

# 1. Use the bootstrap to build a null distribution and calculate a p-value.

We observe two independent random samples:

$$\begin{aligned} F \rightarrow & \quad \mathbf{z} = (z_1, z_2, \dots, z_n) \text{ independently of} \\ G \rightarrow & \quad \mathbf{y} = (y_1, y_2, \dots, y_m) \end{aligned}$$

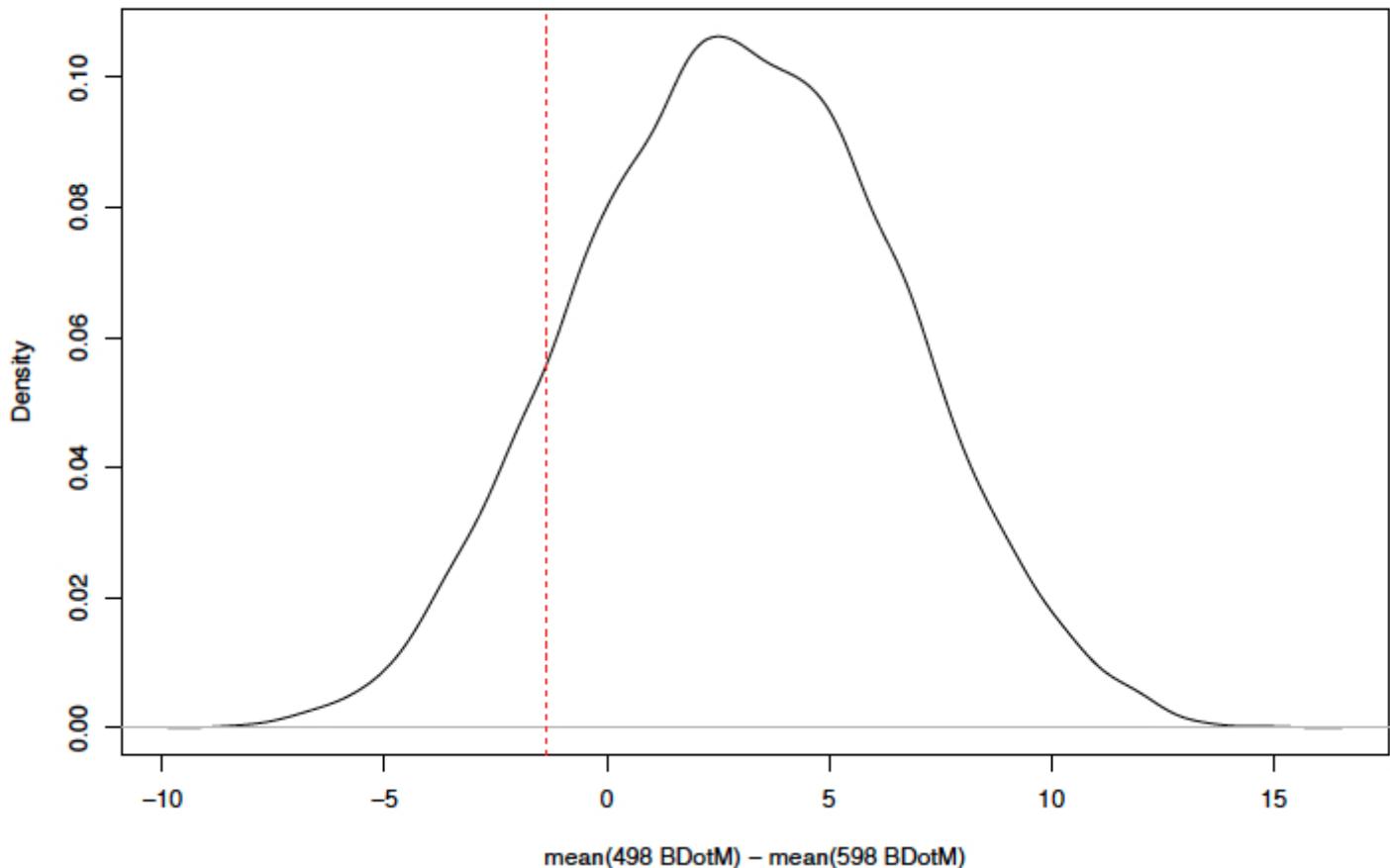
And we wish to test the *null hypothesis* of no difference between F and G,

$$H_0: F = G$$

## Bootstrap Hypothesis Testing $F = G$

- Denote the combined sample by  $\mathbf{x}$ , and its empirical distribution by  $\hat{F}_0$ .
  - Under  $H_0$ ,  $\hat{F}_0$  provides a non parametric estimate for the common population that gave rise to both  $\mathbf{z}$  and  $\mathbf{y}$ .
1. Draw  $B$  samples of size  $n + m$  **with replacement** from  $\mathbf{x}$ . Call the first  $n$  observations  $\mathbf{z}^*$  and the remaining  $m - n$   $\mathbf{y}^*$
  2. Evaluate  $t(\cdot)$  on each sample -  $t(\mathbf{x}^{*b})$
  3. Approximate  $ASL_{boot}$  by  
$$\widehat{ASL}_{boot} = \#\{t(\mathbf{x}^{*b}) \geq t(\mathbf{x})\}/B$$
- \* In the case that large values of  $t(\mathbf{x}^{*b})$  are evidence against  $H_0$

Bootstrap Null Distribution



## 2. Use and interpret the results of a t-test.

```
t.test(days498, days598, alternative="two.sided",
       var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  days498 and days598
## t = -0.45729, df = 28, p-value = 0.651
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
## -7.458096  4.735874
## sample estimates:
## mean of x mean of y
## 13.88889 15.25000
```

statistic: the value of the t-statistic.

conf.int: a confidence interval for the mean appropriate to the specified alternative hypothesis.

estimate: the estimated mean or difference in means depending on whether it was a one-sample test or a two-sample test.

### 3. Vocab: response, predictor, factor, intercept, coefficient, effect size, parameter, residual.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

- ▶  $y$  is the **response**
- ▶  $x_i$  is a **predictor** or **factor**
- ▶  $\beta_0$  is the **intercept**
- ▶  $\beta_i, i > 0$  is a **coefficient**, **effect size**, or **parameter**
- ▶  $\epsilon$  is a **residual**

### 4. Use effect sizes to relate changes in factor levels to changes in the response.

### 5. Use linear models for hypothesis testing.

t test.

Multivariate linear models consider how all factors affect the response.

### 6. Explain the meaning of interactions.

Interactions are modeled as the product of variables.

on/off → interaction when both "on"

$x_1$	$x_2$	$x_1 x_2$
0	0	0
0	1	0
1	0	0
1	1	1

high/low → interaction when both "high" or both "low"

$x_1$	$x_2$	$x_1 x_2$
-1	-1	+1
-1	+1	-1
+1	-1	-1
+1	+1	+1

Higher order interactions are possible but are rare.

## 7. Calculate the number of interactions in a model with n factors.

A model with n factors has  $2^n$  possible terms;  $2^n - n - 1$  of these are interactions

## 8. Explain how transformations affect the relationship between factors and response.

**Log transformation:** the model changes from an additive model to a multiplicative one.

Models with transformed responses are more difficult to interpret. There is a tradeoff between prediction and interpretation.

## 9. Transformations: mean centering, z-scoring, rescaling to compare binary and continuous factors.

**scaling:**  $y = \beta_0 + \beta_1(kx) + \epsilon \rightarrow y = \beta_0 + (k\beta_1)x + \epsilon$

**mean centering:**  $\bar{x} = x - \text{mean}(x)$  when intercept is uninterpretable.

$\beta_1$  remains the increase in response given a unit increase in factor;  $\beta_0$  is the predicted response for an factor is on average value.

**z-scoring:**  $\hat{x} = \frac{x - \text{mean}(x)}{\text{stdev}(x)}$

$\beta_1$  is the change in response based on an increase of one standard deviation in factor.

Why rescale by two standard deviations?

Assume a binary factor takes the value 1 with probability  $p$ . Then the standard deviation of this factor is

$$\sqrt{p(1-p)}$$

Without any additional knowledge, we assume that  $p = 0.5$ . Then the standard deviation of the factor is  $\sqrt{0.5^2} = 0.5$ .

When this binary factor switches from 0 to 1, it is moving two standard deviations. To keep things on the same scale, the continuous variables should be rescaled so a unit change also corresponds to two standard deviations.

When to use scaling?

- leave binary factors unscaled
- mean center and scale continuous factors by 1 stdev.
- if continuous variables only have a few discrete values, used coded factors
- if having both binary and continuous variables, center and scale continuous factors by 2 stdev.

## 10. Apply and interpret the results of a Box-Cox analysis.

A general family of nonlinear transformations are described by the Box-Cox transformation:

$$T(y) = \begin{cases} (y^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

- ▶  $\lambda = 2$  suggests  $y \rightarrow y^2$
- ▶  $\lambda = 1$  suggests no transformation
- ▶  $\lambda = 1/2$  suggests  $y \rightarrow \sqrt{y}$
- ▶  $\lambda = -1$  suggests  $y \rightarrow 1/y$

Box-Cox suggests a common transformation.

## 11. Vocab: run, experiment, experimental unit, replicate, duplicate, background variable, effect, experimental design, confounded factors, biased factors, bias error, random error.

- ▶ **Run or Experiment.** A single action with the change of at least one variable followed by an observation.
  - ▶ **Experimental Unit.** An instance of the item under study that is changed.
  - ▶ **Replicate.** Two or more runs conducted with the same settings on different experimental units.
    - ▶ The responses of replicates vary due to differences inherent in experimental units or the lurking variables.
  - ▶ **Duplicates.** Multiple measurements on the same experimental unit.
    - ▶ Duplicates should always be averaged before analysis.
- 
- ▶ **Factor or Independent Variable.** The thing under study that can be controlled or changed.
  - ▶ **Background or Lurking Variable.** A thing we are unaware of or cannot control.
  - ▶ **Response or Dependent Variable.** The things we measure. Depends on the settings of the factors and background.
  - ▶ **Effect.** The change in response due to a change in factor or background.
    - ▶ **Calculated effects** come from the model.
    - ▶ **Practical effects** come from our knowledge of the system.
- 
- ▶ **Experimental Design.** Collection of experiments planned in advance.
  - ▶ **Confounded Factors.** When a change in factor corresponds with an identical change in another factor.
  - ▶ **Biased Factors.** When a change in a factor coincides to a change in a lurking variable.
  - ▶ **Experimental Error.** Difference between observed response and long run average of all experiments with the same settings.
    - ▶ There is nothing wrong with error.
    - ▶ **Bias error** remains constant or changes consistently.
    - ▶ **Random error** changes unpredictably and averages to zero.

## 12. Explain the differences between continuous, ordinal, and nominal factors.

- ▶ **Numerical** (or **continuous**) variables are modeled by real numbers using a single coefficient.
- ▶ **Ordinal** variables have discrete but *ordered* levels. If the levels are evenly spaced, we model them using integers.
- ▶ **Nominal** (or **categorical**) variables are unordered with no numeric relationship between levels.

## 13. Apply one-hot encoding to nominal factors.

- ▶ In one-hot encoding, a nominal variable with  $k$  levels is modeled with  $k$  binary dummy variables.
- ▶ Only one dummy variable is nonzero ("hot") at a time.
- ▶ Example: DNA  $\in \{A, C, G, T\}$ .

$$\beta_A x_A + \beta_C x_C + \beta_G x_G + \beta_T x_T$$

## 14. Explain why degeneracy arises in models with an intercept and multilevel factors.

redundant constraints. The design matrix with an intercept is not full rank, so there could be many coefficients resulting in the same predictions and residuals.

$$\beta_0 - \Delta, \beta_1 + \Delta, \beta_2 + \Delta, \dots$$

## 15. Define and interpret contrasts.

Contrast is a linear combination of variables whose coefficients add up to 0, allowing comparison of different treatment with multiple conditions.

# Contrasts in the Rothamsted Experiment

- ▶ (A) no fertilizer
- ▶ (B) plowed fertilizer in January
- ▶ (C) broadcast fertilizer in January
- ▶ (D) broadcast fertilizer in April

What if we wanted to make other comparisons?

- ▶ Effect of broadcast vs. plowed: (C & D) = (B)
- ▶ Effect of early vs. late application: (B & C) = (D)
- ▶ Effect of any fertilizer: (A) = (B, C, & D)

There are all *contrasts*, or comparisons between variables. The null hypotheses for each contrast can be written as a linear combination of the model's variables:

$$\frac{1}{2}\beta_C + \frac{1}{2}\beta_D - \beta_B = 0$$

When specifying contrasts, we require that the coefficient sum to zero (hence the 1/2 factors above).

Adjust p-value threshold (Tukey's HSD method) when testing all contrasts.

## 16. Determine if a contrast is estimable.

- its coefficients sum to zero
- a linear combination of the rows of the design matrix

## 17. Understand and apply blocking factors.

To account for differences between each block of runs when the runs cannot be performed under the same condition.

The blocking factor is included as a main effect in the model and adjusts the intercept for each group of runs.

Couldn't the main effects and interactions also be different in each block?

Yes, but there are two issues when correcting this:

1. Adding multiple blocking factors reduces our degrees of freedom, requiring more runs.
2. Some of these effects are confounded with background variables in the block.

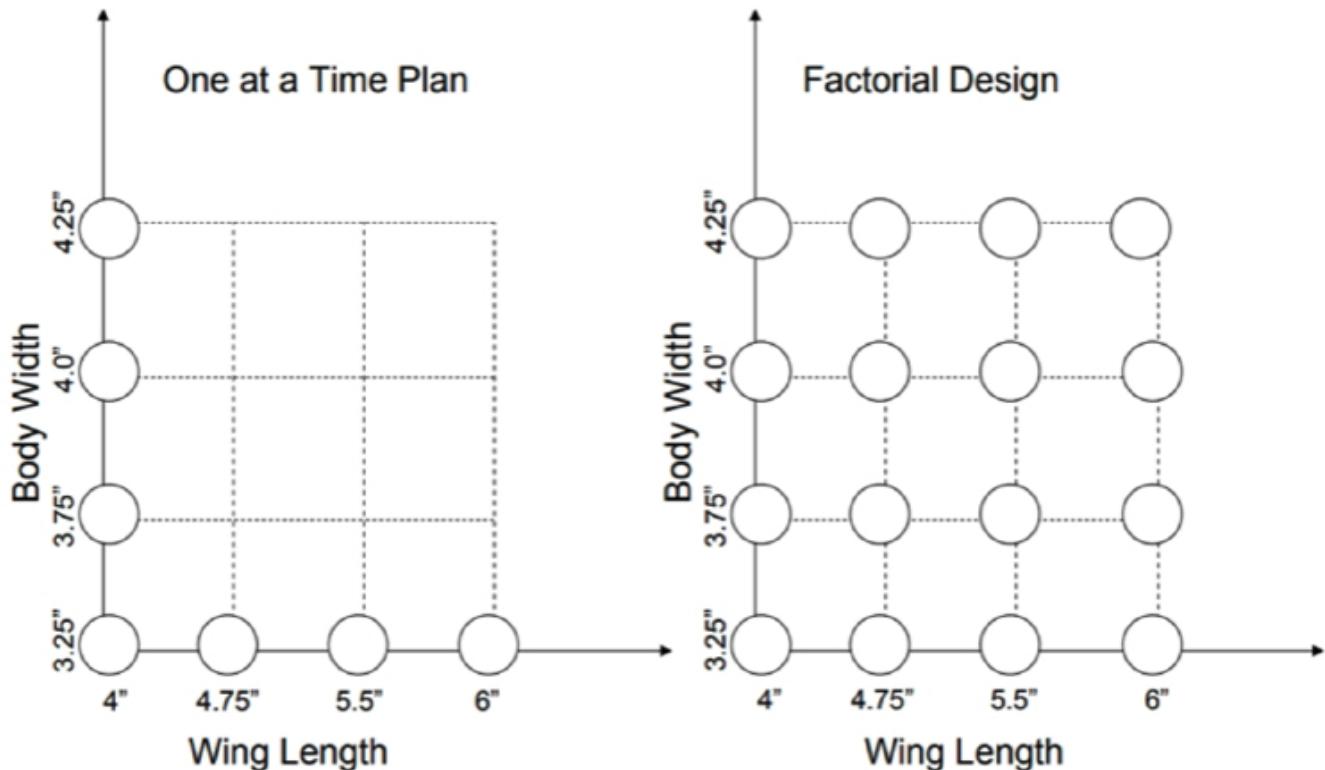
**Our recommendation:** Always include a blocking factor when needed, but assume the blocking effects are limited to the intercept.

## 18. Explain the advantages and disadvantages of factorial designs.

A factorial design studies multiple factors at discrete intervals. It includes runs with every combination of factors set at every level.

**Pro:**

- find better optima



- more efficient
- make better estimates of effect sizes

For OFAT designs:

$\beta_i$  is the effect of moving  $x_i$  from – to +  
**while all other factors stay at –.**

For factorial designs:

$\beta_i$  is the effect of moving  $x_i$  from – to +  
**averaged over all other factors at all levels.**

FD are nested.

A model is solvable if the design matrix is full rank but need extra rows to estimate the model's uncertainty. To estimate the error in FD,

1. perform replicates of some (or all) runs
  2. only estimate a subset of the  $2^n$  coefficients
  3. some combination of 1 & 2
- ▶ For small  $n$  designs we perform replicates since there are already few runs and the interactions are probably significant.
  - ▶ For large  $n$  designs we drop coefficients for higher order terms since we already have lots of runs and the higher-order interactions are most likely zero.

**Cons:**

- the number of runs is prohibitive
- rarely need to higher-order interactions

## 19. Calculate the number of runs for a factorial design.

Imagine an experiment with four factors, each with two levels (-, +). We want three replicates for each level.

### One Factor at a Time Design

- ▶ 3 runs at level (-)
- ▶ 4 factors  $\times$  3 runs at (+) = 12 runs
- ▶ **15 total runs**

### Factorial Design

- ▶  $2^4 = 16 \text{ total runs}$

A factorial design in n variables has  $2^n$  runs, but  $2^{(n-1)}$  replicates at each level. Adding another replicate: OFAT is nk; FD is about k.

## 20. Find the degrees of freedom in a model.

DoF=N-k-1, N is the number of observations, k is the number of variables

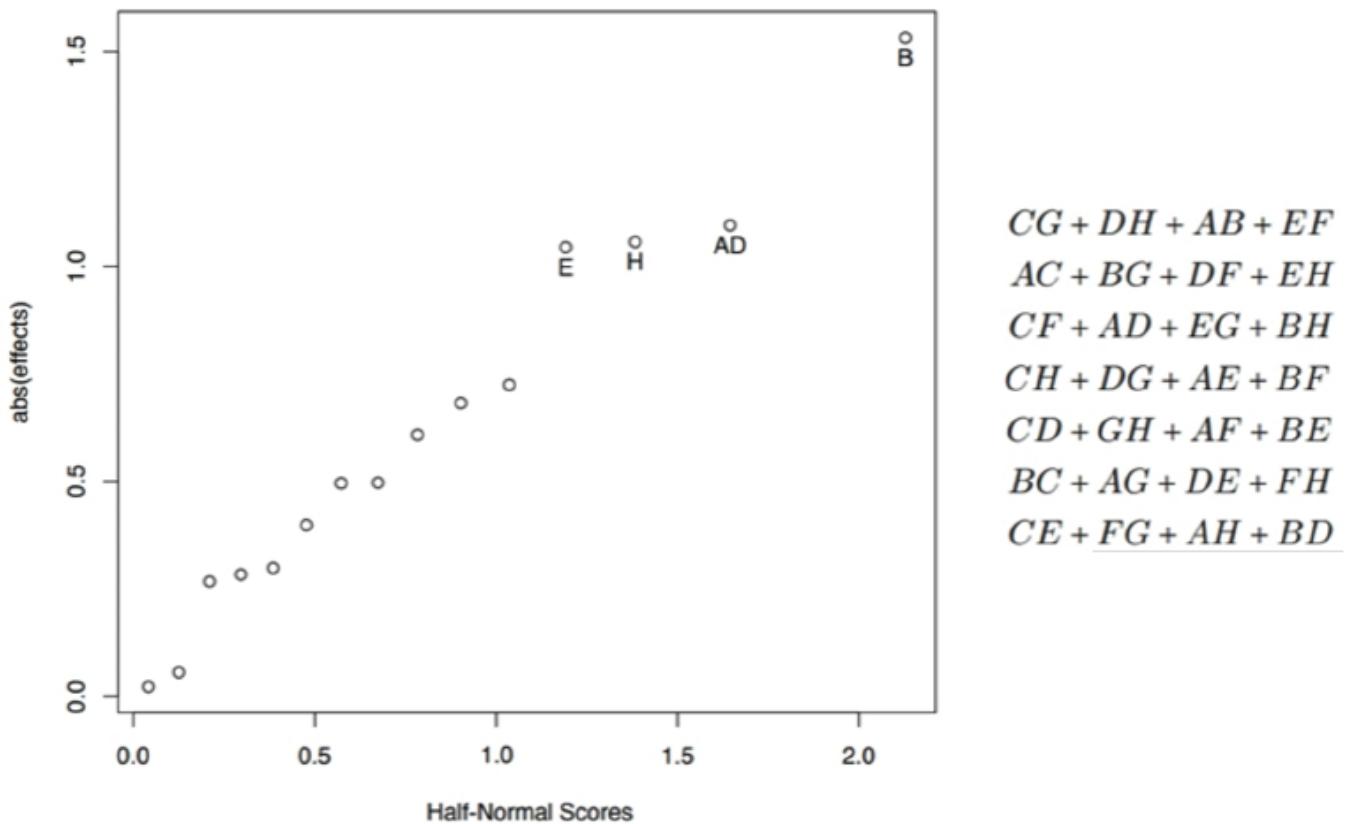
## 21. Explain and interpret half-normal plots.

No DoF to estimate all interactions. No DoF to estimate confidence intervals.

Instead, all factor levels are coded to units -1 and +1, thus the effect sizes are directly comparable.

Assume **practical significance** of an effect is proportional to its magnitude.

Figure 6.5 Half-Normal Plot of Effects from  $2^{8-4}$  Paecilomyces variotii Culture Experiment



- ▶ Assumes the variables are coded so the effects sizes can be compared directly.
- ▶ Provides a ranked list of factors based on *practical significance*.
- ▶ Assuming normality of effect sizes allows estimation of *statistical significance* using z-scores.
- ▶ Great for screening designs; practically insignificant factors are dropped for a follow-up design with replicates or higher resolution.

## 22. Vocab: effect sparsity principle, hierarchical ordering principle, heredity principle.

**Effect sparsity principle:** only a small proportion of the factors in an experiment will have significant effects.

**Hierarchical ordering principle:**

- lower order effects are more likely to be important than higher order effects.

- effects of the same order are equally likely to be important.

**heredity effect:** a model that includes an interaction should also include the corresponding main effect.

## 23. Vocab: practical and statistical significance.

**practical significance:** the magnitude of the effect

**statistical significance:** assume effect sizes normally distributed with mean zero. The z-score of the effect sizes can be compared with a standard normal to find a p-value. It refers to whether the effect exists.

## 24. Fractional Factorial Designs

- Use generators to derive the defining relation.

confounding: factors vary together that cannot estimate effects separately.  $D = ABC$

$$\beta_{D|ABC} = \beta_D + \beta_{ABC} \approx \beta_D.$$

We replace the highest interaction (ABC) with D and fill in the rest of the interactions.

I	A	B	C	AB	AC	BC	D= ABC	AD	BD	CD	ABC	BCD	ABD	ACD	ABCD
+	-	-	-	+	+	+	-	+	+	+	-	-	-	-	+
+	+	-	-	-	-	+	+	+	-	-	+	+	-	-	+
+	-	+	-	-	+	-	+	-	+	-	+	-	-	+	+
+	+	+	-	+	-	-	-	-	-	+	-	+	-	+	+
+	-	-	+	+	-	-	+	-	-	+	+	-	+	-	+
+	+	-	+	-	+	-	-	-	+	-	-	+	+	-	+
+	-	+	+	-	+	-	+	-	-	-	-	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

All of the variables are now confounded:

$$\begin{array}{ll}
 A + BCD & AB + CD \\
 B + ACD & AC + BD \\
 C + ABD & AD + BC \\
 D + ABC & I + ABCD
 \end{array}$$

Generator:  $XX = I$  and  $IX = X$

$$\begin{aligned}
 D &= ABC \\
 (D)D &= (ABC)D \\
 D^2 &= ABCD \\
 I &= ABCD
 \end{aligned}$$

This last statement ( $I=ABCD$ ) is called the *defining relation* for the design with generator  $D=ABC$ .

- Use the defining relation to compute confounding structure.

For A:

$$\begin{aligned}
 A(I) &= A(ABCD) \\
 A &= A^2BCD \\
 &= IBCD \\
 &= BCD
 \end{aligned}$$

For the interaction CD:

$$\begin{aligned}
 CD(I) &= CD(ABCD) \\
 AB &= ABC^2D^2 \\
 &= AB
 \end{aligned}$$

- Compute and interpret the resolution, aberration, and clarity of a design.

**Resolution:** difference in the level of confounding. The length of shortest word in the defining relation.

It measures the degree of the confounding. The resolution R design has no i-level interaction aliased with effects lower than  $R-i$ .

### Resolution III

- ▶ Main effects ( $i = 1$ ) are confounded with secondary ( $3 - 1 = 2$ ) interactions.

### Resolution IV

- ▶ Main effects ( $i = 1$ ) are confounded with tertiary ( $4 - 1 = 3$ ) interactions.
- ▶ TWIs ( $i = 2$ ) are confounded with other TWIs ( $4 - 2 = 2$ ).

A design with resolution R contains a full factorial design for any subset of  $k=R-1$  factors. After the factorial experiment, drop to  $k$  factors and could re-analyze the data for all the interactions.

**Aberration:** the multiplicity of the worst confounding. The number of words with length equal to the resolution.

$$I = ABCDF = ABCEG = DEFG \quad \text{resolution IV, aberration 1}$$

$$I = ABCF = ADEG = BCDEFG \quad \text{resolution IV, aberration 2}$$

Favor design with lower aberration, fewer main effects confounded with low-order interactions.

**Clarity:** # of confounded main effect or two-way interactions

**Clear effects:** the main effect or two-way interaction effect is clear if it is only confounded with higher order terms.

- Use foldover and mirror image designs to clear confounded factors.

## Foldover Designs

Imagine a  $2^{6-3}_{\text{III}}$  design with

$$D = AB, \quad E = AC, \quad F = BC$$

$$\begin{aligned} I &= ABD = ACE = BCF = DEF \\ &= BCDE = ACDF = ABEF \end{aligned}$$

After analysis, we find that both  $B$  and  $D$  are significant.

Since  $D = AB$ , the significance of  $D$  might be due to  $B$  and  $AB$ .

We can *augment* the design by doubling the runs with  $D$  flipped. This clears  $D$  and its interactions.

Run	A	B	C	D	E	F
1	—	—	—	+	+	+
2	+	—	—	—	—	+
3	—	+	—	—	+	—
4	+	+	—	+	—	—
5	—	—	+	+	—	—
6	+	—	+	—	+	—
7	—	+	+	—	—	+
8	+	+	+	+	+	+
9	—	—	—	—	+	+
10	+	—	—	+	—	+
11	—	+	—	+	+	—
12	+	+	—	—	—	—
13	—	—	+	—	—	—
14	+	—	+	+	+	—
15	—	+	+	+	—	+
16	+	+	+	—	+	+

If we combine a Resolution III design with its mirror image (all factors flipped), we have a Resolution IV design with all main effects clear.

If we add a blocking factor we can perform the experimental batches sequentially.

As with foldover designs, mirror image designs are only necessary if more than one main effect is significant.

Estimate main effects clear of any two-way interactions.

## 25. Plackett-Burman Designs

- Construct PB designs for a set number of factors.

Allow run sizes in multiples of 4 regardless of the number of factors. Have no generators or defining relation.

1. Start with the first run from the following table.

Runs	Factor Levels
12	++- + + + - - - + -
20	++- - + + + + - + - - - + + -
24	+++ + + - + - + + - + + - + - - - -

2. Cycle the factor levels by one to get run #2. Repeat for 11, 19, or 23 runs.
3. Set the final run to all low (-).
4. If the number of factors  $k$  is less than the number of runs, select the first  $k$  columns.

- Vocab: complex aliasing, hidden projection property.

**complex aliasing:** PB design is partially correlated.

**hidden projection property:** the complex aliasing of PB designs allow us to fit models with main and TWI terms provided the number of terms is small.

- Explain how to fit a PB design with a linear model.

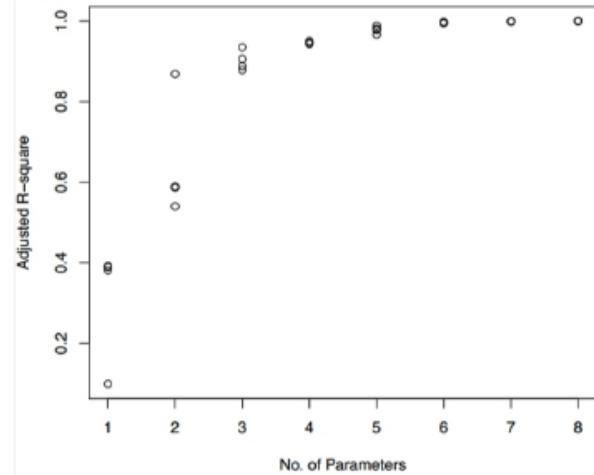
Effects are estimated for all columns with factors and interactions with complex aliasing.

- Interpret the results of subset selection.

Many small models can be built from the 7 main effects and 21 TWIs.  
How many effects should I include, and which ones?

We use *subset selection* to find good models with few terms.

```
> castfr <- castf[, c(1:7, 12)]
> library(leaps)
> modpbr<-regsubsets(y ~ .)^2, data=castfr,
+ method="exhaustive",nvmax=4,nbest=4)
> rs <- summary(modpbr)
> plot(c(rep(1:4,each=4)), rs$adjr2, xlab="No.
+ ylab="Adjusted R-square")
> plot(modpbr,scale="r2")
```



We stop adding effects when the model improvement diminishes.

Here 3 parameters is a good cutoff.

Be mindful of heredity effect: a model that includes an interaction should also include the corresponding main effects.

- ▶ Fractional designs are the **most** efficient method to screen large numbers of factors.
- ▶ Factors are confounded, but the alias structure is known.
- ▶ PB designs are an alternative if
  1. a specific # of runs is needed, or
  2. you don't want a secondary experiment.
- ▶ Factors with >2 levels require OA designs.

## 26. Vocab: mixed-level factorial designs and Orthogonal Arrays.

Fractional factorial design and PB design use two level factors.

Mixed level factorial design are factorial design with multi-level factors.

Orthogonal array design (OA): hand-crafted for mixtures of 2- and 3-level factors.

Factors with > 2 levels require OA designs.

OA is similar to PB, with resolution III, no defining relation, complex aliasing, hidden project. Models with few parameters could be fit directly to the data.

## 27. Interpret the 95% CI for effects in a model.

$$s.e. = \sigma / \sqrt{n}.$$

95% CI for a parameter is 1.96 standard error:

$$95\% CI \text{ of } \beta = [\beta - 1.96s.e., \beta + 1.96s.e.]$$

A parameter estimate is significant if and only if 95% C.I. excludes zero.

## 28. Perform power analysis (standard normal and t-test) on model coefficients.

Assume the standard deviation ( $\delta$ ) will not change in subsequent experiment.

The parameter estimate  $\beta$  will change when new samples are added, since it's the estimate of true parameter values.

Be more conservative in the estimate of n, adding another **0.84s.e.** to the bound ensure the 95% CI for  $\beta$  excludes zero for **80% of the new estimates of  $\beta$ .**

$$\beta - (1.96s.e. + 0.84s.e.) > 0$$

```
power.t.test(n=NULL,  
             delta=...,  
             sd=...,  
             power=...,  
             alternative="one.sided")
```

- ▶ `delta` is the effect size ( $\beta$ )
- ▶ `sd` is the standard deviation
- ▶ `power` is 0.8 for an 80% chance of seeing a significant result
- ▶ `alternative="one.sided"` assumes the effect won't change signs

## 29. Explain the limitations of power analysis.

High chance the estimate is not significant.

Given enough runs, any effect size, no matter how small, will become statistically significant. But statistical significance doesn't imply practical significance. Focus on effect size, not the p-value.

## 30. ANOVA

- Explain the decomposition of the sum of squares for a model.

$$SS_{total} = SS_{explained} + SS_{residual}$$

$$SS_{total} = \sum_i (y_i - mean(y))^2$$

$$SS_{residual} = \sum_i (y_i - predicted(y_i))^2$$

- Compute SStotal, SSexplained, SSresidual, and the degrees of freedom for each.

For SS\_total, DoF = (# data points)-1, 1 means the mean value.

For SS\_explained, DoF=# parameters

For SS\_residual, DoF= (# data points) - (# parameters) -1

- Compute the F statistic for an entire model and an individual factor.

F-statistic is the ratio between the explained variance and the residual variance after adjusting for the DoF.

$$\text{For entire model, } F = \frac{SS_{explained}/DoF(SS_{explained})}{SS_{residual}/DoF(SS_{residual})}$$

F-statistic follows F-distribution. Use F distribution convert the F-statistic into p-value.

For single factor, apply ANOVA (analysis of variance).

Let's find the explained variance for a model with only handedness:

```
model_hand <- lm(distance ~ hand)
sst - ss(residuals(model_hand))
```

```
## [1] 10.125
```

Now let's compare this to the residuals of the entire model:

$$F = \frac{10.125/1}{5.875/(8 - 3 - 1)} = 6.894$$

- ▶  $p$ -values on effect sizes tell us if the effect size is nonzero.
- ▶ A significant effect size does not mean the effect matters.
- ▶ ANOVA can tell us which variables explain a significant fraction of the variance in our data.
- ▶ Significance is relative to the unexplained variance in the model.