Using Multiple Regression to Develop ANOVA Power Formulae

Dale G. Shaw and David R. McCormack
University of Northern Colorado

Multiple regression may be used to examine the relationship between a single dependent variable and a set of several independent variables. The ANOVA power tables presented by Cohen (1988) can be considered such a data set. In these tables, the power of a balanced one-factor ANOVA design may be considered the dependent variable which is predicted by four independent variables: sample size, alpha level, number of groups, and effect size. Cohen's 66 pages of tables provide 15,526 power values for various combination of values of the four independent variables.

This article presents regression equations fitted to Cohen's ANOVA power tables in an effort to obtain simple yet accurate formulae for estimating the power of an ANOVA design. Simple equations were sought because power analysis is presently receiving limited attention in research planning (Cohen, 1988). Having a simple, easy-to-use formula which estimates a design's power might lead to improved designs. Obviously accurate power estimates are desirable, but the criterion of accuracy is less stringent than might be supposed. The researcher who is planning an ANOVA design does not usually require a power estimate to the nearest percentage point as Cohen's tables provide. For example, if a design's power were estimated to be .88 with a ±.06 margin of error, the experiment could proceed with with reasonable confidence of having high power even though the exact power is unknown.

In contrast to the simplicity criterion which required subjective judgment, the accuracy criterion was quantifiable. We used 2 indicators of accuracy for evaluating and comparing regression models: R^2 (the proportion of variance "explained" by the formulae) and RMSE (the root mean square error). We sought equations with $R^2 > .95$ and RMSE $\leq .03$. Regrettably, these two criteria for desirable formulae, simplicity and accuracy, conflicted with each other. The simplest formulae were not the most accurate and the most accurate formulae were not simple.

Linear Formulae

The first attempt to model ANOVA power was a simple linear model in which the dependent variable was Cohen's (1988) ANOVA power values. The independent variables were α, u, n, and f, described in Table 1. Cohen's tables provided 15,526 power values between .01 and .99, which served as data points on a hyper-surface. As expected, this first model failed to meet the accuracy criterion (R^2 = .4320). General knowledge of power curves as well as inspection of Cohen's tables suggested the surface was curvilinear rather than linear. To accommodate the curvature, and still keep the models composed of fairly simple terms, the predictor set was increased from 4 to 24 variables by including the square, the cube, the square root, the natural logarithm (ln), and the natural logarithm of the natural logarithm (ln(ln)) of each basic predictor. The last of these new predictors was undefined for some data points because ln(ln(1)) is undefined. Therefore, the basic variables were modified: α was multiplied by 1000, f by 100, and u by 10. R^2 values for various models created from the 24 variables did not exceed .95.

The search for a better model progressed by imposing a restriction on the data set. This was justified because a user of the resulting formulae probably would not need accuracy for very high or very low power values. A very low power, whether .10 or .25, indicates the proposed design is probably not worthy of further consideration. On the other hand, a very high power, whether .95 or .99, suggests a design worthy of further consideration. Because a research planner probably needs only limited accuracy at either end of the power range, the data points of these asymptotic tails of the power data (which offer the greatest difficulty in fitting a linear model) were eliminated. The greatest R^2 value of the new models (.9524 using the predictor set ln(f), ln(n), $\sqrt{u}$, $\sqrt{\alpha}$, ln(ln(n)), and u^2) was observed when power was restricted to the interval [.25, .95]. Thus, a decision was made to continue the search for linear formulae using only the reduced data set.

The next step was to increase the set of predictors by including the products of pairs of the 24 predictors so that interactive effects of the predictor variables could be accommodated. For each basic predictor, a set of 6 predictors had already been included, such as f, f^2, f^3, ln(f), ln(ln(f)), and $\sqrt{f}$. When each of the 6

f predictors was paired with each of the 6 n predictors, 36 predictors were possible. When all 4 basic predictor variables were considered, a total of 216 product pairs were added to the former 24 predictors, creating a set of 240 predictors. Regression by the forward, stepwise, and all-possible techniques was employed in search of terms that explained large portions of the variance in p. When the residuals of models based on these predictors were plotted, three somewhat parallel curves were observed. This prompted separation of the data set into three sets, one for each α level. R^2 values greater than .98 were obtained for each α level considered separately. Such formulae marginally satisfied the accuracy criterion but did not meet the simplicity criterion in which one formula incorporating all α levels was desired.

In the interest of simplicity, all terms containing logarithms were eliminated from the model. This reduced the possible predictor set from 240 to 112 predictors. Similar R^2's were attained without the complexity of the logarithmic terms. While marginally acceptable R^2 levels were obtained for specific α levels, the R^2 values obtained for general formulae were not deemed acceptable.

Several recurring predictors were observed in the formulae for the separate α levels, and it was hoped some form of α could be entered as a factor with these predictors to develop formulae that were acceptable for all α levels. Especially encouraging was the pair $f\sqrt{n}$ and f^2n, the second being the square of the first. Because $f\sqrt{n}$ continued to be prominent throughout the experimentation, η (eta) was defined u^2 to simplify future predictor notations. Experimentation with powers of η and α, along with various other terms from the current models, failed at this point to obtain acceptable formulae, however.

As stated earlier, the above models included only power values in the interval [.25, .95]. Continuing the search for formulae which would be simpler and more accurate, another reduction of the data set was tested. Because the user of a k-group ANOVA design is only rarely concerned with a comparison of more than five groups, the data for six or more groups was removed. Because u = k-1, this reduction meant only data points with values of u in the interval [1, 4] were used in the continuing search. This restriction of the data set permitted a model which included all α levels and which attained an R^2 =.9656 for the predictor set η, $f\eta^2$, $\sqrt{u\alpha}$, $u\alpha$, $u^2\alpha^3$, and n^2f^3. Individual models for specific α levels attained R^2 >.99. The best model (for α =.01 using the predictors η, η^2, $\sqrt{u}$, f, f^2, and $f^2\eta$) attained R^2 =.9962.

Many other possibilities were explored in the search for good models. Just as the separation of α levels had been explored, a separation of f levels was tested. Models in which p was replaced with ln(p), exp(p), or a trigonometric function of p were tried. None of these experiments yielded any improvement when compared to those models already reported.

To balance the two criteria, simplicity and accuracy, a compromise was required. Having chosen an accuracy requirement of RMSE < .03, it appeared the best general formula (given here in a factored form) contained six predictors and seven constants:

$$p = -.034\eta^3 + \left(.240 - .720\sqrt{\alpha}\right)\eta^2 + \left(2.178\sqrt{\alpha} + .043u\right)\eta - (.192f + .268) \quad (1)$$

The accuracy of Formula 1 was attained by considering only data points with power in [.25, .90] and u in [1, 4]. While Formula 1 is not as simple as originally hoped, its simplicity was deemed reasonable, considering the magnitude of the problem. The simplest possible linear combination of the four basic predictors would require five constants for the four terms plus an intercept term. That simple model, however, demonstrated very poor accuracy. Formula 1 requires only seven constants and it provides good accuracy, so it may be considered reasonable by potential users.

To reduce the number of terms (and constants) required in the model, consideration was given to re-entering logarithmic predictors into the model. Many combinations were tried using all possible regressions on various predictor sets. The model

$$p = .058 + .149\ln(\alpha) + (.355 + .045u)\eta + .197\ln(n)\sqrt{f} \quad (2)$$

emerged after much experimentation. Formula 2 has only five constants, compared to seven in Formula 1. The four predictors are, however, more complex than the six predictors of Formula 1.

The two formulae presented above appeared comparable in simplicity. To test the accuracy of Formulae 1 and 2, their residuals were analyzed (see Table 2) and the two formulae were again found to be comparable. The residual plots indicated a high degree of accuracy had been attained, but each plot exhibited a curvature which invited further exploration using a cubic function of p. Thus a two stage estimation procedure was considered. Stage One was either of the above two formulae. Stage Two then entered the resulting p into the model $P = b_0 + b_1p + b_2p^2 + b_3p^3$. The new P gave a better power prediction but the improvement was judged too minimal to warrant the application of Stage Two. The first formulae were already less simple than desired and it was felt the application of a second stage formula would probably not be attractive to any user.

Non-Linear Formulae

Because power data is not linear, non-linear models such as $p = b_0\left(1 - \exp\left(b_1 u^{e_1} a^{e_2} n^{e_3} f^{e_4}\right)\right)$ and $p = b_0 + b_1 u^{e_1} a^{e_2} + b_2 n^{e_3} f^{e_4}$ were tested using computer iterations to determine the b and e parameter values which most closely fit the surface. Although many models failed to converge to a set of parameters, the above models did each converge with RMSE < .04. These results were less satisfactory, however, than the results from the linear Formulae 1 and 2 already reported. Finally, the logistic model was considered because its graph approximates a power curve in shape, being asymptotic to zero and one. The logistic model, unlike linear models, might allow use of the full data set and might model ANOVA power well.

The logistic model is based on a sigmoidal curve with an equation similar to $Y = \frac{1}{1+e^{-X}}$. With $P = Y$ and a transformed $p' = x$, $P = \frac{1}{1+e^{p'}}$. Solving, $p' = \ln\left(\frac{1-P}{P}\right)$. Although P is sigmoidal, the transformed p' is linear. This p' was regressed on various sets of predictors using the linear model $p' = \beta_0 + \sum_{j=1}^{k} \beta_j X_j$. The resulting coefficients were then substituted yielding a model similar to $P = \frac{1}{1+e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_j}}$.

For initial trials, the five predictors (α, u, n, f, and η) were tested. The standard deviation of the residuals was .0858 with the data set limited to power in the interval [.25, .95] but when the entire data set was allowed, the standard deviation of the residuals was .0768. The logistic model performed as well, if not better, with an unrestricted data set. As experimentation continued, the predictors which had been discovered in the search for linear models were found to be helpful in the search for logistic models.

Although logistic regression produced pleasing power estimates in the asymptotic tales, a disturbing feature of these models was the wide range of the prediction errors indicated by the minimums and maximums of the residuals. Even in the best model, one residual was as large as .39. Through analysis of the data, the source of these extreme residuals was found to be cases of very low n and large u. When the restrictions $n \geq 10$ and u in [1, 4] were placed on the data, similar to the restrictions used in developing the linear models, a better fit was obtained:

$$p = \frac{1}{1+2.81\alpha^{-.72}u^{(.31-.27\eta)}\eta e^{[.91f-(2.31+.17u)\eta]}} \quad (3)$$

In Formula 3, all errors were within ± .05 and the standard deviation of the errors was .0150, a very pleasing result when considering the accuracy criterion of acceptable formulae. However, the simplicity criterion was challenged by this model. Any logistic model is by nature complex when compared to a linear model.

The Formulae Compared

The three formulae produced by this study each have features which may be attractive to users. Formula 1 is simple but lengthy. Formula 2 is more compact, but it includes logarithmic terms. Formula 3 is the most accurate, but it is also the most complex. In addition to these basic comparisons, the user might consider the tables of residuals associated with the formulae. Tables 3, 4, and 5 show the standard deviations of the residuals as well as the minimum and maximum residuals under various restrictions of the predictor variables. As an example, a design of five groups (u = 4) and five subjects per group (n = 5) is described in the next to last line of each table. If Formula 3 is chosen, the standard deviation of the residuals is .0125. Assuming normality of the residuals, 95% of the predicted power values would be within ± 1.96(.0125) = ± .0245. For the worst case, the predicted power value could be as much as .0984 too great or .0320 too small. (Power = predicted power + error.)

The superiority of the logistic Formula 3 of Table 5 is obvious, shown by the smaller numbers throughout. In addition, the logistic formula is based on the entire data set with power in [.01, .99]. The linear formulae were developed using only the data with power values in [.25, .90]. Of course, the accuracy of Formula 3 was gained at the expense of simplicity.

The user's choice of one of these three formulae will depend upon the user's desires and purposes. If the user desires the simplest formula, one of the linear formulae (Formula 1 or Formula 2) should be chosen. If greater accuracy is desired, the logistic formula (Formula 3) should be chosen. The user desiring accurate predictions in the tail regions of the model should always choose the logistic formula. A user may use a linear formula several times to test possible models and then, having narrowed the choices, use the logistic formula to make a final model selection. With computer spreadsheets, it is also possible for the user to consider the results of all three formulae simultaneously when proposing various ANOVA designs.

Although the formulae provide good power estimates in most cases, a user never knows whether

the estimate obtained in a particular case is highly accurate or only marginally accurate. Reference to plots of residuals can provide further insight for interpreting the power predictions calculated from the formulae. Figures 2, 3, and 4 show residuals plotted against the predicted power for each of the three formulae. To illustrate, consider Figure 2. Under Formula 1, if the predicted power is .75, the plot shows the residuals vary from -.04 to .06. Thus the actual power of the design is .71 to .81.

Although the user of the formulae may not always have the residual plots available, the formulae can still be used effectively if the user understands the general shape of the residual plots. The user of the linear Formulae 1 and 2 must be aware that the power will be over predicted when power is high, and under predicted when power is low. This is especially clear in Figure 1 where the full data set of power in [.01, .99] is plotted. When such predictions are obtained from the formulae, the user must interpret the results as "high" or "low" power respectively, without stating a specific power value. An example of extreme power predictions is the case of α=.05, n=500, u=1, and f=.4 (prediction = 15.602, error = -14.612). This error results from the dramatic negative effect of the factor η^3 for large values of n. Power for n=500 is expected to be very high, clearly outside of the [.25, .90] power range. Computed power estimates which fall into the range for which the formulae were developed will be reasonable power estimates, but the user is warned that any extreme power predictions of Formulae 1 or 2 should be ignored.

The shape of the residual plot of Formula 3 (Figure 4) is very different from the shape of the plots for the linear formulae, the logistic formula being more accurate in the tail regions than in the central regions. A comparison of the scales of the plots, however, demonstrates that the increased tail accuracy is not at the expense of accuracy in the central regions. Formula 3 meets or exceeds the performance of the other two formulae even in the central regions.

The formulae developed by this study offer a new way to compute the power of ANOVA designs. These formulae resulted from a directed "trial and error" search among those predictors which seemed reasonable. Certainly, the study did not exhaust all possible predictors of power. Thus, other researchers may discover better (simpler and/or more accurate) formulae than those presented here. This may be done with the tool of regression, as used in this study, or by some other method not yet considered.

References

Borenstein, M., Cohen, J., Rothstein, H. R., Pollack, S., & Kane, J. M. (1990). Statistical power analysis for one-way analysis of variance: a computer program. *Behavior Research Methods, Instruments, and Computers, 22,* 271-282.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hilldale, NJ: Lawrence Erlbaum Associates.

Dixon, W. J., & Massey, F. J. (1969). *Introduction to statistical analysis* (3rd ed.). New York: McGraw-Hill.

Donaldson, T. S. Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. *Journal of the American Statistical Association, 63,* 660-676.

Fox, M. (1956). Charts of the power of the F-test. *Annals of Mathematical Statistics, 27,* 484-497.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research,* 42, 237-288.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology (2nd ed.).* Englewood Cliffs, NJ: Prentice-Hall.

Goldstein, R. (1989). Power and sample size via MS/FC-DOS computers. *The American Statistician, 43,* 253-260.

Haase, R. F. (1986). A BASIC program to compute statistical power for atypical values of alpha. *Educational and Psychological Measurement, 46,* 629-632.

Lehmer, E. (1944). Inverse tables of probabilities of errors of the second kind. *Annals of Mathematical Statistics, 15,* 388-398.

Lenth, R. V. *PowerPack (Version 2.22)* (no date given) [Computer program]. 3061 Hastings Avenue, Iowa City, IA 52240. (319) 337-8549.

Levin, J. R. (1975). Determining sample size for planned and post hoc analysis of variance comparisons. *Journal of Educational Measurement, 12*, 99-108.

McCormack, D. R. (1993). Formulae for estimating the power of one-factor ANOVA designs. (Doctoral dissertation, University of Northern Colorado, Greeley).

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models.* (2nd ed.). Homewood, IL: Irwin.

NONCDIST *(Quality control and industrial experiments)* (no date given) [Computer program]. Lionheart Press, P. O. Box 379, Alburg, VT 05440. (514) 933-4918.

O'Brien, R. G. (1988). Review of PowerPack. The *American Statistician, 42*, 266-270.

Pearson, E. S., & Hartley, H. O. (1951). Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. *Biometrika, 38*, 112-130.

Rogers, W. T., & Hopkins, K. D. (1988). Quick power estimates incorporating the joint effects of measurement error and a covariate. *Journal of Experimental Education, 57*, 86-94.

Seber, G. A., & Wild,C. J. (1989). *Nonlinear regression.* New York: Wiley.

Shaw, D. G., & McCormack, D. R. (1992) Estimating the power of a t-test. Unpublished manuscript, University of Northern Colorado, Greeley.

Srivastava, A. B. L. (1959). Effect of non-normality on the power of the analysis of variance test. *Biometrika, 46*, 114-122.

Statistical Power Analysis (version 1.0) (no date given) [Computer program]. Lawrence Erlbaum Associates, Inc., 365 Broadway, Hillsdale, NJ 07642. (201) 666-4110.

Tang, P. C. (1938). The power function of the analysis of variance with tables and illustrations of their use. *Statistical Research Memoirs, 2*, 126-149.

Table 1. Independent Variables of the Linear Regression Model

Variable	Description	Values or Range
α	significance level	.01, .05, .10
u = k - 1	numerator degrees of freedom for a k-group ANOVA	1 - 24
n	per group sample size	2 - 1000
f	Cohen's effect size	.05 - .80

Table 2. Comparison of Residuals of Formula 1 and Formula 2

Formula	Mean	Standard Deviation	Minimum	Maximum
1	.0006	.0254	-.1387	.0617
2	.0006	.0251	-.0901	.1080

Table 3. Residuals of Linear Formula 1, p in [.25, .90].

$$p = -.034\eta^3 + \left(.240 - .720\sqrt{\alpha}\right)\eta^2 + \left(2.178\sqrt{\alpha} + .043u\right)\eta - (.192f + .268)$$

u Values	n Values	St Dev	Min	Max
All	All	.1094	-.7782	.0617
[1, 8]	All	.0340	-.2301	.0617
[1, 8]	n≥5	.0313	-.1466	.0617
[1, 8]	n≥10	.0306	-.1414	.0608
[1, 4]	All	.0254	-.1387	.0617
[1, 4]	n≥5	.0242	-.0979	.0617
[1, 4]	n≥10	.0235	-.0897	.0608

Table 4. Residuals of Linear Formula 2, p in [.25, .90].

$$p = .058 + .149\ln(\alpha) + (.355 + .045u)\eta + .197\ln(n)\sqrt{f}$$

u Values	n Values	St Dev	Min	Max
All	All	.1167	-.7375	.1080
[1, 8]	All	.0350	-.2027	.1080
[1, 8]	n≥5	.0338	-.1221	.1080
[1, 8]	n≥10	.0335	-.1221	.1080
[1, 4]	All	.0250	-.0901	.1080
[1, 4]	n≥5	.0247	-.0876	.1080
[1, 4]	n≥10	.0243	-.0876	.1080

Table 5. Residuals of Formula 3, p in [.01, .99].

$$p = \frac{1}{1 + 2.81\alpha^{-.72}u^{(.31-.27\eta)\eta}e^{\left[.91f-(2.31+.17u)\eta\right]}}$$

u Values	n Values	St Dev	Min	Max
All	All	.0445	-.5195	.0320
[1, 8]	All	.0160	-.2156	.0320
[1, 8]	n≥5	.0126	-.0984	.0320
[1, 8]	n≥10	.0115	-.0516	.0295
[1, 4]	All	.0143	-.1276	.0320
[1, 4]	n≥5	.0125	-.0984	.0320
[1, 4]	n≥10	.0113	-.0516	.0295

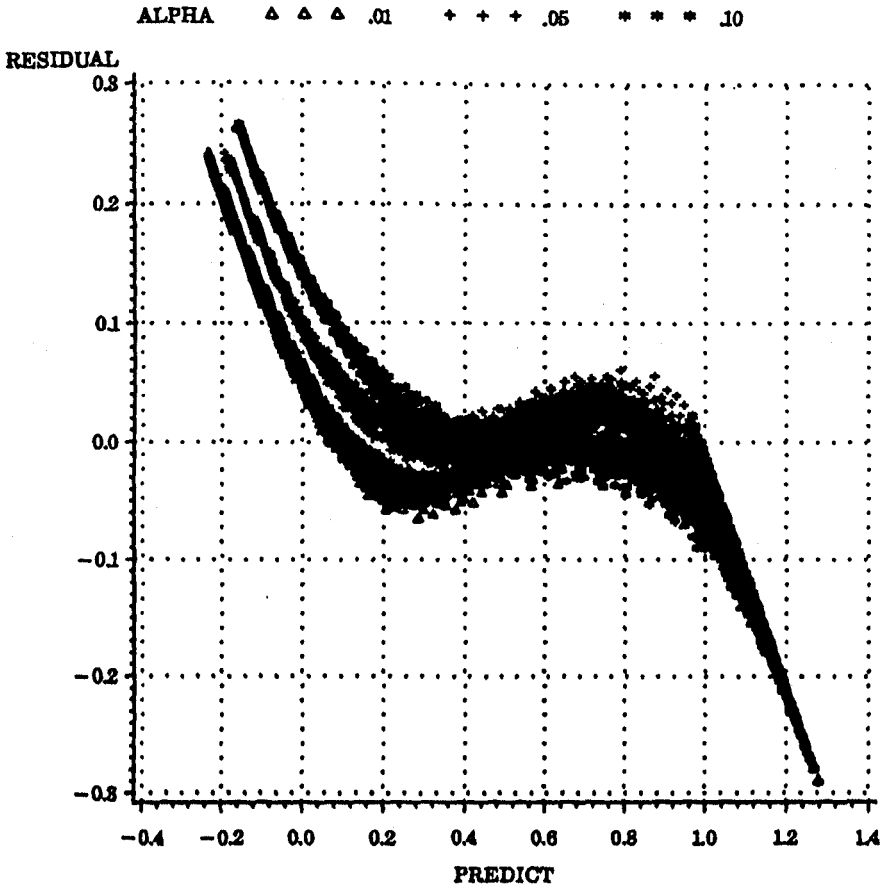

Figure 1. Formula 1, p in [.01, .99], u in [1, 4], n

$$p = -.034\eta^3 + \left(.240 - .720\sqrt{\alpha}\right)\eta^2 + \left(2.178\sqrt{\alpha} + .043u\right)\eta - (.192f + .268)$$

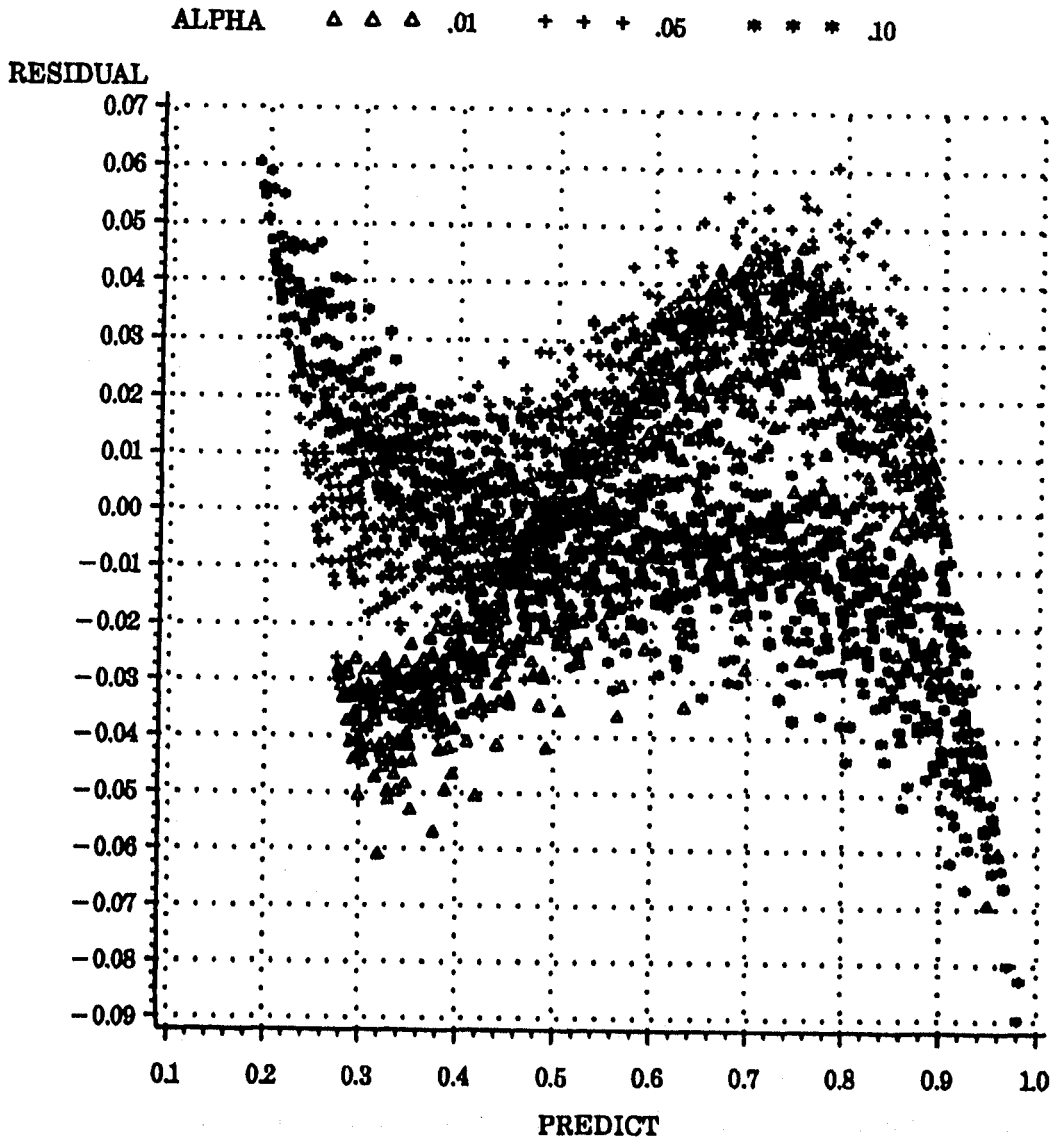

Figure 2. Formula 1, p in [.25, .90], u in [1, 4], n

$$p = -.034\eta^3 + \left(.240 - .720\sqrt{\alpha}\right)\eta^2 + \left(2.178\sqrt{\alpha} + .043u\right)\eta - (.192f + .268)$$

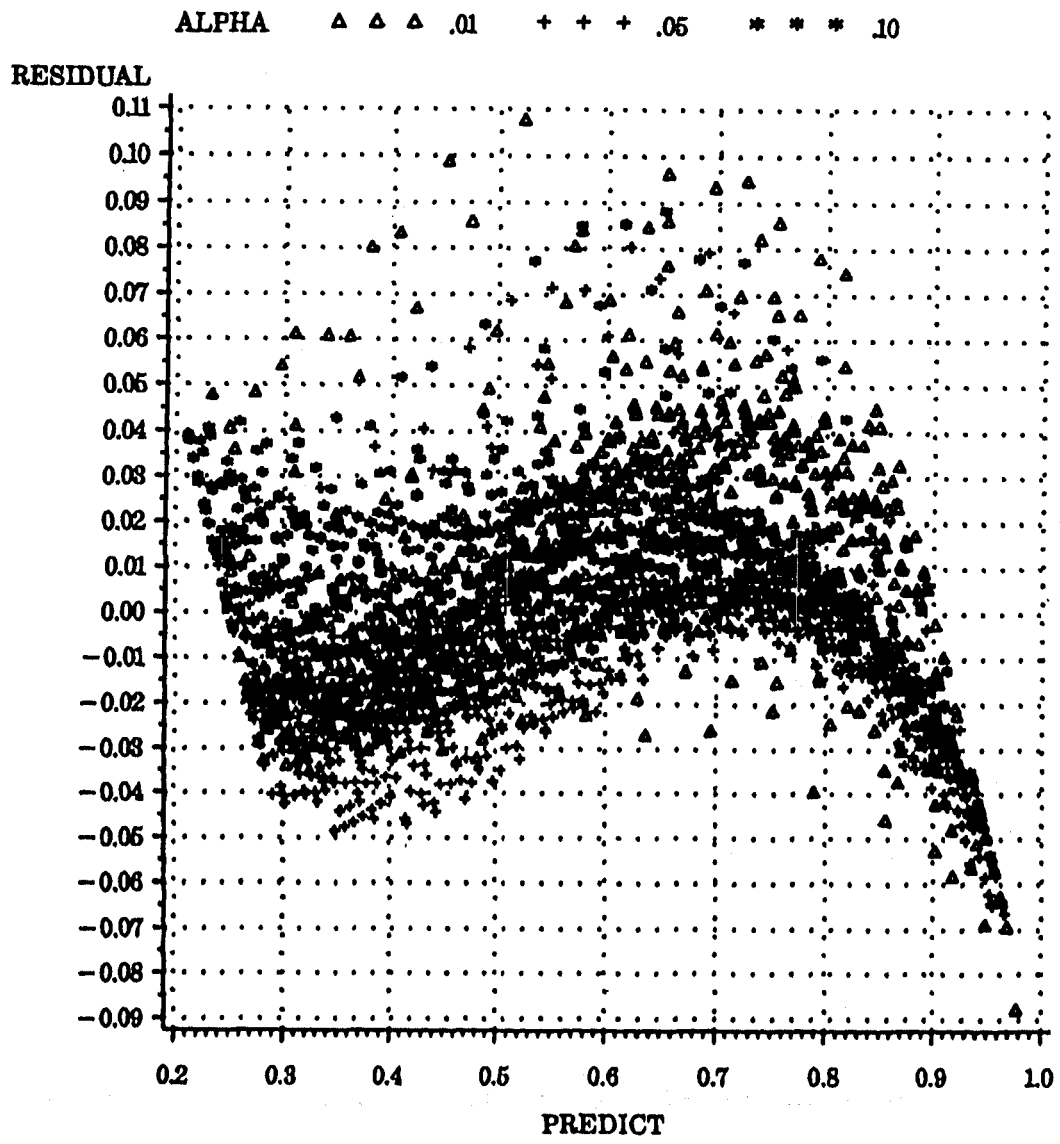

Figure 3. Formula 2, p in [.25, .90], u in [1, 4], n

$$p = .058 + .149\ln(\alpha) + (.355 + .045u)\eta + .197\ln(n)\sqrt{f}$$

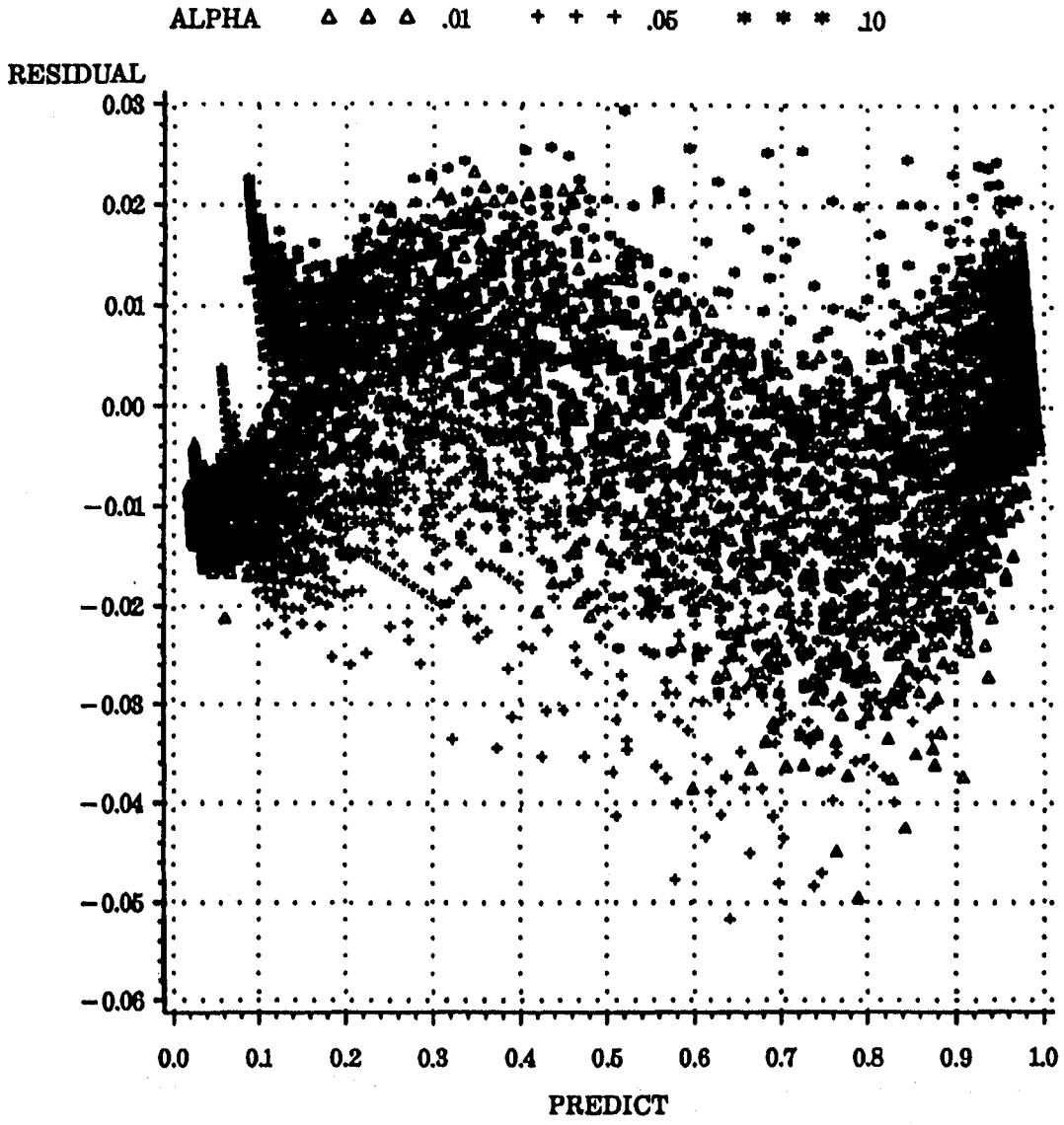

Figure 4. Formula 3, p in [.01, .99], u in [1, 4], n

$$p = \frac{1}{1 + 2.81\alpha^{-.72}u^{(.31-.27\eta)\eta}e^{[.91f-(2.31+.17u)\eta]}}$$