Interaction

Gary W. Oehlert

School of Statistics
University of Minnesota

November 25, 2014

One of my pet peeves is an analysis where someone says "These factors interact," and then leaves it at that.

I suppose it's worse if someone says "These factors have main effects," and then doesn't examine them further, but that is a less common case.

Let's take a little time and look at a couple more ways of thinking about interaction.

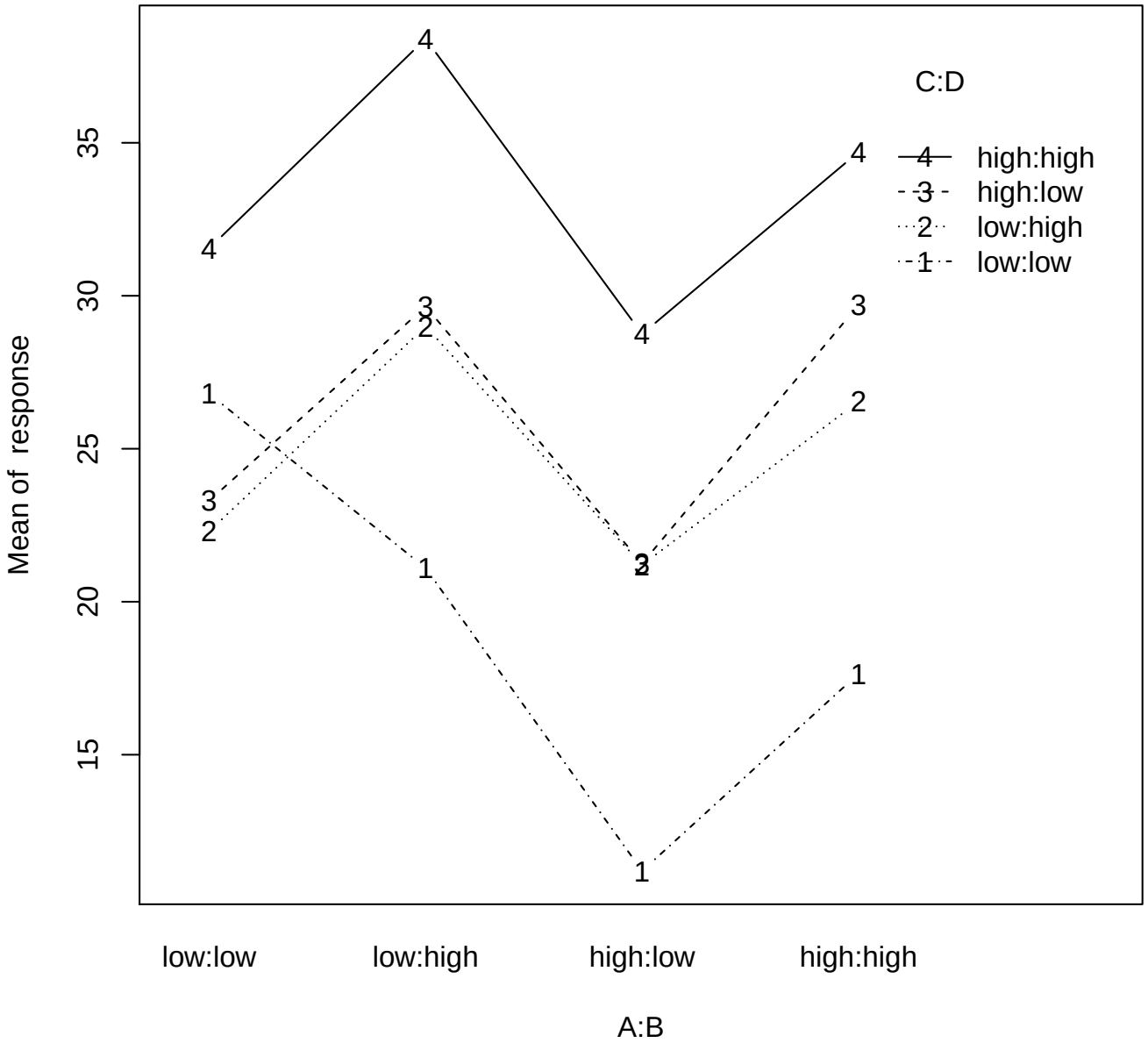
One cell interaction

Sometimes much or all of the interaction in a data set is due to a single treatment (cell) being off the pattern predicted by lower order terms.

For example, you can have data where all cells except one fit a main effects only model, but all main effects and interactions will look big.

Example 9.2 of the text is a 2^4 factorial with $n = 2$. Everything is significant.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	120.90	120.90	117.4511	8.871e-09	***
B	1	204.02	204.02	198.1979	1.970e-10	***
C	1	472.78	472.78	459.2896	3.288e-13	***
D	1	335.40	335.40	325.8336	4.621e-12	***
A:B	1	18.00	18.00	17.4863	0.0007050	***
A:C	1	24.85	24.85	24.1421	0.0001559	***
B:C	1	27.38	27.38	26.5987	9.541e-05	***
A:D	1	15.12	15.12	14.6934	0.0014664	**
B:D	1	10.81	10.81	10.5027	0.0051192	**
C:D	1	6.48	6.48	6.2951	0.0232492	*
A:B:C	1	11.52	11.52	11.1913	0.0041075	**
A:B:D	1	34.03	34.03	33.0601	2.985e-05	***
A:C:D	1	50.00	50.00	48.5732	3.161e-06	***
B:C:D	1	22.11	22.11	21.4803	0.0002754	***
A:B:C:D	1	13.78	13.78	13.3880	0.0021183	**
Residuals	16	16.47	1.03			


Mean of response
35
30
25
20
15
C:D
4 high:high
3 high:low
2 low:high
1 low:low
low:low
low:high
high:low
high:high
A:B

Everything looks parallel, except for the cell with all factors at the low level. That cell seems to be higher than we would expect.

Looking at the residuals shows that it's not an outlier issue; both data points are higher than expected.

Refit with a dummy variable indicating the interacting cell.

```
           Sum Sq Df  F value    Pr(>F)
           Df Sum Sq Mean Sq  F value    Pr(>F)
A           1 120.90  120.90 117.4511 8.871e-09 ***
B           1 204.02  204.02 198.1979 1.970e-10 ***
C           1 472.78  472.78 459.2896 3.288e-13 ***
D           1 335.40  335.40 325.8336 4.621e-12 ***
alllow      1 217.35  217.35 211.1485 1.229e-10 ***
A:B         1   0.05    0.05   0.0438   0.83684
A:C         1   0.30    0.30   0.2916   0.59666
B:C         1   0.77    0.77   0.7438   0.40120
A:D         1   0.20    0.20   0.1916   0.66744
B:D         1   1.35    1.35   1.3117   0.26892
C:D         1   4.79    4.79   4.6517   0.04658 *
A:B:C       1   3.00    3.00   2.9120   0.10725
A:B:D       1   0.34    0.34   0.3288   0.57435
A:C:D       1   5.47    5.47   5.3115   0.03492 *
B:C:D       1   0.49    0.49   0.4760   0.50012
Residuals  16  16.47    1.03
```

Almost all of the interaction SS is in that one single cell, although CD and ACD are still marginally significant. (Their p-values are almost the same if you use Type II tests.)

```
Model 1: response ~ A + B + C + D + alllow
Model 2: response ~ A + B + C + D + alllow + C:D
Model 3: response ~ A + B + C + D + alllow + A:C:D
Model 4: response ~ A * B * C * D + alllow
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     26 33.210
2     25 29.241  1    3.9691 3.8558 0.06719 .
3     22 22.034  3    7.2076 2.3340 0.11262
4     16 16.470  6    5.5636 0.9008 0.51808
```

Anova goes with the additive plus alllow, but AIC or BIC would add one or more interactions.

Clearly, we could also have a “few cell” interaction instead of a one cell interaction.

One cell interactions are easier to understand than to find.

Few cell interactions are harder still to find (although still fairly easy to understand).

Polynomial Models

We have seen that polynomial models (or more general functional models) can be helpful when a factor is quantitative.

For example, for a single factor with quantitative levels z_i, we looked at models like

$$\mu_i = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \beta_3 z_i^3 + \beta_4 z_i^4$$

We can extend this to polynomial models in factorial designs and polynomial models of interaction.

First, some notation. Suppose that we have a two-factor model with factors A and B quantitative (like temperature and pressure).

Let w_i be the quantitative levels associated with factor A.
Let z_j be the quantitative levels associated with factor B.

$$\begin{aligned}\mu_{ij} &= \mu + \\ &\lambda_{1,0}w_i + \lambda_{2,0}w_i^2 + \cdots + \lambda_{a-1,0}w_i^{a-1} + \\ &\lambda_{0,1}z_j + \lambda_{0,2}z_j^2 + \cdots + \lambda_{0,b-1}z_j^{b-1} + \\ &\lambda_{1,1}w_iz_j + \lambda_{2,1}w_i^2z_j + \lambda_{1,2}w_iz_j^2 + \ldots \lambda_{a-1,b-1}w_i^{a-1}z_j^{b-1}\end{aligned}$$

Polynomial terms separately in each factor (main effects) plus all their cross-product terms (interactions).

This polynomial model can describe all of the means that our usual $\alpha_i + \beta_j + \alpha\beta_{ij}$ sorts of models can.

However, we will frequently find that we do not need all of the terms and can describe the means with a simpler model.

Use Hierarchical Models! That is, if you use the (r,s) term, you should also use all terms with first subscript $\leq$ r and second subscript $\leq$ s.

Don't look at coefficients until you have selected your reduced model.

Why am I happy to remove non-significant model terms here when I usually don't?

One major use for polynomial models is to predict at w and z values that were not used in the experiment.

Leaving non-significant polynomial terms in the model makes predictions worse, so we want to remove them in this situation.

Suppose now that we also have a categorical factor C. We can interact categorial and polynomial factors.

The AC interaction could be modeled as:

$$\lambda_{1,0,k} w_i + \lambda_{2,0,k} w_i^2 + \cdots + \lambda_{a-1,0,k} w_i^{a-1}$$

where $\sum_k \lambda_{i,j,k} = 0$.

We have an order a-1 polynomial in w for every level k of factor C, but it's a different polynomial (has different coefficients) for every level of C.

Tukey and Mandel interactions

These models for interaction are most useful in cases where we have a single replication and thus no estimate of pure error.

What these models do is use a few df from the overall interaction to attempt to capture some particular kinds of interaction, with the remainder of the df and SS of interaction used as surrogate error.

If this type of interaction is present and we model it, then the surrogate error is a better surrogate error.

Remember how we said that interaction effects $\alpha\beta_{ij}$ were not the same as $\alpha_i\beta_j$?

The Tukey one degree of freedom for non-additivity assumes

$$\alpha\beta_{ij} = \lambda\alpha_i\beta_j/\mu$$

This is actually a very common kind of interaction.

To test for the presence of a Tukey one degree of freedom effect, follow these steps:

1. Fit the additive model with terms A and B (rows and columns).
2. Get the squared predicted values, divide them by 2, then divide that by our estimate of μ.
3. Fit a new model that includes A, B, and the rescaled squared predicted values (RSPV).
4. Testing whether we need the rescaled squared predicted values tests for the need of the Tukey interaction.

Alternatively, one can produce $\widehat{\alpha}_i\widehat{\beta}_j/\widehat{\mu}$ directly from the results of the lm, and then use that instead of the rescaled squared predicted values.

```
m <- lm(y~A+B)
mA <- model.effects(m,"A")[A]
mB <- model.effects(m,"B")[B]
compval <- mA*mB/coef(m)[1]
m2 <- lm(y~A+B+compval)
```

Tukey interaction is also called transformable nonadditivity, because it is a kind of interaction that you can decrease by a power family transformation.

If $\widehat{\lambda}$ is the estimated coefficient for the RSPV (you get the same thing using the comparison values), then $1 - \widehat{\lambda}$ is the estimated power to transform the data.

The RSPV approach works for finding a transformation with more factors and more complicated base models. The comparison value approach can be generalized, but it is much more work.

Mandel generalized the Tukey model.

$$\begin{aligned}
&\text{Row-model:} && \mu_{ij} = \mu + \alpha_i + \beta_j + \zeta_j \alpha_i \\
&\text{Column-model:} && \mu_{ij} = \mu + \alpha_i + \beta_j + \xi_i \beta_j \\
&\text{Slopes-model:} && \mu_{ij} = \mu + \alpha_i + \beta_j + \zeta_j \alpha_i + \xi_i \beta_j \quad .
\end{aligned}$$

Row model says every mean is a linear function of row effects, but slope and intercept differ by column, and vice versa for column model. Slopes puts everything in.

Recalling mA and mB from above

```
row.model <- lm(y~A+B+B:mA)
column.model <- lm(y~A+B+A:mB)
slopes.model <- lm(y~A+B+B:mA+A:mB)
```