SUBJECT: ANOVA: Equations for Linear and Quadratic Contrasts

Data are often examined for linear and quadratic trends when the levels of the treatment variable are quantitative (numeric). In ANOVA, this is often done with linear and quadratic contrasts (see Biometrics Information pamphlets 12, 16, and 23). When a linear or quadratic trend is found, it is often desirable to have the corresponding equation.

The easiest way to obtain these equations is to run a regression on the treatment levels (X) for the linear case, and on the treatment levels (X) and the squares of the treatment levels (X^2) for the quadratic fit. In general, the statistics associated with these regressions (especially the R^2) should be ignored. If the ANOVA is balanced then regressions on the individual data will provide the same equation as will regressions on the treatment means. In this case, regressions on the individual data provide a check since the model sums of squares should be the same as the contrast sums of squares. For the quadratic case, the sums of squares for the linear and quadratic contrasts would have to be added to make this check.

If the ANOVA is unbalanced, it may be preferable to use treatment means so that each mean has an equal influence on the fit. Otherwise means with larger sample sizes will be fit "closer" than will other treatment means. This is a choice between an unweighted versus a weighted regression with sample size as the weight.

The use of the equation for the linear case is straight-forward. For the quadratic case, it may be desirable to rewrite the equation so that estimates of the height and position of the "peak" or "trough" can be easily obtained. The regression program will output an equation of the form:

$$Y = A + BX + CX^2.$$

But the position of the peak or trough is more easily obtained when the equation is rewritten in the form: $Y = a(X - b)^2 + c.$

A little algebra shows the following relations between the two sets of parameter estimates:

$$a = C \qquad\qquad b = \frac{B}{-2C} \qquad\qquad c = A - \frac{B^2}{4C}$$

The following data will demonstrate how to use these equations:

Treatment Code	X-value	Replication Values					Treatment Means
A	8	18.4	24.0	24.8	25.3	24.5	23.40
B	13	28.6	23.8	26.5	22.0	25.0	25.18
C	18	31.1	25.9	28.9	29.7	29.7	29.06
D	23	27.0	27.9	27.3	28.3	28.4	27.78
E	28	32.1	31.9	31.4	27.4	30.8	30.72
F	33	30.7	28.1	29.1	27.5	29.4	28.96

The contrast coefficients used to calculate polynomial sums of squares and tests are presented below as part of the corresponding SAS statements in `PROC GLM`:

```
****************        TREATMENTS:     A     B     C     D     E     F :
*******************       X-VALUES:      8    13    18    23    28    33 ;

CONTRAST 'LINEAR'          TREAT      -5    -3    -1     1     3     5 ;
CONTRAST 'QUADRATIC'       TREAT       5    -1    -4    -4    -1     5 ;
CONTRAST 'CUBIC'           TREAT      -5     7     4    -4    -7     5 ;
```

The resulting ANOVA table is:

Source of Variation	Degrees of Freedom	Sums of Squares	F-value
Treatment	5	186.014	9.40
Contrasts:			
Linear	1	132.933	33.57
Quadratic	1	27.413	6.92
Cubic	1	0.954	0.24
Error	29	95.028	

The results of the regressions on the individual data are:

MODEL	Degrees of Freedom	Sums of Squares	Parameter Estimates A	B	C
Linear	1	132.933	22.463	0.247	
Quadratic	2	160.345	17.139	0.875	-0.0153

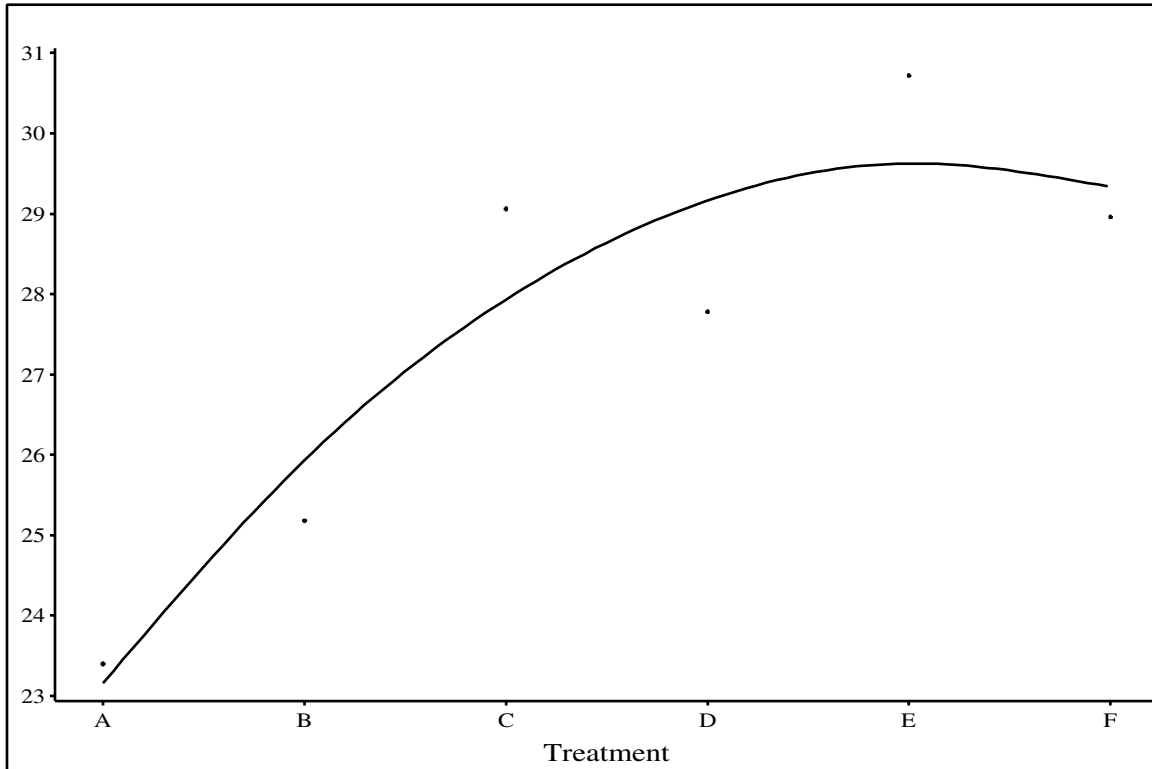The quadratic equation is: $Y = 17.139 + 0.875\ X - 0.0153\ X^2$

The parameter estimates for the other version of the quadratic equation is:

$$a = -0.0153 \qquad\qquad b = \frac{0.875}{(-2)(-0.0153)} = -28.54$$

$$c = 17.139 - \frac{0.875^2}{(4)(-0.0153)} = 29.625$$

so that $Y = -0.0153(X - 28.54)^2 + 29.625$

Hence the peak[1] occurs at $X = 28.54$ with a value of $Y = 29.625$, and the equation could be plotted on the graph of Y versus X. For example:

Note that you should keep as many decimal places as possible during the calcuations and only round the values at the end.

CONTACT: Wendy Bergerud
                    387-5676

─────────────────────────────NEW PROBLEM─────────────────────────────

Suppose that you had obtained the quadratic equation:
$$Y = 5X^2 - 120X + 720.$$
Rewrite this equation and determine where the peak occurs.

---

[1]Since $a < 0$, this equation describes a peak instead of a trough.

Calculating the F_{max}-test and pooled variance:

Treatment	Mean	SD	n	Variance	df	Var*df
1	15.2	20	10	400	9	3600
2	16.5	25	8	625	7	4375
3	18.7	23	13	529	12	10051
				SUM:	28	14323

Hence $F_{max} = \dfrac{625}{400} = 1.56$ and the pooled variance is $\dfrac{14323}{28} = 511.54$. The number of treatments, k, is 3 and the largest degrees of freedom is 12. Since the critical F_{max} value is 4.16 at the probability level of 0.05, one does not reject the homogeneity of variance assumption and uses the pooled variance of 511.54.

```
/* EXAMPLE.SAS */

libname lib 'd:\usr\examples';
TITLE 'TESTING FOR A POLYNOMIAL FIT TO TREATMENTS';
TITLE2 'AND OBTAINING THE CORRESPONDING EQUATION';

DATA LIB.NTEMP;
  INFILE 'EXAMPLE.DAT';
  INPUT TRT $ X @;
    DO REP = 1 TO 5;
       INPUT Y  @;
       X2 =  X * X;
       X3 = X2 * X;
       OUTPUT;
    END;
RUN;

PROC PRINT; RUN;

PROC  GLM  DATA=LIB.NTEMP;
  CLASS TRT;
  MODEL Y = TRT;
************  TREATMENTS:  A  B  C  D  E  F:
**************  X-VALUES:  8 13 18 23 28 33;
  CONTRAST 'LINEAR'    TRT -5 -3 -1  1  3  5;
  CONTRAST 'QUADRATIC' TRT  5 -1 -4 -4 -1  5;
  CONTRAST 'CUBIC'     TRT -5  7  4 -4 -7  5;
RUN;

PROC REG DATA=LIB.NTEMP;
  MODEL Y = X  ;
  MODEL Y = X X2;
  MODEL Y = X X2 X3;
RUN;

/* End of Program */
```