Graphical Methods for Selecting Effects in Factorial Models

Gary W. Oehlert

School of Statistics
University of Minnesota
gary@stat.umn.edu

May 15, 2009
Work with Pat Whitcomb

Introduction

ANOVA is the primary tool for analyzing designed experiments.

ANOVA gives us a mechanism to determine which terms are unlikely to be due to random error.

We do model selection by choosing the terms with small p-values, or perhaps by combining ANOVA with AIC or BIC.

Everything is straightforward as long as there is replication.

A single replication means 0 degrees of freedom for error, and no estimate of σ^2.

Many analytic/computational approaches have been proposed for this situation, but that is not our goal today.

Today, I'd like to look at Graphical Methods.

Daniel (1959) made the following observation and suggestion.

Single replication from a 2^k design, and standard assumptions hold.

Estimated main effects and interactions will be Gaussian distribution with constant variance.

Estimated effects have mean 0 in the null case and non-zero mean otherwise.

Non-null effects will look like outliers in a (half) normal plot of the estimated effects.

Filtration rate experiment of Montgomery (2001, example 6-2). Factors are temperature, pressure, concentration, and stirring rate. Each factor has two levels.

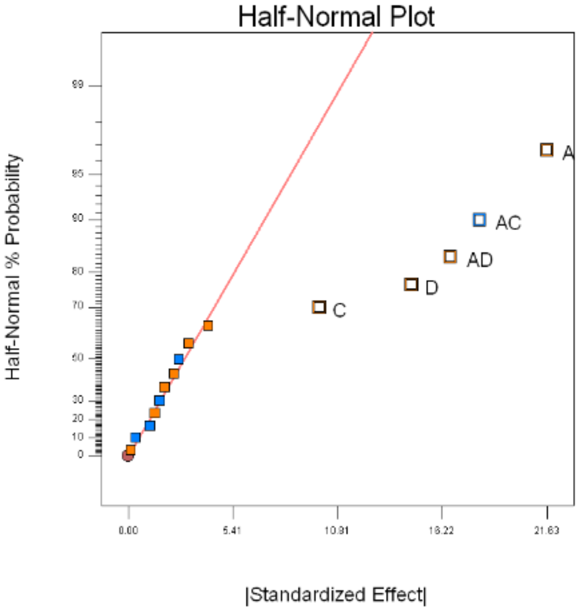
The estimated effects, in standard order, are

T	21.625	P	3.125	T.P	0.125
C	9.875	T.C	-18.125	T.P	2.375
T.P.C	1.875	S	14.625	T.S	16.625
P.S	-0.375	T.P.S	4.125	C.S	-1.125
T.C.S	-1.625	P.C.S	-2.625	T.P.C.S	1.375

We can now make normal and/or half-normal plots to detect outliers by eye.
This plot is from Design-Expert 7.0.

Design-Expert® Software
Filtration Rate

Shapiro-Wilk test
W-value = 0.974
p-value = 0.927
A: Temp
B: Pres
C: Conc
D: Stir
Positive Effects
Negative Effects



Possible Extensions

These are tremendously useful plots for selecting effects in models.

- Very easy to use and understand.
- Provide a degree of protection against multiple testing.

What else would we like to do?

We would like to have plots/procedures that:

1. can also work for replicated designs.
2. can provide more quantitative error rates.
3. work with factors with more than two levels.
4. work with unbalanced designs.
5. assist in choosing sensible models.
6. ???

(1) Whitcomb and Larntz(1992) showed how to handle replication by plotting points that represent pure error.

(2) There are literally dozens of papers on providing quantitative error rates, e.g. Lenth's pseudo-standard error. Hamada and Balakrishnan (1998) provide a nice review as of a few years ago.

That leaves us with (3), (4), and (5) to consider today.

More than two levels

We seek an extension to the Daniel half-normal plot that will

- be identical to the usual plot for a 2^k
- permit factors with more than two levels.

Existing methods for more than two levels either require that:

1. the entire design be split into single degree of freedom effects with equal variance (so that the usual half-normal plot will work), or
2. all effects have the same degrees of freedom (which can be more than 1).

Compute effects the hard way.

Term i has SS_i. In 2^k design

$$\text{effect}_i \propto \sqrt{SS_i}$$

Using $\sqrt{SS_i}$ gives us a rescaled normal plot, but doesn't help for more than 2 levels.

If we knew σ^2, we could test the terms using the SS_i and χ^2.

Suppose that we have a provisional value for the standard deviation of the errors; call it $\tilde{\sigma}$. Compute the provisional p-value

$$\tilde{p}_i = (1 - \chi^2_{\nu_i}(SS_i/\tilde{\sigma}^2))$$

Express provisional p-value as a half-normal percent point:

$$\tilde{z}_i = \Phi^{-1}(\tilde{p}_i/2)$$

Rescale by $\tilde{\sigma}$:

$$\tilde{q}_i = \tilde{\sigma}\tilde{z}_i$$

When $\nu_i = 1$

$$\tilde{q}_i = \sqrt{SS_i}$$

For 2^k, q_i is proportional to effects.

However we choose $\tilde{\sigma}$, the half-normal plot of the $\tilde{q}_i$s looks just like the Daniel plot.

But ... we can compute $\tilde{q}_i$ for any ν_i.

$\tilde{q}_i$ accounts for the different amounts of information present in terms with different degrees of freedom.

Where do we get $\tilde{\sigma}$?

Timid: select terms for model and pool all unselected model terms into error to get $\tilde{\sigma}^2$.

Bold: plot a line with slope $1/\tilde{\sigma}$ on our half-normal plot and try different values of $\tilde{\sigma}$ until we find one that seems to match the apparently inactive terms.

In either case, designate terms with large $\tilde{q}_i$ as active and terms with small $\tilde{q}_i$ as error. (Our procedure is probably liberal from a testing perspective.)

Concrete faults

Example. Faulting occurs when the concrete lifts up at the joint. Four factor experiment on faulting (measured as the difference in height at the joint and the pavement 30 cm away):

W, the water/cement ratio (.33 .42, .55)
D, presence or absence of dowels in the joints
B, thickness of the base (12.5, 25, or 38 cm)
S, resilient modulus of the subgrade (5K, 10K, or 20K psi).

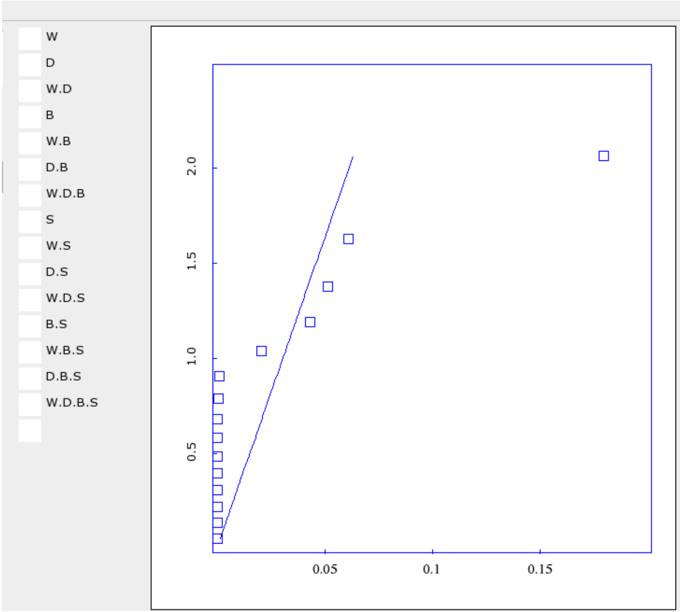
This is a 3 by 2 by 3 by 3 factorial with a single replication and 54 runs.

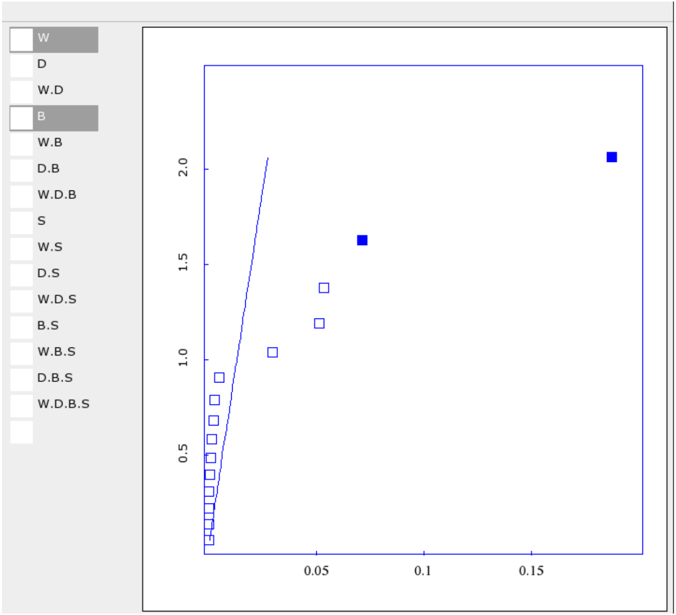
Consider first the timid method (pool unselected terms). Also, add line with slope $1/\tilde{\sigma}$.

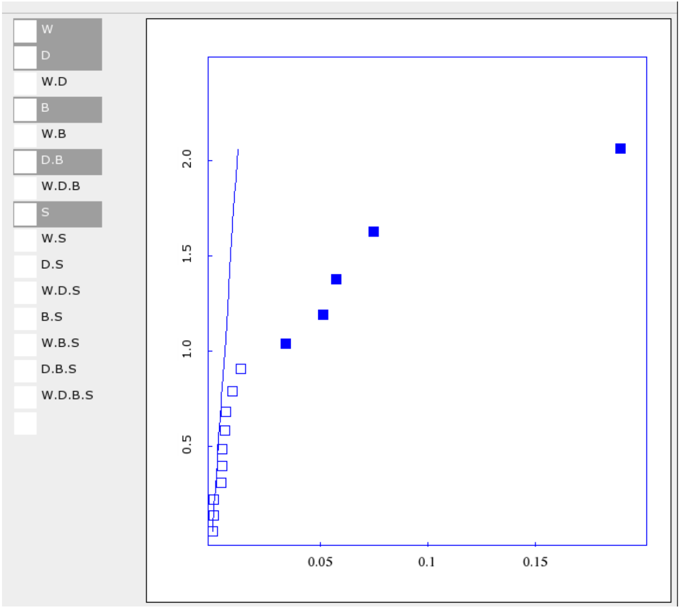
The first is the plot with no terms selected.

Next we select the two largest. The line is now steeper because $\tilde{\sigma}$ is reduced. The “line” of points is no longer vertical.

Finally, we select all five “outliers.” The unselected points hug the line fairly closely, but there are a couple below (or to the right) that might be worth investigation.


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15

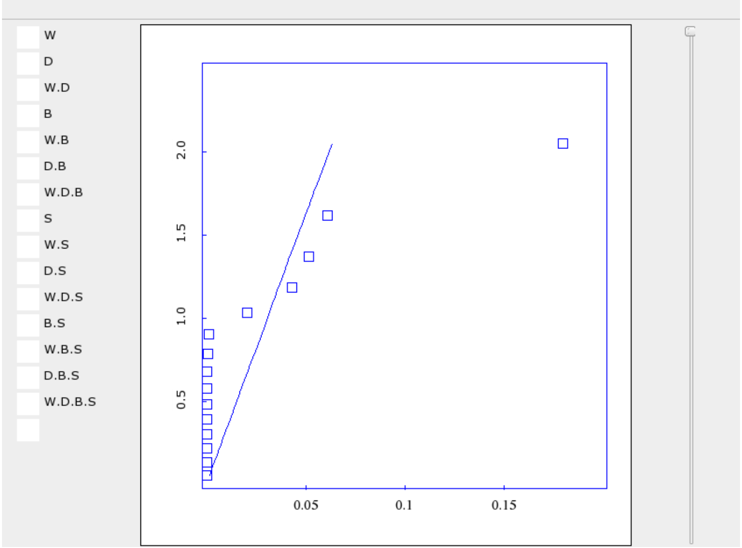

W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15

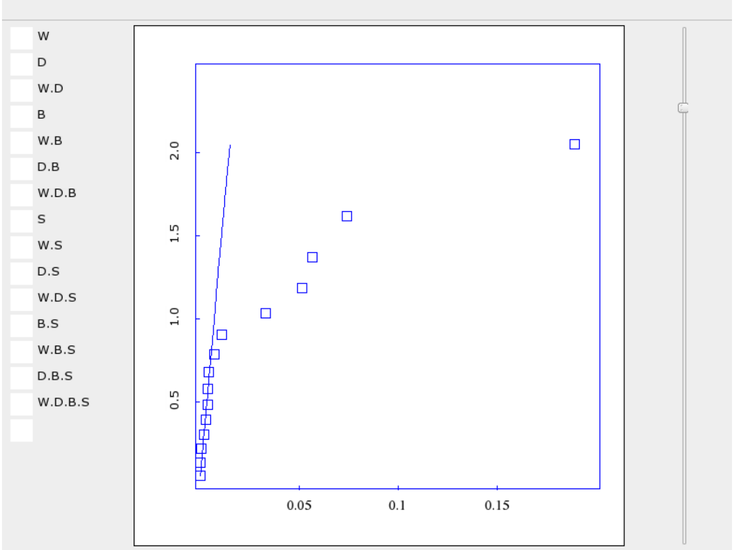

W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15

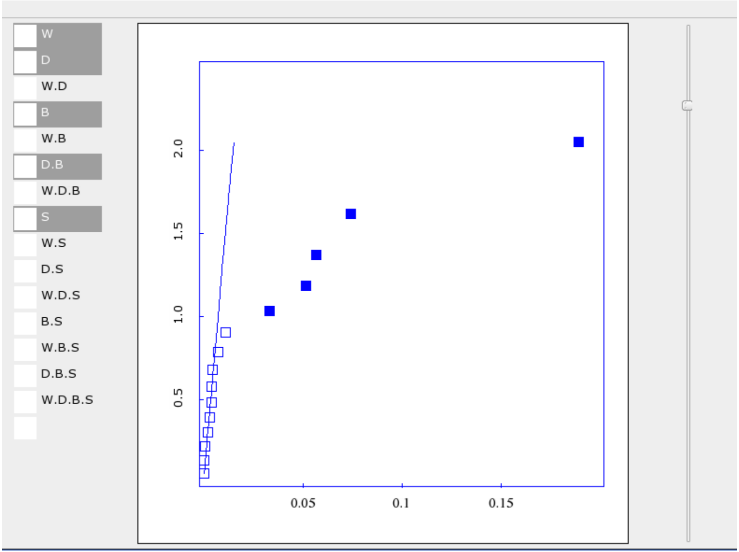
Now be bold. This implementation has a slider on the right that controls the value of $\tilde{\sigma}$. The maximum value for $\tilde{\sigma}$ is the mean square for error when no terms are selected. Each step down reduces $\tilde{\sigma}$ by two percent.

The first plot shows 0 terms selected and the maximum $\tilde{\sigma}$.

The second plot shows 0 terms selected and a more reasonable choice for $\tilde{\sigma}$.
The third plot shows 5 terms selected and the same $\tilde{\sigma}$.


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15

Unselected terms tend to sag below the line.

Suppose that there are M terms. We are plotting

$$q_{(j)}, h_{(j),M}$$

where $h_{(j),M}$ is the jth smallest half-normal score for a sample of size M.

The null terms will be linear for

$$q_{(j)}, h_{(j),n}$$

where n is the number of null terms.

Can we fix this?

Plotting

$$q_{(j)}, h_{(j),n}$$

pairs shows linearity and keeps the faith with $1/\tilde{\sigma}$ line, but is ugly and non-monotone.

Plotting

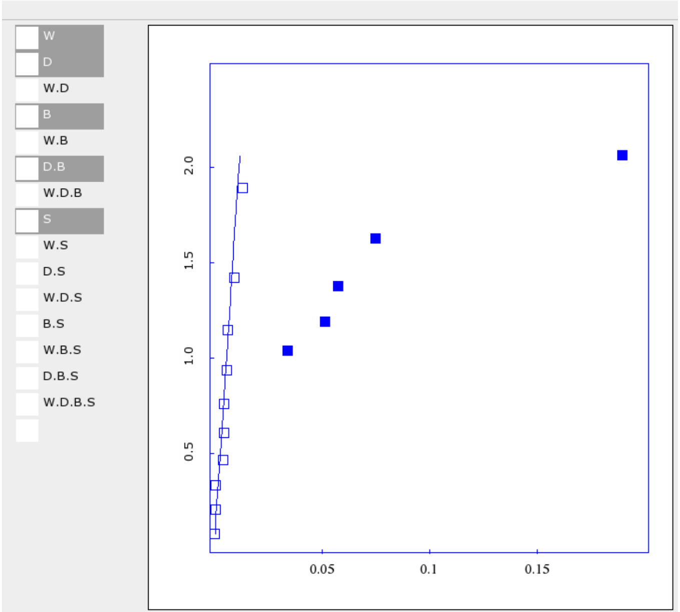
$$q_{(j)}, h_{(j),n} h_{(n),M}/h_{(n),n}$$

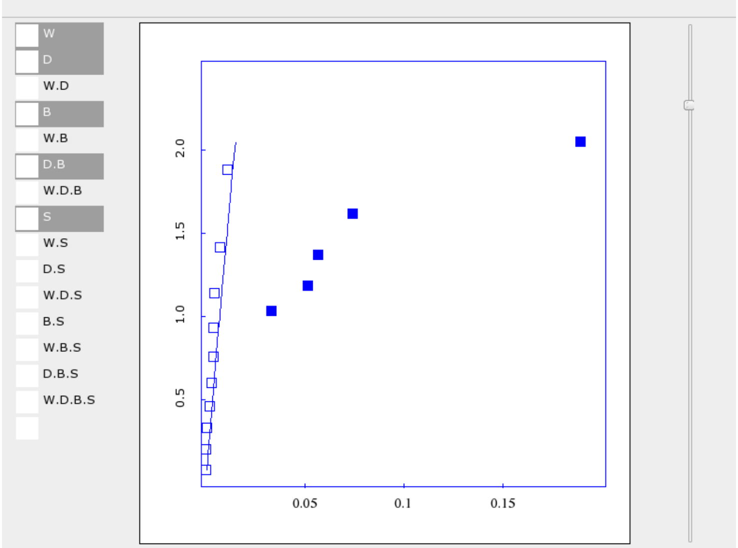
shows linearity and keeps plot monotone, but loses connection with $1/\tilde{\sigma}$ line.

Latter probably better for naive users.

Here are two plots. The first uses $\tilde{\sigma}$ computed from the unselected terms and rescales the unselected effects. The unselected effects now hug the line closely.

The second plot uses the slider to select the same value of $\tilde{\sigma}$ we used before, but again rescales the unselected effects.


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15


W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15

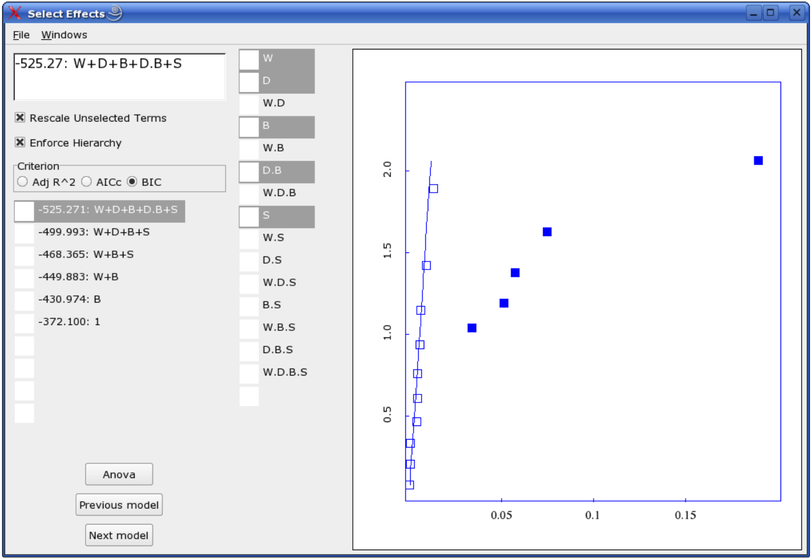
Hierarchy and Model Selection

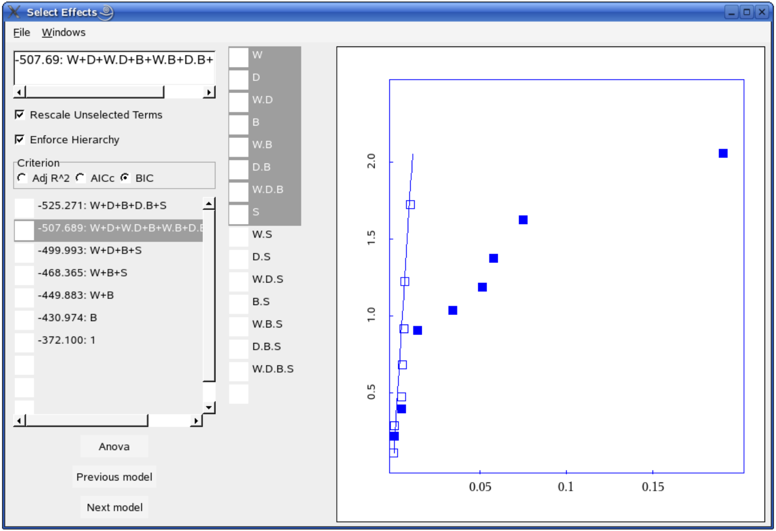
Hierarchical models are generally appropriate. Help users by enforcing hierarchy (optionally).

Display selection criteria such as AIC or BIC to help users select models. I tend to favor BIC, but it’s not a clear cut choice.

Two examples. The first plot just shows the five largest terms selected.

The second plot adds the next largest term (WDB), but we must also add WD and WB to maintain hierarchy. Doing this, the BIC says that the previous model was better.


Select Effects
File Windows
-525.27: W+D+B+D.B+S
Rescale Unselected Terms
Enforce Hierarchy
Criterion
Adj R^2 AICc BIC
-525.271: W+D+B+D.B+S
-499.993: W+D+B+S
-468.365: W+B+S
-449.883: W+B
-430.974: B
-372.100: 1
Anova
Previous model
Next model
W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15


Select Effects
File Windows
-507.69: W+D+W.D+B+W.B+D.B+
Rescale Unselected Terms
Enforce Hierarchy
Criterion
Adj R^2 AICc BIC
-525.271: W+D+B+D.B+S
-507.689: W+D+W.D+B+W.B+D.B
-499.993: W+D+B+S
-468.365: W+B+S
-449.883: W+B
-430.974: B
-372.100: 1
Anova
Previous model
Next model
W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
W.D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15

Unbalanced designs

For unbalanced designs, SS only defined relative to some other model. So what SS do we use in our plot? What is the base model for computing SS?

The SS for a term <u>in</u> the model (selected) is the SS for <u>removing</u> that term from the model.

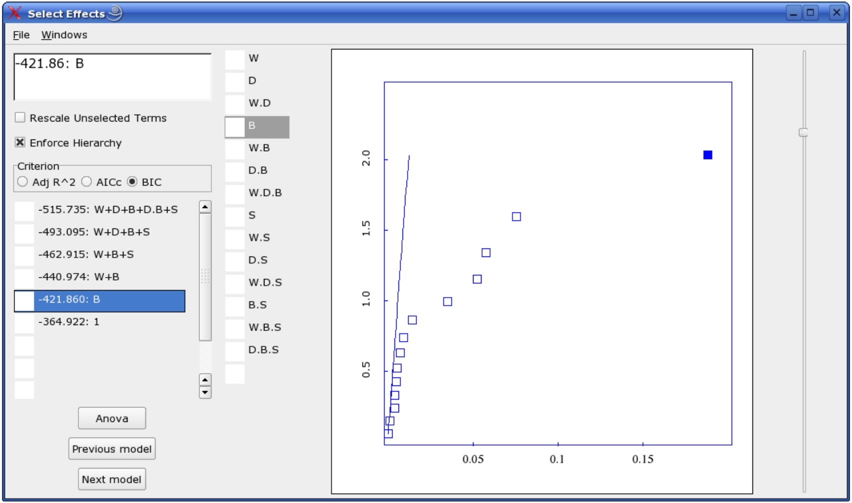
The SS for a term <u>not in</u> the model (unselected) is the SS for <u>adding</u> that term to the model.

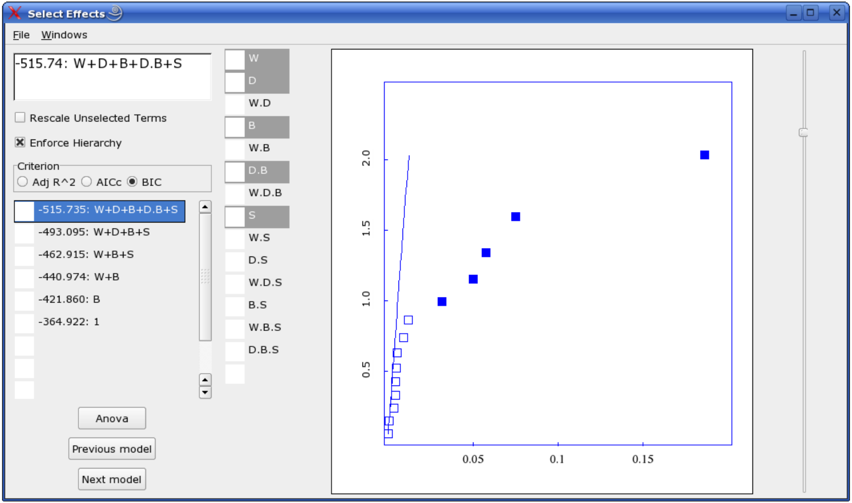
These are effectively the sums of squares used in stepwise regression.

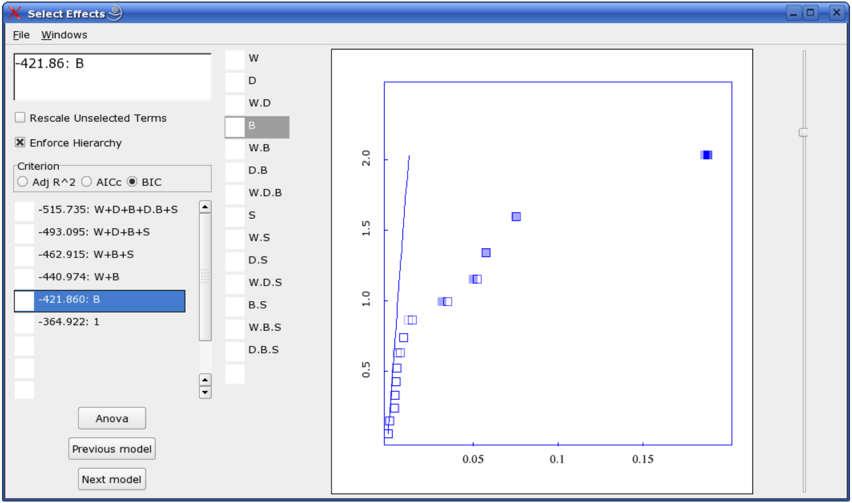
So we recompute $\tilde{q}_i$ each time model changes.

Set the first response to missing to make the design is unbalanced. Turn off effect rescaling to make things easier to compare. Use the “slider” version so that $\tilde{\sigma}$ can be held constant.

The first plot is for a single selected term, and the second plot is for five selected terms. It’s a bit difficult to see, but the points move slightly.


Select Effects
File Windows
-421.86: B
Rescale Unselected Terms
Enforce Hierarchy
Criterion
Adj R^2 AICc BIC
-515.735: W+D+B+D.B+S
-493.095: W+D+B+S
-462.915: W+B+S
-440.974: W+B
-421.860: B
-364.922: 1
Anova
Previous model
Next model
W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
0.5
1.0
1.5
2.0
0.05
0.1
0.15


Select Effects
File Windows
-515.74: W+D+B+D.B+S
Rescale Unselected Terms
Enforce Hierarchy
Criterion
Adj R^2 AICc BIC
-515.735: W+D+B+D.B+S
-493.095: W+D+B+S
-462.915: W+B+S
-440.974: W+B
-421.860: B
-364.922: 1
Anova
Previous model
Next model
W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15


Select Effects
File Windows
-421.86: B
Rescale Unselected Terms
Enforce Hierarchy
Criterion
Adj R^2 AICc BIC
-515.735: W+D+B+D.B+S
-493.095: W+D+B+S
-462.915: W+B+S
-440.974: W+B
-421.860: B
-364.922: 1
Anova
Previous model
Next model
W
D
W.D
B
W.B
D.B
W.D.B
S
W.S
D.S
W.D.S
B.S
W.B.S
D.B.S
2.0
1.5
1.0
0.5
0.05
0.1
0.15

Conclusions

1. Graphical techniques for selecting model terms are highly useful when there is only a single replication, and convenient to use even when there is pure error.
2. More than single degrees of freedom can be easily accommodated in the plots, but you must get used to points moving.
3. Enforcement of hierarchy and feedback on model quality (via BIC or similar) helps us home in on an appropriate model quickly.
4. Unbalanced data can be accommodated by carefully defining SS.

Akaike, Hirotugu (1974). “A new look at the statistical model identification,” IEEE Transactions on Automatic Control 19, 716723.

Daniel, Cuthbert (1959). “Use of half-normal plots in interpreting factorial two-level experiments,” Technometrics 1, 311–341.

Hamada, M. and Balakrishnan, N. (1998). “Analyzing unreplicated factorial experiments: A review with some new proposals (Disc: p29-42)” Statistica Sinica 8, 1–23

Hurvich, Clifford M. and Tsai, Chih-Ling (1991). “Bias of the corrected AIC criterion for underfitted regression and time series models” Biometrika 78, 499–509

Montgomery, Douglas C. (2001). Design and Analysis of Experiments, 5th Edition, Wiley: New York.

Schwarz, G. (1978). "Estimating the Dimension of a Model," Annals of Statistics 6, 461–464.

Whitcomb, Patrick J. and Larntz, Kinley (1992). "The role of pure error on normal probability plots" ASQC Technical Conference Transactions, 1223-1229 American Society for Quality Control: Milwaukee.