

Assessing the effect of a treatment when subjects are growing at different rates

Charles J. Kowalski^a, Emet D. Schneiderman*^b, Stephen M. Willis^b

^a*Department of Biologic and Materials Sciences, Dental School, and The Center for Statistical Consultation and Research, The University of Michigan, Ann Arbor, MI 48109, USA*

^b*Department of Oral and Maxillofacial Surgery and Pharmacology, Baylor College of Dentistry, PO Box 660677, Dallas, TX 75266-0677, USA*

Received 22 December 1993; accepted 20 January 1994

Abstract

The analysis of covariance is often used in the context of premeasure/postmeasure designs to compare treatment and control groups in both randomized [1] and nonrandomized [2] studies. The intent is to adjust the difference between the changes in the 2 groups for any difference which might exist at baseline, i.e., for any difference between the premeasures in the 2 groups. An important assumption underlying the use of the analysis of covariance is that the slopes of the lines for the regression of the postmeasure on the premeasure in the 2 groups are equal. In this paper we describe a program which can be used to test the hypothesis of equal slopes; and performs an alternative analysis which does not depend on this assumption. This is done in the context of comparing treatment and control groups with respect to a measurement subject to natural maturation as in [3]. Equal slopes in this context means equal growth rates; unequal slopes implies that the 2 groups are growing at different rates. The method, known as the Johnson-Neyman procedure [4] is, however, more general than this, and can be used in any two-sample comparison where an alternative to the usual analysis of covariance is deemed appropriate. The procedure identifies a 'region of significance' which is especially useful in practice. This region consists of a set of values of the premeasure for which the treatment and the control groups are significantly different with respect to the postmeasure.

Keywords: Growth; Analysis of covariance; Nonparallel regressions; Region of significance; PC program

1. Introduction

We consider the situation in which treatment (T) and control (C) groups are measured at comparable times before and after a treatment (TX) is administered, resulting in a data set like the one shown in Table 1.

We let \bar{Z}_i denote the mean premeasure in group i ($i = 1, 2$), \bar{X}_i the corresponding mean of the measurement post-TX. $\bar{D}_i = \bar{X}_i - \bar{Z}_i$ is then the mean change. s_{iZ} and s_{iX} are the within-group standard deviations, and $r_i(Z, X)$ the correlations. \bar{Z} denotes the overall (weighted) mean of the measurement pre-TX.

This data set is equivalent to the one used in [3] to illustrate how the effect of the TX could be

* Corresponding author.

Table 1

Summary statistics for the weights (g) of rats in T (thiouracil + water) and C groups (water only)

	Pre (Z)	Post (X)	Correlation
T ($n_1 = 10$)	Mean = 54.7 S.D. = 4.692	Mean = 76.3 S.D. = 7.917	$r = 0.751$
C ($n_2 = 10$)	Mean = 54.0 S.D. = 5.437	Mean = 78.5 S.D. = 9.641	$r = 0.945$

assessed using either simple (unadjusted) mean differences of the form

$$\bar{D}_1 - \bar{D}_2 \quad (1)$$

or adjusted differences

$$\bar{D}_1' - \bar{D}_2' \quad (2)$$

where

$$\bar{D}_i' = \bar{D}_i - \hat{\beta}(\bar{Z}_i - \bar{Z}) \quad (3)$$

and $\hat{\beta}$ is obtained using the analysis of covariance (ANCOVA). The numbers in Table 1 refer to the weights (g) of 20 rats. The TX involved the use of the drug thiouracil, which was added to the drinking water of the rats in this group. The intent was to see if this drug inhibited normal growth.

The ANCOVA model as described in [1, p 194] and applied in [3] is

$$X_{ij} = \mu_i + \beta (Z_{ij} - \bar{Z}) + e_{ij} \quad (4)$$

where μ_i is the mean level of response in group i when $Z_{ij} = \bar{Z}$ and the e_{ij} are normally distributed errors (residuals) with mean 0 and variance σ^2 . This model assumes that β is the same in both groups. A more general model is

$$X_{ij} = \mu_i + \beta_i (Z_{ij} - \bar{Z}) + e_{ij} \quad (5)$$

which has the same structure as (4), but allows β to be different in the 2 groups.

In this paper we describe, illustrate and make available a PC program which can be used to (a) fit the model (5), (b) test the hypothesis $H: \beta_1 = \beta_2$, and (c) determine a so-called 'region of

significance,' i.e., values of Z for which the T and C groups differ significantly.

2. Fitting the model

The least squares estimators of the parameters in (5) are [1, p 203]

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(Z_{ij} - \bar{Z}_i)}{\sum_{j=1}^{n_j} (Z_{ij} - \bar{Z}_i)^2} \quad (6)$$

and

$$\hat{\mu}_i = \bar{X}_i - \hat{\beta}_i(\bar{Z}_i - \bar{Z}) \quad (7)$$

In terms of the summary statistics in the table,

$$\hat{\beta}_i = \frac{r_i(Z, X)s_{iX}}{s_{iZ}}$$

so

$$\hat{\beta}_1 = 0.751 * 7.917 / 4.692 = 1.2672$$

and

$$\hat{\beta}_2 = 0.945 * 9.641 / 5.437 = 1.6757.$$

Similarly, since $\bar{Z} = 54.35$,

$$\begin{aligned} \hat{\mu}_1 &= \bar{X}_1 - \hat{\beta}_1(\bar{Z}_1 - \bar{Z}) \\ &= 76.3 - 1.2672(54.7 - 54.35) = 75.8565 \end{aligned}$$

and

$$\hat{\mu}_2 = \bar{X}_2 - \hat{\beta}_2(\bar{Z}_2 - \bar{Z})$$

$$= 78.5 - 1.6757(54 - 54.35) = 79.0865.$$

An unbiased estimator of σ^2 is given by [Ref. 1, p 204]

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^2 \sum_{j=1}^{n_i} [X_{ij} - \hat{\mu}_i - \hat{\beta}_i(Z_{ij} - \bar{Z}_i)]^2}{n_1 + n_2 - 4}$$

$$= \frac{\Sigma \Sigma (X_{ij} - \bar{X}_i)^2 - \Sigma \hat{\beta}_i^2 \Sigma (Z_{ij} - \bar{Z}_i)^2}{n_1 + n_2 - 4} \quad (8)$$

In terms of the summary statistics,

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_{1X}^2 + (n_2 - 1)s_{2X}^2 - \hat{\beta}_1^2(n_1 - 1)s_{1Z}^2 - \hat{\beta}_2^2(n_2 - 1)s_{2Z}^2}{n_1 + n_2 - 4}$$

so in our example

$$\hat{\sigma}^2 = \frac{9 \cdot 7.917^2 + 9 \cdot 9.641^2 - 1.2672^2 \cdot 9 \cdot 4.692^2 - 1.6757^2 \cdot 9 \cdot 5.437^2}{16} = 20.9650$$

3. Testing $H_0: \beta_1 = \beta_2$

We can test $H_0: \beta_1 = \beta_2$ by comparing

$$F = \frac{\sum_{i=1}^2 (\hat{\beta}_i - \hat{\beta})^2 \sum_{j=1}^{n_j} (Z_{ij} - \bar{Z}_i)^2}{\hat{\sigma}^2} \quad (9)$$

to $F_{1-\alpha}(1, n_1 + n_2 - 4)$, the $(1-\alpha) \times 100$ th percentile of the F -distribution with 1 and $n_1 + n_2 - 4$ degrees of freedom, where

$$\hat{\beta} = \frac{(n_1 - 1)r_1(Z, X)s_{1Z}s_{1X} + (n_2 - 1)r_2(Z, X)s_{2Z}s_{2X}}{(n_1 - 1)s_{1Z}^2 + (n_2 - 1)s_{2Z}^2} \quad (10)$$

In our example,

$$\hat{\beta} = \frac{9 \cdot 0.751 \cdot 4.692 \cdot 7.917 + 9 \cdot 0.945 \cdot 5.437 \cdot 9.641}{9 \cdot 4.692^2 + 9 \cdot 5.437^2}$$

$$= 1.5013$$

so

$$F = \frac{(\hat{\beta}_1 - \hat{\beta})^2(n_1 - 1)s_{1Z}^2 + (\hat{\beta}_2 - \hat{\beta})^2(n_2 - 1)s_{2Z}^2}{\hat{\sigma}^2}$$

and

$$F = \frac{(1.267 - 1.501)^2 \cdot 9 \cdot 4.692^2 + (1.676 - 1.501)^2 \cdot 9 \cdot 5.437^2}{20.954} = 0.9039$$

Since $F_{0.95}(1, 16) = 4.49$, there is no reason to suspect the slopes are not equal in this example.

Our program prints the P -value corresponding to the observed value of $F = 0.9039$; in this case, $P = 0.3559$. Our program (described below) also plots the separately fitted regressions of X on Z for the 2 groups, as shown in Fig. 1. It is seen that, as indicated by the test, the slopes of these lines are approximately equal, i.e., the lines are reasonably parallel.

4. The region of significance

At a given value of Z , the difference between the predicted X -values in the 2 groups is

$$D(Z) = [\bar{X}_1 + \hat{\beta}_1(Z - \bar{Z}_1)] - [\bar{X}_2 + \hat{\beta}_2(Z - \bar{Z}_2)] \quad (11)$$

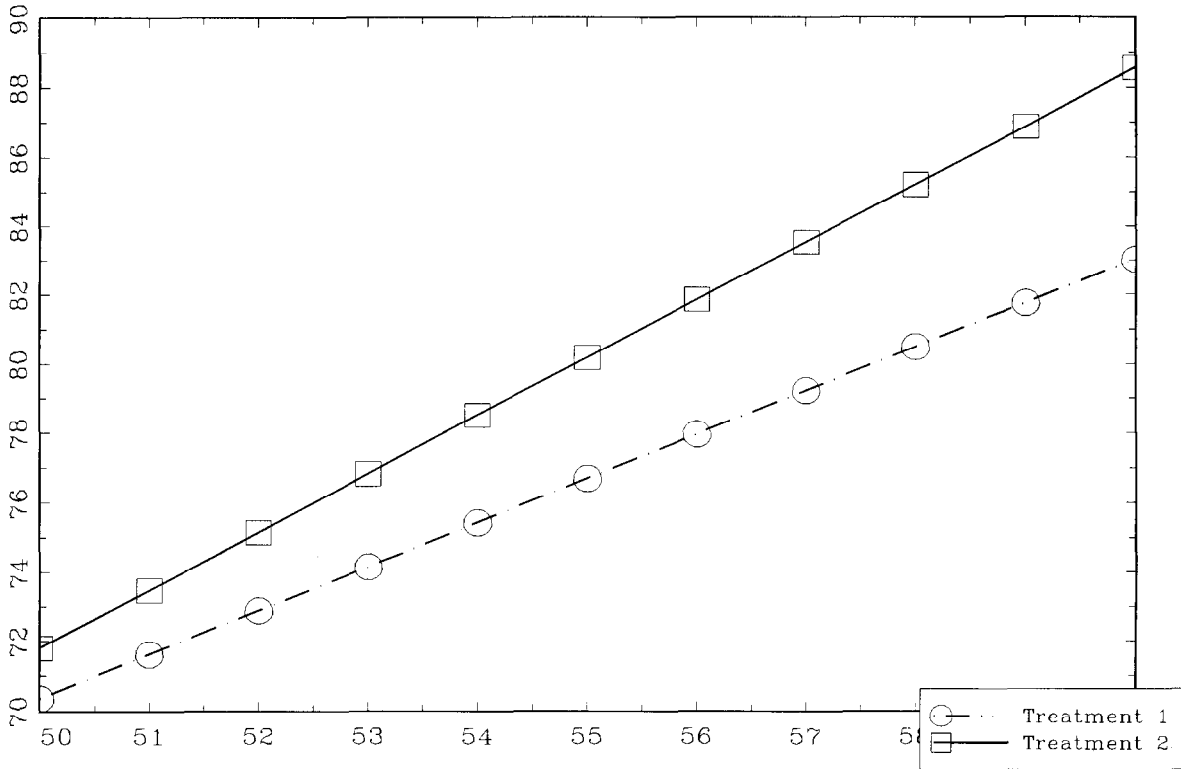


Fig. 1. Separately fitted regressions of postmeasure on initial value for the data in Table 1 (thiouracil data).

and the estimated standard error of $D(Z)$, say $SE[D(Z)]$, is [1, p 205]

$SE[D(Z)]$

$$= \hat{\sigma} \left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{(Z - \bar{Z}_1)^2}{(n_1 - 1)s_{1Z}^2} + \frac{(Z - \bar{Z}_2)^2}{(n_2 - 1)s_{2Z}^2} \right]^{1/2} \quad (12)$$

The equations

$$D(Z) \pm \sqrt{2F_{1-\alpha}(2, n_1 + n_2 - 4)} SE[D(Z)] \quad (13)$$

then define a pair of hyperbolas bounding $D(Z)$.

In our example,

$$D(Z) = 18.999 - 0.409Z$$

$$\sqrt{2F(2,16)} = \sqrt{7.26} = 2.694$$

and

$SE[D(Z)]$

$$= \text{sqrt}(20.954) \left[\frac{1}{10} + \frac{1}{10} + \frac{(Z - 54.7)^2}{9 \cdot 4.692^2} + \frac{(Z - 54)^2}{9 \cdot 5.437^2} \right]^{1/2}$$

We then have, e.g., at $Z = 50$, $D(Z) = -1.451$, $SE[D(Z)] = 2.791$, and the 95% confidence interval at this point is $(-8.978, 6.072)$. Since this interval contains zero, there is no significant difference (at $\alpha = 0.05$) between the predicted values of X in the 2 groups at $Z = 50$. The results for several values of Z are shown below and plotted in Fig. 2.

Z	D(Z)	SE[D](Z)	Confidence interval
50	-1.45	2.791	-8.978-6.072
51	-1.86	2.520	-8.655-4.932
52	-2.27	2.298	-8.465-3.924
53	-2.68	2.140	-8.477-3.089
54	-3.09	2.060	-8.641-2.467
55	-3.50	2.069	-9.074-2.083
56	-3.90	2.165	-9.741-1.933
57	-4.31	2.337	-10.61-1.988
58	-4.72	2.570	-11.65-2.208
59	-5.13	2.849	-12.81-2.552
60	-5.54	3.163	-14.06-2.988

model (5) and produce the above 'region of significance' when the preliminary test of $H_0: \beta_1 = \beta_2$ is accepted, since the ordinary ANCOVA model (4) should yield tighter confidence intervals in this situation. We should note, however, that there is nothing 'wrong' in doing this. Some investigators will prefer not to make the assumption that $\beta_1 = \beta_2$ (whether the preliminary test is significant or not) in which case the above is the procedure of choice. In the next section we describe the structure of our program and illustrate its use with another example; one in which differences between the groups exist, which should reveal the considerable descriptive power of the method.

It is seen that for $Z = 50(1) 60$, since zero is always included in the confidence bands, there is no significant difference between the groups at any of these points.

We might remark that it is not usual to fit the

5. The program; another example

The program is designed to accept either data in the form of summary statistics or individual measurements. The summary statistics may be in

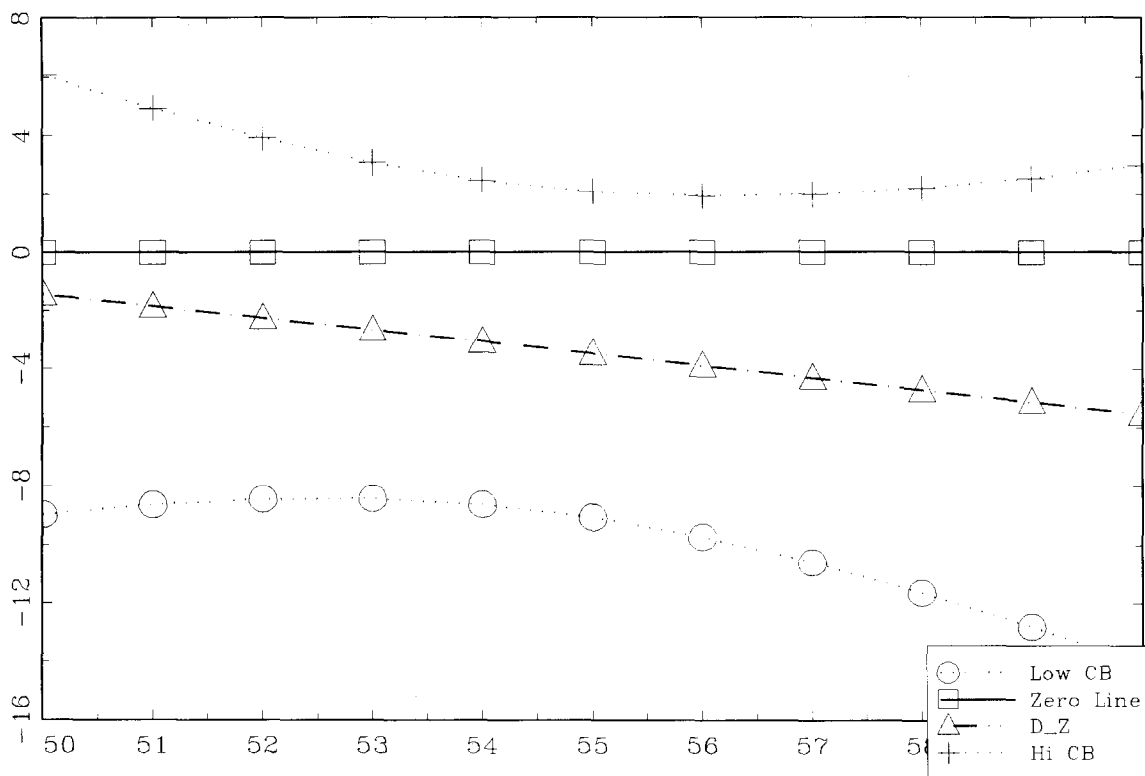


Fig. 2. Region of significance for the data in Table I (thiouracil data).

Z, X form as in the above illustration; or they may be in Z, D form, where $D = X - Z$. Examples of both these types are given in [1,3]. The program is invoked by the command *gsruni txdgr* (Different Growth Rates). The user is first asked if the data are in the form of summary statistics, or if the (raw) measurements are contained in a file. If summary statistics is chosen, the user is asked if they are in Z, X or Z, D form, and the program prompts the user for the appropriate values. If, e.g., Z, X is chosen, the user will have to supply the values of the following quantities:

- NUMBER IN TX GROUP
- PRE-TX MEAN IN TX GROUP
- PRE-TX S.D. IN TX GROUP
- POST-TX MEAN IN TX GROUP
- POST-TX S.D. IN TX GROUP
- PRE-POST CORRELATION IN TX GROUP
- NUMBER IN CONTROL GROUP
- PRE-TX MEAN IN CONTROL GROUP
- PRE-TX S.D. IN CONTROL GROUP
- POST-TX MEAN IN CONTROL GROUP
- POST-TX S.D. IN CONTROL GROUP
- PRE-POST CORRELATION IN CONTROL GROUP

If the user indicates that the data are in a file, he/she is prompted for the name and location of an ASCII (or GAUSS) file containing a group indicator variable in column 1 (1 = TX, 2 = C) and the values of the pre- and post-TX measurements in columns 2 and 3. An example is given in [3].

We use the example given in [1, p 187]. There the data are presented in the form of sample sizes, means, and standard deviations for Z, X and D in the 2 groups. This can be converted to a form

accepted by our program by using the relationship, for any random variables U and V,

$$r(U, V) = \frac{s_U^2 + s_V^2 - s_{U-V}^2}{2s_U s_V} \tag{14}$$

The data may now be presented as shown in Table 2.

The output appears in the form

```
MU1 = 0.5320
MU2 = 0.4032
BETA1 = 0.8593
BETA2 = 0.4008
BETA = 0.6677
SIGMA^2 = 0.0400
F = 10.1367 (P = 0.0018)
```

The plot of the separately fitted regressions of X on Z in the 2 groups is shown in Fig. 3, and the plot of D(Z) and its confidence bands are given in Fig. 4. The user has control over the range of Z-values for which the regressions and D(Z) are plotted. After the data are read in, the user is prompted for the initial and final values of Z and the increment to be used. In this example, following [1, p 207], we specified $Z = 0(0.2)1.4$, i.e., start at Z = 0 and increment by steps of 0.2 up to Z = 1.4.

In this case, it seems clear from Fig. 3, as is reflected in the F-test, that the regressions of X on Z are not parallel. From Fig. 4, one can note that differences between the groups exist, and it is seen that these differences are confined to values of Z > 0.5. To complete the interpretation of the output, we must realize that the response measure in this example is an index of gingivitis for which

Table 2
Summary statistics for the gingival index in a treated and control group

	Pre (Z)	Post (X)	Correlation
T ($n_1 = 74$)	Mean = 0.6065 S.D. = 0.2514	Mean = 0.5514 S.D. = 0.3054	$r = 0.7074$
C ($n_2 = 64$)	Mean = 0.5578 S.D. = 0.2293	Mean = 0.3927 S.D. = 0.1988	$r = 0.4623$

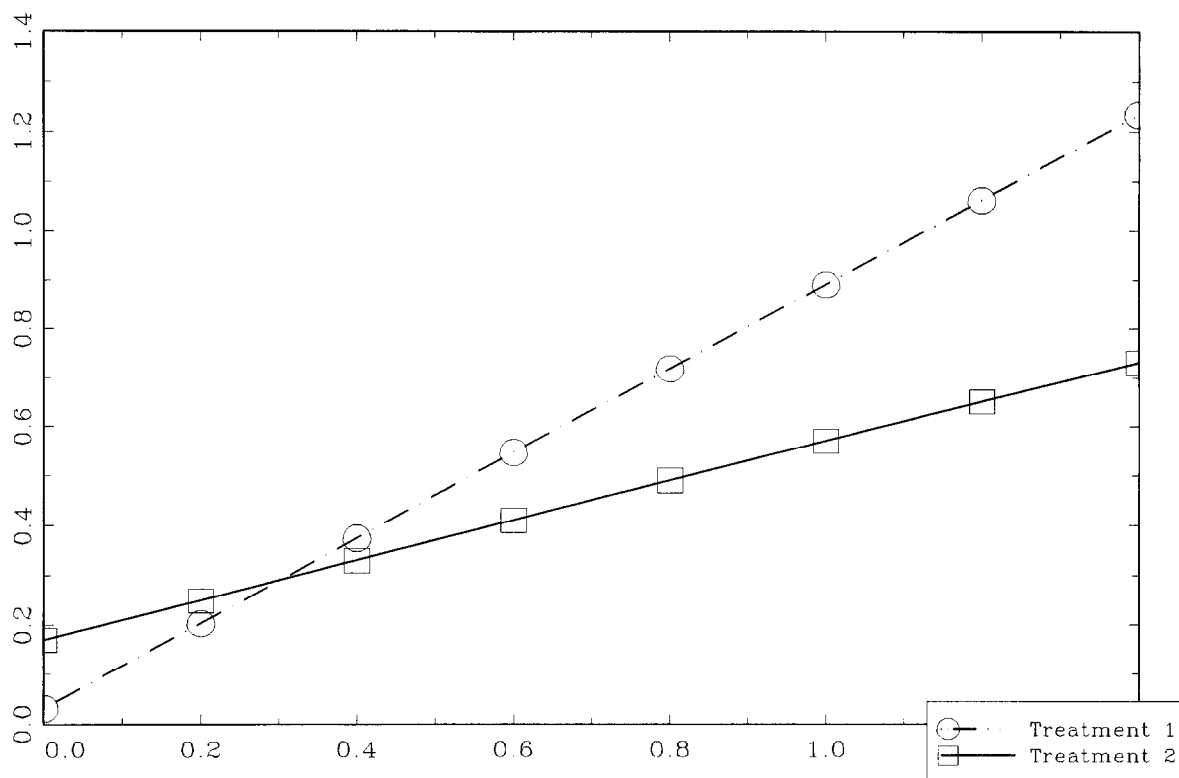


Fig. 3. Separately fitted regressions of postmeasure on initial value for the data in Table 2 (gingival index data).

large values are 'bad', and that the 'control' group is really a competing TX group, call it TX2. The plot shows that $D(Z) = \hat{X}_1 - \hat{X}_2$, where \hat{X}_i is the expected postmeasure in group i (cf Equation (11)), is significantly greater than zero for $Z > 0.5$ which means that TX2 is superior to TX1 for values of $Z > 0.5$. For values of $Z < 0.5$, TX2 is not significantly better (it might even be worse).

6. Discussion

We have considered the comparison of treatment and control (or 2 treatment) groups for pre- and post-TX designs where subjects in the groups may be growing at different rates, so that adjustment procedures based on ANCOVA (4) are not applicable. The more general model (5) was applied to produce 'regions of significance' which show clearly the values of Z for which differences exist. Two examples were considered. In the first,

which was used primarily to illustrate the computations, no significant difference between the slopes of the individual regressions was found, but the point was made that use of the more general model could still be considered useful in situations where the investigator is not willing to assume that the slopes are, in fact, equal [4]. In the second example, the region of significance was non-empty, and the plot of $D(Z)$ vs. Z showed clearly those values of Z for which TX1 was superior. As explained in [1, p 206], this reflected on the design of the study; and has implications for the design of future studies of the same general type. With respect to the study under consideration, it was noted that several patients with very low — even zero — premeasures were included in the study. These individuals had no gingivitis to treat, and one would expect that their change scores would be close to zero. Only patients with at least moderate amounts of gingivitis could change, and

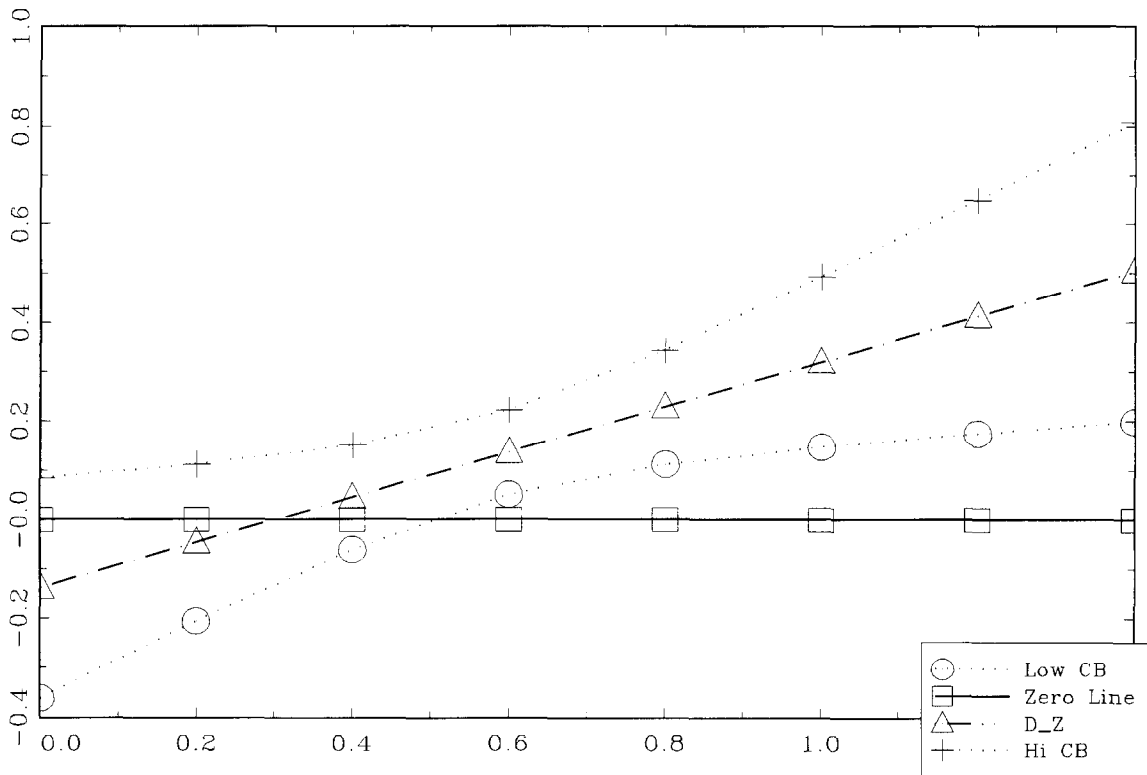


Fig. 4. Region of significance for the data in Table 2 (gingival index data).

it was only for these that the TX effect was apparent. This is an illustration in a particular case of the importance of establishing inclusion/exclusion criteria which will best serve the aims of the study. Liberal inclusion criteria will broaden generalizability, but may compromise a study's ability to detect an important therapeutic effect [1, p 207].

Acknowledgement

Supported by DE 08730 from the National Institute for Dental Research.

Appendix

A set of PC programs, including this and related procedures, can be obtained on 5.25" or 3.5" diskettes (please request type) by sending \$25 to defray the cost of handling and licensing fees.

These programs require a 80386- or 80486-based personal computer (PC) running the MS-DOS operating system (version 5.0 or higher is recommended, although versions as low as 3.3 will suffice). 80386 computers must also be equipped with a 80387 math coprocessor. At least 4 MB of memory is required, and must be available to GAUSS386i, i.e., not in use by memory resident programs such as Windows. EGA or VGA graphic capabilities are required to display the color graphics; VGA or SVGA is suggested to display optimally the graphic results. Runtime modules are supplied with the programs so that no additional software (i.e., compiler or interpreter) is required to run these programs. One can create and edit ASCII data sets for use by these programs using the full screen editor supplied with MS-DOS version 5.0. The programs are written and compiled using GAUSS386i, version 3.0, require no

additional installation or modification, and are run with a single command. When requesting the programs, address inquiries to the corresponding author and make checks payable to Baylor College of Dentistry.

References

- [1] Fleiss JL: *The Design and Analysis of Clinical Experiments*, Wiley, New York, 1986.
- [2] Anderson S, Auquier A, Hauck WW, Oakes D, Vandaele W and Weisberg HI: *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*, Wiley, New York, 1980.
- [3] Kowalski CJ, Schneiderman ED and Willis SM: PC program for assessing the effect of a treatment when subjects are growing: the randomized parallel groups design, *Int J Biomed Comput*, 37 (1994) 83–91.
- [4] Potthoff RF: Johnson-Neyman technique. In *Encyclopedia of Statistical Sciences* (Eds: S Kotz and NL Johnson), Vol. 4, Wiley, New York, 1983.