

Case Study -- Lead Recovery Data

Interest in the amount of lead in household paint resulted in research performed by NIST in 2001 that was sponsored by the U.S. Department of Housing and Urban Development (HUD). The determination of the amount of lead is often done using field-portable ultrasonic extraction-anodic stripping voltammetry (UE/ASV). Several studies have been performed during the past 10 years to assess the measurement reliability of field-portable UE/ASV, and this case study will examine the data from one such study.

In that study, there was interest in determining the factors that affect “lead recovery”, expressed as a percentage of the known amount of lead in 112 paint specimens. Sonicators were used for the ultrasonic lead extraction, with sonicator power (low and high) being one of five factors, each at two levels, that were used in the experiment. The other four factors were sonication temperature, sonication time, specimen mass, and specimen particle size. The response variable, as stated, was the lead recovery percentage.

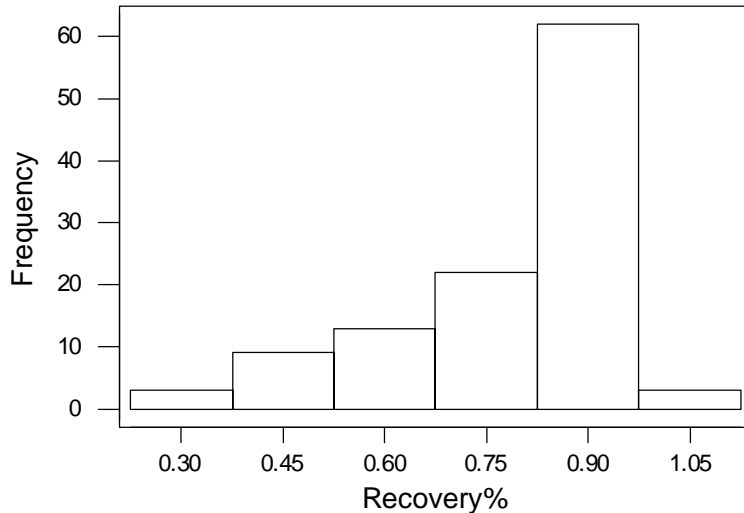
The factors will be denoted by the letters *A*, *B*, *C*, *D*, and *E*, with *A* = Sonicator (power), *B* = Temperature, *C* = Time, *D* = Mass, and *E* = Size.

It is worth noting that although there were two particle sizes (large and small) specified for the experiment, specimen particle size is essentially a random variable, as it is apparently not possible to grind a specimen so that the particle size is exactly equal to a nominal value. Thus, specimen particle size really wasn't fixed (rather, it was more or less dichotomized), nor was specimen mass. This necessitates alternative methods of analysis and a comparison of the results obtained using each analysis method. This will be discussed further in subsequent sections.

Seven specimens could be handled by the sonicator on a given run, so for reasons of efficiency this number was used per run. This does create a potential problem from a statistical standpoint, however, as $2^5 = 32$ and 7 is not a factor of 32. Consequently, the experiment was designed by splitting the 32 runs into two halves, with one half replicated 3 times and the other half replicated 4 times, producing 112 observations to correspond to the 112 specimens. This unbalanced nature of the design does not create any problems as long as the pair of 16 runs is chosen the way that we would normally split a 2^5 , with one of the splits used as a 2^{5-1} design, and as long as we don't estimate interactions of order 2 (i.e., 3 factors) and higher. If, however, each set of 16 is selected in such a way as to inadvertently cause a correlation between two or more factors in each set of 16, then the design will be nonorthogonal when the two sets are joined together because of the unequal number of replicates and there will be non-zero correlations involving main effects and two-factor interactions. Indeed this is what happened, but the correlations for terms that are of interest are so small (all $\pm .071$) as to be inconsequential. Therefore, we will not be concerned with this very slight departure from an orthogonal design.

I Transforming the Response Variable

Since we have an experiment with replicates, we can perform an analysis using Analysis of Variance (ANOVA). Such an analysis is based on the assumption of a normal distribution for the error terms in a ANOVA model, which translates into an assumption of normality for the response variable. Percentages will not be normally distributed, however, and in fact will be highly skewed, as can be seen below for the 112 observations.

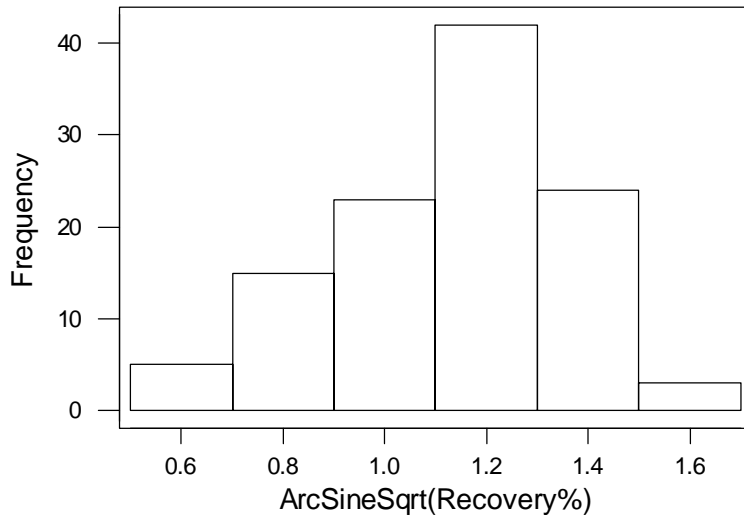


In addition to the non-normality problem, we could expect to have a problem with the error terms not having a constant variance. More explicitly, the sample variance computed from the replicates for each of the 32 combinations can be expected to differ considerably.

Even though the binomial distribution doesn't strictly apply here because we do not have Bernoulli trials, etc., we can still use that distribution as a reference point in guiding our analysis.

Accordingly, it would be reasonable to transform the data using the arcsin (\sqrt{x}) transformation in an effort to effect both approximate normality and homogeneity of variance. (Using this transformation does create a minor problem, as measurement error caused two of the values to be recorded as 100% plus a fraction, and the transformation is undefined for those values. So 99.999% was used instead for those two percentages.)

The histogram for the transformed values after the transformation, using the same number of classes as used in the preceding histogram, is shown below. Although some skewness remains, the improvement is obvious. Furthermore, we should not expect to see perfect symmetry in a histogram any more than we should expect to encounter a data set that is exactly normally distributed, especially with a small number of classes. For the moment we will not be overly concerned with this degree of asymmetry.



It isn't worthwhile to test for nonhomogeneity of variance when there are only a few observations per treatment combination and when the number of observations is not constant over the treatment combination.

We might still investigate the within-cell variability, however, at least as part of an exploratory data analysis. From such an investigation there is one unusual observation that stands out, as one cell has the following three values for lead recovery percentage: 30.8, 76.2, and 79.3. If we viewed these data with an eye

toward a “2 out of 3 decision”, then we would conclude that the 30.8 should be investigated. Certainly there is no other cell for which one observation differs from the other two observations by anything even close to this difference. We will see if this observation that stands out in a univariate analysis will also appear aberrant in subsequent analyses involving multiple variables/dimensions.

II Analysis

Since interactions involving more than three factors rarely exist in the physical world, we might begin the analysis by fitting a hierarchical model that includes all three-factor interactions, with the response variable being the transformed response given in the previous section. The initial output is given below

Estimated Effects and Coefficients for arcsinesqrt(recovery%)(coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		1.1416	0.01013	112.74	0.000
A	0.0196	0.0098	0.01007	0.97	0.333
B	0.1297	0.0648	0.01015	6.39	0.000
C	0.0899	0.0450	0.01013	4.44	0.000
D	-0.0156	-0.0078	0.01015	-0.77	0.444
E	-0.2992	-0.1496	0.01013	-14.77	0.000
AB	-0.0105	-0.0053	0.01013	-0.52	0.605
AC	0.0106	0.0053	0.01007	0.53	0.600
AD	-0.0127	-0.0064	0.01013	-0.63	0.531
AE	0.0149	0.0075	0.01007	0.74	0.460
BC	-0.0197	-0.0098	0.01015	-0.97	0.335
BD	0.0078	0.0039	0.01015	0.38	0.702
BE	0.1544	0.0772	0.01015	7.60	0.000
CD	0.0403	0.0202	0.01015	1.99	0.050
CE	0.0841	0.0421	0.01013	4.15	0.000
DE	0.0403	0.0202	0.01015	1.99	0.050
ABC	-0.0066	-0.0033	0.01013	-0.33	0.744
ABD	0.0185	0.0092	0.01015	0.91	0.365
ABE	0.0456	0.0228	0.01013	2.25	0.027
ACD	-0.0246	-0.0123	0.01013	-1.21	0.229
ACE	-0.0200	-0.0100	0.01007	-0.99	0.324
ADE	0.0126	0.0063	0.01013	0.62	0.535

BCD	-0.0105	-0.0052	0.01015	-0.52	0.607
BCE	0.0177	0.0089	0.01015	0.87	0.385
BDE	-0.0215	-0.0107	0.01015	-1.06	0.294
CDE	0.0080	0.0040	0.01015	0.40	0.693

Analysis of Variance for arcsinesqrt(recovery%) (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	3.17905	3.18758	0.637517	56.39	0.000
2-Way Interactions	10	0.93123	0.92891	0.092891	8.22	0.000
3-Way Interactions	10	0.12967	0.12967	0.012967	1.15	0.338
Residual Error	86	0.97235	0.97235	0.011306		
Lack of Fit	6	0.05128	0.05128	0.008547	0.74	0.617
Pure Error	80	0.92107	0.92107	0.011513		
Total	111	5.21230				

Unusual Observations for arcsinesqrt(recovery%)

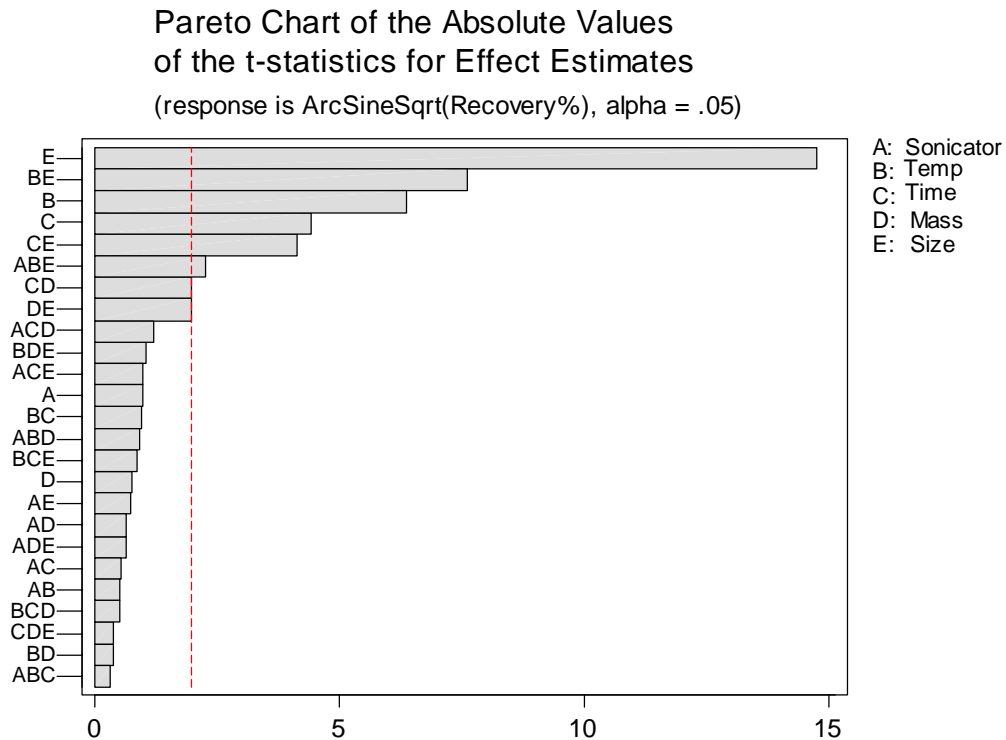
Obs	asinsqrt(Rec%)	Fit	SE Fit	Residual	St Resid
5	1.57080	1.32450	0.05458	0.24630	2.70R
8	1.12876	1.32450	0.05458	-0.19574	-2.14R
29	1.57080	1.36541	0.04859	0.20539	2.17R
47	0.98030	1.22079	0.05458	-0.24049	-2.64R
48	1.02778	1.26347	0.04859	-0.23569	-2.49R
75	1.09845	0.90928	0.05458	0.18917	2.07R
79	0.58834	0.90928	0.05458	-0.32095	-3.52R
97	1.01766	1.27232	0.04859	-0.25466	-2.69R

R denotes an observation with a large standardized residual

When we look at the list of “unusual observations” (defined as the standardized residual exceeding 2 in absolute value), we see that observation #79 stands out. This is the observation with a recovery percentage of 30.8 that was pointed out in the preceding section.

Rather than look at the list of estimated effects and try to determine which effects are significant, it would be better to use a Pareto chart of standardized effect estimates (i.e., *t*-statistics) as a visual aid.

Given below is the Pareto chart, with the dotted line serving as a threshold value (obtained from using a significance level of .05), with bars that extend to the right of the line indicating significance.



We observe that a 3-factor interaction is significant, but since it is very close to the dotted line, whether the interaction is declared significant or not will depend upon what other terms are in the model. We would expect a moderate number of effects for a 2^5 design. Daniel (1976, p. 75) estimates that four significant effects is about average for a 2^4 design, so we might expect about eight for a 2^5 design because we have almost twice as many effects to estimate with the 2^5 design. Of course the number of effects that are significant will depend on how well the factors in the experiment are selected.

Because three effects, the *ABE*, *CD*, and *DE* interactions, are borderline in terms of significance, we need to determine if they are significant when effects that are clearly not significant are not included in the model.

We might make this determination by using a variable selection approach such as stepwise regression, using, say, the 10 largest effects from the Pareto chart as candidate terms. Doing so results in the following effects being selected in the order indicated, using the letter designation as in the Pareto chart: *E*, *BE*, *B*, *C*, *CE*, *ABE*, *CD*, and *DE*. This is also the model that stands out in terms of the C_p statistic when all possible subsets of models are examined. The R^2 value is 79.18, so the model explains 79.18% of the variability in the lead recovery percentage values. If we wanted to know what this is on the original scale, we would have to convert the fitted values back to the original scale and then compute the square of the correlation between those values and the response values on the original scale.

Other statistics for the model with the selected terms are given below and it can be observed that all effects are significant at the .05 level. (The slight lack of orthogonality of the design is reflected in the fact that the standard errors are not all the same, as they are when an orthogonal design is used.)

Estimated Effects and Coefficients for arcsinesqrt(recovery%) (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		1.1409	0.009699	117.63	0.000
B	0.1304	0.0652	0.009749	6.69	0.000
C	0.0916	0.0458	0.009699	4.72	0.000
E	-0.2964	-0.1482	0.009699	-15.28	0.000
BE	0.1529	0.0764	0.009749	7.84	0.000
CE	0.0843	0.0422	0.009699	4.35	0.000
ABE	0.0455	0.0228	0.009699	2.35	0.021
CD	0.0430	0.0215	0.009749	2.20	0.030
DE	0.0403	0.0201	0.009749	2.07	0.041

Thus, we identify eight effects as being real, which is about what we would expect. Since five of these effects are interactions, we are presented with a challenge in trying to interpret the data. We note that factor *A* is a component of the 3-factor interaction but does not appear in the model as either a main effect or in a two-factor interaction. Thus, if we use this model we will be using a non-hierarchical model.

The model does not violate the principle of *effect heredity*, however, which was introduced by Hamada and Wu (1992). This principle, which is somewhat the reverse of the principle of hierarchical modeling, states that a two-factor interaction should be included in a model only if the interaction contains at least one factor identified as having a significant main effect. Lin (1998-99) challenged this concept, noting that Box and Draper (1987) contains numerous real data sets that violate the principle of effect heredity. (Such datasets should be analyzed using conditional effects, as the violation could be caused by large interactions.)

We will first focus attention on factor *D*, which appears in two of the two-factor interactions, but does not appear in the model as a main effect. If we add factor *D* to the model that we have selected, we obtain the results given below.

Estimated Effects and Coefficients for arcsinesqrt(recovery%) (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		1.1409	0.009716	117.42	0.000
B	0.1293	0.0646	0.009792	6.60	0.000
C	0.0916	0.0458	0.009716	4.72	0.000
D	-0.0155	-0.0077	0.009767	-0.79	0.430
E	-0.2964	-0.1482	0.009716	-15.25	0.000
BE	0.1540	0.0770	0.009792	7.86	0.000
CE	0.0843	0.0422	0.009716	4.34	0.000
ABE	0.0455	0.0228	0.009716	2.34	0.021
CD	0.0430	0.0215	0.009766	2.20	0.030
DE	0.0404	0.0202	0.009767	2.07	0.041

When there are large interactions (“large” means an interaction effect in which, for a two-factor interaction, the interaction effect is at least 1/3 or so of the smaller of the two main effects), it is necessary to look at conditional effects. That is, since the CD interaction is the larger of the two two-factor interactions involving factor D , we should look at the effect of factor D at each level of factor C . It is known that these conditional effects are obtained as $D \pm CD$, so the conditional effects of D are $-0.0155 \pm 0.0430 = -0.0585$ and 0.0275 . Notice that the first number is larger in absolute value than three of the interactions that are judged to be significant. These numbers are not directly comparable, however, because the standard error of a conditional effect is larger than the standard error of an unconditional effect. Therefore, a conditional effect would have to be divided by $\sqrt{2}$ so as to make a conditional effect comparable to an unconditional effect in two-level designs. Since $-0.0585/\sqrt{2} = -0.0414$, which is greater than the (significant) DE interaction, factor D is not unimportant. Rather, it simply appears to be unimportant when only unconditional effects are examined because the conditional effects are opposite in sign and add to a small number

There are various ways in which we could view the ABE interaction. One view is recognizing that the BE interaction differs for each level of A . Graphically, this means that the interaction profiles for BE will be noticeably different for each level of A . Quantitatively, the BE conditional interaction effects are $BE \pm ABE = 0.1540 \pm 0.0455 = 0.1085$ and 0.1995 .

Since the A effect was not significant, we can proceed as before and insert factor A in the model, after first removing factor D . The results are given below.

Estimated Effects and Coefficients for arcsinesqrt(recovery%) (coded units)

Term	Effect	Coef	SE Coef	T	P
------	--------	------	---------	---	---

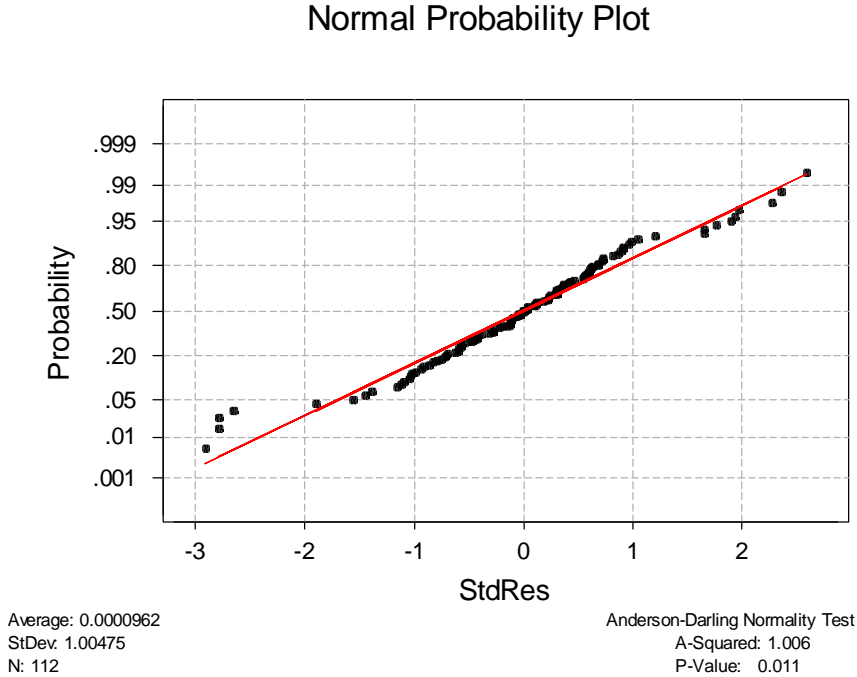
Constant		1.1409	0.009704	117.57	0.000
A	0.0183	0.0091	0.009704	0.94	0.348
B	0.1304	0.0652	0.009754	6.68	0.000
C	0.0916	0.0458	0.009704	4.72	0.000
E	-0.2964	-0.1482	0.009704	-15.27	0.000
BE	0.1529	0.0764	0.009754	7.84	0.000
CE	0.0843	0.0422	0.009704	4.35	0.000
ABE	0.0455	0.0228	0.009704	2.35	0.021
CD	0.0430	0.0215	0.009754	2.20	0.030
DE	0.0403	0.0201	0.009754	2.07	0.041

The sonicator power effect is small, but we need to look beyond the 0.0183 number because the *ABE* effect is significant. Analogous to the preceding calculations, we obtain the conditional main effects for *A* as $A \pm ABE = 0.0183 \pm 0.0455 = 0.0638$ and -0.0272 . These numbers require some explanation since *ABE* is a 3-factor interaction rather than a 2-factor interaction. The *BE* interaction is the average of two average response values, which one average computed when both factors are at the same levels, and then the other average computed when the factors are at opposite levels. If we computed the effect of *A* under each of these two scenarios, we would obtain 0.0638 and -0.0272 as the conditional effects. Of course we don't actually have to do this, but this is what underlies the arithmetic. Since $0.0638/\sqrt{2} = 0.0451$, which exceeds two of the significant unconditional effects, factor *A* is not unimportant.

III Checking Assumptions

Before we rely very heavily on the results in the preceding section, we need to check the assumptions. As stated previously, it is difficult to check the assumption of homogeneity with only 3 or 4 observations per treatment combination. We observe a problem when we look at a normal probability plot of the standardized residuals,

however, as the p -value for the Anderson-Darling test for non-normality is .011, as is shown below.



It is obvious from the graph, however, that this small p -value is caused by a few poorly fit observations, and we recall from Section I that observation #79 looked very suspicious because it differed greatly from the other two observations at the same combination of factor levels. When that observation is excluded from the analysis, the p -value is .062. We should keep in mind that with 111 or 112 observations we are going to have pretty good power in detecting even fairly small departures from normality, but such a degree of non-normality will not undermine the analysis to an appreciable extent.

(We should also keep in mind that we are never doing a true test for nonnormality when we test for nonnormality because ANOVA models do not have a single error term. Rather there is an error term for each treatment combination, and

we certainly can't test for non-normality with only 3 or 4 observations, any more than we can practically test equality of variances with such small numbers.)

We might exclude observation #79 from the analysis, but we really wouldn't want to do that unless there was evidence that something went wrong and the observation is not a good data point. The scientists involved in the experiment had no such evidence, so we will use all of the observations and simply recognize that we do have a minor problem with non-normality that is caused by a few data points. (Of course we could downweight one or more data points that look suspicious, but we will not pursue that here.)

IV Replicates or Multiple Readings?

The analysis in the preceding section presupposes that the multiple observations per treatment combination constitute true replicates, as would be the case if an experiment were repeated by “starting over” for each replicate. Since most replicated experiments are probably not performed this way --- and indeed this experiment wasn't performed this way --- the question arises as to whether or not the experiment was performed in such a way as to permit the type of analysis to be performed as if there were true replicates and appropriate randomization was performed for each replicate.

If this were not the case, it would be better to use the average value at each treatment combination, or at least to do the analysis with the averages in addition to the analysis with the replicates.

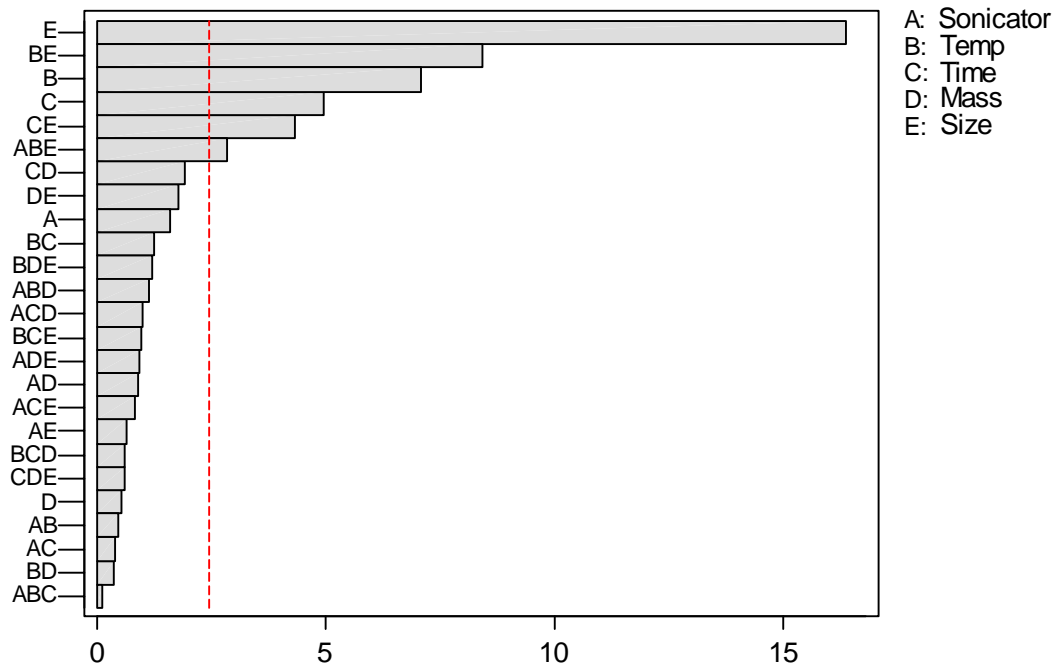
As stated previously, particle size is very much a random variable, and particle mass is also not fixed. If factors cannot be fixed, then a replicated experiment cannot

be performed, because to replicate an experiment means to use the same factor levels in each replicate. Clearly that isn't going to happen if at least one factor is random, as is the case here. Consequently, it is highly desirable to perform an analysis using the averages and compare the results.

As before, the response variable is the arcsin of the square root of the lead recovery proportions. The Pareto chart of the t -statistics of the effects estimates is given below and we see little difference between this chart and the chart that results from using all of the observations. One noticeable difference is that the CD and DE interactions fall below the .05 line, whereas they were right at the line when all of the observations are used.

Pareto Chart of the Absolute Values of the t -statistics for Effect Estimates

(Response is ArcSineSqrt(Recovery%), alpha = .05)



Of course that doesn't mean that the effects won't be judged significant when a variable selection approach is used. In this instance, however, the use of stepwise regression does result in those interactions not being selected. It seems clear why the *CD* and *DE* interactions are not significant, as recall they were not selected when observation #79 was deleted in the analysis using the replicates.

Thus, it seems apparent that the significance of those two interactions was strongly affected by one or more extreme observations, which of course have a lesser effect when averages are used.

The statistics for the model with the six terms are given below.

Estimated Effects and Coefficients for arcsinesqrt(recovery%)(coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		1.1341	0.009079	124.92	0.000
B	0.1269	0.0635	0.009079	6.99	0.000
C	0.0888	0.0444	0.009079	4.89	0.000
E	-0.2937	-0.1469	0.009079	-16.18	0.000
BE	0.1509	0.0755	0.009079	8.31	0.000
CE	0.0774	0.0387	0.009079	4.26	0.000
ABE	0.0507	0.0254	0.009079	2.79	0.010

Analysis of Variance for arcsinesqrt(recovery% (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	6	1.13281	1.13281	0.188802	71.58	0.000
Residual Error	25	0.06594	0.06594	0.002637		
Lack of Fit	9	0.02084	0.02084	0.002315	0.82	0.606
Pure Error	16	0.04510	0.04510	0.002819		
Total	31	1.19875				

It is worth noting that the R^2 value for this model is .945, which is much higher than the R^2 value (.795) that is obtained using all 112 observations. (The R^2 values are slightly higher when they are converted back to the original scale and are .963 and .817, respectively.)

The difference between the .945 and the .795 is undoubtedly due in part to the fact that R^2 generally declines as the number of observations is increased. It is also likely due in part, however, to the few observations of the 112 that were not well-fit by the model, including the one suspicious observation (#79).

V Omitted Factors

In a previous study, three factors that were not examined in this study -- operator, substrate, and overlayer -- were found to be significant. One might conjecture the extent to which R^2 would have been increased if they had been included in this study, but we also need to consider other possible effects of their omission. Since particle size has the largest effect, an obvious question is whether or not the operators performed almost identically in grinding the specimens to what has been “large” and “small”.

For example, was the particle size for observation #79 considerably larger than it was supposed it be, which might explain why the recovery percentage was so small? Another possible explanation is that the specimen could have had a thick-oil overlayer that might have made it difficult to grind the particle down to the desired size. If so, there would then be an operator effect, in a manner of speaking. It would be a different type of effect from what one speaks of in experimental design, however, as the effect would be to cause incorrect levels of another factor, which is different from the way that we generally view extraneous factors.

Because of these and other uncertainties associated with the experiment, it is desirable to also analyze the data as having come from an unreplicated experiment, as well as to look for unusual observations. Although it might be difficult or impossible to after the fact discover any problems caused by extraneous factors when data on

those factors were not recorded, suspicions might lead to future experiments being designed differently. Therefore, in the next section we use a formal search for unusual observations.

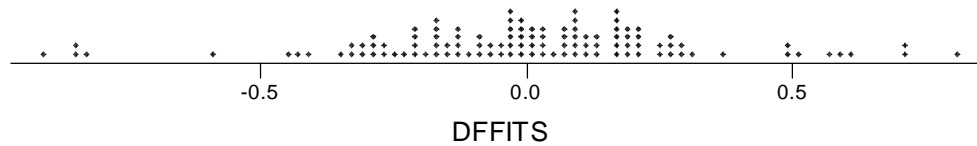
VI Unusual Observations

Outliers and unusual observations in general can only be classified as such relative to a model. This doesn't mean that we should wait until we have a model to look for unusual observations, as bad data points could send us in the wrong direction in our search for a good model, and indeed observation #79 was identified as being unusual before a model was fit.

Before a model with the resultant coefficients is used and interpreted, a final search for unusual observations should be conducted. Various statistics might be used in such a search, such as Cook's- D statistic, DFFITS, etc. These statistics have been most often associated with regression analysis, and we should recognize that with a 2^5 design we are not going to have any X -outliers (i.e, points that are outlying in the factor space). Therefore, our search should be for Y -outliers and influential data points, and our search will be for the model with be for the model with the constant term plus the terms B , C , E , BE , CD , CE , DE , and ABE .

The dotplot for the DFFITS values (using all 112 data points) is shown below

Dotplot for DFFITS

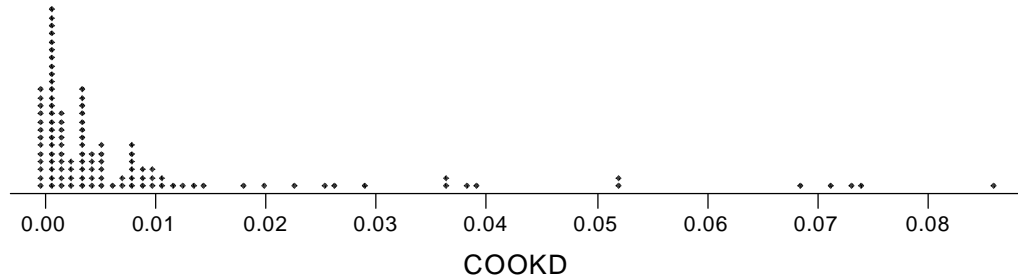


The absolute value of the benchmark for DFFITS proposed by Jensen (2000) is $2\sqrt{h_{ii}/(1-h_{ii})}$, with the h_{ii} being the leverage values, which are the diagonal elements of the “hat matrix” in regression. The h_{ii} are all approximately .08 for these data, so using that number produces benchmarks of ± 0.59 . There are four DFFITS values between -0.825 and -0.912 and three are between 0.701 and 0.806 . It is of particular interest to see if any of the extreme values occur at the same treatment combination, with the four extreme observations on the low end are numbers 79 (as expected), 97, 48, and 47, in descending order in terms of absolute value. These are all at different treatment combinations, but it would be of interest to try and determine why two of the values are consecutive observations. On the high end, of the three observations that stand out as being extreme (numbers 5, 29, and 75), the first two had recorded values of 100.2% and 100.6% (i.e., there was at least slight measurement error), and the three observations all occurred at different treatment combinations. Measurement error might be the explanation for all three values on the

high end, and since the four observations on the low end are even more extreme (and are all greater in absolute value), a cause of their extreme DFFITS values should be sought.

Although Cook's- D is a frequently used diagnostic for identifying unusual observations, it is of very limited value in replicated designs. This can be explained as follows. The statistic can be written as the product of four factors, one of which is $h_{ii}/(1 - h_{ii})$. Factorial designs are equileverage designs with $\sum_{i=1}^n h_{ii} = p$, with p denoting the number of parameters in the model. Thus, with 9 parameters, as in the current model, we would have each $h_{ii} = 9/n$ if there were no replicates. If there are k replicates, then each $h_{ii} = 9/(kr)$, with r here denoting the number of distinct treatment combinations. Thus, the h_{ii} can be quite small when for replicated designs n is at least of moderate size, so that $h_{ii}/(1 - h_{ii})$ will also be quite small. One suggested threshold value for Cook's- D is 1.0, but even highly unusual data points will have values of the statistic far less than 1.0 for replicated designs. Indeed as the graph below shows, the largest value is quite small (actually 0.0857), and this is the value for observation #79.

Dotplot for COOKD



VII Determination of Optimum Operating Conditions

Since the objective is of course to maximize lead recovery, it is obviously desirable to have a small particle size, with particle size having the most pronounced effect. Therefore, if we condition on small particle size and use only the 56 observations with small particle size, it doesn't make any significant difference which time and temperature we use, but there is a preference for a small particle mass. (Of course this makes sense because the smaller the size, the smaller the mass.) Note that this is not in accord with the conclusion that time and temperature were important when all 112 observations were used. Thus, the conditional effects of time and temperature for the small particle size are not significant. Since the unconditional effects are significant, this means that time and temperature have pronounced effects for the large particle size(s).

The results are given below.

Estimated Effects and Coefficients for arcsinesqrt(recovery%) using only small particle size (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		1.28908	0.01361	94.73	0.000
A	0.00203	0.00101	0.01361	0.07	0.941
B	-0.02474	-0.01237	0.01375	-0.90	0.373
C	0.00730	0.00365	0.01361	0.27	0.790
D	-0.05594	-0.02797	0.01375	-2.03	0.047

SUMMARY

We have analyzed the data that resulted from an investigation of lead recovery with a 2^5 design used to investigate the effect of 5 factors. We have noted some nuances that slightly complicated the analysis, but these are not at all different from what very often transpires when experiments are performed. The analysis showed that a follow-up experiment would be desirable to investigate certain possible effects that were previously noted in NISTIR 6834, with their possible effect indicated here. Textbook examples unfortunately often fail to convey various complications that can occur when experimentation is performed.

Within the past several years there has been interested demonstrated in the literature in trying to determine the effects when experiments are not performed in a textbook manner (e.g., there are restrictions on randomization, etc.). The issue of replicates versus multiple readings must also be considered and in this study we noted that it was impossible to achieve true replicates, primarily because the particle size could not be fixed.

REFERENCES

Box, G. E. P. and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces*. New York: Wiley.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. New York: Wiley.

Hamada, M. and C.F.J. Wu (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, **24**, 130-137.

Jensen, D. R. (2000). The use of studentized diagnostics in regression. *Metrika*, **52**, 211-223.

Lin, D. K. J. (1998-99). Spotlighting interaction effects in main effect plans: A supersaturated design approach. *Quality Engineering*, **11**(1), 133-139.

Rossiter, W. J., Jr., B. Toman, M. E. McKnight, and M. B. Anaraki (2002). Factors affecting ultrasonic extraction of lead from laboratory-prepared household paint films. NISTIR 6834, prepared for the U.S. Department of Housing and Urban Development, Office of Healthy Homes and Lead Hazard Control.