Practical Assessment, Research, and Evaluation

From Gain Score t to ANCOVA F (and vice versa)

Thomas R. Knapp

William D. Schafer

Follow this and additional works at: https://scholarworks.umass.edu/pare

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

From Gain Score t to ANCOVA F (and vice versa)

Thomas R. Knapp, *The Ohio State University*
and
William D. Schafer*, University of Maryland*

Although they test somewhat different hypotheses, analysis of gain scores (or its repeated-measures analog) and analysis of covariance are both common methods that researchers use for pre-post data. The results of the two approaches yield non-comparable outcomes, but since the same generic data are used, it is possible to transform the test statistic of one into that of the other. We derive a formula that can be used to accomplish a conversion between the two and give an example. Such a result could be helpful to meta-analysts, where the outcomes in different research reports may be of either of the two types, yet need to be synthesized. Suggestions for additional research that can improve the usefulness of the formula are offered.

A common theme in the methodological research literature consists of contributions regarding the superiority of one type of data analysis over another for the same research design. Some well-known examples are the use of planned orthogonal contrasts rather than multiple traditional t tests when there are more than two groups (see, for example, Kirk, 1968); the application of non-parametric analyses (Mann-Whitney, Friedman, etc.) to non-normal data; and the use of the concordance coefficient rather than multiple pairwise rank correlations (Kendall & Gibbons, 1990). The best discussions of such alternatives also tell the reader how to carry out the more defensible analysis.

A recent article (Hedges, 2007) discussed in considerable detail how one might correct an analysis that inappropriately used the individual as the unit of analysis when an aggregate (classroom, school, etc.) should have been used instead. A subsequent article (Schochet, 2008) compared the power of analyses based upon individuals with the power based upon aggregates.

In the spirit of both of those articles we would like to revisit a long-standing controversy regarding analyses of the data for the true experimental pretest-posttest control group design (Design #4 in Campbell & Stanley, 1963) and offer a way of converting from one analysis to another. The controversy, which is well known, revolves primarily around the use of an independent t test on the "gain" scores (gain defined as posttest minus pretest) vs. an analysis of covariance (ANCOVA) with group as the principal independent variable, with the posttest score as the dependent variable, and with the pretest score as the covariate.

The existence of the controversy implies that literature in which the very common two-group, pre-post is used will differ in the analyses used. This poses a problem for anyone who is working on a quantitative synthesis (meta-analysis) of the research in the field. Thus, a way to convert the analyses from one to the other would be convenient.

We first discuss the gain-score vs. ANCOVA controversy. Then we present a conversion formula and give an example. Finally, we discuss how the formula might be used in practice. The formula itself is derived in an appendix.

The controversy

The claim that gain score analysis (GSA) and ANCOVA can yield disparate results was forcefully made by Lord (1967) and has come to be known as "Lord's Paradox" (see also Locascio & Cordray, 1983). In Lord's hypothetical example there was a large treatment effect when using ANCOVA but no treatment effect whatsoever when using GSA.

There are several arguments for and against the use of a t test on the gain scores. The principal arguments favoring the t test approach are its relative simplicity, its fewer assumptions and calculations, and its ubiquitous use. Nothing is more straightforward than comparing the mean change from pretest to posttest for an experimental group with the mean change from pretest to posttest for a control group in order to get some evidence regarding the effect of an experimental treatment. There is no need even to carry out the regression of posttest scores on pretest scores. Such analyses have been conducted for years.

Some methodologists advocate the use of repeated-measures analysis of variance (ANOVA) for this same design (see Huck & McLean, 1975). The principal problem is the confusion regarding the three F-ratios: one for the main effect of treatment, one for the main effect of time, and one for the treatment-by-time interaction. The most relevant F is for the treatment-by-time interaction, and it turns out that it is mathematically equivalent to the square of the t for the gain scores. Because of the mathematical equivalence, we will not discuss the repeated measures approach further.

The principal arguments against the gain score approach (and for ANCOVA) are that the simplicity is deceiving. Gain scores may not be very reliable, power is usually greater for ANCOVA, the gain scores are negatively correlated with the pretest, and the assumptions of linearity and homogeneity of regression are interesting in and of themselves and may be tested. More subtle questions can be addressed if the assumptions are found not to be tenable.

There is an important difference between the research question that is implied by the use of the t test and the research question that underlies the use of ANCOVA. For the former, the question is: "What is the effect of the treatment on the change from pretest to posttest?" For the latter the question is: "What is the effect of the treatment on the posttest that is not predictable from the pretest (i.e., conditional on the pretest)?"

While methodologists (e.g., Elashoff, 1969) overwhelmingly recommend randomization of participants to treatments, research opportunities often do not allow for it. Failing randomization, use of pretests is sometimes recommended as a way of exerting statistical control over pre-existing differences. Lord's paradox is an example, the group membership variable

in his data is gender. Thus, it is common to see GSA t tests and ANCOVAS on pre-post data.

No matter whether the preference is for a t test on the gain scores or for an ANCOVA, it is of some interest to be able to convert the bases for the corresponding statistical inferences to one another. For example, in meta-analysis an effect size is needed for each study that is comparable across studies. An effect size may be found by knowing the test statistic and the sample sizes (see Cohen, 1988), so a method to convert from one test statistic to another may be of value to a meta-analysis. The remainder of this paper is devoted to an explanation of how that can be accomplished, using for illustrative purposes a set of data taken from Rogosa (1980).

The conversions

The transformation from the independent-samples t to the ANCOVA F can be accomplished by first defining $F_t = t^2$ as the F equivalent of t. It is well-known that F for 1 and ν degrees of freedom is equal to the square of t for ν degrees of freedom, where in our context $\nu = n_E + n_C - 2$, n_E is the number of subjects in the experimental group, n_C is the number of subjects in the control group, and $n_E + n_C = n_T$.

In order to convert F_t to the covariance F_c one may use the following formula (proof provided in the Appendix):

F_c (for df = 1 and n_T-3) = A F_t (for df = 1 and n_T-2), where

$$A = \frac{\left[\frac{n_T - 3}{n_T - 2}\right]\left[\left(\frac{(n_T - 1)s_{YT}^2(1 - r_{XYT}^2)}{(n_T - 2)s_{YW}^2(1 - r_{XYW}^2)}\right) - 1\right]}{\frac{(n_T - 1)(s_{YT}^2 + s_{XT}^2 - 2r_{XYT}s_{YT}s_{XT})}{(n_T - 2)(s_{YW}^2 + s_{XW}^2 - 2r_{XYW}s_{YW}s_{XW})} - 1}$$

The variances s^2 for the pretest (X) and the posttest (Y), and the correlations r_{XY} between the pretest and the posttest are taken within group (W) and total-across-groups (T) as indicated by the respective subscripts. Since the two within-group variances are unlikely to be identical, nor are the two within-group correlations, they have to be "pooled."

In order to convert from an ANCOVA F_c to a gain score t one may divide the F_c by A and take the square root of the F_t with one more df for t.

An example

Consider these data taken from Table 1 of Rogosa (1980), where Group 1 = experimental, Group 2 = control:

ID	Group	Pre	Post	Gain
1	1	0.28	2.23	1.95
2	1	0.97	4.99	4.02
3	1	1.25	3.37	2.12
4	1	2.46	8.54	6.08
5	1	2.51	8.40	5.89
6	1	1.17	3.70	2.53
7	1	1.78	7.93	6.15
8	1	1.21	2.43	1.22
9	1	1.63	5.40	3.77
10	1	1.98	8.44	6.46
11	2	2.36	3.25	0.89
12	2	2.11	5.30	3.19
13	2	0.45	1.39	0.94
14	2	1.76	4.69	2.93
15	2	2.09	6.56	4.47
16	2	1.50	3.00	1.50
17	2	1.25	5.85	4.60
18	2	0.72	1.90	1.18
19	2	0.42	3.85	3.43
20	2	1.53	2.95	1.42

Some descriptive statistics:

	Group	n	MEAN	VARIANCE (div.by n_i-1)
Pre(X)	1	10	1.524	0.476
	2	10	1.419	0.489
	Total	20	1.472	0.460
Post(Y)	1	10	5.543	6.703
	2	10	3.874	2.876
	Total	20	4.708	5.272
Gain(Y-X)	1	10	4.019	4.028
	2	10	2.455	2.082
	Total	20	3.237	3.538

Group 1 regression:
Correlation of Pre and Post = .882
Squared correlation = .778
The regression equation is
Post = .497 + 3.311 Pre

Group 2 regression:
Correlation of Pre and Post = .542
Squared correlation = .294
The regression equation is
Post = 2.010 + 1.313 Pre

Total sample regression:
Correlation of Pre and Post = .705
Squared correlation = .497
The regression equation is
Post = 1.200 + 2.385 Pre

S_{XW} = .695
S_{YW} = 2.189
r_W = .730
b_W = 2.298

The adjusted posttest means are 5.424 (group 1) and 3.996 (group 2).

For those data the experimental group had a greater mean gain (4.02) than the control group (2.46). An independent-samples t test shows that the difference in mean gain is statistically significant at the .05 level, one-tailed (t = 2.00, df = 18; F_t = 4.00), but not two-tailed. The ANCOVA for the same data yields an F_c of 4.27 (df = 1 and 17), which is also statistically significant at the .05 level, one-tailed, but not two-tailed. (Since the within-sample variances were not identical, nor were the within-sample correlations, "pooling" was necessary.) The conversion factor A for transforming from one analysis to the other for these data is found to be (rounding to two decimal places for convenience):

$$A = \frac{\left(\frac{17}{18}\right)\left[\frac{19(5.27)(1-.50)}{18(4.79)(1-.53)} - 1\right]}{\frac{19[5.27+.46-2(.71)(2.30)(.68)]}{18[4.79+.48-2(.73)(2.19)(.69)]} - 1}$$

= 1.07 (rounded)

Other approaches to the problem

GSA and ANCOVA are not the only defensible alternatives for analyzing the data for a two-group pretest-posttest design. One reasonably well-known approach is the Johnson-Neyman technique (Johnson & Neyman, 1936; Johnson & Fay, 1950; Schafer, 1981) that is to be preferred to ANCOVA when the two within-group population regression lines are not assumed to be parallel. In the above example the sample regression lines have slopes of 3.31 and 1.31, indicating that the homogeneity of regression assumption for ANCOVA might be violated; the homogeneity of regression test yielded $F_{1,16}$ = 4.38, p=.053, non-significant two-tailed, but statistically significant, one-tailed. Since the Johnson-Neyman technique is not germane to our purpose, we did not follow up.

Another approach is the less well-known and computationally more complex method advocated by Rogosa (1980) and applied to the above data, in which the treatment effect is determined as a function of the distance between the within-group regression lines for the two samples.

All general linear model analyses such as the t test, ANOVA, ANCOVA, and regression can be considered as special cases of canonical correlation analysis (see, for example, Knapp, 1978) through the use of Wilks' lambda, which is a function of F. Therefore, all of the conversions could be expressed in terms of that statistic.

Researchers may find our formula helpful in converting from one analysis to the other. Further work might be designed to enhance the formula's utility in practice. For example, one difficulty that researchers may face is lack of information in papers and articles prepared by others about the results needed to calculate the terms in the formula. Suggestions for estimates of the terms given results (or lack of them) that may be contained in research reports could be proposed and evaluated, either though study of available literature and/or by simulation. Another direction for further work would be to reflect more complicated designs, with more groups, more covariates, more factors, or any combination of these. Finally, study of ways to convert the significance testing results to effect sizes along with their standard errors would ensure that fewer research reports are lost to meta-analysts because of incompatibilities between the choices made by different researchers and the information they report. Some headway has already been made in that regard. For example, Rudner, Glass,

Evartt, and Emery (2002) provide formulas for converting a t or an F to a common "d" scale, where d is the difference between two means divided by the pooled estimate of the standard deviation of the dependent variable.

References

Campbell, D.T., and Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally. [This classic chapter has been re-published as a separate paperback book several times over the years, and is often temporarily out of print.]

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

Cortina, J.M., & Nouri, H. (2000). *Effect size for ANCOVA designs*. Thousand Oaks, CA: Sage.

Elashoff, J.D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6* (3), 383-401.

Hedges, L.V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics, 32* (2), 151-179.

Huck, S.W., & McLean, R.A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin, 82* (4), 511-518.

Johnson, P.O., & Fay, L.C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika, 15* (4), 349-367.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs, 1*, 57-93.

Kendall, M.G., & Gibbons, J.D. (1990). *Rank correlation methods* (5th ed.). London: Arnold, 1990.

Kirk, R.E. (1968). *Experimental design: Procedures for the behavioral sciences*. Monterey, CA: Brooks/Cole.

Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin, 85*, 410-416.

Locascio, J.J., & Cordray, D.S. (1983). A re-analysis of "Lord's Paradox". *Educational and Psychological Measurement, 43*, 115-126.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68* (5), 304-305.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88* (2), 307-321.

Rudner, L., Glass, G.V, Evartt, D.L., & Emery, P.J. (2002). *A user's guide to the meta-analysis of research studies*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Knapp and Schafer, Gain Score t and ANCOVA F

Schafer, W.D. (1981). Letter to the Editor. *The American Statistician, 35* (3), 179.

Schochet, P.V. (2008). Statistical power for random assignment evaluations of educational programs. *Journal of Educational and Behavioral Statistics, 33* (1), 62-87.

Appendix

The lemma and the theorem that follow are concerned with the F statistic in the analysis of covariance for two groups and one covariate. The lemma shows that the covariance F is a direct function of the ratio of the variance about the regression line for the total sample to the variance about the within-group regression line. The theorem demonstrates the mathematical relationship between gain score F and covariance F.

Lemma: Let n_E and n_C be the number of individuals in Group 1 (the experimental group) and Group 2 (the control group), respectively. Let s_{YT}^2 be the variance of the posttest scores for the total group of $n_E + n_C$ ($=n_T$) observations, and let s_{YW}^2 be the within-group variance of the posttest scores. Let r_{XYT} and r_{XYW} be the Pearson product-moment correlation coefficients between the pretest (X) and the posttest (Y) for total group and within-group, respectively. Then, under the usual assumptions for the analysis of covariance,

$$F_c = (n_T - 3)\left(\frac{(n_T - 1)s_{YT}^2\,(1 - r_{XYT}^2)}{(n_T - 2)s_{YW}^2\,(1 - r_{XYW}^2)} - 1\right)$$

Proof:

$F_c = \left(\frac{aBSSY}{df_B}\right)/\left(\frac{aSSWY}{df_W}\right)$, where aBSSY and aWSSY are the adjusted sums of squares for between and within groups, respectively, and df_B and df_W are the corresponding numbers of degrees of freedom.

$= (n_T - 3)\left(\frac{aBSSY}{aWSSY}\right)$, since $df_W/df_B = (n_T-3)/1 = n_T-3$.

$= (n_T - 3)\left(\frac{aTSSY-aWSSY}{aWSSY}\right)$, where aTSSY is the adjusted total sum of squares.

$= (n_T - 3)\left[\left(\frac{aTSSY}{aWSSY}\right) - 1\right]$

$= (n_T - 3)\left(\frac{(n_T-1)s_{YT}^2\,(1-r_{XYT}^2)}{(n_T-2)s_{YW}^2\,(1-r_{XYW}^2)} - 1\right)$

[since $(n_T-1)s_{YT}^2 = $ TSSY and $(1-r_{XYT}^2)$ is the adjustment, and since $(n_T-2)s_{YW}^2 = $ WSSY and $(1-r_{XYW}^2)$ is <u>its</u> adjustment]

The products of the variances and the residual r-squares are recognizable as the variances around the total and within regression lines.

Theorem: Let $F_t = t^2$ be the gain score F and let F_c be the covariance F for the same data. Then, if s_{XT}^2 and s_{XW}^2 are the variances of the pretest scores for total and within-group, respectively, then

$$F_c = F_t\,\frac{(n_T - 3)\left(\dfrac{(n_T - 1)s_{YT}^2\,(1 - r_{XYT}^2)}{(n_T - 2)s_{YW}^2\,(1 - r_{XYW}^2)} - 1\right)}{(n_T - 2)\left(\dfrac{(n_T - 1)(s_{YT}^2 + s_{XT}^2 - 2r_{XYT}\,s_{YT}s_{XT})}{(n_T - 2)(s_{YW}^2 + s_{XW}^2 - 2r_{XYW}\,s_{YW}s_{XW})} - 1\right)}$$

Proof:

$F_t = t^2 = \dfrac{(M_1-M_2)^2}{s_p^2\left(\left(\frac{1}{n1}\right)+\left(\frac{1}{n2}\right)\right)}$, where M_1 and M_2 are the mean gains for the two groups and s_p^2 is the "pooled" variance.

Knapp and Schafer, Gain Score t and ANCOVA F

$= BMS/WMS$, i.e., the mean square between groups divided by the mean square within groups for the gain scores

$$= \left(\frac{BSS}{df_B}\right) \Big/ \left(\frac{WSS}{df_W}\right)$$

$$= \left(\frac{df_W}{df_B}\right) \Big/ \left(\frac{TSS-WSS}{WSS}\right)$$

$= (n_T - 2)(\frac{TSS}{WSS} - 1)$, where TSS and WSS are the total and within sums of squares for the gain scores, respectively.

$= (n_T - 2)\left(\frac{(n_T-1)s_T{}^2}{(n_T-2)s_W{}^2} - 1\right)$, where $s_T{}^2$ and $s_W{}^2$ are the variances of the gain scores for total and within, respectively.

$$= (n_T - 2)\left(\frac{(n_T-1)s_{(Y-X)T}{}^2}{(n_T-2)s_{(Y-X)W}{}^2} - 1\right)$$

$$= (n_T - 2)\left(\frac{(n_T-1)(s_{YT}{}^2 + s_{XT}{}^2 - 2r_{XYT}\,s_{YT}s_{XT})}{(n_T-2)(s_{YW}{}^2 + s_{XW}{}^2 - 2r_{XYW}\,s_{YW}s_{XW})} - 1\right)$$

From the above lemma,

$$F_c = (n_T - 3)\left(\frac{(n_T-1)s_{YT}{}^2\,(1-r_{XYT}{}^2)}{(n_T-2)s_{YW}{}^2\,(1-r_{XYW}{}^2)} - 1\right)$$

$$F_c/F_t = \left(\frac{(n_T-3)\left(\frac{(n_T-1)s_{YT}{}^2\,(1-r_{XYT}{}^2)}{(n_T-2)s_{YW}{}^2\,(1-r_{XYW}{}^2)} - 1\right)}{(n_T-2)\left(\frac{(n_T-1)(s_{YT}{}^2 + s_{XT}{}^2 - 2r_{XYT}\,s_{YT}s_{XT})}{(n_T-2)(s_{YW}{}^2 + s_{XW}{}^2 - 2r_{XYW}\,s_{YW}s_{XW})} - 1\right)}\right)$$

Therefore,

$$F_c = F_t \frac{(n_T-3)\left(\frac{(n_T-1)s_{YT}{}^2\,(1-r_{XYT}{}^2)}{(n_T-2)s_{YW}{}^2\,(1-r_{XYW}{}^2)} - 1\right)}{(n_T-2)\left(\frac{(n_T-1)(s_{YT}{}^2 + s_{XT}{}^2 - 2r_{XYT}\,s_{YT}s_{XT})}{(n_T-2)(s_{YW}{}^2 + s_{XW}{}^2 - 2r_{XYW}\,s_{YW}s_{XW})} - 1\right)}$$

The expression that multiplies F_t is the conversion factor A referred to in the text.

Knapp and Schafer, Gain Score t and ANCOVA F

Citation

Knapp, Thomas R. and Schafer, William D. (2009). From Gain Score t to ANCOVA F (and vice versa) *Practical Assessment, Research & Evaluation*, 14(6). Available online: http://pareonline.net/getvn.asp?v=14&n=6.

Authors

Thomas R. Knapp
Professor Emeritus
The Ohio State University
e-mail: tknapp5 [at] juno.com
http://www.tomswebpage.net

William D. Schafer
Affiliated Professor (Emeritus)
University of Maryland -- College Park
e-mail: wschafer [at] umd.edu
http://www.education.umd.edu/EDMS/fac/Schafer/Bill.html