

# BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. 61

DATE: March 14, 2003

SUBJECT: ANOVA: Power of Linear Contrasts — Interpreting the Non-centrality Parameter

Determining the power of contrasts requires some knowledge of the contrast's non-centrality parameter. For a linear contrast, this parameter measures the departure of the slope of the line that best fits the observed treatment means from the null hypothesis of no trend at all (a horizontal line; see Figure 1). Using a balanced, completely randomized design for illustration, this pamphlet will discuss how we can interpret this parameter for a meaningful power analysis of a linear contrast.

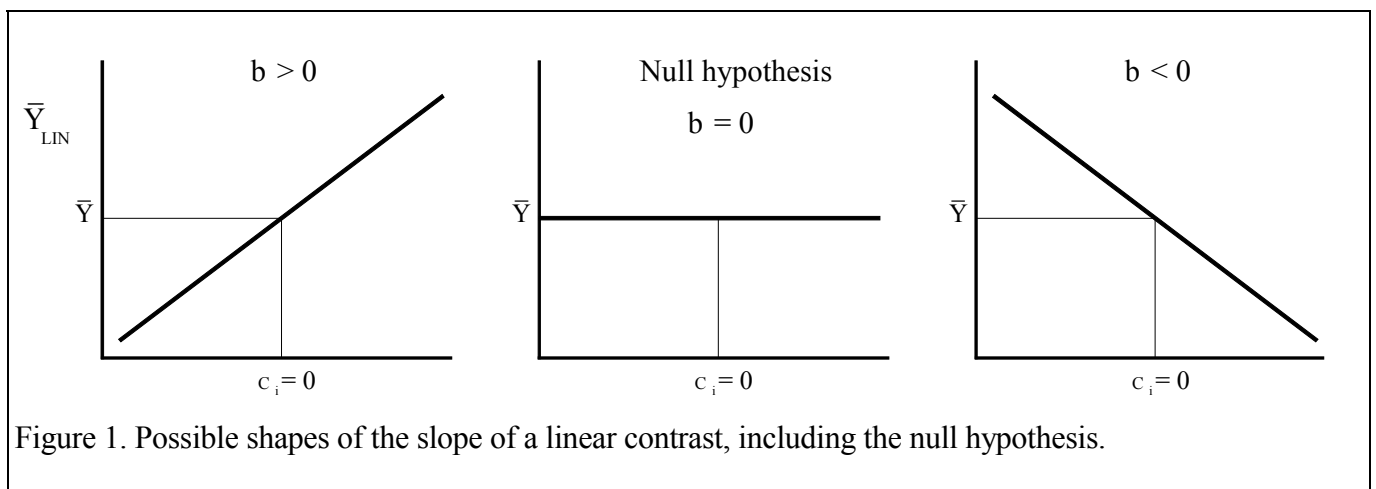


Figure 1. Possible shapes of the slope of a linear contrast, including the null hypothesis.

The linear contrast fits a line between the  $a$  observed treatment means,  $\bar{Y}_i$ , and each treatment's quantitative value (the independent variable),  $X_i$ . Each mean has a sample size of  $n$  experimental units, and  $i$  indexes the treatment levels,  $i = 1, 2, \dots, a$ . Contrast analysis converts the treatment levels,  $X_i$ , into contrast coefficients,  $c_i$ , which are then used in the calculations. The treatment level means predicted by the linear contrast model<sup>1</sup> are:

$$\bar{Y}_{LINi} = \bar{Y} + b c_i, \quad (1)$$

where  $\bar{Y}$  is the grand mean of the response variable and  $b$  is the slope of the line. This slope can be written either in terms of the contrast coefficients or in terms of the original independent variable. While this pamphlet will work with the contrast coefficients, it is often easier to interpret the slope,  $b_x$ , scaled for  $X_i$  instead of the slope,  $b$ , scaled for the contrast coefficients. Using the conversion factor,  $m$ , described in Appendix 1, this is easily calculated by:

$$b_x = m b. \quad (2)$$

The non-centrality parameter,  $nc$ , is:<sup>2</sup>

$$nc = (n \times SSM) / \sigma^2 = \left( n b^2 \sum_{i=1}^a c_i^2 \right) / \sigma^2 \quad (3)$$

since  $SSM = b^2 \sum c_i^2$ , and  $\sigma^2$  is the background or error variance. Thus, all else being equal and

<sup>1</sup> This model assumes that the independent variable,  $X_i$ , and sample size,  $n$ , are fixed and known without error. Usually  $\sigma^2$  is unknown and must be estimated by the mean square error from a previous study.

<sup>2</sup> These equations are developed in Appendix 2.

regardless of whether the slope is descending or ascending, the non-centrality parameter changes as the square of the slope changes. The power changes non-linearly with changes in the non-centrality parameter, but does increase with increasing values of the non-centrality parameter.

**Example:** A researcher plans to use a completely randomized design<sup>3</sup> with treatments of 0, 50, 100, 150, and 200 g/ha of herbicide on seedlings with a current growth rate of about 9 cm/yr. A year after herbicide application, a linear decrease in conifer growth is to be tested (assuming no differences in mortality between the treatments).<sup>4</sup>

**Scenario 1:** The researcher decides that a decrease to 1 cm/yr for the highest level (200 g/ha) should be detectable with 80% power. To determine the required sample size we must first translate this specific alternate hypothesis into its associated non-centrality parameter. The contrast coefficients,  $c_i$ , the  $X_i$ -values, and the expected means  $\bar{Y}_{LINi}$  for this alternate hypothesis are:

$i$ :	1	2	3	4	5	
Treatment level, $X_i$ :	0	50	100	150	200	
Linear coefficients, $c_i$ :	-2	-1	0	1	2	$\sum c_i^2 = 10$
Expected $\bar{Y}_{LINi}$ :	9	7	5	3	1	$\bar{Y} = 5$

The expected  $\bar{Y}_{LINi}$  were calculated by dividing the values between 9 and 1 cm/yr into 5 equally spaced values. Accordingly, the grand mean is  $\bar{Y} = (9 + 1)/2 = 5$  and the expected slope,  $b$ , the vertical change divided by the horizontal change, is  $b = (\bar{Y}_{LIN5} - \bar{Y}_{LIN1})/(c_5 - c_1) = (9 - 1)/(-2 - 2) = -2$ .

Thus the straight line equation for this alternate hypothesis is  $\bar{Y}_{LINi} = \bar{Y} + b c_i = 5 - 2 c_i$ . For this example, the conversion factor  $m = 0.02$  so that  $b_x = (0.02)(-2) = -0.04$  (cm/yr per g/ha of herbicide).

Recall that the non-centrality parameter can be written as  $n SSM / \sigma^2$  where  $SSM = b^2 \sum c_i^2$ . For the example,  $SSM = (-2)^2(10) = 40$ . If the estimate of  $\sigma^2$  is  $MSE = 50$ , then we can calculate the power for some sample sizes. Some of the values output by the SAS program in Appendix 3 are:

Power for the linear contrast with a slope of -2  
Or a change in growth rate from 9 to 1 cm/yr

Sample size (n)	Error df (dfe)	Non-centrality parameter (nc)	Sums of squares of the means (SSM)	Critical F-value for contrast	Power for contrast
7	30	5.6	40	4.17	0.62940
8	35	6.4	40	4.12	0.69145
9	40	7.2	40	4.08	0.74487
10	45	8.0	40	4.06	0.79035
11	50	8.8	40	4.03	0.82871

<sup>3</sup> We ignore all subsampling issues in this discussion. The response variable is some overall measure of each experimental unit's response to the treatment (e.g., the average height of seedlings in each experimental unit).

<sup>4</sup> This is not an optimal design if the researcher *knows* that there is a linear trend in the response. If that were indeed the case, then a design with  $X$ -values only at the extremes of the  $X$ -range of interest would be optimal. An intermediate approach would assign more experimental units to these end points, while still having some units assigned to intermediate  $X$ -values.

From this table it is clear that a sample size of 10 or 11 meets the requirements, namely that the study should have a power of about 80% to detect a slope of  $b = -2$  using the contrast scale. This is equivalent to  $b_x = -0.04$  using the original scale of  $X$  or a decrease of 1 cm/yr growth for each 25 g/ha of herbicide.

As we consider different alternate hypotheses we change the slope of the line. By default, this line pivots about its centre at  $(c_i, \bar{Y}_{LINi}) = (0, \bar{Y})$  or  $(X_i, \bar{Y}) = (\bar{X}, \bar{Y})$ . If necessary, we can change the pivot point to a more meaningful place. The next example shows how important this can be.

**Scenario 2:** Suppose that seven experimental units per treatment level are the most our resources will allow. Then we might ask: At what size of linear contrast will we get sufficient power? We can determine this by looking at the power for a range of slopes. Some of the values output by the SAS program in Appendix 3 are:

Power for the linear contrast with a sample size of 7

Slope of the line (B)	Slope in terms of X (Bx)	Error df (dfe)	Non-centrality parameter (nc)	Sums of squares of the means (SSM)	Critical F-value for contrast	Power for contrast
-2.00	-0.040	30	5.6000	40.000	4.17	0.62940
-2.25	-0.045	30	7.0875	50.625	4.17	0.73114
-2.50	-0.050	30	8.7500	62.500	4.17	0.81653
-2.75	-0.055	30	10.5875	75.625	4.17	0.88260
-3.00	-0.060	30	12.6000	90.000	4.17	0.92975

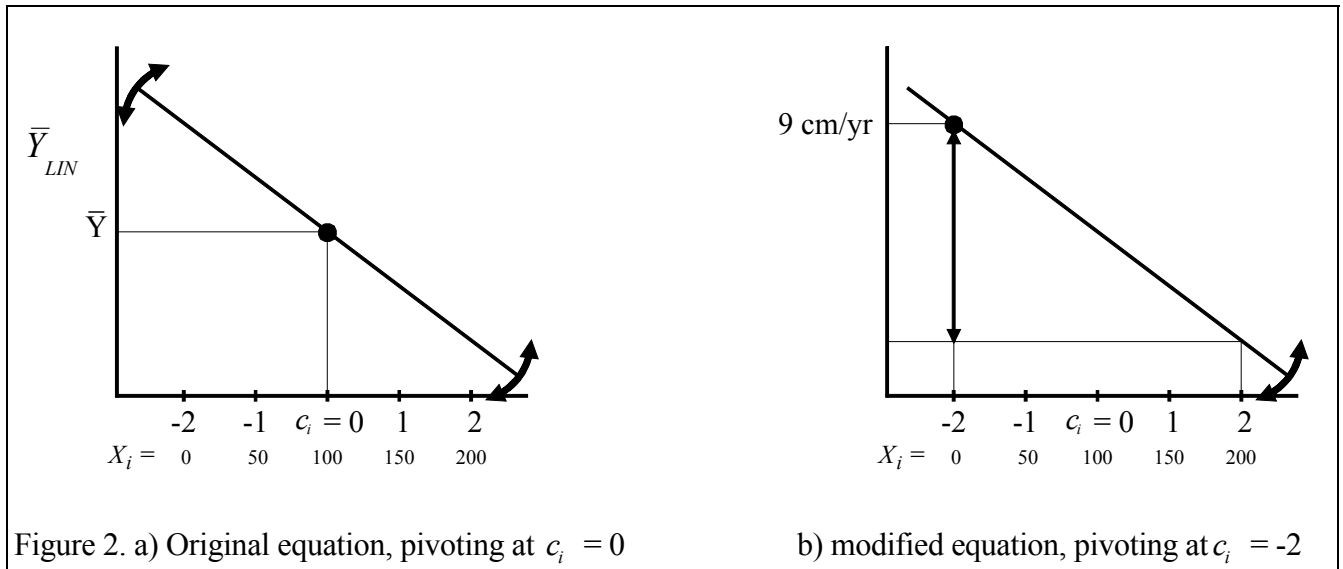
This table shows that we can get about 80% power if  $b = -2.50$ . We can interpret this by using equation (1). At  $c_5 = 2$  (or  $X_5 = 200$ ), the predicted growth rate is  $\bar{Y}_{LIN5} = 5 + (-2.5)(2) = 0$  cm/yr. This is a possible value since it is not negative. So we might conclude that our study would be powerful enough for any slopes this steep or steeper, that is, for  $b \leq -2.50$ .

While  $\bar{Y}_{LINi} = 5 - 2.5 c_i$  looks like a reasonable alternate hypothesis, notice that this equation also predicts that when no herbicide is applied the growth rate is  $\bar{Y}_{LIN1} = 5 + (-2.5)(-2) = 10$  cm/yr. This does not match our expected growth rate of 9 cm/yr. The problem here is that equation (1) describes a line that pivots about the point  $c_i = 0$  (or  $X_i = \bar{X}$ ) and  $\bar{Y}_{LINi} = \bar{Y}$  (see Figure 2a). Instead we want an equation that remains fixed at the no herbicide level ( $c_1 = -2$  and  $\bar{Y}_{LIN1} = 9$ ) and shows us what happens at the 200 g/ha level ( $c_5 = 2$ ) as we change the slope (see Figure 2b). This would match our understanding of the situation and make it easier for us to interpret possible slope values.

Changing the pivot point requires determining the growth rate at  $c_5 = 2$  as a function of the growth rate at  $c_1 = -2$ , namely,  $\bar{Y}_{LIN5}$  (at  $c_5 = 2$ ) =  $\bar{Y}_{LIN1}$  (at  $c_1 = -2$ ) +  $b(c_5 - c_1)$ . Thus  $\bar{Y}_{LIN5} = 9 + 4b$ .

Now,  $\bar{Y}_{LINi}$  will always be 9 at  $c_i = -2$  as we expect of our model. Further, it is easier to see that at the 200 g/ha level  $\bar{Y}_{LIN5} = 9 + (4)(-2.5) = -1$ . This is not a sensible value because we would not expect a growth rate less than zero. Thus, we must conclude that our experiment is not powerful enough for the

kinds of effects that we are looking for. A change in growth rate from 9 cm/yr to zero at 200 g/ha ( $c_5 = 2$ ) is the largest, most reasonable reduction in growth rate and occurs at slope  $b = -2.25$ . The corresponding power for this slope is only 73%. The researcher will have to decide if this power is adequate for the purposes at hand.



Since the values for  $\bar{Y}_{LINi}$  run into a “wall” at zero, we might also want to consider whether a linear model is appropriate for this situation. A non-linear model that allows a slow approach to the “wall” might make more biological sense.

**Determining the best fit line from the observed contrast sums of squares:** Suppose we have the final ANOVA results with a good-fitting linear contrast. We can calculate the slope<sup>5</sup> and line for the linear model using the sums of squares of the contrast,  $SS_{LIN}$ . Recall that  $n = 7$  and  $\sum c_i^2 = 10$ . If the observed  $SS_{LIN} = 315$  and the observed  $MSE$ <sup>6</sup> is 65 then:<sup>7</sup>

$$\hat{b} = \pm \sqrt{SS_{LIN} / \left( n \sum_{i=1}^a c_i^2 \right)} = \pm \sqrt{315 / (7 \times 10)} = \pm 2.12.$$

The standard error for  $\hat{b}$  is  $SE(\hat{b}) = \sqrt{\frac{MSE}{n \sum c_i^2}} = \sqrt{\frac{65}{7 \times 10}} = 0.96$ . If the observed grand mean is  $\bar{Y} = 5.1$

(and not the 5.0 we modelled), then the observed equation is  $\bar{Y}_{LINi} = \bar{Y} + \hat{b} c_i = 5.1 - 2.12 c_i$ .

<sup>5</sup> Many computer programs for analysis of variance can calculate the slope and its standard error. Example SAS code for PROC's GLM and MIXED is presented at the end of Appendix 3.

<sup>6</sup> Since we observed an  $MSE$  of 65 instead of 50, our earlier a priori power calculations were overly optimistic.

<sup>7</sup> This equation for the slope and its standard error are developed in Appendix 4. Notice that the slope could be either positive or negative but that in this context, we expect a negative slope and so use the negative value.

These can be converted to the  $X$ -scale using the conversion factor  $m$ :

- $\hat{b}_x = m \hat{b} = 0.02(-2.12) = -0.0424$  and its  $SE(\hat{b}_x) = m SE(\hat{b}) = 0.02 \times 0.96 = 0.019$ .
- The equation<sup>8</sup> on the  $X$ -scale is  $\bar{Y}_{LINi} = (\bar{Y} - \hat{b}_x \bar{X}) + \hat{b}_x X_i$ . Thus the observed equation is  $\bar{Y}_{LINi} = (5.1 + 0.0424 \times 100) - 0.0424 X_i = 9.34 - 0.0424 X_i$ .

The fitted means and their standard errors from the linear contrast are:<sup>9</sup>

$i$ :	1	2	3	4	5
Treatment level, $X_i$ :	0	50	100	150	200
Linear coefficients, $c_i$ :	-2	-1	0	1	2
Fitted $\bar{Y}_{LINi}$ :	9.34	7.22	5.10	2.98	0.86
Standard error:	2.36	1.67	1.36	1.67	2.36

**Reporting the results:** Power analyses should be performed BEFORE the experiment is conducted. Final reports can include a discussion of the power of experiments based on results from Scenario 1 and/or 2 but not the power of the observed slope calculated post hoc. In other words, reports can discuss the power for a specific alternate hypothesis for the slope and/or the power for a specific sample size.

#### Summary of equations useful for the power analysis of a linear contrast:

- The treatment level means of the linear contrast are fitted by  $\bar{Y}_{LINi} = \bar{Y} + b c_i$ , where  $\bar{Y}$  is the grand mean of the response variable, and  $b$  is the slope of the line using the contrast coefficients  $c_i$ .
- The conversion factor between the original scale in the  $X$ -variable and the contrast scale is  $m = \{\text{Last}(c_i) - \text{First}(c_i)\} / \{\text{Last}(X_i) - \text{First}(X_i)\}$ .
- The contrast coefficients are linearly related to the independent variable  $X_i$  by  $c_i = m (X_i - \bar{X})$ .
- The slope on the original  $X$ -scale has a simple relationship to the contrast slope:  $b_x = mb$ .
- The equation on the original  $X$ -scale is  $\bar{Y}_{LINi} = (\bar{Y} - \hat{b}_x \bar{X}) + \hat{b}_x X_i$ .
- The non-centrality parameter,  $nc$ , can be written in several different ways, two of which are:

$$nc = \left( nb^2 \sum_{i=1}^a c_i^2 \right) / \sigma^2 = (n \times SSM) / \sigma^2 \text{ where } SSM = \sum_{i=1}^a (\bar{Y}_{LINi} - \bar{Y})^2 = b^2 \sum c_i^2.$$

#### Prepared by:

Wendy Bergerud, B.C. Ministry of Forests, Research Branch.  
Phone: 250-387-5676; e-mail: wendy.bergerud@gems8.gov.bc.ca.

#### References:

Bergerud, W.A. 1993. Calculating contrast F-tests when SAS will not. B.C. Min. For., Res. Br., Victoria, B.C. Biometrics Inf. Pam. No. 45.

<sup>8</sup> See Appendix 1.

<sup>9</sup> The required equation for the standard errors are developed in Appendix 4.

- Bergerud, W.A. and V. Sit. 1992. Power analysis workshop notes. B.C. Min. For., Res. Br., Biometrics Section, Victoria, B.C.
- Keppel, G. 1973. Design and analysis: A researcher's handbook. 1st ed. Prentice-Hall Inc., Englewood Cliffs, N.J.
- Maxwell, S.E. and H.D. Delaney. 1990. Designing experiments and analyzing data: A model comparison perspective. Wadsworth Publ. Co., Belmont, Calif.

### Appendix 1. Converting the slope to the scale of the independent variable

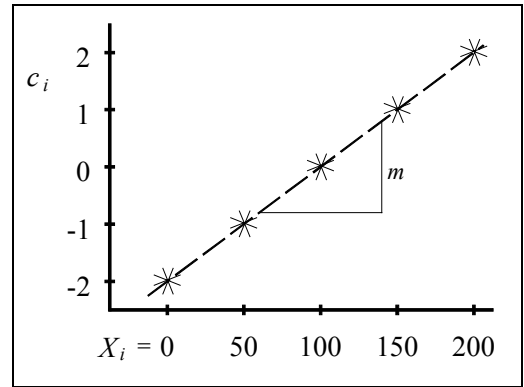
The contrast coefficients,  $c_i$ , and the independent variable,  $X_i$ , have a linear relationship as shown by the figure on the right using values from the example. The slope between these two is the conversion factor,  $m$ . Thus:

$$m = \{\text{Last}(c_i) - \text{First}(c_i)\} / \{\text{Last}(X_i) - \text{First}(X_i)\}.$$

For the example,  $m = \{2 - (-2)\} / \{200 - 0\} = 4/200 = 0.020$ .

Showing that this is the conversion factor is straightforward. First, note that  $c_i = m(X_i - \bar{X})$ . Since

$\bar{Y}_{LINi} = \bar{Y} + b c_i$ , then  $\bar{Y}_{LINi} = \bar{Y} + bm(X_i - \bar{X}) = (\bar{Y} - mb\bar{X}) + mb X_i = (\bar{Y} - \hat{b}_x \bar{X}) + \hat{b}_x X_i$ . Hence the slope on the  $X$ -scale is  $b_x = mb$ .



### Appendix 2. The relationship between the non-centrality parameter and the slope

To show the relationship between the non-centrality parameter and the slope, we must first recall<sup>10</sup> that the non-centrality parameter is  $nc = (n \times SSM) / \sigma^2$ , where  $n$  is the constant sample size for each treatment in a balanced ANOVA,  $SSM$  is the sums of squares of the means for the alternate hypothesis, and  $\sigma^2$  is the variance usually estimated by  $MSE$  (the error term for the treatment source of variation under consideration). The  $SSM$  for the non-centrality parameter of a linear contrast can be calculated by using the expected means for the contrast,  $\bar{Y}_{LINi}$  (where  $i$  indexes the treatment levels), and their grand mean  $\bar{Y}$ :

$$SSM = \sum_{i=1}^a (\bar{Y}_{LINi} - \bar{Y})^2.$$

Now the linear relationship between the  $\bar{Y}_{LINi}$  and the contrast coefficients,  $c_i$ , is  $\bar{Y}_{LINi} = \bar{Y} + b c_i$ , where  $b$  is the slope on the contrast scale. Thus,  $SSM = \sum (\bar{Y}_{LINi} - \bar{Y})^2 = \sum (\bar{Y} + b c_i - \bar{Y})^2 = b^2 \sum c_i^2$  since  $b$  is constant for all treatment levels. Thus,  $nc = (nb^2 \sum c_i^2) / \sigma^2$ .

If desired, we can write this in terms of  $X_i$  rather easily. From Appendix 1,  $b = b_x / m$  and  $\sum c_i^2 = \sum [m(X_i - \bar{X})]^2$  so  $nc = [n b_x^2 \sum (X_i - \bar{X})^2] / \sigma^2$ .

<sup>10</sup> From the discussion in the *Power Analysis Workshop Notes* (Bergerud and Sit 1992).

**Appendix 3. Example SAS Program**

```

/* Program to calculate the Power for a Linear Contrast          */
/* For a simple completely randomized design with no subsampling */

title 'Power for a Linear Contrast';
data ylin;
  mse = 50; t = 5;      ** Setting Mean Square Error value and No. of treatments;
  dfh = t-1;  alpha = 0.05;
  do n = 5 to 20;      ** A range of sample sizes;
    do b = 0 to -5 by -0.25; ** A range of slope values;
      bx = 0.02 * b;    ** Slope for X-values;
      ssm = b*b*10;     ** The SS of the coefficients is 10;
      dfe = t*(n-1);   ** Error degrees of freedom;
      nc = n*ssm/mse;  ** Non-centrality parameter;
      fcritlin = finv(1-alpha, 1, dfe, 0); ** Critical F-value for contrast;
      power = 1-probf(fcritlin, 1, dfe, nc);
      output; end; end;
label  b='Slope of_the line_(B1c)'  n='Sample_Size_(n)'
       dfe='Error_df_(dfe)'          nc='Non-centrality_Parameter_(nc)'
       ssm='Sums of Squares_of the Means_(SSM)'
       power='Power for_Contrast'    bx = 'Slope in_terms of X_(Bx)'
       fcritlin='Critical_F-value_for contrast';  format fcritlin 5.2;
run;

*** SAS Output Table for Scenario 1;
proc print split='_'; where b = -2; id n; var dfe nc ssm fcritlin power;
  title 'Power for the linear contrast with a slope of -2';
  title2 'Or a change in growth rate from 9 to 1 cm/yr';  run;
proc plot vpercent=70; where b = -2;
  plot power*n='+' / vref = 0.8, 0.9; run;      *<=== Output not shown;

*** SAS Output Table for Scenario 2;
proc print split='_'; where n = 7; id b bx; var dfe nc ssm fcritlin power;
  title 'Power for the linear contrast with a sample size of 7'; run;
proc plot vpercent = 70; where n = 7;
  plot power*b='+' / vref = 0.8, 0.9; run;      *<=== Output not shown;

/**** Example code to obtain slope estimates. For the X-value slope the X-mean
must be subtracted from each value. The divisor is the sum of the squared
values listed after x in the estimate statement;          ****/

proc glm or mixed; class x; model y = x;
  contrast 'linear' x -2 -1 0 1 2;      *<=== F-test only, no slope estimate;
  estimate 'contrast slope b' x -2 -1 0 1 2 / divisor = 10;
  estimate 'X-value slope bx' x -100 -50 0 50 100 / divisor = 25000;
run;

```

**Appendix 4.** Using the sums of squares,  $SS_{LIN}$ , to obtain the slope and standard errors for the fitted means of a linear contrast within a balanced completely randomized design without subsampling

The observed slope  $\hat{b}$  of the best fit line to the treatment means can be calculated from the data by the standard formula<sup>11</sup> for a slope where the  $c_i$  take the place of the usual  $X_i$ :

$$\hat{b} = \left( \sum_{i=1}^a c_i \bar{Y}_i \right) / \sum_{i=1}^a c_i^2 \quad (4.1)$$

For a balanced one-way ANOVA the sums of squares for the linear contrast,  $SS_{LIN}$  (Keppel 1973) is calculated by:  $SS_{LIN} = n \left( \sum_{i=1}^a c_i \bar{Y}_i \right)^2 / \sum_{i=1}^a c_i^2 = n \left( \sum_{i=1}^a c_i \bar{Y}_i \right) \times \left( \sum_{i=1}^a c_i \bar{Y}_i \right) / \sum_{i=1}^a c_i^2$ .

Equation (4.1) can be rearranged to show that  $\left( \sum_{i=1}^a c_i \bar{Y}_i \right) = \hat{b} \times \sum_{i=1}^a c_i^2$ , the first half of the above equation, while the second half is simply  $\hat{b}$ . Therefore,  $SS_{LIN} = n \hat{b}^2 \sum_{i=1}^a c_i^2$ , which can be rearranged to produce:

$$\hat{b}^2 = SS_{LIN} / \left( n \sum_{i=1}^a c_i^2 \right). \quad (4.2)$$

Note that  $\hat{b}$  is the slope with respect to the contrast coefficients, and not the original treatment levels  $X_i$ . Also, the sign of the slope (positive or negative) cannot be determined from equation (4.2) but must be determined from the context. The slope's standard error is given by  $SE(\hat{b}) = \sqrt{\frac{MSE}{n \sum c_i^2}}$ . Using the conversion factor  $m$  we can get the  $X$ -scale versions of the slope:  $\hat{b}_x = m \times \hat{b}$ , and its standard error:  $SE(\hat{b}_x) = m \times SE(\hat{b})$ .

The best fit means for the linear contrast can be calculated using equation (1):  $\bar{Y}_{LINi} = \bar{Y} + \hat{b} c_i$ . Since  $\bar{Y}$  and  $\hat{b}$  are independent,  $SE^2(\bar{Y}_{LINi}) = \text{Var}(\bar{Y}_{LINi}) = \text{Var}(\bar{Y}) + c_i^2 \text{Var}(\hat{b}) = \frac{MSE}{an} + \frac{c_i^2 MSE}{n \sum c_i^2}$ .

Rearranging, we obtain:  $SE(\bar{Y}_{LINi}) = \sqrt{\frac{MSE}{n} \left( \frac{1}{a} + \frac{c_i^2}{\sum c_i^2} \right)}$ .

A SAS program to produce the fitted linear means and their standard errors using these equations for the example is:

```
data ylin; do x = 0 to 200 by 50; ci = 0.02*(x-100);
  b = -1*sqrt(315/(7*10)); bx = 0.02 * b;
  ylin = (5.1 - 2*b) + bx * x; se = sqrt((65/7)*(1/5 + ci*ci/10)); output; end;
label x = 'Xi' cd = 'Ci' ylin = 'Fitted Y-value' se = 'Standard Error'; run;
proc print label; id x; var ci ylin se; format ylin se 5.2; title ' '; run;
```

<sup>11</sup> Any textbook discussing simple regression will describe this formula. Maxwell and Delaney (1990) present this formula on page 212.