

## Orthogonal and Non-orthogonal Polynomial Contrasts

We had carefully reviewed orthogonal polynomial contrasts in class and noted that Brian Yandell makes a compelling case for nonorthogonal polynomial contrasts. In the following example, we will revisit both methods and compare analyses.

The **Solution Concentration** data set from *Applied Linear Statistical Models, 5th ed* by Kutner et al, measures concentration of a solution over time.

Concentration (Y)	Time in Hours (X)
0.07	9.0
0.09	9.0
0.08	9.0
0.16	7.0
0.17	7.0
0.21	7.0
0.49	5.0
0.58	5.0
0.53	5.0
1.22	3.0
1.15	3.0
1.07	3.0
2.84	1.0
2.57	1.0
3.10	1.0

A plot of the data set (Figure 1) in **R** shows a sharply nonlinear decreasing trend in concentration over time. Given such a trend, the natural log transformation of the response should fix both the nonlinearity and the increasing error variance. As you can see (Figure 2), the transformation works perfectly, which is likely what the textbook authors had in mind.

It might be more interesting to apply polynomial models to the untransformed data, but given the violation of regression assumptions (unequal error variances), we will go ahead and model the transformed data instead.

A quick inspection of the data set confirms that the independent variable, Time in Hours, is quantitative with equally-spaced levels (in increments of 2.0 hours). In addition, the design is balanced, with  $n = 3$  replications per factor level, so this data set is appropriate for analysis with polynomial contrasts.

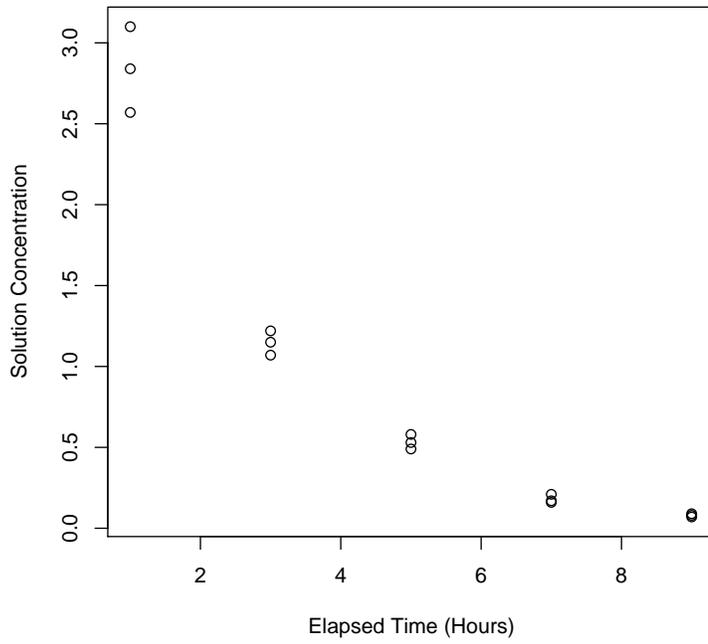


Figure 1: Scatterplot of untransformed response

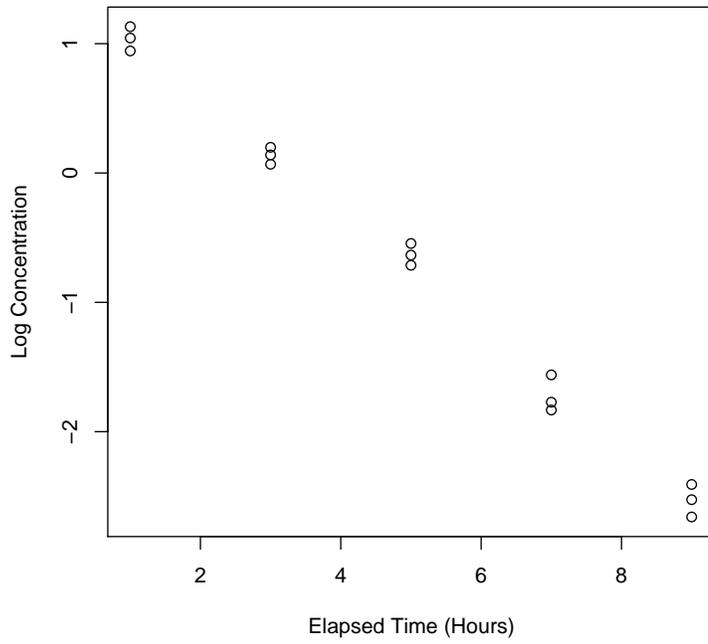


Figure 2: Scatterplot of transformed response

Since the data set has 5 levels, the orthogonal polynomial contrasts would be:

Time (X) in Hours	Linear coefficient	Quad coefficient	Cubic coefficient	Quartic coefficient
1.0	-2	2	-1	1
3.0	-1	-1	2	-4
5.0	0	-2	0	6
7.0	1	-1	-2	-4
9.0	2	2	1	1

Examining the data, interesting hypotheses (in addition to the general ANOVA hypothesis  $H_o : \mu_1 = \dots = \mu_a$ ) would include a test of the linear contrast, the quadratic contrast (given the linear contrast), and the linear lack of fit. I have specified these hypotheses using the CONTRAST command in SAS:

```
data ortho;
input Conc Hours;
LConc=log(Conc);
datalines;
0.07 9.0
0.09 9.0
0.08 9.0
0.16 7.0
0.17 7.0
0.21 7.0
0.49 5.0
0.58 5.0
0.53 5.0
1.22 3.0
1.15 3.0
1.07 3.0
2.84 1.0
2.57 1.0
3.10 1.0
;
proc glm data=ortho;
class Hours;
model LConc=Hours;
contrast 'linear' Hours -2 -1 0 1 2;
contrast 'quadratic' Hours 2 -1 -2 -1 2;
contrast 'linear lof' Hours 2 -1 -2 -1 2,
```

```
Hours -1 2 0 -2 1,
Hours 1 -4 6 -4 1;
```

```
run;
```

The ANOVA table shows that the general ANOVA hypothesis is strongly significant ( $p < .0001$ ). As expected the linear contrast is significant ( $p < .0001$ ). Note that the quadratic contrast can also be considered a test of whether or not a quadratic term could be included given that a linear term is already in the model. Hence, it serves as a hierarchical test of a quadratic model (with both linear and quadratic terms) versus a linear model. We can confirm from an inspection of Figure 2 that this test is not likely to be significant (in fact,  $p = .5715$ ).

ANOVA summary (based on  $MSE = .01128552$  with 10 df)

Test	df	SS	MS	F Value	p-value
General	4	24.35092977	6.08773244	539.43	< .0001
Linear	1	24.29199104	24.29199104	2152.49	< .0001
Quadratic	1	.00386341	.00386341	.34	.5715
Linear LOF	3	.05893874	.01964625	1.74	.2217

From the same figure (Figure 2), it would not be the least bit surprising if the Linear Lack-of-fit Test was not significant, and it is not ( $p = .2217$ ). Hence, the significant result for the general ANOVA hypothesis can be substantially explained by the linear relationship between the independent variable and the mean response.

Yandell rightly points out that most researchers are interested in the linear hypothesis, the quadratic hypothesis (Q|L), and either the linear or quadratic lack-of-fit hypothesis. Since these are essentially sequential tests, he argues that we can simply create appropriate linear and quadratic covariates without worrying about whether or not they are orthogonal. If we declare the independent variables as a factor (with the CLASS statement) and include it as the last term in our model, we can obtain our lack-of-fit tests by default from the Type I analysis. Yandell specifies a Type I analysis explicitly, even though SAS outputs both a Type I and Type III analysis by default—better safe than sorry.

Note that we actually need two separate analyses to obtain the linear lack-of-fit and quadratic lack-of-fit tests, unless we are willing to combine model results by hand.

```
data yandell; set ortho;
/* Linear and quadratic covariates for Yandell approach */
```

```

/* Note how simple they are to construct */
Hlin=Hours;
Hquad=Hlin*Hlin;
/*Yandell approach for quadratic lack-of-fit */
proc glm data=yandell;
class Hours;
model LConc=Hlin Hquad Hours/ss1;
run;
/*Yandell approach for linear lack-of-fit */
proc glm data=yandell;
class Hours;
model LConc=Hlin Hours/ss1;
run;

```

ANOVA summary using Yandell's approach  
(MSE=.01128552 with 10 df)

Test	df	Type I SS	MS	F Value	p-value
Linear	1	24.29199104	24.29199104	2152.49	< .0001
Q L	1	.00386341	.00386341	.34	.5715
Quadratic LOF	2	.05507532	.02753766	2.44	.1371
Linear LOF	3	.05893874	.01964625	1.74	.2217

The Type I analysis of the linear term is the same as the analysis of the linear term in our model based on orthogonal contrasts. Note that the test for the quadratic term (given a linear term is already present in the model) is the same for both analyses. The advantage of the orthogonal approach is that the test of the quadratic effect is the same whether or not the linear term is in the model; its disadvantage is that it is more difficult to set up. The quadratic lack-of-fit isn't an interesting hypothesis, but note that the linear lack-of-fit test is the same as the linear lack-of-fit test in the orthogonal analysis; it has the same advantages and disadvantages in Yandell's formulation as the test of Q|L.