General Factorial Designs

The results of the two-factorial design can be extended to the general case, where there are t_1 levels of factor 1, t_2 levels of factor 2, t_3 levels of factor 3, and so on. Based on this, there will be $t_1 t_2 t_3 \cdots r$ total observations. Remember $r \geq 2$ to determine a sum of squares due to error if all possible interactions are included in the model.

- The degrees of freedom for each main effect are the number of levels for that factor minus one (ex. $dof(SSF1) = t_1 - 1$).
- The degrees of freedoms for every two-way interaction are the product of main effect degrees of freedom for the two factors involved (ex. $dof(SSF1F2) = (t_1 - 1)(t_2 - 1)$).
- The degrees of freedoms for every three-way interaction are the product of main effect degrees of freedom for the three factors involved (ex. $dof(SSF1F2F3) = (t_1 - 1)(t_2 - 1)(t_3 - 1)$).
- and so on.

We can decompose the SSTot into orthogonal components, thus the $dof(SSE)$ can be determined from subtracting the $dof(SSTot)$ from the degrees of freedom from the components of the model. To date, we typically worked the other way — first we determined the ANOVA decomposition and then discussed the degrees of freedom. Let's consider the full three-factor model:

$$\begin{aligned} dof(SSE) = \quad dof(SSTot) \quad & -dof(SSF1) - dof(SSF2) - dof(SSF3) \\ & -dof(SSF1F2) - dof(SSF1F3) - dof(SSF2F3) \\ & -dof(SSF1F2F3). \end{aligned}$$

From the dof we can also determine the appropriate decomposition. Say we are interested in the following model:

$$\begin{aligned} y_{i,j,k,l} &= \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{i,j,k} + \epsilon_{i,j,k,l}, \\ y_{i,j,k,l} &= \bar{y}_{....} + (\bar{y}_{i...} - \bar{y}_{....}) + (\bar{y}_{.j..} - \bar{y}_{....}) + (\bar{y}_{..k.} - \bar{y}_{....}) \\ & +(\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....}) + (\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k.} + \bar{y}_{....}) + (\bar{y}_{.jk.} - \bar{y}_{.j..} - \bar{y}_{..k.} + \bar{y}_{....}) \\ & +(\bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{..k.} - \bar{y}_{....}) \\ & +error. \end{aligned}$$

- Can you figure out the error term?

Example (paper strength): How does the percentage of hardwood pulp concentration, vat pressure, and cooking time affect paper strength?

- Response: paper strength — units are not presented in the text.
- Hard: percentage of hardwood pulp concentration $\{i = 2, 4, 8\}$.
- Cook: cooking time in hours $\{j = 4, 5\}$.
- Pressure: vat pressure $\{k = 400, 500, 650\}$.

Design: A CRD with three factors and two replicates per treatment combination was conducted (r=2).

Analysis: The following model was fit to the data:

$$y_{i,j,k,l} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{i,j,k} + \epsilon_{i,j,k,l}.$$

The ANOVA table:

```
> full <- lm(y~hard*cook*pressure)
> anova(full)

Analysis of Variance Table

Response: y
                   Df  Sum Sq Mean Sq F value    Pr(>F)
hard                2  7.7639  3.8819 10.6193 0.0008996
cook                1 20.2500 20.2500 55.3951 6.745e-07
pressure            2 19.3739  9.6869 26.4992 4.327e-06
hard:cook           2  2.0817  1.0408  2.8473 0.0842597
hard:pressure       4  6.0911  1.5228  4.1657 0.0146262
cook:pressure       2  2.1950  1.0975  3.0023 0.0749564
hard:cook:pressure  4  1.9733  0.4933  1.3495 0.2903053
Residuals          18  6.5800  0.3656
```

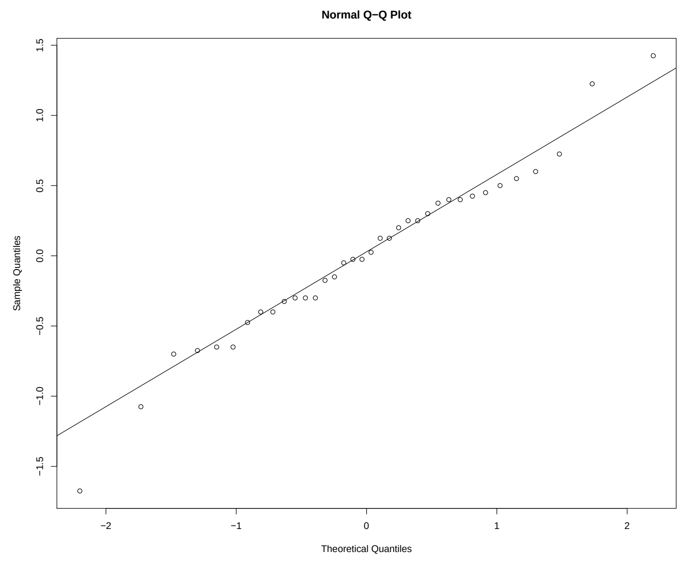
Let's fit the reduced model:

$$y_{i,j,k,l} = \mu + a_i + b_j + c_k + (ac)_{ik} + \epsilon_{i,j,k,l}.$$

```
> sub <- lm(y~hard + cook + pressure + hard:pressure)
> anova(sub)
Analysis of Variance Table

Response: y
              Df  Sum Sq Mean Sq F value    Pr(>F)
hard           2  7.7639  3.8819  7.8668   0.00213 **
cook           1 20.2500 20.2500 41.0366 8.711e-07 ***
pressure       2 19.3739  9.6869 19.6306 6.370e-06 ***
hard:pressure  4  6.0911  1.5228  3.0859   0.03322 *
Residuals     26 12.8300  0.4935
```

Now let's at least do a check of normality using the residuals — of course a full analysis would also conduct a check of the constant variance assumption.


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2
−1
0
1
2