COMPARISON OF ALTERNATIVE REGRESSION MODELS FOR PREDICTING CHANGE

Robert B. Bendel, Washington State University

ABSTRACT:

The consulting statistician frequently encounters problems in which an initial score (pre-test) and a final score (post-test) are observed. This paper contrasts three regression models which use the final score and change score (final score minus initial score) as dependent variables. It has been noted that for most problems the initial score should be used as an independent variable or covariate. When the two regression models which have the same independent variables but different dependent variables are contrasted, the models differ only in their multiple correlation coefficients but the standard error of estimate and other important statistics are the same. Tests of hypotheses conditional on the initial score are also the same for both models. An example is given and related topics encountered in the behavioral and animal sciences are discussed.

INTRODUCTION:

The problem is to see the mathematical relationships between three regression models with a view toward choosing the most appropriate model. Measurements on some variable are taken before (initial score) and after (final score) treatment. Treatment can be either quantitative (a regression problem) or qualitative (an ANOVA/ANCOVA problem). The mathematical aspects will be displayed in the more general context of a regression model but the results also apply to the more popular (special case) ANCOVA model. In the context of psychometrics, say, the problem can be viewed as the regression analogue of the pre-intervention-post design:

Pre-intervention-Post Design

	Pre	Post
Control		
Treatment		

Question researcher asks is: Is there more change in the treatment group than in the control group?

Statistical hypothesis: Test for equality of gains for the two or more groups.

Assumptions: At this point, we assume that all classical assumptions are satisfied, including random assignment to treatment groups.

DEFINITIONS:

IS = initial score
FS = final score
G = FS-IS = gain score
R^2 = coefficient of determination
= square of the multiple correlation coefficient
$s_{y \cdot x}$ = standard error of estimate
= square root of the estimated variance about the regression line
RSS = residual sum of squares
edf = degrees of freedom for RSS

PROBLEM:

Which of the three regression models do we choose and what differences are there in the regression statistics, R^2 and $s_{y \cdot x}$?

Model	Dependent Variable	Independent Variables
FS	FS	IS + $\underset{\sim}{x}_2$
G	G = FS-IS	IS + $\underset{\sim}{x}_2$
$\overline{\text{IS}}$	G = FS-IS	$\underset{\sim}{x}_2$ only

MODEL EXAMPLE:

For the G model, let $\underset{\sim}{x}_2$ represent the independent variables excluding IS. Then, for example, $\underset{\sim}{x}_2$ could represent age and a treatment variable, trt. The G model could then be represented by the equation

$$G = \beta_0 + \beta_1 \times \text{IS} + \beta_2 \times \text{Age} + \beta_3 \times \text{trt}.$$

We may be interested in predicting gain or we may be interested in seeing how treatment affects gain after adjustment for IS and age.

MODEL PREFERENCE:

Choose either the FS or G model since $s_{y \cdot x}$ is the same for both models. R^2 is generally different. The $\overline{\text{IS}}$ model is usually inadequate since IS is frequently related to G or FS.

THEOREM:

The residuals, RSS and $s_{y \cdot x}$ are the same for the FS and G models.

IDEA OF PROOF:

$$\text{Var (FS-IS}|\text{IS)} = \text{Var (FS}|\text{IS)}$$

That is, the variance about the population regression line is the same for the FS and G models since both have the same independent variables and both vary the same at each value of IS. To illustrate, consider the (IS, FS) data (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (2, 5) consisting of three FS values at each of the two values of IS. A plot of IS versus FS and a separate plot of IS versus G indicate that $s_{y \cdot x}$ is equal to one for both plots whereas $R^2 = .273$ for the (IS, FS) data and $R^2 = 0$ for the (IS, G) data.

PROOF:

Basic idea is to show that the regression coefficients for $\underset{\sim}{x}_2$ are the same for both models and that the regression coefficient for IS satisfies the condition $b = 1 + g$ (This has been noted by Werts and Linn, 1970). Without loss of generality, let $\underset{\sim}{x}_2$ consist of one independent variable x_2. Then

$$RSS = \min \Sigma(FS - b_0 - b_1 \times IS - b_2 x_2)^2$$

for FS model and

$$RSS = \min \Sigma(FS - IS - g_0 - g_1 \times IS - g_2 x_2)^2$$
$$= \min \Sigma(FS - g_0 - (1 + g_1) \times IS - g_2 x_2)^2$$

for G model. Assuming there are no singularity problems, the Gauss-Markov theorem says that the least squares estimates of the regression coefficients are unique so,

$$b_0 = g_0$$
$$b_1 = 1 + g_1$$
$$b_2 = g_2$$

from which it follows that the residuals, $FS - b_0 - b_1 \times IS - b_2 x_2$, are the same for the FS and G models.

Hence, RSS and

$$s_{y \cdot x} = (RSS/edf)^{1/2}$$

are also the same since edf = n - #parameters estimated is the same for both models.

EXAMPLE FROM PSYCHOLOGY (Mental Retardation):

- Score = Adaptive Behavior (AB)
- IS = initial score, AB at time 1
- FS = final score, AB at time 2
- $\underset{\sim}{x}_2$ = vector of independent variables, e.g., IQ, age, treatment = environmental factor score.

See Table 1. Note the following:

a) $.88 = -.12 + 1.0$ since $b_1 = g_1 + 1$

b) the standard errors are the same for the FS and G models.

c) the t values are the same (except for IS); for x_{23}, $t = 5.83$; so a test for significance of the partial regression coefficient of trt after adjustment for IS, age, etc. is highly significant.

d) $s_{y \cdot x}$ is the same for the FS and G models.

e) R^2 is different; R^2 is usually higher for the FS model since the variance of FS is higher than the variance of G whenever IS and G are positively correlated - as is usually the case.

Note: In general, it can be shown that tests of hypotheses conditional on IS are the same for the FS and G models. To see this, let H_0: $\underset{\sim}{\beta}_q = \underset{\sim}{0}$ where $\underset{\sim}{\beta}_q$ is a vector of regression coefficients which does not include IS. Then the F test can be written as

$$F = (RSS' - RSS)\, edf/RSS\, (edf' - edf)$$

where RSS' and edf' denote the RSS and edf under the null hypothesis. Since IS is "included" in RSS and RSS' and edf and edf' are the same for both the FS and G models, it follows that tests of hypothesis, H_0: $\underset{\sim}{\beta}_q = \underset{\sim}{0}$, are identical for both the FS and G models.

EXAMPLE FROM ANIMAL SCIENCE:

- IW = initial weight of steer
- FW = final weight of steer
- G = FW-IW = gain in weight
- x_2 = treatment coded as 1, 2, 3, 4, which is simply a ranking of the amount of concentrate in the diet. For ANCOVA and ANOVA of gain scores, x_2 is used as a qualitative grouping variable.

Table 2 shows the results for comparing ANOVA of Gain, repeated measures (RM) ANOVA, ANCOVA with FW and G as the dependent variables and multiple regression using the treatment variable as a quantitative (1, 2, 3, 4) variable. Note that the ANOVA of Gain and the time by treatment interaction in the RM ANOVA test the same hypothesis (that the gain is the same for each treatment) so that the F value is the same as that in the ANOVA of Gain. In comparing the ANCOVA models, the same results are true for the ANCOVA models as are true for the regression models, i.e., that $s_{y \cdot x}$ is the same, R^2 is generally different, and that the regression coefficient for IW is one more for the FW model than for the G model.

When all the assumptions are met including random assignment to treatment groups and the covariate and independent variables are measured without error, it has been established (Bock, 1975) that ANCOVA is more powerful than ANOVA of Gain scores (and repeated measures since the same F value is obtained when testing the treatment by time interaction in a repeated measures ANOVA).

DIFFICULTIES INVOLVED WHEN THE ASSUMPTIONS ARE VIOLATED:

The educational, psychological, and sociological literature contain many papers discussing the use of gain scores and ANCOVA when the assumptions are not met. The paper by Cronbach and Furby entitled, "How should we measure "change" or should we?" is a classic and Lord's Paradox (Lord and Novick, 1968; Bock, 1975) is also a controversial paper. In short, it is felt that there is some agreement that ANCOVA can be used with caution when there are intact groups (no random assignment to treatment). See Elashoff (1969), Kenny (1975), and Alwin and Sullivan (1975). Also, when the covariate is measured with error, there is a general agreement that some form of adjustment should be made, but as Cochran (1968) discusses, this could depend on the assumed model (Is there a linear regression of FS on "true" IS or on IS measured with error?). Werts and Linn (1970) and Bergman (1971) discuss alternative models to use when dealing with change. The problems with some of these models is that they require an estimate of the reliability of the covariate and/or independent variable. The reliability, $\tilde{R}$, is defined as the ratio of the variance of the true value to the variance of the observed value. To be specific, let $X = x + e$, where X is the observed score, x is the true score, and e is the error of measurement. Assuming that x and e are independent, it follows that

$$\tilde{R} = \sigma_x^2/\sigma_X^2 = \sigma_x^2/(\sigma_x^2 + \sigma_e^2).$$

CONCLUSION:

Use either the FS or G models, but there is some controversy and some unanswered problems when the assumptions are violated.

REFERENCES:

Alwin, D. F. and Sullivan, M. J., 1975. Issues of design and analysis in evaluation research. Sociological Methods and Research, 4(August): 77-100.

Bergman, L. R., 1971. Some univariate models in studying change. Reports from the Psychological Laboratories. The University of Stockholm Supplement Series 10.

Bock, R. D., 1975. Multivariate Statistical Methods in Behavioral Research. McGraw-Hill, New York.

Cochran, W. G., 1968. Errors of Measurement in Statistics. Technometrics, 13 (No. 4):637-666.

Cronbach, L. J. and Furby, L., 1979. How should we measure "change" - or should we? Psychological Bulletin, 74 (No. 1):68-80.

Elashoff, J. D., 1969. Analysis of covariance: a delicate instrument. American Educational Research Journal, 6 (No. 3):383-401.

Kenny, D. A., 1975. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin, 82 (No. 3): 345-362.

Lord, F. M. and Novick, M. R., 1968. Statistical Theories of Mental Test Scores. Addison-Wesley, Reading, Mass.

Porter, A. C., 1973. Analysis strategies for some common evaluation paradigms. Occasional Paper No. 21, College of Education, Michigan State University.

Werts, C. E. and Linn, R. L., 1970. A general linear model for studying growth. Psychological Bulletin, 73 (No. 1):17-22.

Table 1.
An Example From MR Comparing the Three Regression Models

Model:	Coefficient FS	Coefficient G	Coefficient $\overline{IS}$	Standard Error FS	Standard Error G	Standard Error $\overline{IS}$	t-value FS	t-value G	t-value $\overline{IS}$
IS	.88	-.12	-	.062	.062	-	14.20	-2.02	-
x_{21}	-.10	-.10	-.14	.057	.057	.053	-1.74	-1.74	-2.64
x_{22}	-1.05	-1.05	-.56	.640	.640	.602	-1.64	-1.64	-.93
x_{23}	1.57	1.57	1.23	.269	.269	.211	5.83	5.83	5.82
x_{24}	1.22	1.22	1.07	.460	.460	.457	2.65	2.65	2.34
x_{25}	-2.33	-2.33	-2.328	1.189	1.189	1.200	-1.96	-1.96	-1.94

Model:	R	$s_{y \cdot x}$	edf	SS(TOTAL)	RSS	F for testing R
FS	.932	11.16	205	193078	25523	225.9
G	.528	11.16	205	35374	25523	13.2
$\overline{IS}$	.514	11.24	206	35374	26032	14.8

Table 2.
Comparison of Alternative Models for
Assessing Differences in Treatment Gains for Angus Steers

Model	edf	R^2	$s_{y \cdot x}$	b_{IW}, s_b, t^b	b_t, s_b, t^c	F^d
ANOVA of Gain	28	.71	55.39	-	-	22.4
RM ANOVA	28	-	-	-	-	22.4
ANCOVA [G]	27	.72	55.40	.18, .18, 1.0	-	21.1
ANCOVA [FW]	27	.82	55.40	1.18, .18, 6.4	-	21.1
REGR [G]	29	.52	69.28	.28, .22, 1.3	-59, 11, -5.4	28.7
REGR [FW]	29	.70	69.28	1.28, .22, 5.8	-59, 11, -5.4	28.7

[b]The three numbers, b_{IW}, s_b, t, represent the regression coefficient for IW, the standard error and the t value.

[c]The three numbers, b_t, s_b, t, represent the regression coefficient for the treatment variable, the standard error, and the t value.

[d]The value of F given in the table represents the F value associated with the main (treatment) hypothesis of interest. For the ANOVA of Gain, the hypothesis is the equality of mean treatment gains; for the RM ANOVA, it is the time by treatment interaction; for the ANCOVA models, it is the equality of the adjusted treatment means; for the regression models, the hypothesis is testing for significance of the treatment partial regression coefficient.