall below a reorder point s, and to order an amount D that would then take the inventory to level S. A dynamic programming algorithm to find such optimal policies is known but would be too slow (with current computers) and impractical for any reasonable sized inventory system.

To circumvent this limitation, (Ehrhardt, 1979) and (Ehrhardt and Mosier, 1984) attempted to find an efficient approximation to give the solutions quickly (in modern parlance, a meta-model). They did this by specifying a general expression for the values of interest (s and S) and carrying out a large-scale computer experiment, setting a range of values for each of the parameters of the system and finding the optimal values for each case using the slow, accurate method.

We will study only part of this investigation, the determination of a power approximation to the order size, $D = S - s$, which corresponds to our previous Y $[Q]$. Ehrhardt starts the development of his approximation from Robert's model, which here is the same as the Wilson Lot size formula. He generalizes it to include four additional factors. First the mean, M_r, and standard deviation, s_r $[Q]$ of demand each review period, r. We distinguish here M_r from the demand rate, M $[QT^{-1}]$, with $M_r = rM$. In his analysis, the length of the review period, r $[T]$ is taken as 1 throughout so although they are of different dimensions there is no numeric difference between M_r and M (until scale changes take place). Ehrhardt then adds the penalty costs, p $[\$Q^{-1}T^{-1}]$ and the leadtime, L. We will assume that L is a number of review periods and is therefore dimensionless [1]. Ehrhardt uses $(1 + L)$ instead of L since a full analysis shows that this is needed to allow for extra security because reviews occur only every period. Dimensionally there is, of course, no difference. He arrives at the following general model for D (with a similar model for s).

$$D = k\,(M_r)^a \left(\frac{K}{h}\right)^b (1 + L)^c (s_r)^d \left(\frac{p}{h}\right)^e \tag{6}$$

Ehrhardt then generates 288 cases by varying each of these parameters (except h which was fixed at 1) and three demand probability distributions in a grid design. Using the optimal value of D determined from a full iterative method for each of the 288 cases, he finds the best fitting values of the exponents using multiple log-regression, arriving at the following expression:

$$D = 1.463\,(M_r)^{0.364}\left(\frac{K}{h}\right)^{0.498}(s_L)^{0.138} \tag{7}$$

where $s_L = s_r(1+L)^{0.5}$ $[Q]$ is the standard deviation of demand in $(1+L)$ periods. This expression is given externally to the model and reflects an assumption of statistical independence of demand between review periods. The factor p/h was found to have a negligible effect.

Later Ehrhardt discovered that this formula is dimensionally incorrect. A change of units in M or M_r, for example, would require a change in the size of the constant 1.463. Ehrhardt also notes that the coefficient of M "is significantly lower than the Wilson value of 0.5". Note that if we move the M into the MK/h form, the formula becomes

$$D = 1.463\left(\frac{MK}{h}\right)^{0.498}\left(\frac{s_L}{M_r}\right)^{0.138} M_r^{0.004} \tag{8}$$

In the later paper correcting the power approximation, (Ehrhardt and Mosier, 1984) first forbid a zero variance from forcing a zero value for D. They do this by changing the function of s_L to $1+(s_L/M)^2$ using similar arguments to those used for the lot-size model in the previous section to derive the form of the function of h/p. In this instance DA helps to constrain the parameter to be dimensionless but is of little help beyond this.

Second they constrain the regression so that $a = 1-b$ to force the model to be dimensionally correct for changes in units of demand. The regression is now limited to fitting one constant and two exponents.

The resulting Revised Power Approximation for D can be written as the following:

$$D = 1.30\left(\frac{KM}{h}\right)^{0.506}\left(1+\left(\frac{s_L}{M}\right)^2\right)^{0.116} M^{-0.012} \tag{9}$$

Despite these efforts, this is still not dimensionally correct. One can see that the constant must change by a small amount if the time unit is changed.

However, the regression is very close to the DA solution and any change in the constant is probably within the errors of estimation even for large changes in time unit. The DA solution, arrived at quite simply, is

$$D = a\left(\frac{MK}{h}\right)^{0.5} f\left(\frac{s_r}{M_r},\ (1+L),\ \frac{p}{h}\right) \tag{10}$$

The exponent of 0.5 for the first term gives us the Wilson Number. The dimensional technique cannot predict the forms of the function or the value of the constant, a,(though we would be surprised if it were far from 1.4). If we assume simple power functions of the different dimensionless products, we note that Ehrhardt finds the exponent of s_r/M_r to be about 0.138 and absorbs at least part of the $(1+L)$ term if s_L is to be used instead of s_r. Our analysis, of course, cannot say whether there is any further lead-time effect, though the formula allows it as a possibility.

Nevertheless the dimensional analysis shows that the true exponent for the first term is 0.5 (and not 0.498 or 0.506) and the correct form is the same as that given in equation (10). The function involving s_L/M is acceptably dimensionless if it is corrected to s_L/M_r.

Dimensional considerations also suggest a different experimental design to find a power law model. This would include regressing a range of values of the Wilson Number, hD^2/MK against values of the other dimensionless groups $s_L/M_r, (1+L)$, and p/h or, if further information is available, as in this case, against functions of these such as $1+(s_L/M_r)^2$. Apart from starting with at least part of the correct model for D, this would have the advantage of either reducing the number of points to be analyzed, or, while keeping the same number, increasing the range of situations to be fitted. It would also give a much clearer view of the target functions to be approximated, essentially breaking the main formula down into sensible components.

From these examples we see that because it contains such a varied and interesting collection of dimensionally complicated variables, particularly the range of stock holding and penalty costs, inventory theory gives us good opportunities to use DA to structure our problems in useful and illuminating ways.

2.4 Queues

While it is a particularly rich source of DA applications, inventory theory is not the only part of OR where the technique has made its mark. Consider a simple queue system.

Queues are stochastic models of congestion systems. Customers (or jobs, messages, packets, etc) arrive at random into a service system or network and may have to queue for service. The time taken to serve a customer may be stochastic and the system may involve a number of service channels in series or in parallel. There can be a network of such queues. There may be different queue disciplines which control how waiting customers are chosen for service. The cost of such a system is some combination of the costs of providing service (paying bank tellers, capital cost of equipment) and the cost of waiting for customers or delay in processing.

The average number of customers in a queue system, L, can be considered either as an integer and hence dimensionless [1] or measuring the average number of customers [C]. In contrast to Naddor, I will take the second approach and set the dimensions of the arrival rate, A, to be [C/T]. The mean time in the system. W, has dimensions [T]. The only dimensionless expression that can associate these three variables is (AW/L) and we are led to a solution form:

$$f\left(\frac{AW}{L}\right) = \text{const}, \tag{11}$$

with both the function and the constant unspecified. The simplest function of this form is given

$$L = aAW \tag{12}$$

where the dimensionless constant, a, is unknown. A more traditional analysis shows that $a = 1.0$ and hence gives us the well-known Little's law for queue systems.

We can add further factors to our simple model: s, the number of servers [S], and the service rate for each server, R [$CT^{-1}S^{-1}$]. We then obtain the additional dimensionless group $\rho = (A/sR)$, the traffic intensity, and a general functional form of:

$$f\left(\frac{AW}{L}, \frac{A}{sR}\right) = \text{const}. \tag{13}$$

Thus the simplest form of function for average waiting time is

$$W = \frac{L}{A} f\left(\frac{A}{sR}\right) \tag{14}$$

Adding more variables to the model, let T_s [T] be the mean service time and σ^2 its variance [T^2]. This introduces two new dimensionless groupings, a scale factor W/T_s and σ^2/T_s^2. By dropping the group involving W and including the new groups, we obtain a facsimile of the Pollaczek-Khintchine formula:

$$L = A\,T_s\,f\left(\frac{A}{sR},\ \frac{\sigma^2}{T_s^2}\right) \tag{15}$$

We can extend these models to consider optimization of queue systems. Once again we can assume different types of costs: cost of service (e.g. rental cost rate for server units), of customer waiting time, and of the system.

2.5 A note on simulation

One area where DA is useful is in designing simulation experiments which are a favorite tool of OR workers. For example, real congestion systems soon get too complicated to be handled by queue theory. In those cases the adage is "When all else fails, use simulation".

One dimensional idea that has proved valuable is used less than it should be. Queue models all contained the traffic intensity $\rho = (A/sR)$, as a major dimensionless variable. This is the arrival rate of jobs or customers divided by the service rate. Where $\rho > 1$ the system is overloaded; even if it is less than 1 the system can form queues. In carrying out a simulation where we are asked to vary both the arrival and service rates, it follows that this is the appropriate variable to use instead of the arrival rate. In other words, we can reduce the number of independent variables in the simulation experiment by one by using the mean service time as the time unit.

2.6 Logistics Networks

(Daganzo, 1987) gives another example of the utility of DA to find general solutions in OR. He analyzed a logistic network where the problem was to find the optimal time between deliveries, the best shipment size, and the number of stops, in a large scale transportation network shipping from a delivery warehouse to a number of delivery points.

Treating the number of destination points is as a continuous variable he found that the optimal decision variables are simple functions of two dimensionless constants, N_0, the number of destinations in the delivery area, and K,a dimensionless constant, which is essentially, like many dimensionless numbers, a ratio of two factors, the inventory cost per item and the marginal unit delivery cost

3 DA in Statistical Modelling

Moving to the more general field of statistics, it seems that that it can also be used in but only to a limited extent. We can only use DA where the initial problem involves continuous dimensioned variables. For example, searching through (Hand et al., 1994), a collection of data sets used to test different statistical methods, only a minority are found to be suitable for DA. The remainder often use made-up indices or counts of events.

Despite this, it is possible to find a few examples of datasets that have been extensively analyzed and re-analyzed by statisticians using a range of powerful techniques but which yield to a simple DA. (Vignaux and Scott, 1999) studied studied the famous, and over-analyzed, problem of (Hocking, 1976)'s automobile data using DA. But here I only report on two examples, both published in (Hand et al., 1994), the analysis of Black-cherry Trees and of the Stopping distance of vehicles.

3.1 Black-cherry Trees

The Black Cherry Tree data gives the volume (cubic feet), height (feet) and diameter (inches) of a sample of 31 trees from the Allegheny National Park in the USA. It was first 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.

Traditionally, and in the literature, statisticians go about analyzing this by gradually improving a regression fit starting with a simple linear model and gradually adding power terms.

The data set was analyzed using S-Plus by (Everitt, 1994) who examined the residuals from the simple linear regression of v against d and h which "showed some evidence that a quadratic term [involving d^2] may be needed in the model." He calculated a confidence interval for the normal probability plot which showed that the residuals "show little departure from normality." The analysis was then repeated with an added term in d^2 with a larger R^2 value and confirmation that the coefficient of the quadratic terms was significant.

In another analysis of the dataset, (Cook and Weisberg, 1983) tested the data for heteroscedasticity and concluded that the variance depended on h. (Atkinson, 1982) discussed different statistical diagnostics and transformations of the data. He concluded that a transformation of v to $v^{1/3}$ and a linear fit with the other variables was appropriate and pointed out that this was dimensionally correct. In the extensive discussion of the paper Professor P Sprent suggested out that a forester would think of the trunk shape as something very like a cone and proposed a regression model of the form $v = ahd^2$.

Let us look at this using the standard method of DA. We are given that the volume, v, is connected by some unknown function of height, d, and diameter, h. Write this in the form

$$f(v, d, h) = 0. \tag{16}$$

For our model we can choose the pair of dimensionless variables π_1 and π_2, where

$$\pi_1 = \frac{v}{h^3}, \quad \pi_2 = \frac{d}{h}. \tag{17}$$

Another useful pair might be v/d^3 and d/h and this might, on reflection, turn out to be somewhat simpler to fit. But we retain the original choice as an example.

$$v = ah^3 g\left(\frac{d}{h}\right). \tag{18}$$

A simple regression shows that the relationship actually looks like

$$v = ah^3(d/h)^2 = ahd^2. \tag{19}$$

A more controversial form of DA, called "vectorised" Dimensional or "directional" Analysis allows one to ascribe different fundamental dimensions to different directions of length (labeled 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 height, 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 $[L_x^{0.5} L_y^{0.5}]$. Volume, v would have dimensions $[L_z L_x L_y]$ rather than L^3. (Massey, 1978; Massey, 1986) believes this approach is invalid but, if it is valid, and many analysts find is 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:

$$\pi_1 = \frac{v}{hd^2} = \text{constant}. \quad (20)$$

This equation is the same as that arrived at earlier by a more roundabout route. The dimensional analysis gave this result instantly. Given equation(20), we only have to determine the value of the constant. (It is about 0.302 whereas for a pure cone shape the constant is $\pi/12 = 0.2618$)

3.2 Stopping distance

Another well-known much analyzed standard data set is that for Stopping distance. The data provided in (Hand et al., 1994, dataset 438) include d $[L]$, a stopping distance (feet) and v $[LT^{-1}]$, (miles per hour). Examining the dimensions it is clear that at least one other dimensional constant must be involved to construct a dimensionally homogeneous relationship.

A little thought suggests a combination of variables that will do the job. The mass, m [M], (weight) of the vehicle must be involved since stopping distance requires the removal of momentum. This momentum is destroyed by some form of braking force. F $[MLT^{-2}]$.

Using these four variables d, v, m, F a dimensional analysis concludes that the following relationship must hold:

$$\frac{Fd}{v^2 m} = \text{const}. \quad (21)$$

Of course, we are not given m or F and their ratio will form a dimensional constant (the breaking force to mass ratio) as part of any regression coefficients fitted. These fitted constants will therefore not be dimensionless as in a "pure" DA (and may change their values if the units of measurement are changed). Equation (21) suggests that graphing d against v^2 should give a straight line (and it does: Figure 1(b)).

What is more, plotting the semi-dimensionless distance, d/v^2, against v, for example, should give a horizontal line corresponding to the constant term in Equation (21). Actually, it will be a dimensionless constant term multiplied by the dimensional quantity, m/F $[T^2L^{-1}]$. Such a graph is shown in Figure 1(c) where is clear that the result is not horizontal at all, except at higher values of v.

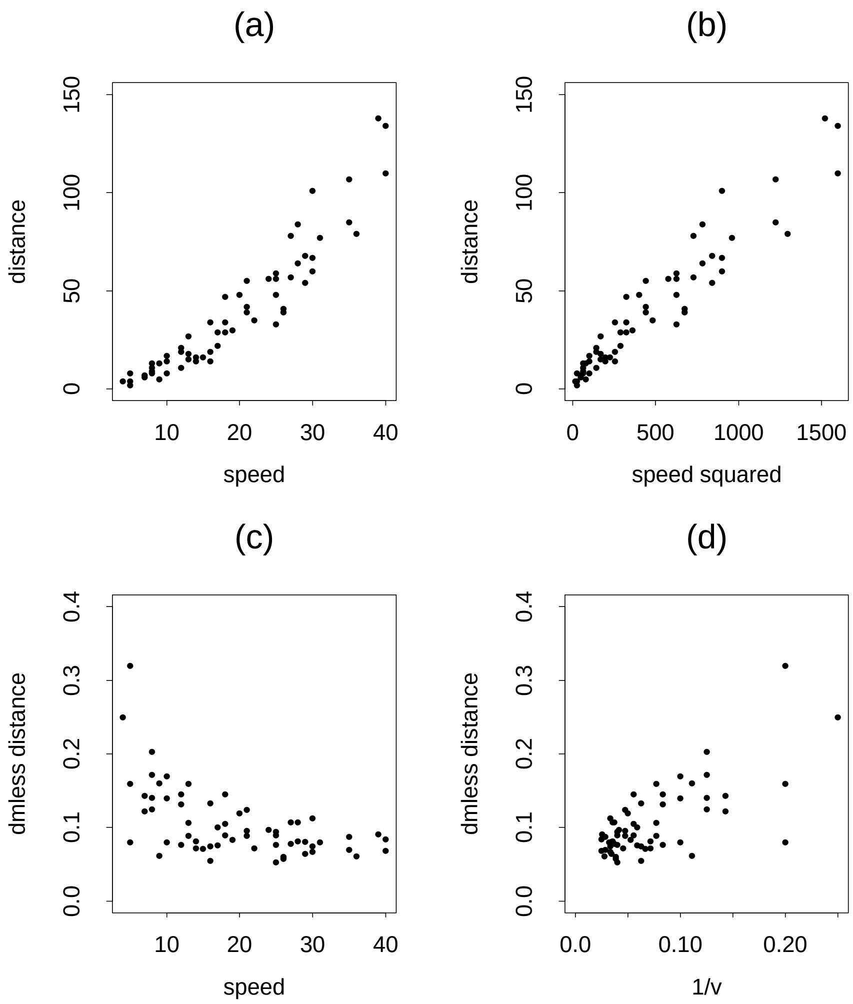

Figure 1: Results from graphing the Stopping data. (a) shows a straightforward plot of d against v; (b) shows a plot of d against v^2; (c) shows a plot of semi-dimensionless distance, d/v^2 against v; (d) shows a plot of semi-dimensionless distance d/v^2 against v^{-1}. Dimensional analysis suggests this should be linear.

So what has gone wrong? There must be another dimensionless term and hence another dimensional variable missing from our analysis. What can it be? Some thought suggests the driver reaction time, τ $[T]$, as a suitable contender. Adding this to the list of variables for the dimensional analysis, we find one alternative model:

$$\frac{dF}{v^2m} = f(\frac{F\tau}{vm}). \tag{22}$$

This suggests, perhaps, that $f()$ would have a constant term (corresponding to the horizontal line in Figure 1(c)) plus a term in v^{-1}. This would give the low speed effect shown in the figure. Figure 1(d) shows the expected linear graph though it has considerable scatter. The intercept gives the constant term in the original DA, the slope the term corresponding to the reaction time.

At last, here was real data that had previously been extensively analyzed by a number of workers using an array of powerful statistical techniques but without having the advantages of DA. Applying DA in a standard way magically produces an, apparently correct, partial solution. To document the analysis, tracing the references back through a sequence of authors, eventually, we unearth the original data, in (Hald, 1952). It is turns out that the data are not real at all; they had been invented as an illustration of linear regression.

3.3 A Note on Probability

Considering the application of DA to statistical problems immediately raises to our attention the central role of probability. Probability whether defined from a frequentist (a ratio of counts of events) or a Bayesian view (a relative likelihood) is, of course, dimensionless.

In contrast, probability density is dimensioned. For example, a random fault rate, λ, in reliability theory is the probability per unit time of an event, with dimensions $[T^{-1}]$. Similarly, random variables just have their own dimensions, the time to failure for a component, t, may be random but has dimensions $[T]$.

So randomness and probability should cause few problems in applying DA. There are a few strange effects, for example in dealing with discrete random variables. One distribution commonly used in inventory theory for an SKU's demand in a period is the Poisson distribution. For a Poisson random variable with mean μ, we find that the variance of the distribution is also μ!

3.4 Concluding Remarks on DA and Statistics

DA will be useful to statisticians only as far as the models that they are trying to fit have the potential for a dimensional analysis: they have a collection of continuous dimensioned variables in a background of several fundamental dimensions. The statistical nature of the problems have little effect. Thus statistical applications in the biological area should prove a good source (see (Stahl, 1961) and later papers). DA cannot gain much leverage in the pure discipline of statistics because it is not physical enough.

However there is a serious side to this. Part of a statistical model will include a hypothesis for the observational error distribution, (often Gaussian). This is defined in connection with the original mathematical or physical model. Applying DA will rearrange the variables into fewer dimensionless variables. Further, the error structure of the new model will be different from the original. It is then more difficult to compare the accuracy of fitting the two models, one using the original and the other using the DA dimensionless variables. Compared with the original model, the new one no longer has errors independent of the other variables.

4 Conclusion

As we have seen, there are useful applications of DA in Operations Research. The part of OR with perhaps the richest collection of dimensioned variables, inventory theory, is naturally the most fruitful for applying DA. But there are other areas, including congestion or queue theory where value can also be obtained. Modest gains in efficiency in simulation experiments are also possible.

In statistics the outlook is not quite so positive. Essentially DA is useful only when statistical analysis is associated with a model of the world that is open to DA. We need a number of fundamental dimensions and a collection of dimensioned variables to work with. In the published statistical literature many data series do not satisfy these conditions. Nevertheless some examples can be found and, where conditions are right, DA has it usual power.

References

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

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

Daganzo, C. F. (1987). The break-bulk role of terminals in many-to-many logistic networks. *Operations research*, 35(4):543–.

de Jong, F. J. (1967). *Dimensional Analysis for Economists*, volume 50 of *Contributions to Economic Analysis*. North-Holland.

Ehrhardt, E. (1979). The power approximation for computing $(s,\ s)$ inventory policies. *Management Science*, 25(8):777–786.

Ehrhardt, E. and Mosier, C. (1984). A revision of the power approximation for computing $(s,\ s)$ policies. *Management Science*, 30(5):618–622.

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.

Flagle, C., Huggins, W., and Roy, R. (1967). *Operations Research and Systems Engineering*. John Hopkins Press.

Hald, A. (1952). *Statistical theory with engineering applications*. New York : Wiley.

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.

Harris, F. W. (1913, February). How many parts to make at once. *Factory: the Magazine of Management*, 10(2):135–136, 152.

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

Massey, B. S. (1978). Directional analysis? *Int J of Mechanical Engineering Educ*, 6(1):33–36.

Massey, B. S. (1986). *Measures in science and engineering: their expression, relation and interpretation.* Ellis Horwood, Chichester.

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

Naddor, E. (1966b). *Inventory Systems.* John Wiley & Sons, New York.

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.

Sasieni, M., Yaspan, Y., and Friedman, L. (1959). *Operations Research.* John Wiley & Sons.

Silver, E. A. (1983). An example of indifference curves in inventory control. *IIE Transactions*, 15(1):79–80.

Sivazlian, B. D. (1971). Dimensional and computational analysis in stationary (s,s) inventory problems with gamma distributed demand. *Management Science*, 17(6):B307–311.

Stahl, W. R. (1961). Dimensional analysis in mathematical biology. I. General discussion. *Bull. Math. Biophys.*, 23:355–376.

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

Vignaux, G. A. and Scott, J. L. (1999). Simplifying regression models using dimensional analysis. *Australian and New Zealand Statistics*, 41 (1):31–42.