A SAS® PROGRAM FOR CALCULATING AND DISPLAYING REGRESSION LINES IN TWO-WAY ANCOVA MODELS WITH UNEQUAL SLOPES

Joseph J. Andary, Bristol-Myers Squibb Company

ABSTRACT

When the assumption of parallel slopes is tested in an ANCOVA, the regression slopes (the interactions with the covariate) compared by the SAS® system must be correctly calculated and graphed so that the nature of any heterogeneity in slopes can be investigated. In a one-way ANCOVA, this is straightforward since the least-squares method can be used and the resulting regression lines could be graphed by using PROC GPLOT. However, in a two-way (or higher order) unbalanced design, only the slopes in each cell of the design (three way interaction with covariate) can be displayed by a least-squares approach. Regression slopes for each factor (two-way interaction, e.g., covariate by factor A) adjusting for unequal sample sizes must be calculated and displayed by an alternative approach.

This paper addresses the calculation and display of the regression lines that the SAS® system compares in two-way ANCOVA models with unequal slopes. The procedure used by the SAS® system for calculating adjusted slopes and intercepts is outlined. Finally, a SAS program that calculates and graphs these regression lines is presented.

INTRODUCTION

In an Analysis of Covariance (ANCOVA) situation, the assumption of a common regression slope (parallelism) is crucial. When this assumption is violated, the difference between any two levels of a factor cannot be summarized by a single number. The parallelism assumption is tested by the addition of interaction terms in the ANCOVA model. These interactions must be correctly calculated and displayed so that the nature of any heterogeneity in slopes can be investigated.

The purpose of this paper is to demonstrate the appropriate display of regression lines (covariate by factor interactions) in two-way (or higher-order) ANCOVA models. The procedure presented in this paper uses SAS/GRAPH® software and is appropriate for both balanced and unbalanced designs.

MOTIVATING EXAMPLE

Consider the following example from a clinical trial of 78 patients. Two treatments (factor T) were compared. The study was conducted at three different centers (factor C). Efficacy assessments were performed at baseline (the covariate) and after eight weeks of double-blind treatment (the response variable Y). In this example, there are unequal sample sizes in each cell of the design. The sample sizes in each center are presented in Table 1. A listing of the data used in this example is provided in Appendix 1.

TABLE 1
Sample Sizes in Each Center

Center	Sample
1	34
2	9
3	35

Two-way ANCOVA fitting a common regression slope across centers and treatments (the reduced model) was planned. To test the appropriateness of the reduced model, a two-way ANCOVA fitting different regression slopes (the full model) in each center and treatment group was performed. The sources of variation for both ANCOVA models are displayed in Table 2.

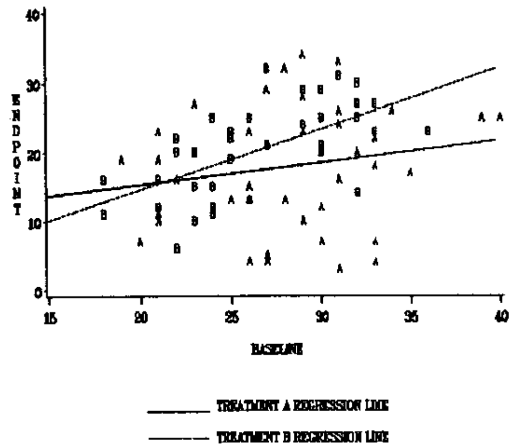
TABLE 2
Analysis of Covariance Models

Full Model	Reduced Model
Center (C)	Center
Treatment (T)	Treatment
C X T	C X T
Baseline (B)	Baseline
B X C	
B X T	
B X C X T	

The three additional interaction terms in the full model involving the covariate baseline, test the assumption of parallelism. If these three terms are not statistically significant, then the parallelism assumption is reasonable and the common slope ANCOVA (reduced model) can be performed.

In a clinical trial, the baseline x treatment (B X T) interaction is of particular interest. This interaction tests the hypothesis that the slope describing the linear relationship between the response variable and the covariate is the same for each treatment. Intuitively, one could test this hypothesis by performing a least-squares regression of the response variable on the covariate separately for each treatment. The resulting regression slopes can then be examined. Figure 1 graphs the B X T interaction for the example data set by using least-squares regression analysis.

FIGURE 1
REGRESSION OF RESPONSE VARIABLE ON THE COVARIATE
LEAST-SQUARES METHOD



The least-squares regression lines graphed in Figure 1 are

Endpoint = 8.78 + 0.33 (baseline) for treatment A, and
Endpoint = -3.15 + 0.89 (baseline) for treatment B.

These regressions indicate that the assumption of equal slopes is not met. For small values (21 or less) of the covariate baseline, treatment B is associated with lower endpoint response; however, treatment A produces lower responses than treatment B when a larger covariate value (at least 22) is observed.

In contrast, the results of a two-way ANCOVA using PROC GLM of the SAS® system are presented in Table 3.

TABLE 3
Results of Two-Way ANCOVA

Source	P-Value
Center (C)	0.18
Treatment (T)	0.75
C X T	0.35
Baseline (B)	0.26
B X C	0.17
B X T	0.94
B X C X T	0.54

In this unbalanced design, the test of B X T interaction using the SAS® system is clearly different than the results observed by ordinary, least-squares regression. The differences in these two methods of testing parallelism are summarized in the next section.

TESTING FOR EQUAL SLOPES

The least-squares regressions shown in Figure 1 do not account for the unequal sample sizes in each study center (see Table 1). Since these regressions were performed by simply pooling across the three study sites, the weight of each of the three centers in the analysis was a function of their sample sizes. Thus, Center 2 received less weight in the least-squares regression than Centers 1 and 3 because of the small sample size in that center.

In contrast, the SAS® system addresses the issue of unequal sample sizes. The p-value for B X T interaction (0.94) shown in Table 3 represents a comparison of the two regression slopes that would be expected if there had been equal sample sizes in each center.

The procedure used by the SAS® system to test for parallelism (B X T interaction) is similar to the method used to calculate adjusted (least-squares) means. The slopes and intercepts from a least-squares regression within each cell (center and treatment) of the design, are averaged to calculate adjusted slopes and adjusted intercepts. The resulting regression lines are used to test the hypothesis of parallelism.

For example, to determine the regression lines that the SAS® system compared in the test for B X T interaction shown in Table 3, a least-squares regression within each center and treatment is performed. The resulting slopes and intercepts are averaged by treatment group. This procedure is demonstrated in Table 4.

TABLE 4
Regressions Within Each Cell

Treatment	Center	Intercept	Slope
A	1	-9.03	0.81
	2	41.90	-0.76
	3	-8.16	1.14
Mean (Treatment A)		8.24	0.40
B	1	28.03	-0.11
	2	22.58	0.09
	3	-8.72	1.07
Mean (Treatment B)		13.96	0.35

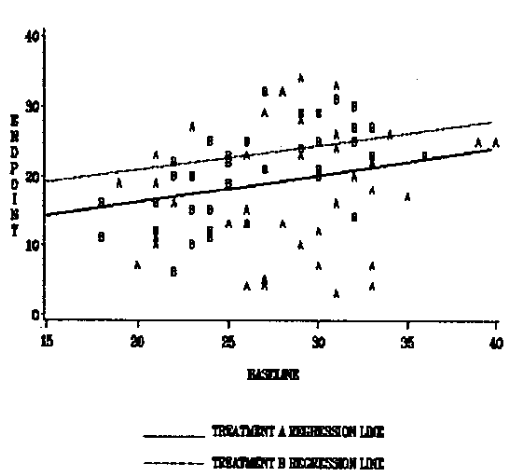
The six slopes (one for each center and treatment) generated by least-squares regression are compared by the SAS® system in the test for B X C X T interaction (p = 0.54 in Table 3). Averaging these slopes and intercepts by treatment gives the three centers equal weight (1/3) in the calculation of the resulting regression lines. Thus, slopes and intercepts obtained by this method provide an estimate of the regression line for each treatment that would be expected if the design had been balanced.

The two regression lines adjusted for the unequal sample sizes in each center are

Endpoint = 8.24 + 0.40 (baseline) for treatment A, and
Endpoint = 13.96 + 0.35 (baseline) for treatment B.

These regression lines are graphed in Figure 2.

FIGURE 2
REGRESSION OF RESPONSE VARIABLE ON THE COVARIATE
ADJUSTED SLOPE METHOD



The slopes of the two regression lines are nearly identical. This result is consistent with the p-value (0.94) for B X T interaction generated by the SAS® system (see Table 3). For all levels of the covariate baseline, treatment A is associated with a lower endpoint response than treatment B.

In summary, two methods of displaying covariate interactions in two-way (or higher order) ANCOVA models have been presented. One method, the least-squares regression approach, does not account for the unequal sample sizes that often occur in clinical trials. The second method, the algorithm used by the SAS® system, addresses the issue of unequal sample sizes. The two methods could produce dramatically different results. It is recommended that statisticians display these interactions by using both methods. If either method produces evidence of interaction, then the nature of the heterogeneity in slopes should be investigated.

COMPUTER ALGORITHM

The SAS® algorithm to calculate and display regression lines adjusted for unequal sample sizes is outlined below. A complete listing of the program can be found in Appendix 2.

1. The SAS® regression procedure RSQUARE is used to perform least-squares regression analysis within each cell of the design (BY CENTER AND TREATMENT). PROC RSQUARE was chosen since it allows the user to output the regression coefficients to a SAS® data set.

2. The SAS® procedure MEANS is used on the data set output by PROC RSQUARE to calculate the mean slopes and mean intercepts for each treatment group.

3. The YHATS macro in Appendix 2 is used to calculate two points on the adjusted regression line for each treatment. The two calculated points correspond to the minimum and maximum covariate values that the user intends to plot. The resulting SAS® data set will have two observations for each treatment group.

4. Finally, PROC GPLOT is used to graph the adjusted regression lines calculated in steps two and three. This is accomplished by a SYMBOL statement with the I = RL option.

REFERENCES

Neter, J., Wasserman, W. (1974). **Applied Linear Statistical Models.** Richard D. Irwin, Inc., Homewood, Illinois.

Huitema, B. (1980). **The Analysis of Covariance and Alternatives.** New York, John Wiley and Sons.

SAS Institute, Inc. 1985. **SAS User's Guide: Basics, Version 5 Edition.** SAS Institute, Inc. Cary, North Carolina.

SAS Institute, Inc. 1985. **SAS User's Guide: Statistics, Version 5 Edition.** SAS Institute, Inc. Cary, North Carolina.

SAS Institute, Inc. 1985. **SAS/GRAPH User's Guide, Version 5 Edition.** SAS Institute, Inc. Cary, North Carolina.

Author Contact:
Joseph J. Andary
Dept. 703
5 Research Parkway
P. O. Box 5100
Wallingford, CT 06492-7660
(203) 284-6891

APPENDIX 1
LISTING OF DATA USED IN THE EXAMPLE

OBSERVATION NUMBER	STUDY CENTER	TREATMENT GROUP	RESPONSE VARIABLE	COVARIATE
1	1	A	26	34
2	1	A	3	31
3	1	A	20	32
4	1	A	34	29
5	1	A	17	35
6	1	A	15	26
7	1	A	7	33
8	1	A	13	25
9	1	A	4	26
10	1	A	5	27
11	1	A	10	29
12	1	A	29	27
13	1	A	18	33
14	1	A	7	30
15	1	A	12	30
16	1	A	22	33
17	1	A	25	39
18	1	A	25	40
19	1	A	16	31
20	1	B	25	26
21	1	B	23	33
22	1	B	30	32
23	1	B	25	32
24	1	B	27	33
25	1	B	20	30
26	1	B	14	32
27	1	B	25	24
28	1	B	25	30
29	1	B	23	36
30	1	B	29	29
31	1	B	31	31
32	1	B	27	32
33	1	B	24	29
34	1	B	21	30
35	2	A	4	33
36	2	A	23	26
37	2	A	28	29
38	2	A	26	31
39	2	A	24	31
40	2	A	13	26
41	2	B	21	30
42	2	B	22	22
43	2	B	32	27
44	3	A	11	21
45	3	A	33	31
46	3	A	23	29
47	3	A	16	22
48	3	A	27	23
49	3	A	20	23
50	3	A	7	20
51	3	A	20	23
52	3	A	23	21
53	3	A	10	21
54	3	A	4	27
55	3	A	32	28
56	3	A	33	31
57	3	A	19	21
58	3	A	19	19
59	3	A	13	28
60	3	B	20	23
61	3	B	29	30
62	3	B	6	22
63	3	B	10	23
64	3	B	23	25
65	3	B	16	18
66	3	B	16	21
67	3	B	11	18
68	3	B	22	25
69	3	B	12	24
70	3	B	19	25
71	3	B	15	23
72	3	B	15	24
73	3	B	12	21
74	3	B	13	26
75	3	B	20	22
76	3	B	21	27
77	3	B	11	24
78	3	B	16	21

APPENDIX 2

```
***********************************************************
* SAS PROGRAM TO CALCULATE "ADJUSTED" SLOPES AND          *
* INTERCEPTS USED TO TEST FOR PARALLEL REGRESSION         *
* LINES                                                   *
***********************************************************

% MACRO REGR(DSET,Y,X);

***********************************************************
* PERFORMING A LEAST-SQUARES REGRESSION WITHIN EACH       *
* CELL (CENTER AND TREATMENT GROUP) OF THE DESIGN         *
* USING PROC RSQUARE. OUTPUT THE RESULTS TO A SAS         *
* DATASET CALLED COEFFS.                                  *
***********************************************************

PROC SORT DATA=&DSET ; BY STDY TREAT;

PROC RSQUARE DATA=&DSET  OUTEST=COEFFS INCLUDE=1 B; BY
STDY TREAT;
MODEL &Y = &X ;
TITLE '1991 SUGI MEETINGS';
TITLE3 "REGRESSION OF &Y ON &X.";
```

```
TITLE4 'FOR EACH CENTER AND TREATMENT GROUP';

PROC PRINT DATA=COEFFS;
PROC SORT DATA=COEFFS; BY TREAT STDY;

*********************************************************
*  USING PROC MEANS TO CALCULATE THE AVERAGE SLOPE      *
*  AND THE AVERAGE INTERCEPT FOR EACH TREATMENT         *
*  GROUP. AVERAGING SLOPES AND INTERCEPTS ADJUSTS FOR   *
*  UNEQUAL SAMPLE SIZES IN EACH STUDY CENTER.  OUTPUT   *
*  THE RESULTS TO A SAS DATASET CALLED RXREGS.          *
*********************************************************

PROC MEANS NOPRINT MEAN; BY TREAT;
   VAR INTERCEP &X.;
   OUTPUT OUT=RXREGS MEAN=B0 B1;

PROC PRINT DATA=RXREGS;
TITLE '1991 SUGI MEETINGS';
TITLE3 'ADJUSTED SLOPES AND INTERCEPTS FOR EACH TREAT-
MENT GROUP';

%MEND REGR;

%MACRO YHATS(MIN,MAX,Y,X);

*********************************************************
*  THE MACRO YHATS CALCULATES TWO POINTS ON THE         *
*  ADJUSTED REGRESSION LINE FOR EACH TREATMENT AND      *
*  OUTPUTS THESE TWO CALCULATED POINTS TO A SAS         *
*  DATASET.  THE TWO CHOSEN POINTS TO BE CALCULATED     *
*  ARE THE MINIMUM AND MAXIMUM COVARIATE (X) VALUES     *
*  THAT THE USER INTENDS TO PLOT.                       *
*********************************************************

DATA DUMMY; SET RXREGS;
  YMIN=B0 + (B1 * &MIN ); XMIN=&MIN.;
  YMAX=B0 + (B1 * &MAX ); XMAX=&MAX.;
  OUTPUT; OUTPUT;

PROC PRINT DATA=DUMMY;
TITLE 'MINIMUM AND MAXIMUM YHATS FOR EACH TREATMENT
GROUP';
PROC SORT DATA=DUMMY; BY TREAT;

DATA MINS MAXS; SET DUMMY; BY TREAT;
  IF FIRST.TREAT THEN OUTPUT MINS; ELSE OUTPUT MAXS;

*********************************************************
*  RENAMING THE CALCULATED Y VALUES AND THE X VALUES    *
*  SUPPLIED BY THE USER AS Y AND X, RESPECTIVELY.       *
*  THIS RENAMING IS NEEDED SO THAT THE CALCULATED       *
*  "FAKE" OBSERVATIONS CAN BE ADDED TO THE ORIGINAL     *
*  DATASET.                                             *
*********************************************************

DATA MINS; SET MINS;
  KEEP TREAT YMIN XMIN;
DATA MINS; SET MINS;
RENAME YMIN=&Y. XMIN=&X.;

DATA MAXS; SET MAXS;
  KEEP TREAT YMAX XMAX;
DATA MAXS; SET MAXS;
RENAME YMAX=&Y. XMAX=&X.;

*********************************************************
*  ASSIGNING DUMMY TREATMENT GROUPS AND SUBJECT         *
*  NUMBERS TO THE CALCULATED OBSERVATIONS SO THAT       *
*  THEY CAN BE ADDED TO THE ORIGINAL DATASET.           *
*********************************************************

DATA FAKE; SET MINS MAXS;
  IF TREAT='A' THEN RX=3;
  IF TREAT='B' THEN RX=4;
PROC SORT; BY TREAT &X ;

DATA FAKE; SET FAKE;
  STDY=9999; SUBJ=_N_;

PROC PRINT;
TITLE '1991 SUGI MEETINGS';
TITLE3 'TWO CALCULATED VALUES FOR EACH TREATMENT GROUP
(N=4)';

%MEND YHATS;

%MACRO RXPLT(DSET,Y,X,MIN,MAX,INT);

DATA PLOT1; SET &DSET ; IF TREAT='A' THEN RX=1; IF
TREAT='B' THEN RX=2;
PROC SORT DATA=PLOT1; BY STDY RX;

*********************************************************
*  CREATING A DATASET CONTAINING THE ORIGINAL           *
*  OBSERVATIONS AS WELL AS THE FOUR COMPUTED            *
*  OBSERVATIONS FROM THE YHATS MACRO.                   *
*********************************************************

DATA PLOTALL; SET PLOT1 FAKE;

*********************************************************
*  PLOTTING THE ADJUSTED REGRESSION LINES AND THE       *
*  VALUES FOR Y AND X FROM THE ORIGINAL DATASET.        *
*********************************************************

PROC GPLOT DATA=PLOTALL;
AXIS3
LABEL = (F=COMPLEX)
VALUE = (F=COMPLEX)
ORDER = 0 TO 40 BY 10;

AXIS4
LABEL = (F=COMPLEX)
VALUE = (F=COMPLEX)
ORDER = &MIN TO &MAX BY &INT.;

LABEL &Y.='ENDPOINT';
LABEL &X.='BASELINE';
  PLOT &Y.*&X.=RX / HMINOR=0 VMINOR=0 NOLEGEND
                    HAXIS=AXIS4 VAXIS=AXIS3;
SYMBOL1 C=BLACK V=A H=0.7;
SYMBOL2 C=BLACK V=B H=0.7;
SYMBOL3 C=BLACK V=NONE W=5 L=1 I=RL;
SYMBOL4 C=BLACK V=NONE W=5 L=2 I=RL;

TITLE1 H=1.2 F=COMPLEX 'FIGURE 2';
TITLE2 H=1.2 F=COMPLEX 'REGRESSION OF Y ON X';
TITLE3 H=1.2 F=COMPLEX 'ADJUSTED SLOPE METHOD';

FOOTNOTE2 H=1.2 F=COMPLEX '_______  DRUG A REGRESSION
LINE';
FOOTNOTE4 H=1.2 F=COMPLEX '-------  DRUG B REGRESSION
LINE';

%MEND RXPLT;

*********************************************************
*  TO EXECUTE THE PROGRAM, CALL THE MACROS REGR,        *
*  YHATS, AND RXPLT IN THAT ORDER.  THE APPROPRIATE     *
*  MACRO CALLS ARE LISTED BELOW.                        *
*********************************************************

%REGR(EXAMPLE,Y,X)
%YHATS(15,40,Y,X)
%RXPLT(EXAMPLE,Y,X,15,40,5)
```