Case Study -- Maldi Time-of-Flight Spectroscopy Data

An "exact" analysis of data from a saturated model cannot be performed, with a saturated model being one for which the number of observations is equal to the number of parameters in the model. This is the case when fractional factorial designs are used (unless the design were replicated, which would offset the advantage of economy that one has in using a fractional factorial).

One way to avoid this problem is to use centerpoints, with a centerpoint defined, as the name implies, as a design point that is in the center of the design. Such a point has all coordinates equal to zero in the coded units, with the coded units for a design for which each factor has two levels being: (raw value minus the average of the raw values)/(half of the range of the raw values).

Centerpoints allow the user to check for lack of fit of the model and to obtain an estimate of the variance of the error term in the model, and they can also be used as a check to see if processes are in a state of statistical control, as is required but not always recognized or checked. When used to estimate the model variance, the analyst assumes that the true variance at the centerpoints is the same as the true variance at each of the design points. That is a rather strong assumption.

Generally only a small number of centerpoints are used, however, so the estimated variance of the error term will have considerable sampling variability. This is discussed further in Section A.

The NIST Maldi Time-of-Flight Spectroscopy dataset of Charles Guttman and Stephanie Wetzel of the NIST Polymers Division can be used to illustrate the type of dilemma that an experimenter may sometimes encounter. The dataset has five factors and three response variables: MEAN, RESOLUTION and SIGNAL-TO-NOISE ratio (S/N).

The data are given below, with each of the five factors given in the coded units described previously, with the subsequent analysis and interpretation to be performed in the coded units.

Row	Mean	Resolution	S/N	DetVoltage	Power	Time	IS/2Voltage	LensVoltage
1	8884.48	767.8	32.6	-1	-1	-1	-1	1
2	8474.13	467.5	673.3	1	1	-1	1	-1
3	8898.94	721.7	28.8	-1	-1	1	1	1
4	8768.65	842.1	825.6	1	1	1	1	1
5	9019.54	496.0	116.6	-1	1	-1	-1	-1
6	8640.24	964.6	321.4	1	-1	-1	1	1
7	8832.85	687.4	232.2	1	-1	1	-1	1
8	8847.35	886.9	140.2	-1	1	1	1	1
9	8846.33	507.7	55.3	-1	1	1	-1	1
10	8799.66	977.7	506.6	1	-1	1	1	-1
11	8827.30	388.4	144.9	-1	1	-1	1	1
12	8976.55	698.2	46.7	-1	-1	-1	1	-1
13	8711.70	695.7	566.6	1	1	1	-1	-1
14	8885.08	577.7	30.6	-1	-1	1	-1	-1
15	8552.11	472.4	285.8	1	1	-1	-1	1
16	8974.98	967.6	316.8	1	-1	-1	-1	-1
17	8807.71	1131.9	741.6	0	0	0	0	0
18	8821.29	975.9	505.8	0	0	0	0	0
19	8802.38	1159.9	573.9	0	0	0	0	0

Table 1. Data and the Factors in the Experiment

The run sequence is almost the same as the order in which the treatment combinations are listed, except that the first centerpoint was the first run, the second centerpoint was the 10th run, and the last centerpoint was the last run.

It is imperative that processes be maintained in a state of statistical control when experiments are performed, and indeed this is a tacit assumption. It seems apparent that the centerpoints were probably used to check this assumption as the first centerpoint was the first experimental run, the last centerpoint was the last experimental run and the other centerpoint was the 10th run and was thus exactly in the middle of the experiment. Thus the centerpoint runs were not randomized. The other 16 runs were randomized, however, as is required in order for the hypothesis tests for the effects to be valid.

This particular placement of centerpoints, which coincides with the recommendation of, for example, Czitrom (2003), can also be used to check for time trends that can undermine the results. Slight changes in materials over time might be viewed as being natural and not signifying an out-of-control process, but those changes might have a significant effect on the results of an experiment.

The design that was used was a 2^{5-1}, a one-half fraction of a 2^5 full factorial design.   Fifteen effects can be estimated with such a design since there are 16-1 = 15 degrees of freedom and these would logically be the five main effects and the $\binom{5}{2}$ = 10 two-factor interactions. The higher-order interactions are not estimable and are confounded with the other effects if the design is properly constructed. (Two effects are said to be "confounded" if they would be estimated using the same linear combinations of the observations.)   For a properly constructed 2^{5-1} design, the main effects are confounded with four-factor interactions and the two-factor interactions are confounded with three-factor interactions. (This is known as a Resolution V design.) The single five-factor interaction is chosen to confound with the average effect of the 16 treatment combinations.

A. Analysis Using the Centerpoints

Although the objectives of the study were to determine the factors that affect the S/N and then to determine the factor settings that maximize the S/N, we will use only the first response variable, MEAN, and see if the centerpoints are helpful in identifying the significant effects, or in serving any other purpose. The analysis is given below.

Estimated Effects and Coefficients for Mean (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		8809.01	1.862	4730.64	0.000
DetVolta	-178.91	-89.45	2.029	-44.08	0.000
Power	-105.71	-52.85	2.029	-26.05	0.000
Time	30.15	15.08	2.029	7.43	0.005
IS/2Volt	-59.28	-29.64	2.029	-14.61	0.001
LensVolt	-54.76	-27.38	2.029	-13.49	0.001
DetVolta*Power	-79.58	-39.79	2.029	-19.61	0.000
DetVolta*Time	87.70	43.85	2.029	21.61	0.000
DetVolta*IS/2Volt	-37.96	-18.98	2.029	-9.35	0.003
DetVolta*LensVolt	13.10	6.55	2.029	3.23	0.048
Power*Time	45.08	22.54	2.029	11.11	0.002
Power*IS/2Volt	6.22	3.11	2.029	1.53	0.223
Power*LensVolt	40.18	20.09	2.029	9.90	0.002
Time*IS/2Volt	68.94	34.47	2.029	16.99	0.000
Time*LensVolt	80.50	40.25	2.029	19.84	0.000
IS/2Volt*LensVolt	64.12	32.06	2.029	15.80	0.001

Analysis of Variance for Mean (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	202416	202416	40483.2	614.48	0.000
2-Way Interactions	10	138666	138666	13866.6	210.48	0.000
Residual Error	3	198	198	65.9		
Curvature	1	7	7	7.5	0.08	0.806
Pure Error	2	190	190	95.1		
Total	18	341280				

The most striking result is that 14 of the 15 estimable effects are declared significant by the *t*-tests. This is due to the fact that there is very little difference between the centerpoint values, relative to differences between the other 16 points. The variance of $\widehat{\sigma}^2$ is $2\sigma^4/3$, assuming a normal distribution for the observations, and in this case $\widehat{\sigma}^2$ = 65.9 since the lack of fit is not significant. Even if we used 65.9 as an estimate of σ^2, we obtain 2895.21 as an estimate of the variance of the estimator, and $\sqrt{2895.21}$ = 53.81. Thus, the standard error of the estimator is of the same order of magnitude as the estimator, which results from the fact that the standard error is $\sqrt{2/3}\,\widehat{\sigma}^2 = 0.82\widehat{\sigma}^2$.

The small variability between the centerpoint values would seem to suggest that the processes were in control, but the small variability between the centerpoint observations relative to the much larger variability among most of the other observations does look somewhat suspicious. Possibly the relevant processes could be out of control in such a way that the observations differ by much less than they would differ if the processes were in control. Although such a scenario would seem unlikely, the analysis of the effects given in the next section does raise questions relative to process stability.

Although the investigators have not suggested that any of the centerpoints are not valid observations, it would clearly be unwise to assume that the variability in the observations on the MEAN response at those centerpoints is a good measure of variability and to conclude that almost all of the effects are real. Even if the factors that are used in an experiment are well-chosen, we would generally not expect more than half the effects to be significant, as we would usually expect that not very many two-factor interactions would be significant.

Therefore, for the response MEAN we need a different approach.

B. Analysis *Without* the Centerpoints

We will analyze the data without the centerpoints, which requires that a different method be used to assess significance of the effects since formal hypothesis tests cannot be performed. Perhaps the most frequently used method is the normal probability plot method due to Lenth (1989), which, as with the other methods, uses a pseudo-error term in assessing effect significance. The pseudo-error term with Lenth's method is computed in the following manner. Let s_0 = 1.5 x median $|c_j|$ with c_j

denoting the estimate of the jth estimable effect. Then the pseudo standard error (PSE) is computed as

$$\text{PSE} = 1.5 \; x \;\; \underset{|c_j| < 2.5 s_0}{\text{median}} \; |c_j|$$

with the PSE essentially computed from a trimmed median of $|c_j|$ values, as the median is computed using only values of $|c_j|$ that are less than $2.5s_0$.

The results are given in the following graph, with A-E denoting the factors in Table 1.

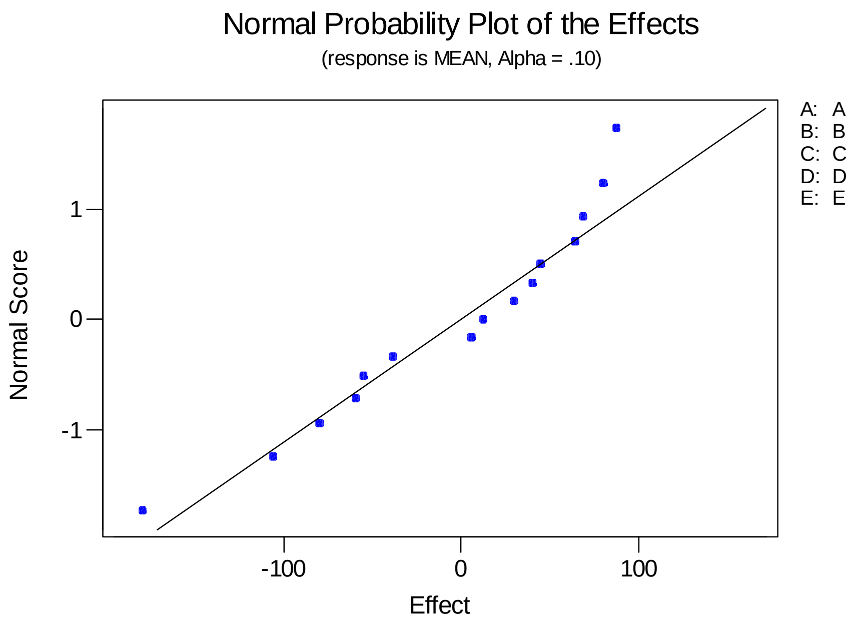

Figure 1. Normal probability plot of effect estimates using Lenth's method

No effects are shown as being significant, even though most of the effects are large. Lenth's method is effective when there is effect sparsity, but fails, as do all of the other approximate methods, when the vast majority of the effects are real. If most of the effects are not real, then the median of the $|c_j|$ values won't be small, which will result in the PSE value being large.

As the table below shows, only two effect estimates could be considered small relative to the other effect estimates.

Term	Effect	Coef
Constant		8808.74
A	-178.91	-89.45
B	-105.71	-52.85
C	30.15	15.08
D	-59.28	-29.64
E	-54.76	-27.38
A*B	-79.58	-39.79
A*C	87.70	43.85
A*D	-37.96	-18.98
A*E	13.10	6.55
B*C	45.08	22.54
B*D	6.22	3.11
B*E	40.18	20.09
C*D	68.94	34.47
C*E	80.50	40.25
D*E	64.12	32.06

Table 2. Effect Estimates and Corresponding Model Coefficients

Even though two of the effect estimates exceed 100 in absolute value, the fact that there are not enough small effect estimates causes even the A and B effects to not be judged significant. The problem is that the median of the $|c_j|$ is 59.28, which is a large value.

So we encounter the two extremes with the two methods of analysis: almost everything judged significant with the first method and nothing judged significant with the second method. Since the average MEAN value is 8,809, the effect estimates are not large relative to the order of magnitude of the response values.

Nevertheless, we would not expect to observe so many interaction effects that are large relative to the corresponding main effects. For example, the C *x* E interaction is much larger in absolute value that the *C* effect and the *E* effect. Similarly, the C *x* D and D *x* E interactions are larger than the corresponding main effects

This state of affairs should cause us to question whether there are any bad data points. The first point we should question is the smallest observation, 8474.13, because it stands out when we look at the scatter plot of MEAN against LENS VOLTAGE, which is given below.

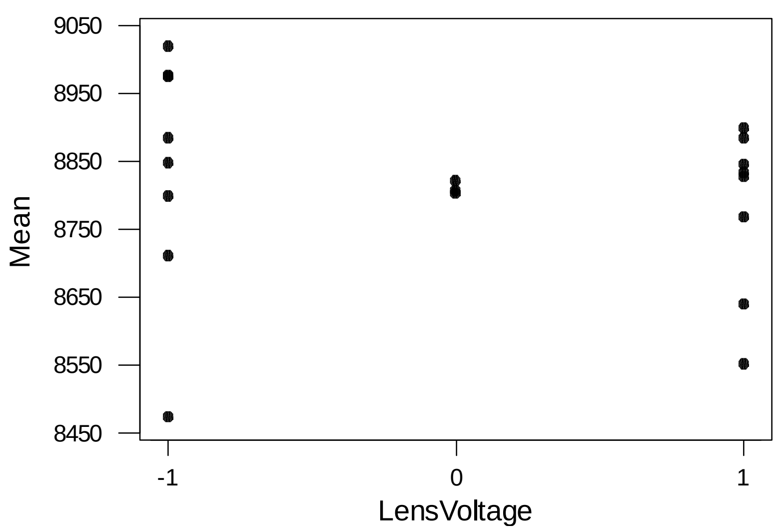

Figure 2. Scatterplot of Mean against Lens Voltage

Notice that the smallest observation is much smaller than the other observations at the low level of the factor. Perhaps that is a valid observation, but it does look somewhat suspicious. Therefore, we might as well run the analysis without the data point (and also of course without the three centerpoints since we also view those as being suspicious because of the small variability between them).

Of course when we delete an observation we lose the orthogonality of the design, which is why we would prefer not to delete observations. By deleting that observation we create a few non-zero correlations, with all of the pairwise correlations between the factors being either plus or minus .071. The Lenth plot is given below.

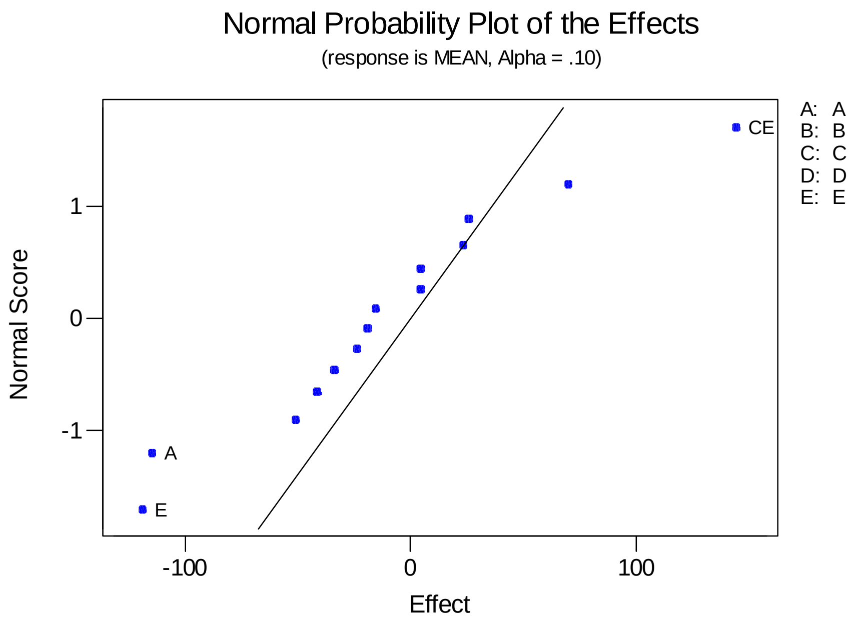

Figure 3. Normal probability plot of effect estimates with smallest observation deleted

At least this time we obtain a somewhat reasonable number of significant effects, although there is a need to look at conditional effects (see case study #2) because the CE interaction is declared significant without factor C being significant. We may also note that the model with only the terms declared significant in the plot would be a non-hierarchical model, and such models are somewhat controversial.

Whether or not this analysis is "right" depends on whether or not we have deleted a bad data point. Clearly we cannot delete data points in an effort to obtain results that look reasonable, however. What we can do, though, in the absence of feedback from the experimenters regarding possible bad data points is to delete points purely on the basis of graphical and numerical evidence and then present the conclusions to the experimenters and indicate which observations were not used. It may be that the experimenters will declare the results regarding significant effects to be in accordance with their intuition, but will still not be able to provide any insight regarding possible bad data points. If so, the result with the deleted points should probably stand, despite the absence of evidence that would support the deletion of particular observations.

C. SUMMARY

It is generally believed that almost all datasets have errors, but the experimenters whose work produced the dataset analyzed herein were not aware of any bad datapoints. This creates a major dilemma regarding the analysis of the data. The small amount of variability of the centerpoints looks a bit suspicious, so estimating σ from the centerpoints would seem to be unwise. Not using the centerpoints and performing an analysis using only the factorial part of the design also creates a problem as none of the effects are judged significant using Lenth's method. (It is likely that other methods would produce comparable results.) If we guess that the smallest observation is a bad data point -- and certainly the scatterplot of MEAN against LENS VOLTAGE would make us suspicious -- then we obtain a reasonable number of significant effects, although perhaps slightly on the low side with two main effects significant out of five and 4 out of 15 overall.

Of course we cannot discard points simply to make the results look reasonable, so it would be best to present all three analyses to the experimenters, along with questions to them that are motivated by the analyses. This might seem to be a very unusual dataset, but reality often does not conform to what we find in textbooks.

REFERENCES

Czitrom, V. (2003). Guidelines for selecting factors and factor levels. in *Handbook of Statistics 22* (R. Khattree and C. R. Rao, eds.), Amsterdam: Elsevier Science B.V.

Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics*, 31, 469-473.