Two-level full factorial designs

What are two level factorial designs?

- These are designs where the factors can only take on two levels
- The designs we will see are all complete, balanced, and unreplicated (i.e. $r = 1$). If there are replicates, we can analyze them with the tools we already know.
- We won't cover them in class, but there are ways to analyze incomplete factorial designs. In the literature, they're referred to as fractional factorial designs... Really useful in practice if we're on a budget

Nomenclature

- These designs are widely used in industrial statistics, and they come with their own nomenclature
- We usually refer to these designs as "2^k designs", where k is the number of factors. The name comes from the fact that if we have, say, 3 factors with 2 levels each, there are 2^3 combinations of 3 factors.
- Since these designs are unreplicated, the sample size of a 2^k the experiment is 2^k
- The levels of the factors are usually encoded as -1 and 1; this is not important (they are just labels)

What's special about these designs?

- These designs have a few peculiarities; we won't cover them in detail here, but you will see them if you take the course on industrial statistics next year
- Recall that, since the design is unreplicated, we can't do F-tests if we include all the interaction terms
- However, in 2^k designs, there's a clever strategy that allows us to identify "important" effects

Daniel plot

- If all the effects were insignificant, they would all be independent draws from a normal distribution centered at zero with the same variance
- This property is a consequence of working with a complete, balanced, two-level design; it isn't true in general
- Given this property, we can do a qq-plot of effects with a normal; if most effects are lined up with the exception of a few effects that stand out, the ones that stand out are the important effects
- The name of this qq-plot of effects is Daniel plot

Example: Spring

Response: number of compressions until a spring breaks

Factors:

- Length: 10 or 15cm
- Girth: 5 or 7mm
- Steel: Type A or B

2^3 design: 8 runs and 3 factors

Example: Read in data

```r
# read in data
molla = read.csv("http://vicpena.github.io/doe/molla.csv")

# create design matrix with FrF2
library(FrF2)
design = FrF2(nruns = 8,  nfactors = 3,
       factor.names = c("Longitud", "Gruix", "Tipus"),
       randomize = FALSE)
# add in response variable
design = add.response(design, molla$Comp)
```

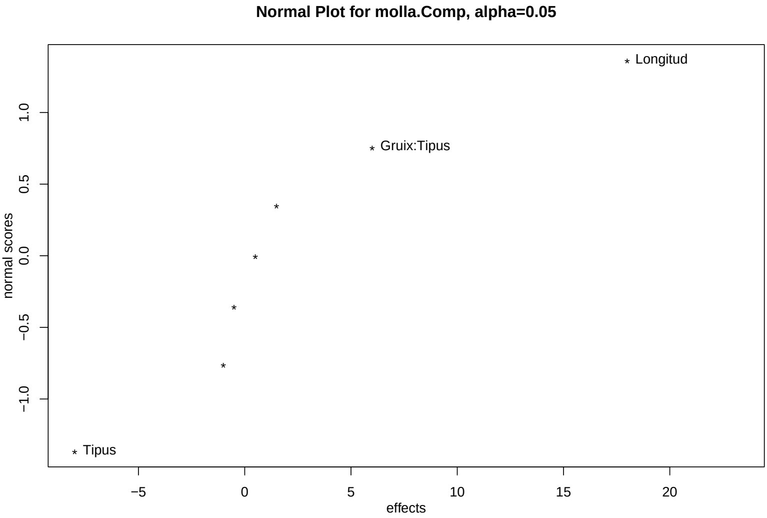
Example: ANOVA table

```r
# since r = 1, can't do any F-testing
mod = aov(molla.Comp ~ Longitud*Gruix*Tipus,
          data = design)
summary(mod)
```

```
                     Df Sum Sq Mean Sq
Longitud              1  648.0   648.0
Gruix                 1    4.5     4.5
Tipus                 1  128.0   128.0
Longitud:Gruix        1    2.0     2.0
Longitud:Tipus        1    0.5     0.5
Gruix:Tipus           1   72.0    72.0
Longitud:Gruix:Tipus  1    0.5     0.5
```

Example: Daniel plot

```
DanielPlot(design)
```



Example: Refitting model

Refit model with “important” terms flagged by `DanielPlot`

```
mod2 = aov(molla.Comp ~ Longitud + Gruix*Tipus,
           data = design)
summary(mod2)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
Longitud      1  648.0   648.0   648.0 0.000133 ***
Gruix         1    4.5     4.5     4.5 0.124027    
Tipus         1  128.0   128.0   128.0 0.001481 ** 
Gruix:Tipus   1   72.0    72.0    72.0 0.003437 ** 
Residuals     3    3.0     1.0                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

Example: `emmip`

```
library(emmeans)
emmip(mod2, Longitud ~ Gruix | Tipus)
```