Replication

BIOE 498/598 PJ

Spring 2022

Why replication?

1. Reduce noise effects.
2. Estimate confidence intervals for effect sizes.
3. Analyze dispersion effects.

Does noise matter?

```r
data <- read.csv("LeafSpring.csv")
data1 <- data[ 1:16, ]
data2 <- data[17:32, ]
data3 <- data[33:48, ]
```

Effect sizes differ for each replicate

Replicate 1

```
##          B     .09937
##          C     .09313
##          Q    -.08688
##        C:Q    -.07563
##        B:Q     .07313
##          E     .07312
##      B:D:Q    -.06687
##          D     .05688
##        C:D     .04313
##        D:Q     .02312
```

Replicate 2

```
##          Q    -.14688
##          B     .12937
##          C     .11063
##        C:Q    -.07688
##        D:Q     .06687
##          E     .06312
##      C:D:Q    -.05063
##        B:Q     .04438
##        B:C     .04187
##        C:D    -.02812
```

Duplicates vs. Replicates

- We use replicates to estimate **confidence intervals** for effect sizes.
- More replicates = narrower confidence intervals
- Treating duplicates as replicates artificially narrows the confidence intervals for two reasons.
  1. Inflating the degrees of freedom fakes additional statistical power
  2. Duplicates typically have less variation than replicates

Duplicates vs. Replicates

- We use replicates to estimate **confidence intervals** for effect sizes.
- More replicates = narrower confidence intervals
- Treating duplicates as replicates artificially narrows the confidence intervals for two reasons.
    1. Inflating the degrees of freedom fakes additional statistical power
    2. Duplicates typically have less variation than replicates
- Statistical significance $\neq$ Practical significance

Sample variance across replicates

If a run is replicated r times with responses y_1, y_2, ..., y_r and mean $\bar{y}$,

$$\text{sample variance} = s^2 = \frac{\sum_i^r (y_i - \bar{y})^2}{r-1}$$

Sample variance across replicates

If a run is replicated r times with responses y_1, y_2, ..., y_r and mean $\bar{y}$,

$$\text{sample variance} = s^2 = \frac{\sum_i^r (y_i - \bar{y})^2}{r-1}$$

For a factorial design with N *unreplicated* runs ($N = 2^k$ for a full factorial or $N = 2^{k-p}$ for a fractional factorial),

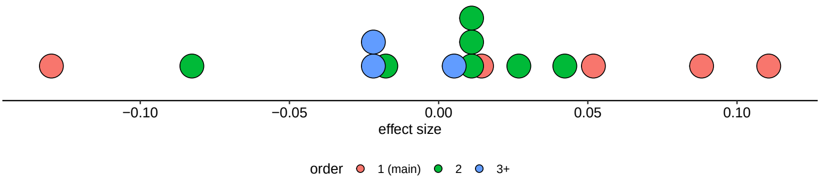
$$\text{standard error of effects} = SE(\beta_i) = \sqrt{\frac{\text{mean}(s^2)}{rN}}$$

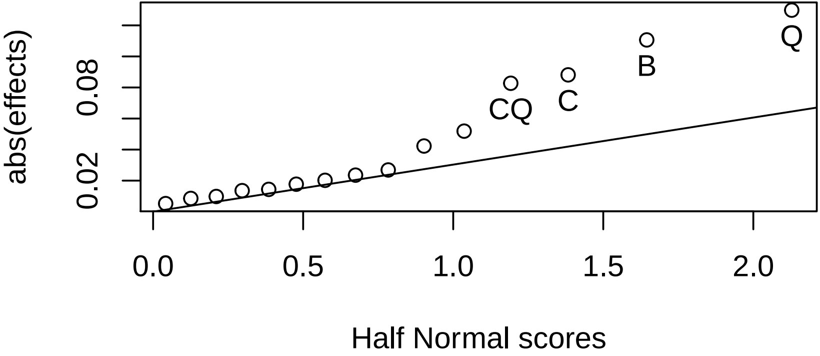
Linear models find the “best fit” effect sizes

```
model <- lm(height ~ B*C*D*E*Q, data=data)
show_model(model, n_coefs=17, show_fit=FALSE)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.636042   0.018571 411.183  < 2e-16 ***
## Q           -0.129792   0.018571  -6.989 6.42e-08 ***
## B            0.110625   0.018571   5.957 1.23e-06 ***
## C            0.088125   0.018571   4.745 4.16e-05 ***
## C:Q         -0.082708   0.018571  -4.454 9.64e-05 ***
## E            0.051875   0.018571   2.793  0.00874 **
## B:Q          0.042292   0.018571   2.277  0.02959 *
## D:Q          0.026875   0.018571   1.447  0.15758
## C:D:Q       -0.023542   0.018571  -1.268  0.21406
## B:D:Q       -0.020208   0.018571  -1.088  0.28465
## C:D         -0.017708   0.018571  -0.954  0.34746
## D            0.014375   0.018571   0.774  0.44458
## E:Q          0.013542   0.018571   0.729  0.47119
## B:D          0.009792   0.018571   0.527  0.60165
## B:C          0.008542   0.018571   0.460  0.64866
## B:C:Q        0.005208   0.018571   0.280  0.78093
## B:E                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Half-normal & dot plots — significance based only on effect size






```
## zscore= 0.0417893 0.1256613 0.2104284 0.2967378 0.3853205 0.4770404
```