

INTERPRETING COVARIATE EFFECTS IN REGRESSION MODELS: SEARCHING FOR SYNERGISM AND ANTAGONISM

John M. Lachin, Sc. D.
Professor of Biostatistics and Epidemiology, and of Statistics
Department of Epidemiology and Biostatistics
The Biostatistics Center
The School of Public Health and Health Services
The George Washington University

The Biostatistics Center
6110 Executive Blvd.
Rockville MD 20852

(originally February 2, 1993)

0.1 INTRODUCTION

Regression models are one of the most widely used methods for statistical analysis because they provide the opportunity to explore the simultaneous effects of multiple covariates on a particular outcome or dependent variable of interest. It is now standard statistical practice to apply a stepwise or other recursive algorithm to fit regression models, whether it be a normal-errors multiple regression model or a likelihood-based linear model such as Cox's proportional hazards regression model for event-time data.

Over the past 30 or more years many papers have been written about the problem of model specification. In a landmark paper, Efroymson (1960) first proposed the now commonly used algorithm for stepwise regression model building. Since then, there have been a number of papers to further refine and evaluate the properties of stepwise algorithms for model selection, and of the reduced models so selected (*c.f.* Cox and Snell, 1974; Hocking, 1976; Miller, 1984, among many). It has been shown that a stepwise algorithm may miss a truly "best" reduced model, and the merits of forward selection and backwards elimination have been debated (Mantel, 1970). Further, the potential bias in estimates of regression coefficients from a reduced model, relative to those of a true full model, has been explored (*c.f.* Hocking, 1976; Miller, 1984), and criteria for model selection based upon minimum bias in the predicted values has been proposed (Mallows, 1973).

In addition, it has been demonstrated that the Type I error associated with tests of model fit and individual effects can be wildly inflated by a stepwise (or any reduced model) algorithm because the resulting p -values for tests of model fit and for individual effects will be far too small. The magnitude of this inflation in Type I error, or devaluation of p -values, has also been explored under special cases (Pope and Webster, 1969; Wilkinson and Dallal, 1981; Freedman, 1987; and Flack and Chang, 1987; among others). Kupper, Stewart and Williams (1976), and Butler (1982, 1984) describe Bonferroni expansion-based procedures for determining a protected p -value or significance level. In general, perhaps the best practice is to restrict tests of significance to those actually associated with the "full model" containing the complete set of relevant covariates. All of the above cited papers explore these problems within the context of the traditional normal-errors linear model. Similar results also apply to the more general family of generalized linear models (*c.f.* McCullagh and Nelder, 1989).

Although it has been argued that the backwards elimination (step-down) procedure is better than the forward selection procedure (Mantel, 1970), the latter is more commonly used in practice dating back to the initial work of Efroymson. Mantel argued that a backwards elimination procedure was preferable because it started from a full model with all of the potential covariates of interest examined simultaneously. Then, by eliminating unimportant variables in turn, it is more likely that such a model will identify synergistic covariate effects. Synergism occurs when two covariates together have a substantial effect within the regression model when neither covariate individually has any substantial effect. Synergism will rarely be detected by a forward model selection

procedure. On the other hand, a forward selection procedure is less susceptible to including two antagonistic covariates in the model. Antagonism occurs when each of two covariates individually, or in combination with other covariates, have a substantial effect within the model, but when employed simultaneously the two covariates tend to cancel each other out.

Regression models are used in many contexts and with many objectives. Efron was an industrial statistician who was principally interested in constructing a parsimonious regression model for actual use in making predictions. In such cases, the objective of the "stepwise" model selection procedure is to identify a reduced model that works about as well as the full model; and it is not necessarily important as to which covariates are included in the final parsimonious model, nor how they got there.

In other cases, the objective of the analysis is to describe the structural relationships between a set of covariates and the dependent variable. Sometimes the analysis is aimed at confirming associations which were suggested from sources other than the data at hand. More often than not, however, the purpose of the analysis is exploratory in an attempt to evaluate and describe the manner in which the covariates inter-react with each other in a regression model which describes their joint association with the dependent variable. In this case, a simple stepwise model selection procedure is far from adequate in exploring the nature of covariate effects.

The purpose of this brief article, therefore, is to describe a simple algorithm which can be used to identify and explain synergism and antagonism among the covariates in an exploratory analysis. In Section 2, some common misinterpretations of regression models are discussed. Section 3 then describes an algorithm for the identification of synergistic and antagonistic effects among the covariates through application to a set of data. Section 4 then further discusses the applicability of this approach to exploratory data analysis.

0.2 MIS-INTERPRETATIONS OF REGRESSION MODELS

Exploratory regression analyses are widespread in virtually all of the natural sciences, especially medical epidemiology. In addition to the common problem of reporting spurious p -values for effects identified by a "stepwise" procedure, two other types of misinterpretations of the data are common. Firstly, the variables selected in a reduced model often are interpreted as "the most important variables." Secondly, it is common practice that when one covariate is significant in the model, in the presence of others, then that covariate is said to have an "independent effect," rather than an additive (or multiplicative) effect. It is a sad commentary on statistical education to note that such widespread interpretations are nonsense.

The first problem, the declaration of seemingly "most important" effects, arises because a regression model is nothing more than a data smoothing technique. However, since the variable selection procedure can not distinguish true

signal from noise, so to speak, no special credence should be placed on the relative importance of selected covariates versus non-selected covariates, without further study. Just as the p -values for effects in a reduced model are spurious, so may also be the actual subset of covariates selected in the final reduced model. This problem was recently addressed by Hauck and Miike (1991) who showed that often there are "close alternative" variables which could be selected at a given step, and that no special credence should be given to the final set of selected variables.

It is usually a sobering experience to take a set of data, split it into two parts at random, and then observe that a stepwise algorithm identifies two different sets of "most important variables", with few if any variables in common among the two sets. Cornfield (1976) describes one such instance where a stepwise procedure identified 10 "important" variables within each half of a large set of data, but the two sets of variables had only two variables in common (Coronary Drug Project Research Group, 1974). Unfortunately, such "split data" validation (Snee, 1977; Picard and Cook, 1984) is rarely performed. Recently, Thall, Simon and Grier (1992) describe a variable selection procedure using cross-validation in an attempt to differentiate truly important covariate effects (the signal) from random covariate effects (the noise). This method, however, is highly computer intensive.

The second problem, the declaration of seemingly "independent" effects, arises from a fundamental lack of understanding, often by statisticians, as to the meaning of a coefficient in a regression model. First, consider the standard multiple linear regression model (identity link with *i.i.d.* normal errors) for a completely balanced designed experiment; *i.e.*, a crossed factorial design with a common cell sample size. Since the vector of values for a given factor (column of the design matrix) is orthogonal to those of all other factors in the model, then the coefficient for that covariate is unique, as is the regression partial sums of squares (or deviance), and these quantities are unchanged in any reduced models. However, this is not the case when the design factors (covariate vectors) are not orthogonal, and the other factors (covariates) have non-zero associated coefficients. Thus, the β_1 and the corresponding deviance are not the same quantities in the model $E(Y) = \beta_0 + \beta_1 X_1$ and in the model $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, when X_1 and X_2 are correlated and $\beta_2 \neq 0$. Further, it has been shown that these quantities are not unique for families of regression models which are neither linear nor exponential, such as the Cox proportional hazards model, even when the covariates are independent in expectation (Gail, Wieand and Piantadosi, 1984).

Therefore, the above simple linear models should be written more informatively as $E(Y) = \beta_{0|(0,1)} + \beta_{1|(0,1)} X_1$ and $E(Y) = \beta_{0|(0,1,2)} + \beta_{1|(0,1,2)} X_1 + \beta_{2|(0,1,2)} X_2$, where $\beta_{1|(0,1)} = \beta_{1|(0,1,2)}$ if the covariates X_1 and X_2 are orthogonal or if $\beta_2 = 0$. In general, any coefficient β_k should always be interpreted as $\beta_{k|\Omega}$ in reference to the particular set of covariates Ω employed in a particular model. When the components of Ω are changed, as in the successive steps of a stepwise algorithm, then both the meaning and usually the value of the coefficient for the k -th covariate change.

Changes in the value of the coefficient for a given covariate reflect changes in the nature of the association with the outcome. However, as one changes the set of covariates Ω , the strength of the association of a covariate with the outcome may change, apart from any change in the coefficient. Herein, it is proposed that the strength of association be measured by the reduction in deviance associated with a given covariate, termed herein the partial deviance for that covariate. In a generalized linear model (McCullagh and Nelder, 1989), the deviance is used to measure the unexplained (or explained) variation associated with a given model, or with a set of nested models. Herein, however, the term "partial deviance" is used to emphasize its relation to a particular covariate in a particular model.

When modelling, often the changes in the partial deviance for each covariate are the most informative aspect of the data, because when examined systematically, they provide a description of how the covariates inter-react. The following section, therefore, provides an algorithm for identifying synergism and antagonism among covariates based on changes in deviance from one model to the next.

0.3 IDENTIFICATION OF SYNERGISM AND ANTAGONISM

The algorithm is best described by application. As a preamble, it is assumed that one has identified the important covariates of interest, based not on data exploration, but based on knowledge of the phenomenon under study. Further it is assumed that the appropriate model has been specified and that the model assumptions assessed.

As an example we use data from the National Cooperative Gallstone Study (NCGS) (Schonefeld, Lachin, et al., 1981). In the NCGS a total of 916 patients with cholesterol gallstones were entered into a randomized clinical trial to assess the effectiveness and safety of a drug treatment for gallstones. Other studies have shown that the baseline or pre-treatment serum cholesterol level (*mg/dl*) is positively related to the effectiveness of bile acid treatment for gallstones, but that such treatment may result in unfavorable increases in post-treatment cholesterol levels. Ironically, however, serum triglycerides (*mg/dl*) rather than cholesterol is a dominant risk factor for the development of gallstones. Thus, an analysis of the cross-sectional, pre-treatment data from these 916 patients is presented in order to describe the association of other patient characteristics with serum cholesterol levels.

For this analysis, the following covariates were selected for exploration based upon the known correlates of serum lipids or of gallstone disease. These included age in years, sex (1 if male, 0 otherwise), body weight (kg), percent of ideal body weight (%) based upon the Metropolitan Life Tables of ideal weights, diastolic blood pressure (mm. Hg), and an indicator variable for whether the patient is a current drinker of alcoholic beverages (1 if yes, 0 if no). All of these variables are readily obtained from a brief patient examination. In addition, the analysis employed serum triglycerides (*mg/dl*) and the biochemical test of overall liver function SGOT (International Units, IU). The latter was included because the

liver serves to regulate blood cholesterol levels, and because alcohol consumption can both raise serum cholesterol and raise SGOT, the latter reflecting loss of liver function.

To search for covariate inter-reactions, it is important that a measure of covariate effect within the regression model be employed. For this purpose, p -values generally can be highly misleading. Rather, a measure of the effect of the covariate in the context of the regression model is desired. In the framework of generalized linear models, the appropriate measure is the partial deviance, or change in deviance (log likelihood) when a given covariate is added to or deleted from a model. In the normal-errors multiple regression model, the deviance is the error sum of squares, SSE, which is proportional to the model $R^2 = (SST - SSE)/SST$, where SST is the total sum of squares. Therefore, the partial deviance for the k -th covariate in a given model is ΔSSE_k , the change in deviance when that variable is removed from the model. This is equal to the partial sum of squares for regression for that covariate (*e.g.*, the SAS Type II sum of squares in PROC REG). This partial sum of squares is proportional to the (semi-) partial $R^2 = \Delta SSE_k/SST$, which is a measure of the proportion of variation explained by that covariate within the multiple regression model.

The attached Table 1 presents the results obtained upon application of the algorithm. The last row presents the partial R^2 for each covariate within the "full" regression model employing all covariates simultaneously. A backwards-elimination procedure was then applied, where at each step the variable with the least influence as measured by the partial R^2 was deleted. The first step model is presented in the second-to-last row, the next step model in the next row etc. The first row of the table then presents the R^2 for regression for each covariate individually based upon a separate univariate regression model for that covariate against the dependent variable. Note that the table is arranged by columns in the order the variables were deleted from the full model. Also, note that the process continued without stopping, *i.e.* until all variables had been eliminated. This was accomplished using the SAS PROC STEPWISE with $SLS = 0.0$.

For now, ignore the values in parentheses in the upper triangular matrix other than the first row. We can now examine the effects of each covariate individually within the respective columns of the lower triangular matrix, reading from top to bottom. For example, examine the effect of sex which has an R^2 of 0.014 in the full model. When percentage ideal body weight is eliminated from this model, the R^2 jumps to 0.036. This indicates an antagonistic effect between sex and percent ideal body weight.

However, this is not the only two-fold change in the partial deviance observed for any covariate in the process of the backwards elimination procedure. Comparison of the R^2 in the last row (full model), and of the R^2 just below the diagonal (immediately before deletion from the model), with the R^2 in the first row (the univariate effects) indicates that there are additional inter-reactions among the covariates which were not detected by the backwards elimination procedure. For example, examine the effects of the indicator variable for alcohol drinkers. In the full model a minor (but significant) effect is observed

($R^2 = 0.0059$) which remains unchanged until it is eliminated from the model in the fifth step. This is the value just below the diagonal. If we compare this value to the effect from the univariate model ($R^2 = 0.0003$), it is clear that alcohol consumption has a synergistic effect with one of the other variables still remaining in the model. To assess this, it was necessary that additional models be run. The partial R^2 for alcohol drinker in these additional models are presented in parentheses in the upper triangular matrix. These values represent the partial R^2 when "alcohol drinker" was added to the model containing the other variables within each successive row. The jump in R^2 (to $R^2 = 0.0034$) occurs when alcohol drinker is added to the model containing triglycerides and sex, which indicates a synergistic effect between alcohol consumption and these variables.

The last row of the table gives a summary of the observed 2-fold antagonistic and synergistic effects found in terms of the indices of the other variables with which each covariate inter-reacts. Note that the variables "alcohol drinker" and "body weight" show antagonistic effects with some variables and synergistic effects with others.

As a companion, Table 2 presents the coefficients for the variables in each reduced model, and in the univariate models. Observed inter-reactions among the covariates indicated by at least 2-fold changes in the coefficients are also cited in the table. Note that this type of analysis fails to detect many of the covariate inter-reactions observed in Table 1. However, Table 2 is important in detecting qualitative inter-reactions with other covariates, as delineated by the changes in the coefficients for "alcohol drinker." Reading from top to bottom, in the univariate case, this covariate has a non-significant negative effect, and the coefficient changes to virtually zero due to the antagonistic effect with triglycerides. However, when sex and age are added to the model, the effect is strongly positive. The changes in the magnitude of the effect are best observed in Table 1, but Table 2 is needed to observe the change in the direction of the effect.

Table 3 presents the nominal p -value associated with each coefficient and the test of each effect. In the univariate analyses, all covariates except "alcohol drinker" and "% of ideal weight" are nominally significant at $p < 0.05$. In the full model, all except "diastolic B.P.", "body weight" and "% of ideal body weight" (variables 6-8) are nominally significant at $p < 0.05$. The "best" reduced model contains the remaining variables (1-5) with these three deleted.

Finally, in a split-half validation, a backwards elimination procedure at the nominal 0.05 level in each half selected the first three covariates triglycerides, age and sex. In the first half, SGOT was also included ($p = 0.013$), and in the second half, "alcohol drinker" ($p = 0.04$) and "diastolic B.P." ($p = 0.017$) were also selected.

Therefore, these data might be interpreted as follows. Seven variables were identified to be assessed individually and in combination as covariates of serum cholesterol. The full model accounted for $R^2 = 0.19$ ($p < 0.0001$). Percent of ideal body weight showed no significant association with cholesterol level either individually or in combination with other variables. Body weight and diastolic

blood pressure, though each was significantly associated with cholesterol individually, contributed little to the multivariate model due to antagonism from the effect of triglycerides (in both cases) and that of sex (diastolic blood pressure). SGOT had a significant weak association individually and in combination with other covariates, with an estimated average increase of 3.6 *mg/dl* of cholesterol per 10 I.U. increase in SGOT in the full model. A slightly greater effect was observed for SGOT individually. The indicator for alcohol drinker showed no effect individually, but showed a significant weak association in combination with other covariates, principally due to synergism from sex and age jointly. In the full model, the mean cholesterol level for drinkers was estimated to be 7.1 *mg/dl* higher than for non-drinkers.

The remaining three covariates (triglycerides, sex and age) were significantly associated with cholesterol individually and in combination. The effect of triglycerides was unaffected in the presence of other covariates, with an estimated average increase of 26.7 *mg/dl* increase in cholesterol per *mg/dl* increase in triglycerides in the full model, and about 27 *mg/dl* in all other models. Individually, age had a significant association with cholesterol, with an estimated average increase of 10 *mg/dl* in cholesterol per 10 year increase in age. However, the effect of age was reduced in the multivariate model due to a antagonism from triglycerides and sex jointly, with an estimated average increase of 7 *mg/dl* in cholesterol per 10 yr increase in age. Sex also had a significant association with cholesterol both individually and in combination, but had an antagonistic effect from percent of ideal body weight. In the full model, males had an estimated mean cholesterol level 17.2 *mg/dl* higher than that among females.

0.4 DISCUSSION

The objective of exploring the data in the manner suggested by the above algorithm is quite different from that of a usual reduced model selection routine. Here the interest is not in determining what appears to be the best reduced model, but rather in explaining the manner in which the covariates inter-react within the context of a regression model. This is useful in communicating the results to the investigators whether or not one decides to recommend using a reduced model or the "full" model.

The notion of inter-reactions among covariates is by no means new. In the social sciences, one type of synergistic covariate inter-reaction is termed suppressor variable action where variable X_1 boosts the effect of variable X_2 by accounting for some of the residual variation in the association of X_2 with Y . Likewise, antagonism is similar to some definitions of "counfounding" as used in many articles on this topic. (For example, see Wickramaratne and Holford (1987) and the reader reactions from Greenland, Holland and Mantel (1989)). The notion of confounding arises when one is trying to assess the presence or absence of a causal association between one critical variable and the dependent variable in the presence of other potential causal factors (potential confounders) in a non-randomized observational study. However, in the types of analyses discussed herein, the objective is to ascertain the inter-reactions among all of

the covariates, not the assessment of a single primary covariate in the presence of others.

The method, however, is truly data exploration and thus is subject to influence both by the signal and by the noise in the data. Unfortunately, just as with a stepwise procedure, it is impossible to separate the effects of noise from the true signal in the data through such an algorithm alone. Additional analyses using model validation techniques could also be applied, such as a split-sample approach. In this case, the data could be randomly divided into two components and then an independent analysis for synergism and antagonism conducted within each of the component halves of the data. When applied to these data, the same patterns of inter-reactions are observed.

It should also be noted that many programs such as the SAS PROC REG present "sequential parameter estimates," which present the changes in regression coefficient values as successive variables specified in the model statement are added to the model. As shown in Table 2, it is also instructive to examine the changes in the regression coefficients themselves from one model to the next. However, this will not be as sensitive an indicator of the magnitude of inter-reactions among the covariates, as shown in the example in Section 3. Thus, an algorithm searching for synergism or antagonism based on the regression coefficients alone may be misguided.

In principle, this algorithm is applicable to the broader family of generalized linear regression models, such as logistic or Poisson regression; and to other likelihood-based models such as the proportional hazards regression model. This would require separate examination of the partial deviance (log likelihood) for each covariate in each reduced model at each step of the algorithm. Unfortunately this would be highly computationally intensive, effectively requiring that one examine all 2^p possible regression models. However, in such models, the Wald test $[\hat{\beta}_k / S.E.(\hat{\beta}_k)]$ and the partial deviance are each asymptotically distributed as chi-square under the null hypothesis $\beta_k = 0$. Therefore, it may be reasonable to employ the Wald test χ^2 value to explain the influence of each covariate in the model, at least as a first approximation. For a related discussion, see Magee (1990) who contrasts R^2 -type measures based on Wald statistics versus those based on likelihood ratio (deviance) statistics. For non-linear models, other measures of explained variation have also been proposed (*c.f.* Schemper, 1990, 1992; Korn and Simon, 1990, 1991; among others).

However, in the normal errors regression model, such as in the example of Section 3, it might be misleading to employ the partial t -test or F -test as the measure of covariate effects because the denominator (the mean square error) fluctuates from one model to the next. For the analyses summarized in section 3, the mean square errors fluctuated from 1537 to 1880.

Finally, we note that the above algorithm is not all inclusive in its exploration of inter-reactions among covariates. It is possible that a synergistic or antagonistic affect of great magnitude could be missed by this algorithm. To perform an all inclusive examination for covariate inter-reactions would require that one examine all $2^p - 1$ models containing an individual covariate with all possible combinations of the remaining covariates. On the other hand, it is un-

likely that any substantial covariate inter-reactions of importance will be missed by the above described algorithm.

1 References

Butler, R.W. (1984). The significance attained by the best fitting regressor variable. *Journal of the American Statistical Association* 79, 341-348.

Butler, R.W. (1982). Bounds on the significance attained by the best fitting regression variable, *Applied Statistics* 31, 290-292.

Cornfield, J. (1976). Recent methodological contributions to clinical trials. *American Journal of Epidemiology* 104, 408-421.

Coronary Drug Project Research Group (1974). Factors influencing long-term prognosis after recovery from myocardial infarction — three year findings on the Coronary Drug Project. *Journal of Chronic Diseases* 27, 267-285.

Cox, D.R., and Snell, E.J. (1974). The choice of variables in observational studies. *Applied Statistics* 23 51-59.

Efroymson, M.A. (1960). Multiple regression analysis. In: *Mathematical Methods for Digital Computers*. Ralston A. and Wilf, H.S., (eds) New York: Wiley.

Flack, V.F. and Chang, P.C. (1987) Frequency of selecting noise variables in subset regression analyses: A simulation study, *The American Statistician* 41, 84-86.

Freedman, D.A., (1983) A note on screening regression equations. *The American Statistician* 37, 152-155

Gail, M.H., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71, 431-444.

Greenland, S., Holland. P.W., and Mantel, N. (1989). Reader Reactions: Confounding in epidemiologic studies. *Biometrics* 45, 1309-1322.

Hauck, W.W. and Miike, R. (1991). A proposal for examining and reporting stepwise regressions. *Statistics In Medicine* 10, 711-715.

Hocking, R.R. (1976). The analysis and selection of variables in linear regression. *Biometrics* 32, 1-44.

Korn, E.L. and Simon, R. (1990) Measures of explained variation for survival data. *Statistics in Medicine* 9, 487-503.

Korn, E.L. and Simon, R. (1991) Explained residual variation, explained risk, and goodness of fit. *The American Statistician* 45, 201-206.

Kupper, L.L., Stewart, J.R. and Williams, K.A. (1976). A note on controlling significance levels in stepwise regression. *American Journal of Epidemiology* 103, 13-15.

Magee, L. (1990) R^2 measures based on Wald and likelihood ratio joint significance tests. *American Statistician* 44, 250-253.

Mallows, C.L. (1973a). Some comments on C_p . *Technometrics* 15, 661-675.

Mantel, N. (1970). Why stepdown procedures in variable selection. *Technometrics* 12, 591-612.

- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. New York: Chapman and Hall.
- Miller, A.J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society A* 147, 389-425.
- Picard, R.R. and Cook, R.D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association* 79, 575-583.
- Pope, P.T. and Webster, J.T. (1972). The use of a F -statistic in stepwise regression equation. *Technometrics* 14, 327-340.
- Schemper, M. (1990). The explained variation in proportional hazards regression. *Biometrika* 77, 216-128.
- Schemper, M. (1991). Further results on the explained variation in proportional hazards regression. *Biometrika* 79, 1, 202-204.
- Schoenfield, L.J., Lachin, J.M., the Steering Committee and the NCGS Group (1981). National Cooperative Gallstone Study: A controlled trial of the efficacy and safety of chenodeoxycholic acid for dissolution of gallstones. *Annals of Internal Medicine* 95, 257-282.
- Snee, R.D. (1977). Validation of regression models: Methods and examples. *Technometrics* 19, 415-428.
- Thall, P.F., Simon, R., and Grier, D.A. (1992). Test-based variable selection via cross-validation. *Journal of Computational and Graphical Statistics* 1, 1, 41-61.
- Wickramaratne, P.J. and Holford, T.R. (1987). Confounding in epidemiologic studies: The adequacy on the control group as a measure of confounding. *Biometrics* 43, 751-765.
- Wilkinson, L. and Dallal, G.E. (1981). Tests of significance in forward selection regression with an F -to-enter stopping rule. *Technometrics* 23, 377-380.