Unreplicated Experiments

When exploring the effects of many factors, you may not have the resources required for replication.

Degrees of freedom for pure error are $(n - 1) \times 2^k$, so $n = 1$ means zero degrees of freedom.

"Sparsity of effects" principle

Most systems are dominated by some main effects and low-order interactions, and most high-order interactions are negligible.

Possible approaches

- assume some high-order interactions are zero, and fit a model that excludes them; degrees of freedom go into error, so testing is possible;
- graphical methods–normal and half-normal probability plots; no formal tests;
- others.

First approach is not recommended, as it depends on knowing that certain effects are negligible.

Example: Filtration rate

Response: filtration rate (gal/h);

Factors:

- A temperature;
- B pressure;
- C concentration of formaldehyde;
- D stirring rate.

Data

filtration.txt:

```
A B C D Rate
- - - -  45
+ - - -  71
- + - -  48
+ + - -  65
- - + -  68
+ - + -  60
- + + -  80
+ + + -  65
- - - +  43
+ - - + 100
- + - +  45
+ + - + 104
- - + +  75
+ - + +  86
- + + +  70
+ + + +  96
```

Analysis (all factors were coded)

```
filtration <- read.table("data/filtration.txt", header = TRUE)
for (j in 1:4)
    filtration[, j] <- coded(filtration[, j])
summary(lm(Rate ~ A * B * C * D, filtration))
summary(aov(Rate ~ A * B * C * D, filtration))
```

Output of lm()

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    70.0625         NA      NA       NA
A              10.8125         NA      NA       NA
B               1.5625         NA      NA       NA
C               4.9375         NA      NA       NA
D               7.3125         NA      NA       NA
A:B             0.0625         NA      NA       NA
A:C            -9.0625         NA      NA       NA
B:C             1.1875         NA      NA       NA
A:D             8.3125         NA      NA       NA
B:D            -0.1875         NA      NA       NA
C:D            -0.5625         NA      NA       NA
A:B:C           0.9375         NA      NA       NA
A:B:D           2.0625         NA      NA       NA
A:C:D          -0.8125         NA      NA       NA
B:C:D          -1.3125         NA      NA       NA
A:B:C:D         0.6875         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
```

Output of aov()

	Df	Sum Sq	Mean Sq
A	1	1870.6	1870.6
B	1	39.1	39.1
C	1	390.1	390.1
D	1	855.6	855.6
A:B	1	0.1	0.1
A:C	1	1314.1	1314.1
B:C	1	22.6	22.6
A:D	1	1105.6	1105.6
B:D	1	0.6	0.6
C:D	1	5.1	5.1
A:B:C	1	14.1	14.1
A:B:D	1	68.1	68.1
A:C:D	1	10.6	10.6
B:C:D	1	27.6	27.6
A:B:C:D	1	7.6	7.6

Exclusion approach

Include only main effects and two-factor interactions:

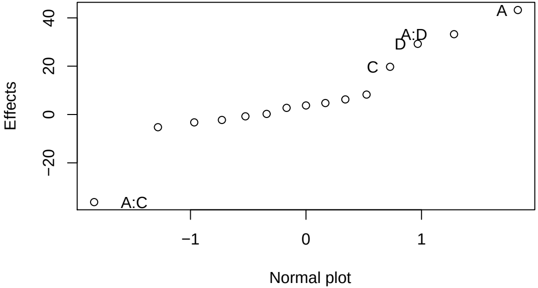
```
summary(aov(Rate ~ (A + B + C + D)^2, filtration))
```

Output

```
             Df  Sum Sq Mean Sq F value    Pr(>F)
A             1 1870.56 1870.56 73.1760 0.0003596 ***
B             1   39.06   39.06  1.5281 0.2712969
C             1  390.06  390.06 15.2592 0.0113371 *
D             1  855.56  855.56 33.4694 0.0021718 **
A:B           1    0.06    0.06  0.0024 0.9624777
A:C           1 1314.06 1314.06 51.4059 0.0008208 ***
A:D           1 1105.56 1105.56 43.2494 0.0012200 **
B:C           1   22.56   22.56  0.8826 0.3906126
B:D           1    0.56    0.56  0.0220 0.8878710
C:D           1    5.06    5.06  0.1980 0.6749089
Residuals     5  127.81   25.56
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

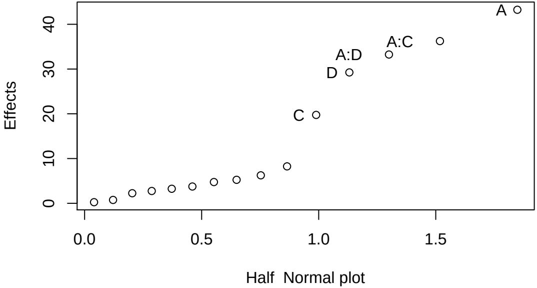
QQ-plot (probability plot) of effects

```
library(gplots)
qqnorm(aov(Rate ~ A * B * C * D, filtration), full = TRUE)
```



Half-normal plot of effects

```
qqnorm(aov(Rate ~ A * B * C * D, filtration))
```



Design Projection

Active effects are identified as A, C, D, AC, AD.

Factor B (pressure) not involved in any active effect.

Ignoring B *projects* the 2^4 unreplicated design onto a replicated 2^3 design.

ANOVA as a replicated 2^3 design in A, C, and D:

```
           Df  Sum Sq Mean Sq F value    Pr(>F)    
A           1 1870.56 1870.56 83.3677 1.667e-05 ***
C           1  390.06  390.06 17.3844 0.0031244 ** 
D           1  855.56  855.56 38.1309 0.0002666 ***
A:C         1 1314.06 1314.06 58.5655 6.001e-05 ***
A:D         1 1105.56 1105.56 49.2730 0.0001105 ***
C:D         1    5.06    5.06  0.2256 0.6474830    
A:C:D       1   10.56   10.56  0.4708 0.5120321    
Residuals   8  179.50   22.44                      
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Estimate of error based on “hidden replication”; not pure error.

Interaction plots:

```
with(filtration, interaction.plot(A, C, Rate))
with(filtration, interaction.plot(A, D, Rate))
with(filtration, interaction.plot(A, C * D, Rate))
```

Other Methods

In any QQ plot, the slope of the line gives an estimate of σ^2.

In the half-normal plot, an estimate could be found from the slope of the line involving the small effects.

Lenth's method is related; implemented in JMP, and in the rsm package for R.