Statistics 5303
Fall 2013
Exam #1 Sketched Solutions

See also the data analyses at the end.

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? If so, how did you fix them?

Residual plots showed no evidence of problems with non-constant variance; residuals were a little short-tailed, but nothing that required concern. (There is banding in the residuals, but that is expected with a discrete response.)

2. How would you describe the effects of temperature and time on the degree of doneness for the wings?

The time and temperature treatments are both highly significant, and the interaction is marginally significant. As would be expected, increasing time or temperature leads to greater doneness of the wings. Looking at the model coefficients suggests that the responses are linear in both factors; fitting with polynomials shows no indication of curvature for either factor. The interaction can be described as the slope in one factor increases as you increase the level of the second factor.

3. Which factor/level combination (or combinations) should be used to obtain the correct degree of doneness?

We can do a t-test with Bonferroni correction comparing each treatment mean to 5. At the 5% level, four treatments are not significantly different from 5; these are: 40 minutes at 400 degrees gives wings that are just right; 50 minutes at 350 degrees is close but slightly over cooked; 45 minutes at 350 and 375 degrees bracket the just right setting.

More generally, doneness increases approximately linearly with the product of time and temperature. The correct level of doneness corresponds to approximately 16,245 degree-minutes. Thus 45 minutes of cooking should be done at 361 degrees.

4. How many experimental units does this experiment have? Justify your answer.

There are 16 experimental units with two measurement units per experimental unit. The randomization and experimental procedure guarantee that the pairs of measurements must receive the same treatment and are thus two measurements on the same experimental unit.

5. Which type of glue (liquid versus stick) should we use for a strong bond? Which quality of glue (brand name versus generic) should we use?

Liquid glue is consistently better than stick glue in terms of force needed to pull the halves apart. This effect is highly significant and is especially large for cardboard (about 5.6 N for paper and 33.7 N for cardboard). There is no evidence of a difference between name brand and generic glues.

6. Should we use a weight to hold the pieces of material together to achieve a strong bond?

No. There is at best marginal evidence that using a weight changes the strength of the bond. The only place where it might be making a difference is for liquid glue on cardboard, but there using a weight makes the bond weaker, not stronger.

7. We have an unbalanced three-way factorial design with factors A, B, and C. I compute both Type II and Type III ANOVAs. Which of the following mean squares will be the same in the two tables (be sure to explain why)?

a. MS_{AB} will, in general, differ between the two. Type II has A+B+C+AC+BC as the base model, and Type III has A+B+C+AC+BC+ABC as the base model.

b. MS_{ABC} will be the same in both tables, because in both cases the base model is all main effects and two factor interactions.

c. MS_{E} will be the same in both tables, because error SS and df are computed after the full model has been fit, and the full model is the same in both cases.

Wings Summary For cooking wings, 40 minutes at 400 degrees is just about right, although longer times at lower temperatures will also produce properly cooked wings.

Data Analysis. The experiment has 36 units, two in each of the factor/level combinations of time (three levels) and temperature (three levels). It is a completely randomized design.

The residuals from an initial full model fit show no evidence of non-constant variance and slightly short tails for the residuals. There is no evidence of any problem with our assumptions.

The anova for the model shows all factors and interactions significant, although the interaction is somewhat marginal.

```
Response: cooked
            Df Sum Sq Mean Sq  F value   Pr(>F)
ftime        2 88.389  44.194 132.5833 1.09e-14 ***
ftemp        2 63.722  31.861  95.5833 5.62e-13 ***
ftime:ftemp  4  4.111   1.028   3.0833  0.03258 *
Residuals   27  9.000   0.333
```

Looking at the model coefficients, we find:

```
ftime
        40         45         50
-1.9722222  0.1111111  1.8611111

ftemp
       350        375        400
-1.5555556 -0.1388889  1.6944444
```

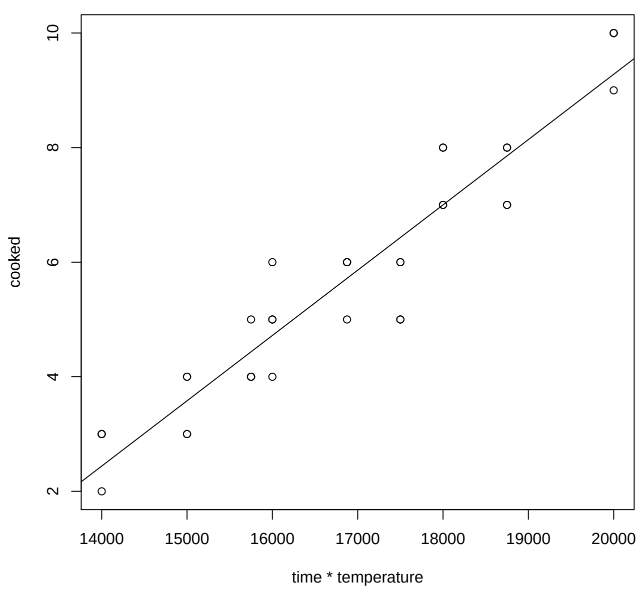
These seem almost perfectly linear in time and temperature. If we fit time and temperature as quantitative factors, then the sums of squares for the two linear terms are 88.167 and 63.375; nearly all of the time and temperature variation is simply linear in the levels of each factor. (We can also show that 4.000 of the 4.111 for the interaction can be described as the linear by linear interaction, although that is jumping ahead.)

The mean responses for the nine treatments are:

```
     ftemp
ftime  350  375  400
   40 2.75 3.50 5.00
   45 4.25 5.75 7.50
   50 5.50 7.50 9.75
```

Clearly, 40 minutes at 400 degrees gives wings that are just right; 50 minutes at 350 degrees is close but slightly over cooked; 45 minutes at 350 and 375 degrees bracket the just right setting. The standard error of these means is .289; using a Bonferroni correction for nine tests and 27 error df, the critical value for a 5% test is 3.01. None of these four combinations is significantly different from a score of 5 using this test.

This pattern suggests that "degree-minutes" are the key. That is, number of minutes times number of degrees seems predictive. If we look at the cooked score versus this product we get:

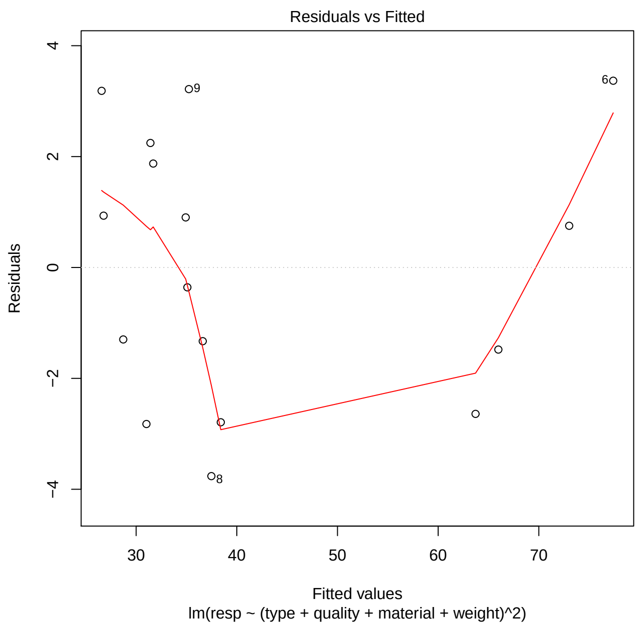


The correlation is over .95 and the line is the simple linear regression of cooked on the time by temperature product with intercept -13.52 and slope .00114. With 5 the desired score, this back calculates to 16,245 degree-minutes. For 400 degrees, this is 40.6 minutes; for 375 degrees, this is 43.3 minutes; for 350 degrees, this is 46.4 minutes. These fit into the table of means about where we would expect them to.

Glue summary. There is strong evidence that type of glue affects the force required to separate the halves, with liquid glue requiring more force than stick glue. There is also very strong evidence that type of material being glued matters, with cardboard requiring a much higher separation force than paper. There is no evidence that glue quality (name brand versus generic) has any effect at all. Surprisingly, there is at best marginal evidence that using a weight affects the force required to separate the halves. While liquid glue is better than stick glue for both materials, liquid glue is far superior to stick on cardboard.

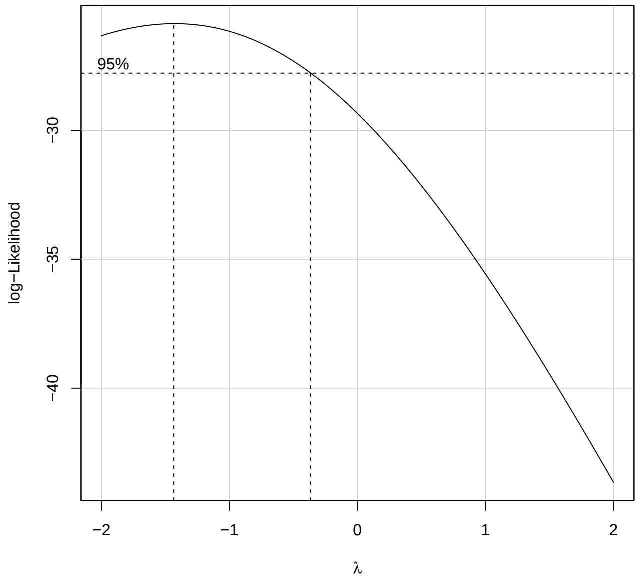
Data analysis. This is a completely randomized 2^4 design with sixteen experimental units and two measurement units per experimental unit. We use the average of the two responses as the response for the measurement unit.

Because there is only a single replication, we must fit a reduced model to have any degrees of freedom for error. We begin by fitting main effects and interactions, pooling the three and four factor interactions into a 5 df surrogate error term. The residual plot from this analysis seems to show some pattern suggesting that we have neglected an important interaction:



The line in the plot probably exaggerates the missing effect, but it is still worrisome.

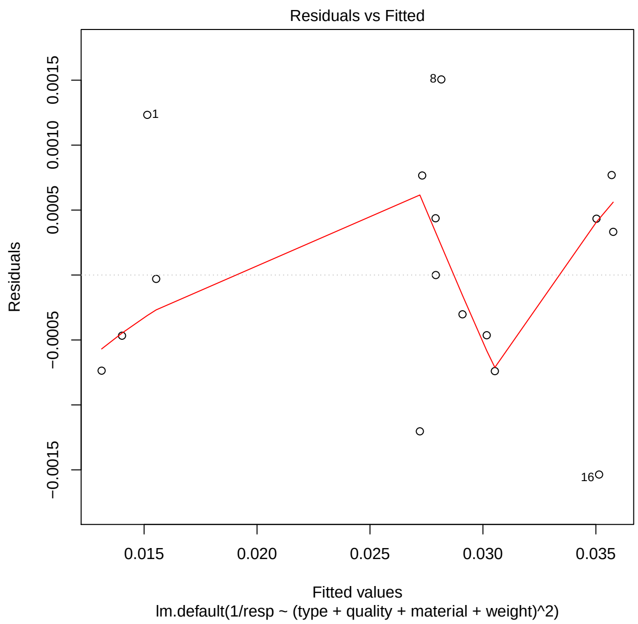
There are two choices at this point. The first is to seek a transformed scale where this interaction is not worrisome, and the second is to go to a model with only the four factor interaction as error, and then pool certain terms into error. We begin with the first approach. Using Box-Cox on the first fit suggests a reciprocal transformation:



The data are now analyzed on the reciprocal scale (reciprocal force, which we might characterize as "ease"). On this reciprocal scale and beginning with the model of main effects and two factor interactions, the only important factors are type of glue, type of material, and their interaction:

```
Response: 1/resp
                 Df     Sum Sq    Mean Sq  F value    Pr(>F)
type              1 0.00036300 0.00036300 166.8199 4.957e-05 ***
quality           1 0.00000005 0.00000005   0.0246  0.881497
material          1 0.00052257 0.00052257 240.1546 2.032e-05 ***
weight            1 0.00000280 0.00000280   1.2874  0.307982
type:quality      1 0.00000010 0.00000010   0.0448  0.840676
type:material     1 0.00005196 0.00005196  23.8786  0.004527 **
type:weight       1 0.00000329 0.00000329   1.5109  0.273690
quality:material  1 0.00000000 0.00000000   0.0005  0.982890
quality:weight    1 0.00000168 0.00000168   0.7738  0.419313
material:weight   1 0.00000000 0.00000000   0.0010  0.976599
Residuals         5 0.00001088 0.00000218
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

In addition, the residual plot is much less suggestive of a missing important interaction:

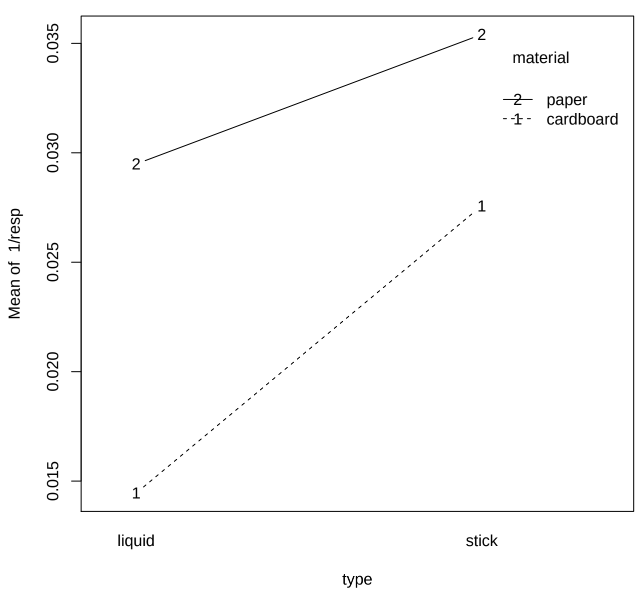


(Normality of residuals looks good too, but is not shown here.)

Here are the important means on the reciprocal scale:

```
        material
type       cardboard      paper
  liquid 0.01445380 0.02948782
  stick  0.02758415 0.03540990
```

as well as an interaction plot:



Clearly, liquid glue has less ease (more required force) for both materials, but the change of ease (or force) is about twice as large for cardboard relative to paper. The means on the original scale are:

```
 type*material effect
       material
type     cardboard    paper
  liquid  70.02625 33.91875
  stick   36.29500 28.26875
```

The alternative to transformation is to fit the model using only the four-factor interaction as surrogate error and seek to pool terms into error. Going down this route we obtain:

```
Response: resp
                        Df  Sum Sq Mean Sq   F value  Pr(>F)
type                     1 1550.88 1550.88 1302.3583 0.01764 *
quality                  1    4.28    4.28    3.5939 0.30901
material                 1 1947.79 1947.79 1635.6604 0.01574 *
weight                   1   51.18   51.18   42.9753 0.09637 .
type:quality             1    6.66    6.66    5.5951 0.25463
type:material            1  788.56  788.56  662.1927 0.02473 *
type:weight              1   62.83   62.83   52.7578 0.08710 .
quality:material         1    3.96    3.96    3.3213 0.31949
quality:weight           1    4.45    4.45    3.7342 0.30401
material:weight          1   31.74   31.74   26.6530 0.12180
type:quality:material    1   14.26   14.26   11.9749 0.17909
type:quality:weight      1    0.62    0.62    0.5224 0.60156
type:material:weight     1   67.92   67.92   57.0345 0.08381 .
quality:material:weight  1    1.33    1.33    1.1178 0.48228
Residuals                1    1.19    1.19
```

Only two of the three-way interactions can be pooled, and when we do that we get:

```
Response: resp
                      Df  Sum Sq Mean Sq   F value    Pr(>F)
type                   1 1550.88 1550.88 1479.8067 3.865e-05 ***
quality                1    4.28    4.28    4.0836 0.1365521
material               1 1947.79 1947.79 1858.5217 2.747e-05 ***
weight                 1   51.18   51.18   48.8308 0.0060160 **
type:quality           1    6.66    6.66    6.3575 0.0860703 .
type:material          1  788.56  788.56  752.4174 0.0001063 ***
type:weight            1   62.83   62.83   59.9462 0.0044807 **
quality:material       1    3.96    3.96    3.7739 0.1473300
quality:weight         1    4.45    4.45    4.2430 0.1315134
material:weight        1   31.74   31.74   30.2846 0.0118108 *
type:quality:material  1   14.26   14.26   13.6065 0.0345476 *
type:material:weight   1   67.92   67.92   64.8055 0.0040035 **
Residuals              3    3.14    1.05
```

In this model, quality is again not significant (except for one marginal three-way, which I am loathe to include with so little evidence for quality in the lower order terms). Weight does show as significant. However, type, material, and their interaction account for about 95% of all the sums of squares, again indicating that these are the dominant effects, even if some of the others might be construed as statistically significant.

Using a weight is thought to improve the bond thereby increasing the required force. However, in this data set, weight only has an effect for liquid glue on cardboard, and there it decreases the separating force: