

BIOMETRICS INFORMATION

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

PAMPHLET NO. # 23

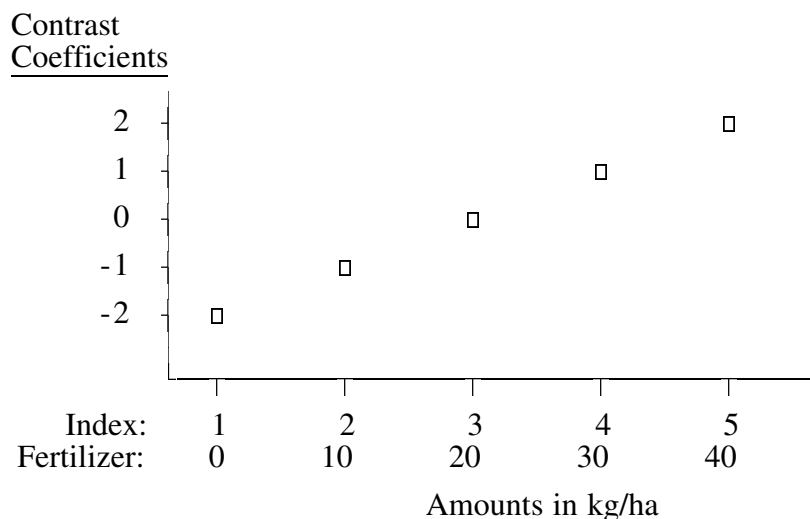
DATE: October 12, 1989

SUBJECT: ANOVA: Contrasts viewed as correlation coefficients

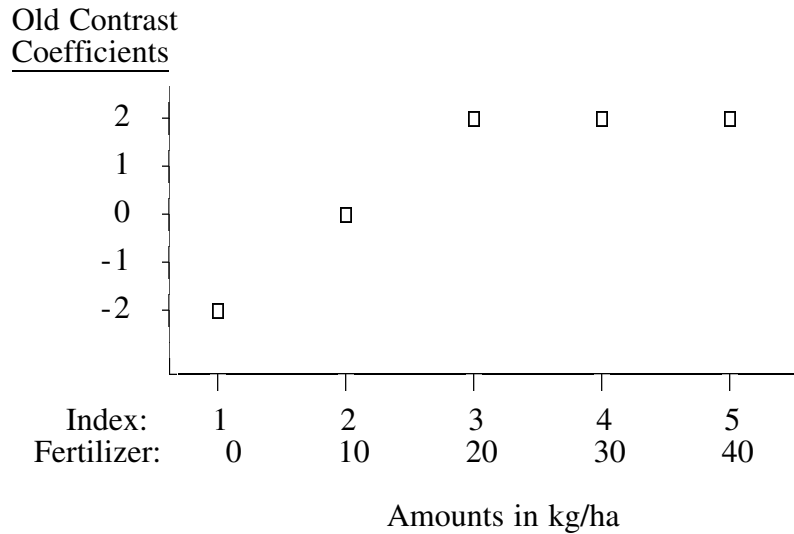
Contrasts can be viewed as correlations between the contrast coefficients, c_i , and the observed means, \bar{Y}_i . If a correlation coefficient, r , is high (i.e. has a value near 1 or -1) then the observed data closely match the pattern displayed by the contrast coefficients. If the correlation is low (i.e. near zero), then the observed data fail to closely match this pattern.

A common example is the linear contrast. Suppose that an experiment has 5 levels of fertilizer, each applied to an equal number, n , of experimental units. In the simplest case the 5 levels or amounts are equally spaced with values such as 0, 10, 20, 30, and 40 kg/ha. The linear contrast coefficients are -2, -1, 0, 1, and 2. A high correlation between these contrast coefficients and the treatment means indicates a strong linear trend in the response.

The pattern being tested by any set of contrast coefficients can be seen by plotting their values against the index (i) or the treatment values. For example:



The real power of this vision of contrasts is that it allows us to pick **any** pattern that we think the data may follow and test it. We are not restricted to the usual or familiar contrasts. For instance, suppose the researcher expected the following pattern from a fertilizer experiment:



This response initially increases (linearly), but then stabilizes at 20 kg/ha. Thus, if the observed response followed this pattern, one would conclude that 30 and 40 kg/ha gave no different response than 20, while 20 was better than 10 (if more is better) and 10 better than none.

The coefficients used in the above graph are not correct for this new pattern since their sum is not zero. This is corrected by subtracting the mean from each coefficient. For example:

The present coefficients: -2 0 2 2 2 have a mean of $4/5$.

So multiply by 5 (to avoid fractions):

to get coefficients: -10 0 10 10 10 with a mean of 4.

Then subtract the mean of 4 so that

the final coefficients: -14 -4 6 6 6 have a mean of 0.

These coefficients can be divided by 2, their greatest

common divisor, to obtain: -7 -2 3 3 3, still with a mean of 0.

TESTS: Contrasts can be tested in the usual way once the above contrast coefficients have been obtained (see Biometrics Information pamphlet #16 and standard textbooks on Analysis of Variance). The correlation coefficient can be calculated on most hand calculators with simple regression capabilities if the treatment means have the same sample size (i.e. a balanced ANOVA). For a one-way ANOVA the Between Sums of Squares (SSB) and the Mean Square Error (MSE) of the ANOVA can also be calculated by hand (see Biometrics Information pamphlet #22).

The hand calculator method is as follows:

1. Determine the correlation coefficient, r , between the contrast coefficients and the treatment means.

2. Calculate the Contrast Sums of Squares (SSC) by multiplying the square of the correlation coefficient by SSB, i.e. $SSC = r^2 SSB$.
3. Calculate the F-test by dividing the SSC by the MSE, i.e.

$$F = SSC/MSE \text{ with } df = 1, df(MSE).$$

Determine the significance as usual.

Example:

Fertilizer Amount:	<u>0 kg/ha</u>	<u>10 kg/ha</u>	<u>20 kg/ha</u>	<u>30 kg/ha</u>	<u>40 kg/ha</u>
Observed Mean:	28	42	50	66	64
Linear contrasts:	-2	-1	0	1	2

The Between Sums of Squares (SSB) is 4000 with $df = 4$.

The Mean Square Error (MSE) is 500 with $df = 10$,
with a sample size of 4 per treatment.

The correlation coefficient is 0.960. Thus $SSC = (0.960)^2(4000)$, and the F-test is given by:

$$F = \frac{(0.960)^2 (4000)}{500} = \frac{3686.40}{500} = 7.37, \text{ } df = 1, 10.$$

CONTACT: Wendy Bergerud
387-5676

PROBLEM FROM BI #22

The calculations are as follows:

<u>Treatment</u>	<u>Data</u>	<u>Mean, \bar{Y}_i</u>	<u>Standard Deviation, S_i</u>	<u>S_i^2</u>
1	5 3 1 0 1	2.000	2.00	4.000
2	7 6 5 8 4	6.000	1.5811	2.500
3	11 9 7 6 7	8.000	2.00	4.000
4	10 6 9 9 6	8.000	1.8708	3.500
	Grand Mean:	6.000	Sum:	14.000
	Std. Dev. S_m :	2.8284		

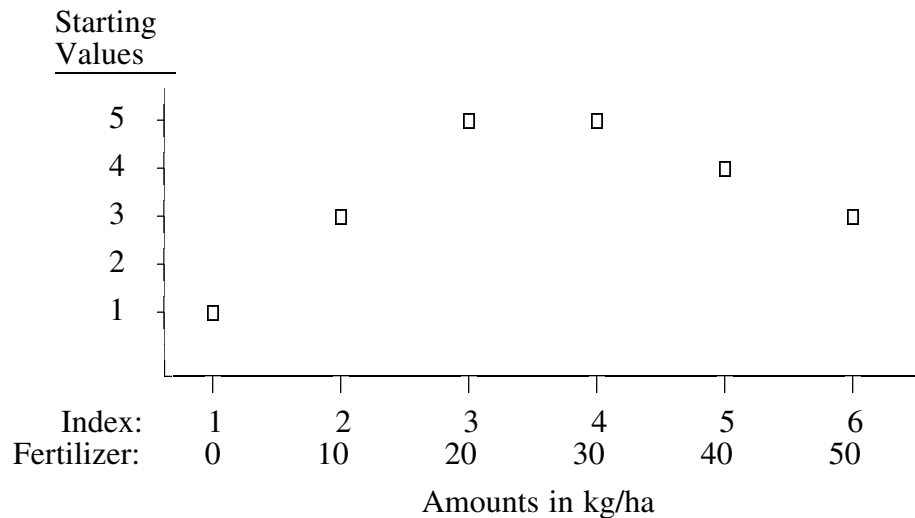
In this case, $a = 4$, $n = 5$, $\sum S_i^2 = 14.000$, and $S_m^2 = 8.000$. Hence:

$$MSE = \frac{\sum S_i^2}{a} = \frac{14.000}{4} = 3.500, \text{ } MSB = nS_m^2 = 5(2.8284)^2 = 40.000,$$

$$\text{and } F = \frac{40.000}{3.500} = 11.43 \text{ with } df = 3, 16.$$

NEW PROBLEMS

1. Calculate the linear contrast using the method described in pamphlet #16. Do you get the same answer?
2. Calculate the second contrast described above for the data given. Try reversing the order of coefficients with the means. What happens? Would the same thing happen with the linear contrast?
3. Calculate the contrast coefficients for the following expected pattern:



Note that although there are two straight line segments in this model, their rate of change (slope) is different, with the initial increase twice the rate of the final decrease. When designing this pattern, and others like it, you must be quite specific. You must specify more than just a linear trend. You must also specify the relationship of the rates of change. For instance, in the example above, the initial increase is twice as fast as the final decrease.
