Statistics 5303
Fall 2003

Exam #1 Sketched Solutions

1. Describe how you checked assumptions and what you decided. Tell me about nonnormality, nonconstant variance, outliers, and so on. Were there any problems that required fixing?

There was no evidence of nonconstant variance. One value looked like a possible outlier on the rankit plot, but its Bonferroni t-value was less than 3, so it wasn't really that outlying.

2. How would you describe the effect of editorials on the voters' perceptions?

The "negative" editorial caused the rating to be about 1.6 points higher. The "positive" editorial did not appear to affect the rating. This can be seen in various contrasts comparing negative editorials to positive editorials, negative editorials to no editorial, and positive editorials to no editorial. It is also apparent from the pairwise comparisons.

3. Is there any evidence that actually seeing (reading) the ad made any difference to the voters' perceptions?

No. The response with the negative editorial was about the same, regardless of whether the ad was available. Similarly for the positive editorial. Again, see the pairwise comparisons results.

4. Describe how you checked assumptions and what you decided. Tell me about nonnormality, nonconstant variance, outliers, and so on. Were there any problems that required fixing?

Data showed considerable nonconstant variance. Box-Cox analysis suggests that a cube-root should work, and indeed it does improve the residual plot. On the cube root scale, the residuals are slightly skewed to the right, but not enough to cause alarm.

5. Is there any simple model that accounts for most of the variation in these data?

Simple is in the eye of the beholder. All main effects and interactions were significant. Both solids and pH were quantitative, and some simplification in the model can be obtained that way. Only linear effects in solids are significant — no quadratic effects of solids are needed. Otherwise, all interaction terms are needed except quadratic pH by linear solids by CaCl.

6. Is there an interaction between solids and pH? If so, how would you describe it?

The effect of solids is linear, but it is increasing for the first two levels of pH, and decreasing for the last level of pH.

7. Suppose that I have planned an experiment with three treatments, 20 units per treatment, and anticipated error standard deviation $\sigma = 10$. Will my power increase more if I spend money to double my sample size

(to 40 units per treatment), or spend money to halve my σ to 5? Explain your answer.

Doubling the sample size will double the noncentrality parameter. Halving σ will cut σ^2 by a factor of four and quadruple the noncentrality parameter. We have enough error degrees of freedom in both cases that the difference in error df will not have much effect on the result. Thus halving σ will give the greater increase in power.

(1) Summary. Voters generally rate the ad as neutral, unless they are told by an editorial that it is negative, in which case they rate the ad as more negative by about 1.6 points. Presence or absence of the actual transcript does not change the result.

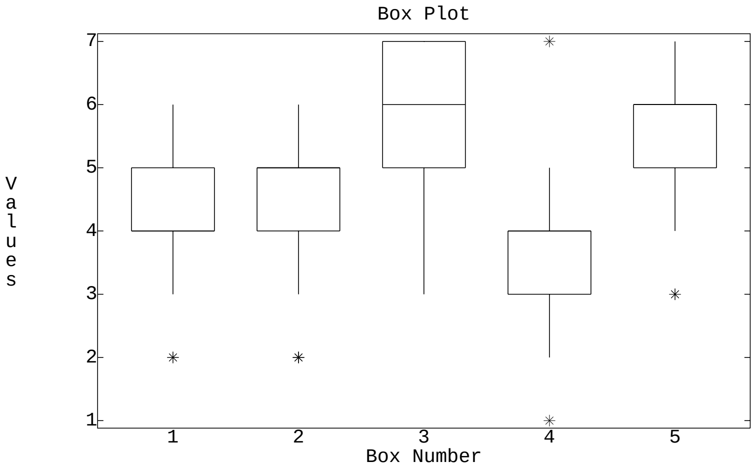
Data Analysis. The experiment is a completely randomized design, with five treatments and 113 units. Four of the treatments are the factor/level combinations of presence/absence of the ad transcript and postive/negative editorial. The fifth treatment is the transcript alone without an editorial.

A one-way ANOVA gives the following results:

	DF	SS	MS	F	P-value
CONSTANT	1	2420.6	2420.6	1645.57763	< 1e-08
ad	4	63.524	15.881	10.79614	2.1231e-07
ERROR1	108	158.87	1.471		

Residuals from this model show no sign of nonconstant variance. The rankit plot shows a potential outlier, but its absolute studentized residual is less than 3, so we do not consider it further.

Based on the ANOVA, there is strong evidence against the null hypothesis that the treatment means are equal. The observed treatment means are 4.26, 4.30, 5.65, 3.62, and 5.42, with standard errors ranging from about .23 to .28. Boxplots of the data tell us most of what we can learn:



Treatments 3 and 5 (which include negative editorials) have higher means, and the other three means are about the same. Pairwise comparisons (using HSD at the .05 level) confirm this impression:

```
|       4    -1.03
|       1   -0.389
|       2   -0.354
  |     5    0.771
  |     3        1
```

We may explore some specific hypotheses more closely using contrasts.

- H_O: presence of the ad transcript does not affect the mean response when an editorial is given; coefficients (0,1,1,–1,–1); SS = 4.56; p-value .08. We fail to reject this null hypothesis.

- H_O: positivity or negativity of the editorial does not affect the mean response when an editorial is given; coefficients (0,1,–1,1,–1); SS = 55.17; p-value $\approx 10^{-8}$. This hypothesis is soundly rejected.

- H_O: The effect of positivity or negativity of the editorial does depend on whether the transcript is available; coefficients (0,1,–1,–1,1); SS = 1.10; p-value $\approx .4$. This hypothesis is not rejected.

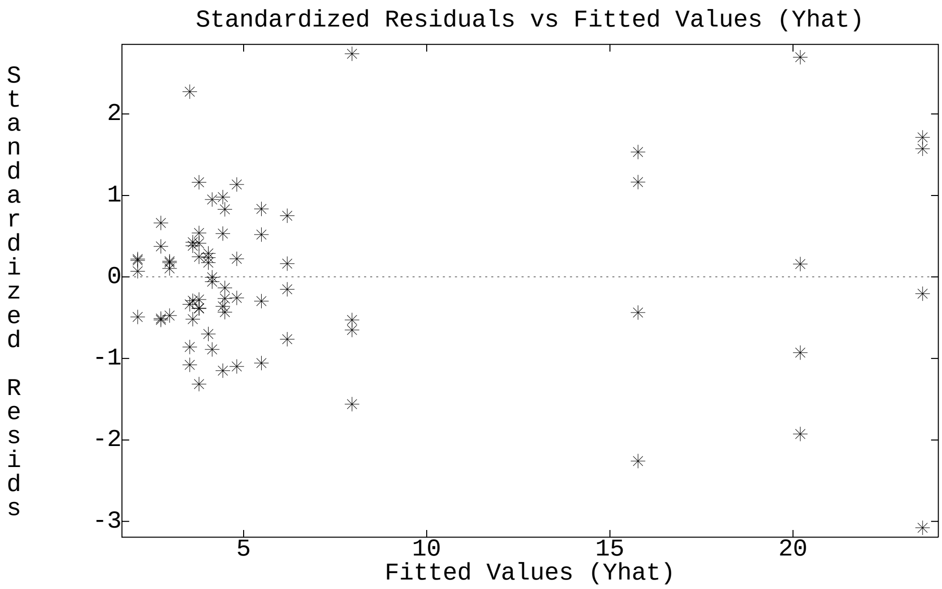
- H_O: the ad alone has the same average response as when a positive editorial is present; coefficients (1,–.5,0,–.5,0); SS = 1.42; p-value $\approx .3$. This hypothesis is not rejected.

- H_O: the ad alone has the same average response as when a negative editorial is present; coefficients (1,0,–.5,0,–.5); SS = 24.11; p-value $\approx .0001$. This hypothesis is soundly rejected.

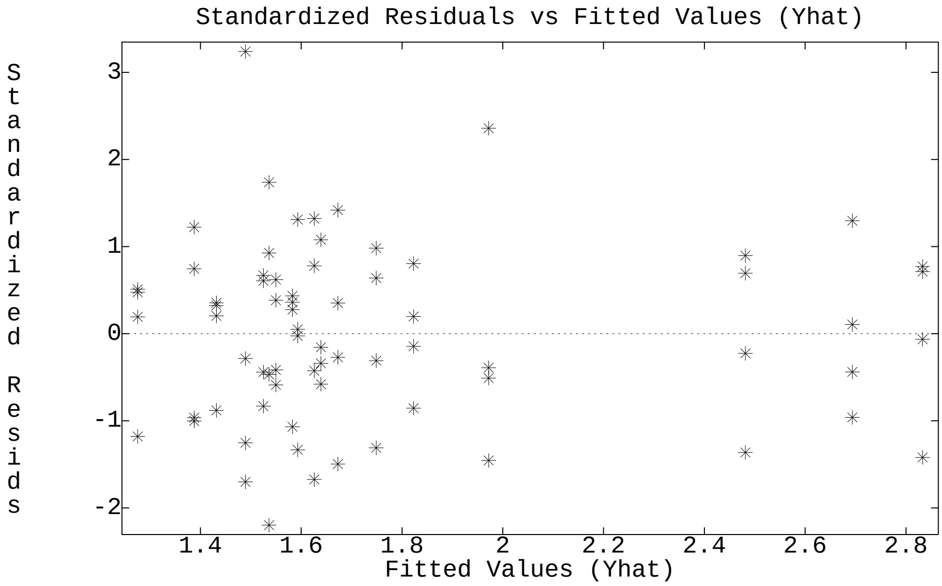
Overall, voters have the same basic response to the ad alone and the positive feedback on the ad; however, voters believe that the ad is negative if they are told that it is negative.

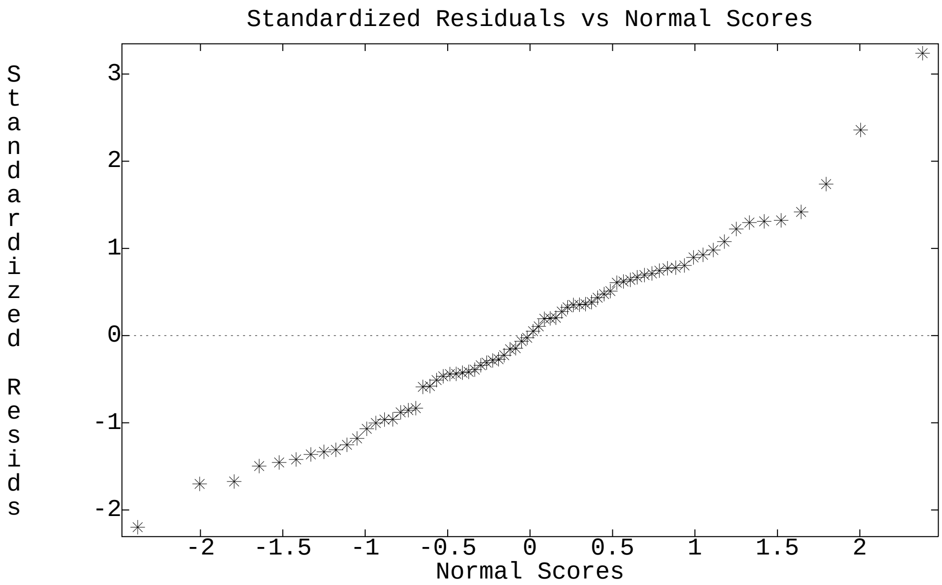
2. Summary. When pH is high and CaCl is low, time is high and decreases linearly with solids. Otherwise, times are lower and increase linearly with solids. The linear effect of pH varies with CaCl, increasing faster at low CaCl, but also beginning at a lower value at low CaCl.

Data analysis. This experiment is a completely randomized design with 72 observations assigned to 18 treatments. The treatments are the factor/level combinations of ph (5.6, 6, 6.6), percent solids (12, 15, 18), and CaCl (0, or .1%). Residuals from a preliminary ANOVA show strong nonconstant variance



Box-Cox analysis suggests a cube root transformation, and the residuals look better on this scale, although they are a little skewed to the right:



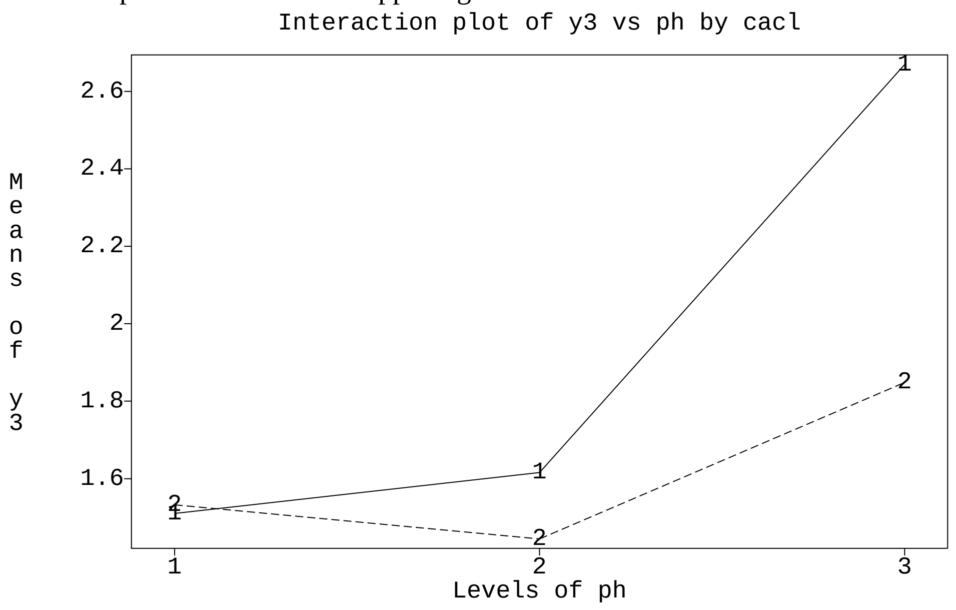


Here is the ANOVA on the cube-root scale

	DF	SS	MS	F	P-value
CONSTANT	1	225.59	225.59	16338.68229	< 1e-08
solids	2	0.13545	0.067725	4.90500	0.011032
ph	2	8.5955	4.2977	311.26566	< 1e-08
solids.ph	4	0.18489	0.046222	3.34767	0.016078
cacl	1	1.8839	1.8839	136.44376	< 1e-08
solids.cacl	2	0.19553	0.097764	7.08058	0.0018584
ph.cacl	2	2.3412	1.1706	84.78064	< 1e-08
solids.ph.cacl	4	0.23682	0.059204	4.28788	0.0043774
ERROR1	54	0.74559	0.013807		

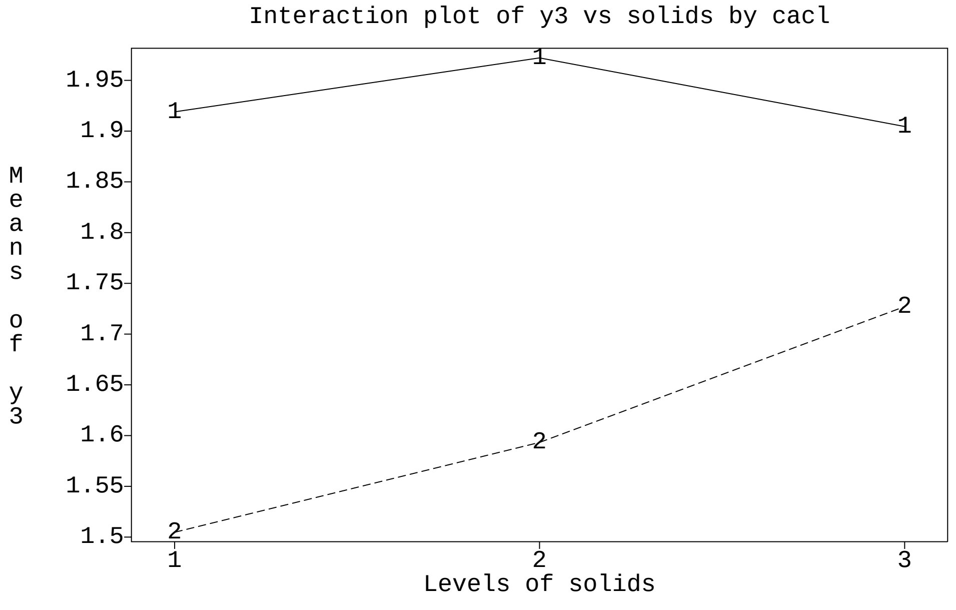
All factors and interactions are significant, although those involving solids tend to be somewhat smaller (less significant).

Almost 95% of the model SS is due to the pH and CaCl factors and their interaction. The following interaction plot reveals what is happening:

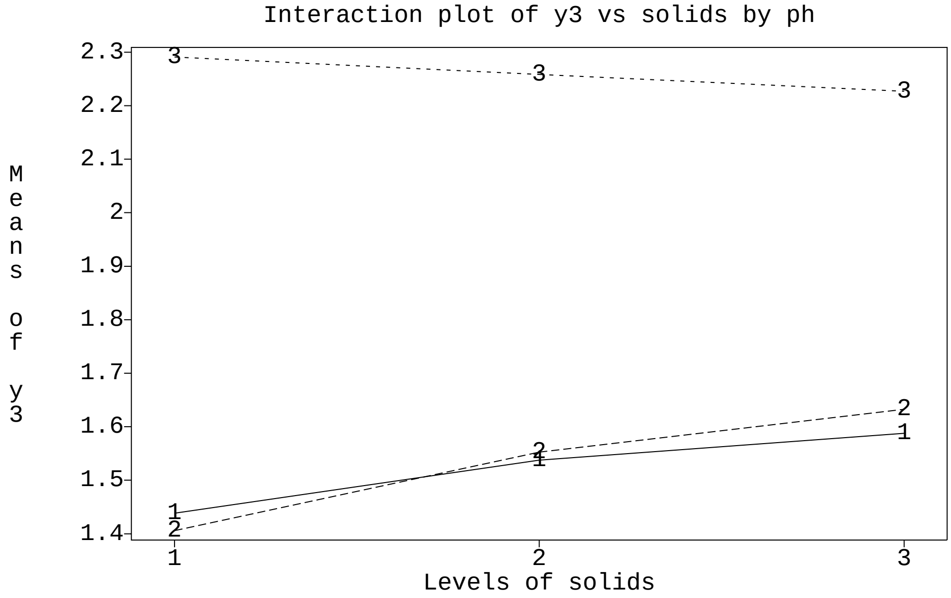


CaCl has no effect for small pH, has opposite effects at intermediate pH, and has very different (though both positive) effects at high pH. The response to pH is curved, but the curvature differs between levels of CaCl.

The other two factor interactions are also fairly obvious in the plots. In the next plot, we see that there is little effect of solids when CaCl is at level 1, but an increasing effect when CaCl is at level 2:



In this plot, we see that solids have an increasing effect at pH levels 1 and 2, but a slight decreasing effect at pH level 3.



Both solids and pH are quantitative, so we may use polynomial modeling. In fact, the interaction plots above suggest that linear terms alone may be adequate for many terms involving solids. Here is a complete ANOVA

	DF	SS	MS	F	P-value
CONSTANT	1	225.59	225.59	16338.68229	< 1e-08
{xph}	1	7.3043	7.3043	529.01892	< 1e-08
{(xph)^2}	1	1.2912	1.2912	93.51241	< 1e-08
{xsol}	1	0.12969	0.12969	9.39297	0.0033966
{(xsol)^2}	1	0.0057582	0.0057582	0.41704	0.52115
{xph*xsol}	1	0.11197	0.11197	8.10924	0.0062168
{xph^2*xsol}	1	0.06967	0.06967	5.04586	0.028797
{xph*xsol^2}	1	0.0020179	0.0020179	0.14615	0.70374
{xph^2*xsol^2}	1	0.0012348	0.0012348	0.08943	0.76605
cacl	1	1.8839	1.8839	136.44376	< 1e-08
{xph}.cacl	1	2.2598	2.2598	163.66791	< 1e-08
{(xph)^2}.cacl	1	0.081371	0.081371	5.89337	0.018562
{xsol}.cacl	1	0.16799	0.16799	12.16666	0.00097576
{(xsol)^2}.cacl	1	0.027539	0.027539	1.99451	0.16361
{xph*xsol}.cacl	1	0.23116	0.23116	16.74210	0.0001438
{xph^2*xsol}.cacl	1	0.0047605	0.0047605	0.34478	0.55953
{xph*xsol^2}.cacl	1	1.2049e-06	1.2049e-06	0.00009	0.99258
{xph^2*xsol^2}.cacl	1	0.0008911	0.0008911	0.06454	0.80043
ERROR1	54	0.74559	0.013807		

No quadratic terms involving solids are significant, and the only significant three factor term is linear by linear by CaCl. The reduced model ANOVA is

	DF	SS	MS	F	P-value
CONSTANT	1	225.59	225.59	17467.93125	< 1e-08
{xph}	1	7.3043	7.3043	565.58209	< 1e-08

{(xph)^2}	1	1.2912	1.2912	99.97553	< 1e-08
xsol	1	0.12969	0.12969	10.04216	0.0023919
{xph*xsol}	1	0.11197	0.11197	8.66971	0.0045731
{xph^2*xsol}	1	0.06967	0.06967	5.39461	0.023551
cacl	1	1.8839	1.8839	145.87408	< 1e-08
{xph}.cacl	1	2.2598	2.2598	174.97983	< 1e-08
{(xph)^2}.cacl	1	0.081371	0.081371	6.30070	0.014738
xsol.cacl	1	0.16799	0.16799	13.00755	0.00062628
{xph*xsol}.cacl	1	0.23116	0.23116	17.89923	7.9515e-05
ERROR1	61	0.7878	0.012915		

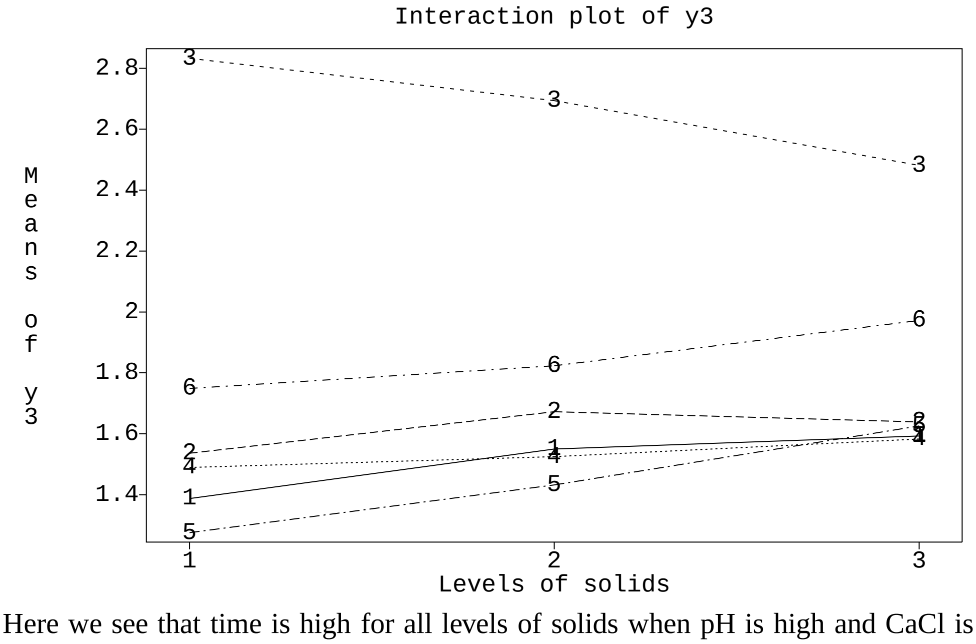
On way to understand this model is that time is a polynomial function of pH and solids, but a different polynomial for the two CaCl levels. In fact, only the coefficient for pH squared time solids is the same for the two levels of CaCl. For the first level of CaCl, the model is

$$y = 104.56 - 36.326\text{pH} + 3.1855\text{pH}^2 - 3.6302\text{solids} + 1.2866\text{pH.solids} - .11299\text{pH}^2\text{.solids}$$

whereas for the second level of CaCl, the model is

$$y = 96.852 - 31.559\text{pH} + 2.5873\text{pH}^2 - 4.2737\text{solids} + 1.3992\text{pH.solids} - .11299\text{pH}^2\text{.solids}$$

A three-factor interaction plot allows us another interesting view



Here we see that time is high for all levels of solids when pH is high and CaCl is low, and that the slope in that case is very negative, whereas the slopes in the other cases are more consistent and positive. This suggests that the pH high, CaCl low data behave very differently.

Let's fit a model that gives a separate slope and intercept for the anomolous data:

	DF	SS	MS	F	P-value
CONSTANT	1	225.59	225.59	16338.68229	0
p3c1	1	11.643	11.643	843.21523	0
xsp3c1	1	0.24705	0.24705	17.89295	9.0969e-05
solids	2	0.38618	0.19309	13.98468	1.2766e-05

ph	2	0.99847	0.49924	36.15742	1.0848e-10
solids.ph	4	0.017147	0.0042869	0.31048	0.86974
cacl	1	0.066976	0.066976	4.85077	0.031923
solids.cacl	2	0.036707	0.018353	1.32926	0.27319
ph.cacl	1	0.11261	0.11261	8.15553	0.0060807
solids.ph.cacl	3	0.065593	0.021864	1.58354	0.204
ERROR1	54	0.74559	0.013807		

If we fit the unusual data separately, then solids does not interact with any other factor. Fitting the significant terms, we get

	DF	SS	MS	F	P-value
CONSTANT	1	225.59	225.59	16580.10811	0
p3c1	1	11.643	11.643	855.67486	0
xsp3c1	1	0.24705	0.24705	18.15734	6.8118e-05
xsol	1	0.38042	0.38042	27.95944	1.601e-06
{xph}	1	0.78677	0.78677	57.82418	1.6078e-10
{(xph)^2}	1	0.2117	0.2117	15.55921	0.00020106
cacl	1	0.066976	0.066976	4.92245	0.030064
cacl.xph	1	0.11261	0.11261	8.27604	0.005452
ERROR1	64	0.8708	0.013606		

The coefficients for this model are:

```
Cmd> coefs()
component: p3c1
(1)       1.7252
component: xsp3c1
(1)    -0.091085
component: xsol
(1)     0.032507
component: cacl
(1)       29.524       32.259
component: cacl.xph
(1,1)      -10.087
(2,1)      -10.571
component: {xph^2}
(1)      0.89238
```

For the unusual data (pH=6.6, CaCl low), the equation is

$$(1.7252 + 29.524 - 10.87 \times 6.6 + .89238 \times 6.6^2) + (.032507 - .091085)\text{solids} = 3.55 - .0586\text{solids}$$

For the data with CaCl low and pH at 5.6 or 6, the equation is

$$29.524 + .0325\text{solids} - 10.087\text{pH} + .8924\text{pH}^2$$

For the data with CaCl high, the equation is

$$32.259 + .0325\text{solids} - 10.571\text{pH} + .8924\text{pH}^2$$