Factorial ANOVA

Testing more than one manipulation

Factorial ANOVA

Today's goal:

Teach you about factorial ANOVA, the test used to evaluate more than two manipulations at the same time

Outline:

- Why Factorial ANOVA?
- Factorial ANOVA in R
- Different types of sums of squares
- Contrasts and simple effects

Why factorial ANOVA?

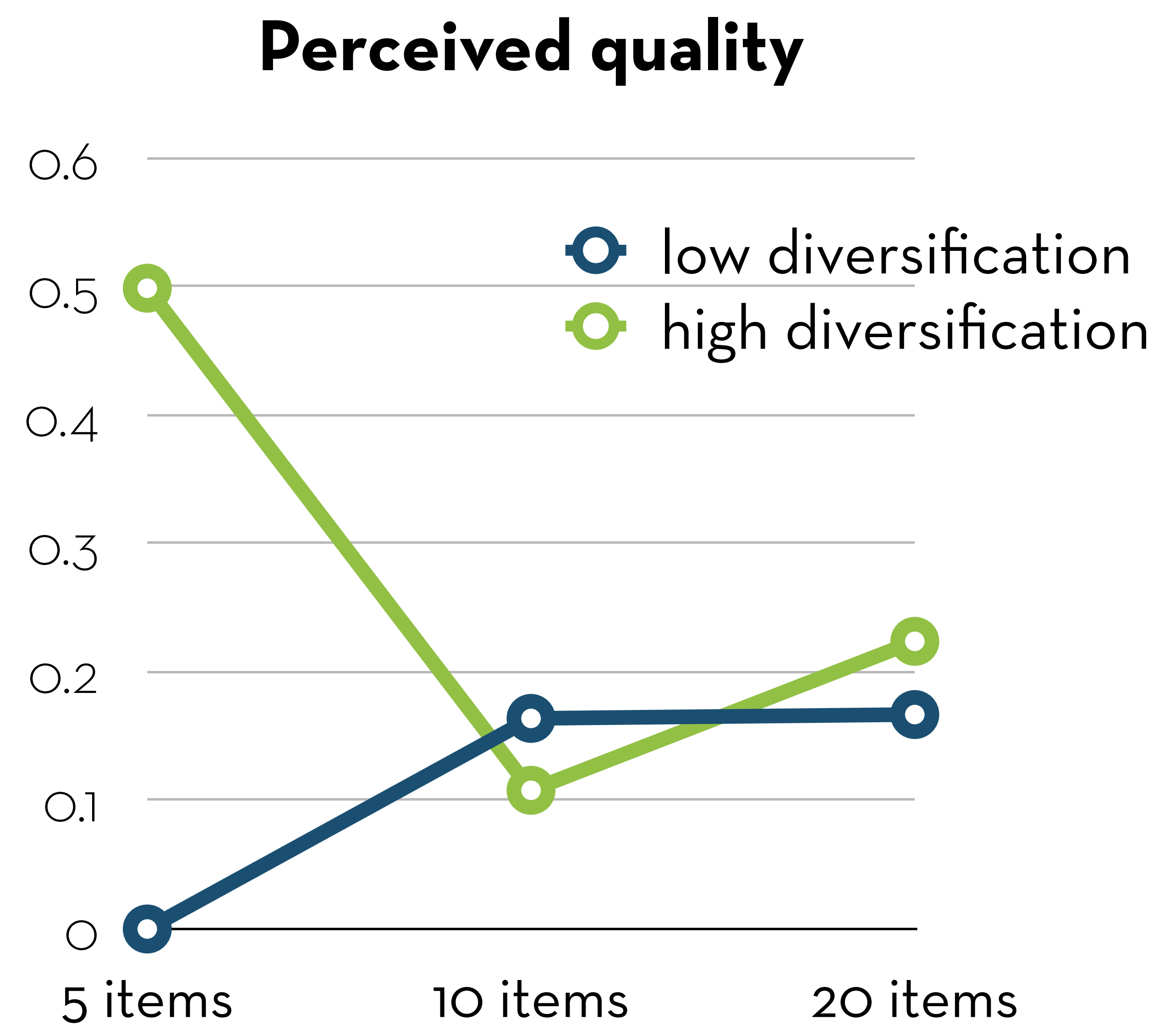
the idea of interaction effects

Factorial ANOVA

Two manipulations at the same time:

> What is the combined effect of list diversity and list length on perceived recommendation quality?

Test for the interaction effect!

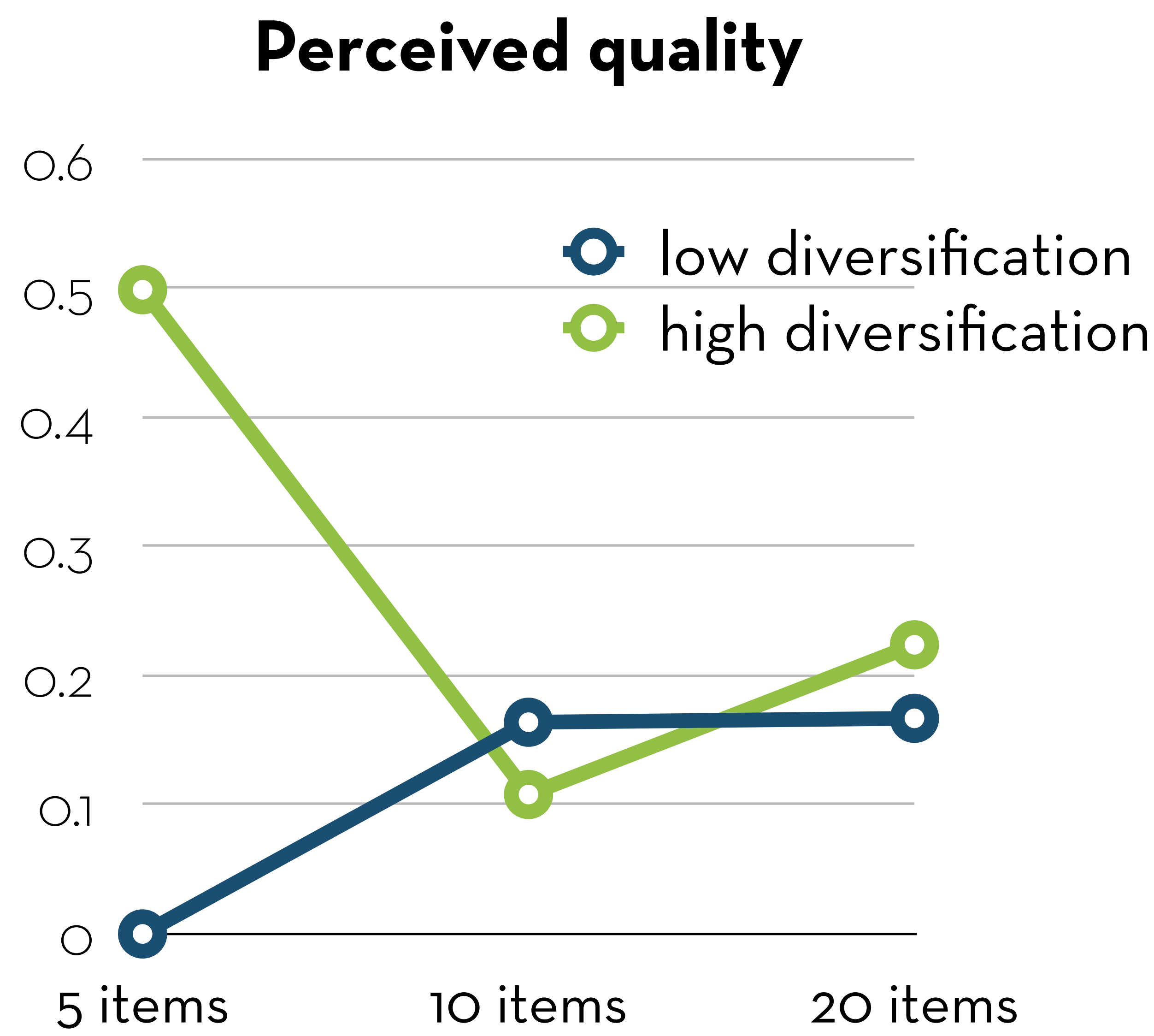


Willemsen et al.: "Understanding the Role of Latent Feature Diversification on Choice Difficulty and Satisfaction", submitted to UMUAI

AB Factorial ANOVA

Interaction effect:

> "5-item lists have a higher perceived quality than 10- or 20-item lists, but only when diversification is high"
>
> "High diversification lists have a higher perceived quality, but only for 5-item lists"



Willemsen et al.: "Understanding the Role of Latent Feature Diversification on Choice Difficulty and Satisfaction", submitted to UMUAI

Explanation

Example: effect of font size (small, large) and background color (blue, white) on readability (0-100)

If there is no interaction effect, we consider a regression model like this:

$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + e_i$

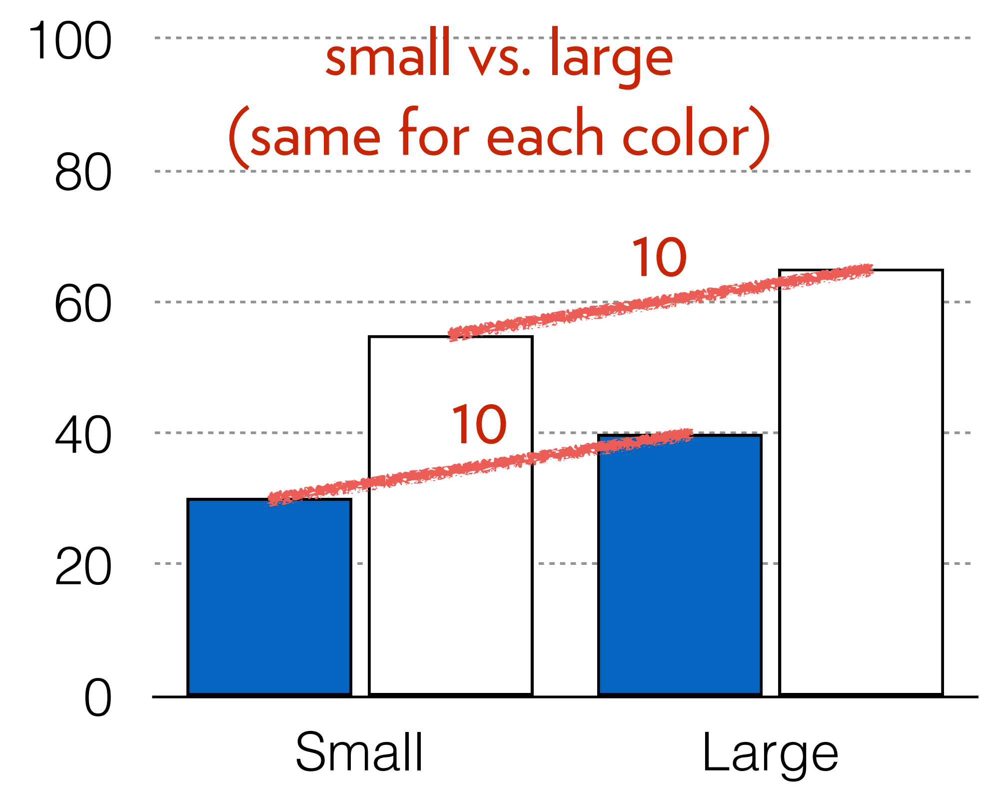
Font size: $X_1 = 1$ for large, $X_1 = 0$ for small

Background color: $X_2 = 1$ for white, $X_2 = 0$ for blue

b_1: difference between small and large (for any color)

b_2: difference between blue and white (for any size)

AB Explanation



Let's say $a = 30$, $b_1 = 10$, and $b_2 = 25$

$Y_i = 30 + 10^*X_{1i} + 25^*X_{2i} + e_i$

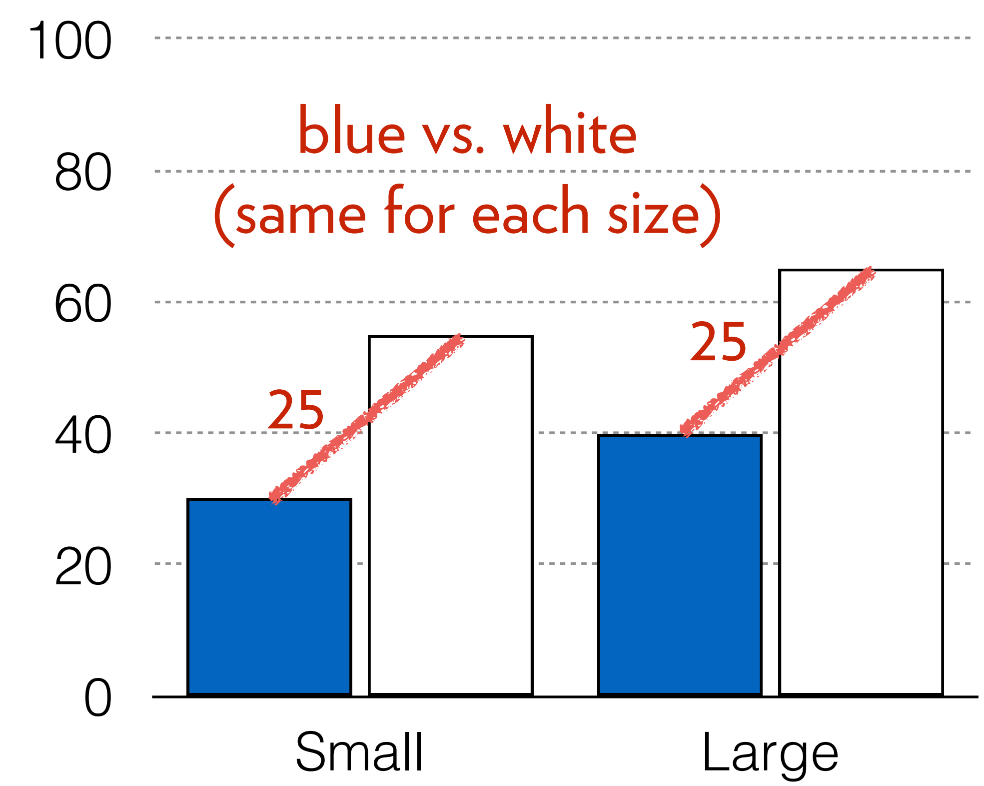
Small, blue: readability = $30+10^*0+25^*0 = 30$

Large, blue: readability = $30+10^*1+25^*0 = 40$

Small, white: readability = $30+10^*0+25^*1 = 55$

Large, white: readability = $30+10^*1+25^*1 = 65$

AB Explanation



Let's say $a = 30$, $b_1 = 10$, and $b_2 = 25$

$Y_i = 30 + 10^*X_{1i} + 25^*X_{2i} + e_i$

Small, blue: readability = $30+10^*0+25^*0 = 30$

Large, blue: readability = $30+10^*1+25^*0 = 40$

Small, white: readability = $30+10^*0+25^*1 = 55$

Large, white: readability = $30+10^*1+25^*1 = 65$

AB Explanation

If there is an interaction effect, we consider a regression model like this:

$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{1i}X_{2i} + e_i$

Font size: $X_1 = 1$ for large, $X_1 = 0$ for small

Background color: $X_2 = 1$ for white, $X_2 = 0$ for blue

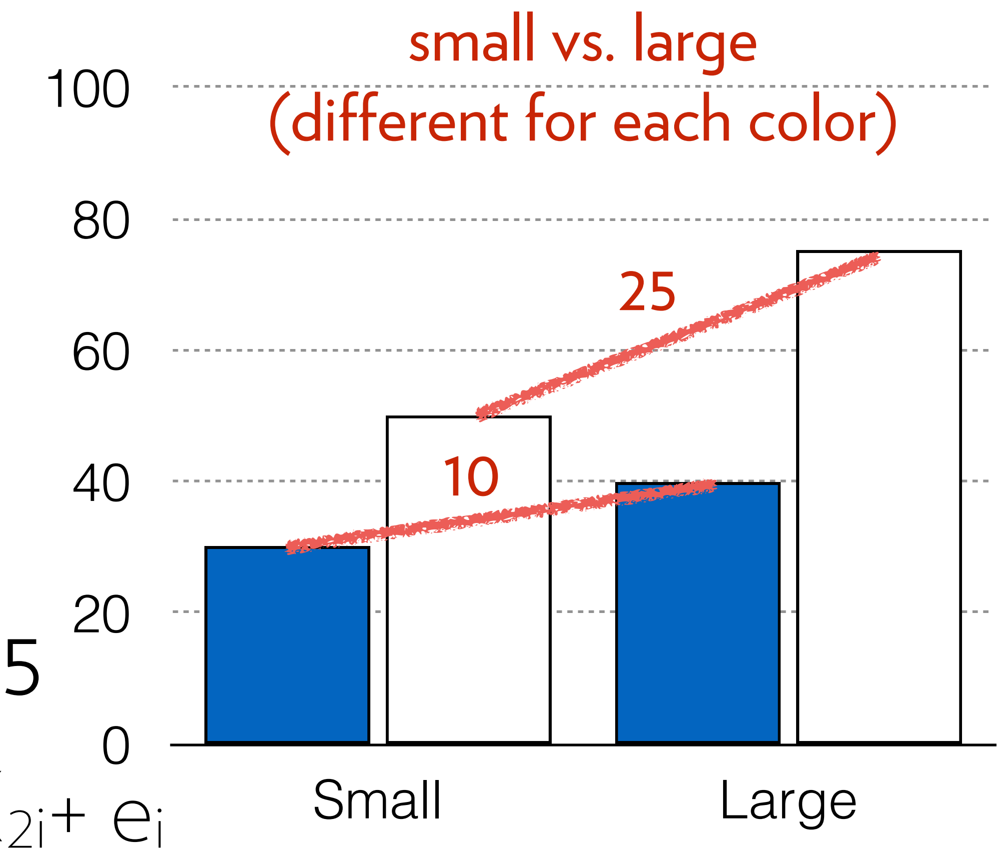
a: value in baseline condition (blue + small)

b_1: difference between small and large (for blue only)

b_2: difference between blue and white (for small only)

b_3: extra difference between small and large for white, or extra difference between blue and white for large

AB Explanation



Let's say $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = 15$

$Y_i = 30 + 10*X_{1i} + 20*X_{2i} + 15*X_{1i}X_{2i} + e_i$

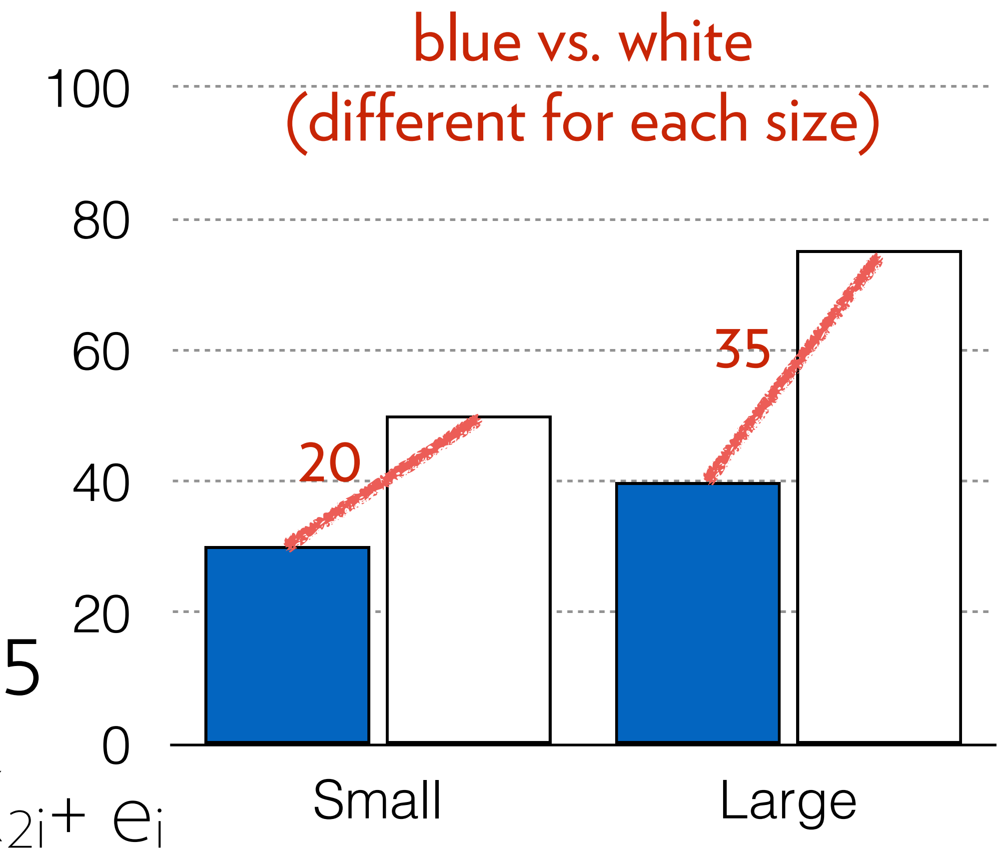
Small, blue: readability = 30+10*0+20*0+15*0*0 = 30

Large, blue: readability = 30+10*1+20*0+15*1*0 = 40

Small, white: readability = 30+10*0+20*1+15*0*1 = 50

Large, white: readability = 30+10*1+20*1+15*1*1 = 75

AB Explanation



Let's say $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = 15$

$Y_i = 30 + 10 * X_{1i} + 20 * X_{2i} + 15 * X_{1i}X_{2i} + e_i$

Small, blue: readability = 30+10*0+20*0+15*0*0 = 30

Large, blue: readability = 30+10*1+20*0+15*1*0 = 40

Small, white: readability = 30+10*0+20*1+15*0*1 = 50

Large, white: readability = 30+10*1+20*1+15*1*1 = 75

Implications

Whether you have a significant interaction depends on the significance of b_3

b_1 and b_2 are uninterpretable without b_3

- Before, b_1 represented the effect of X_1
- Now, there is no single "effect of X_1", because it depends on X_2 (and vice versa)

You can't have b_3 in the model without b_1 and b_2

- Since b_3 is an **additional** effect, it relies on b_1 and b_2

AB Implications

Calculating differences between groups becomes trickier:

- Diff. small and large text for blue background: b_1
- Diff. blue and white background for small text: b_2
- Diff. small and large text for white background: $b_1 + b_3$
- Diff. blue and white background for large text: $b_2 + b_3$

Some involve 2 b's, so you can't check their significance

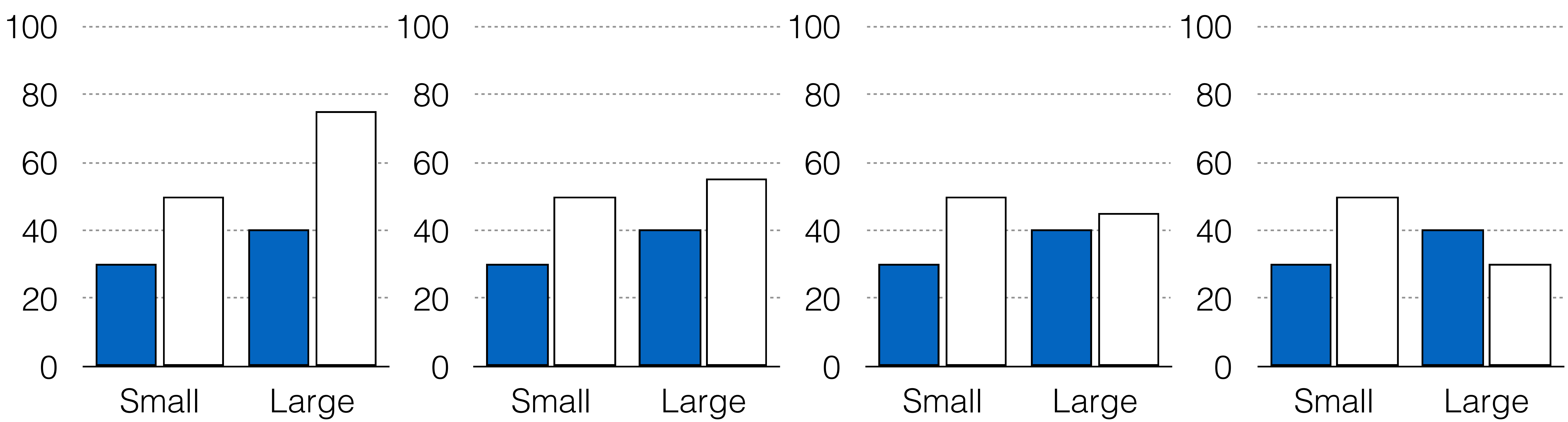
Luckily there are tests for that

Or, you can re-code your dummies!

Explanation

Types of effects:

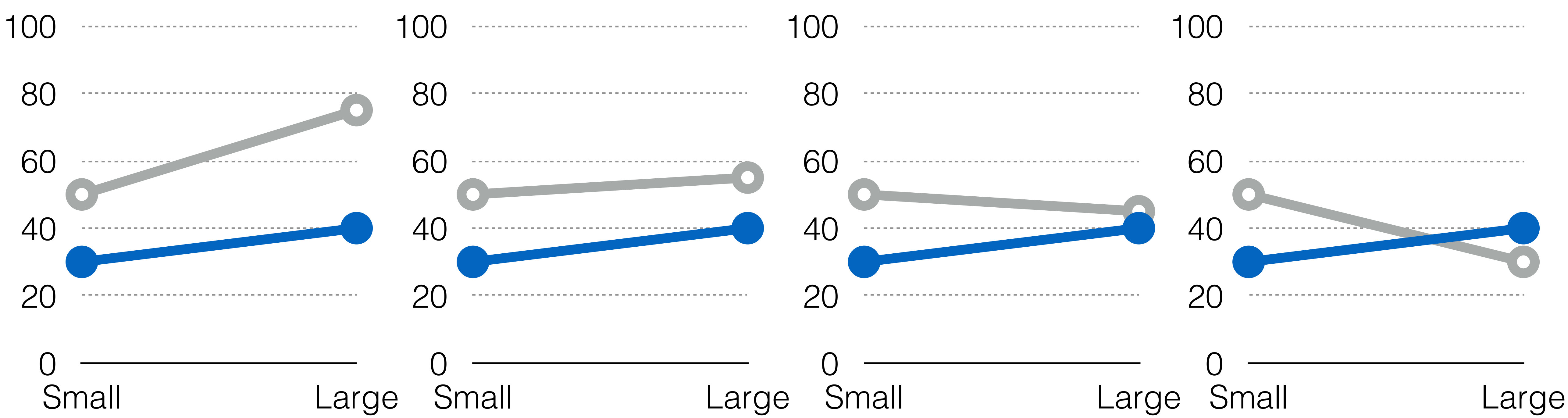
- Super-additive, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = 15$
- Sub-additive, e.g. $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -5$
- Cross-over, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -15$
- Double cross-over, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -30$



Explanation

Types of effects:

- Super-additive, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = 15$
- Sub-additive, e.g. $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -5$
- Cross-over, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -15$
- Double cross-over, e.g.: $a = 30$, $b_1 = 10$, $b_2 = 20$, $b_3 = -30$



Orthogonal spec

We can also build this model orthogonally:

$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{1i}X_{2i} + e_i$

Font size: X_1 = **0.5** for large, X_1 = **–0.5** for small

Background color: X_2 = **0.5** for white, X_2 = **–0.5** for blue

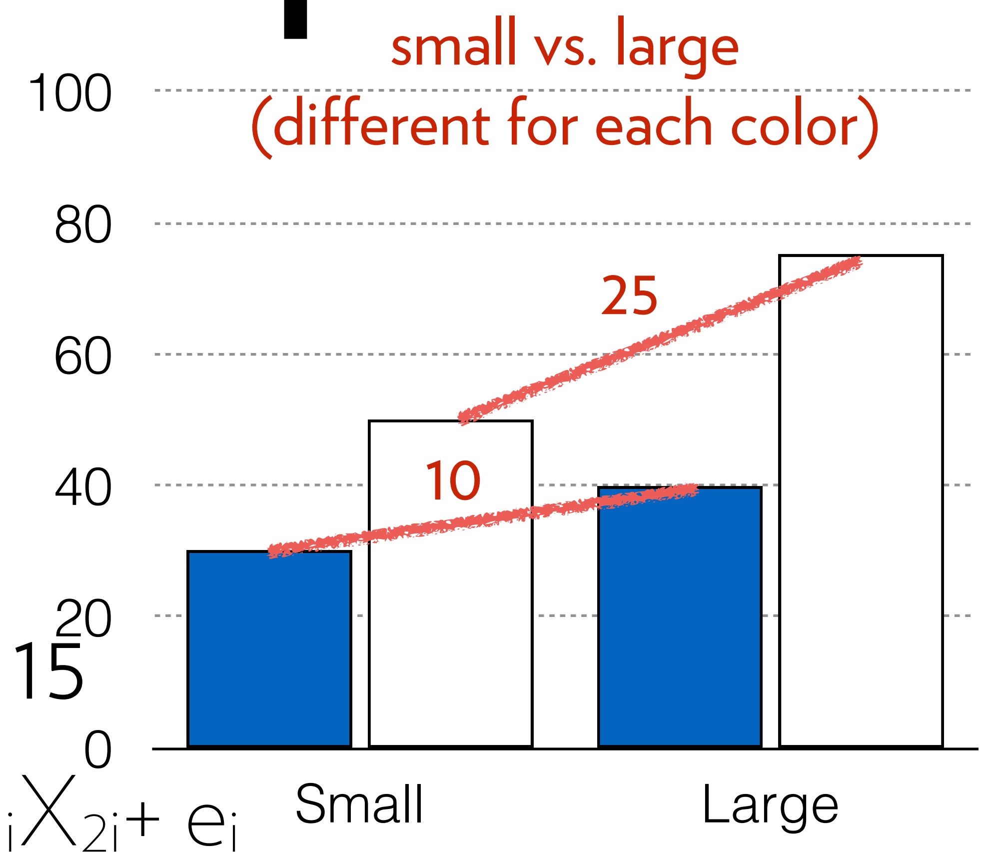
a: **grand mean**

b_1: **average** difference between small and large

b_2: **average** difference between blue and white

b_3: extra difference between small and large for white, or extra difference between blue and white for large

Orthogonal spec



Now, $a = 48.75$, $b_1 = 17.5$, $b_2 = 27.5$, $b_3 = 15$

$$Y_i = 48.75 + 10{*}X_{1i} + 20{*}X_{2i} + 15{*}X_{1i}X_{2i} + e_i$$

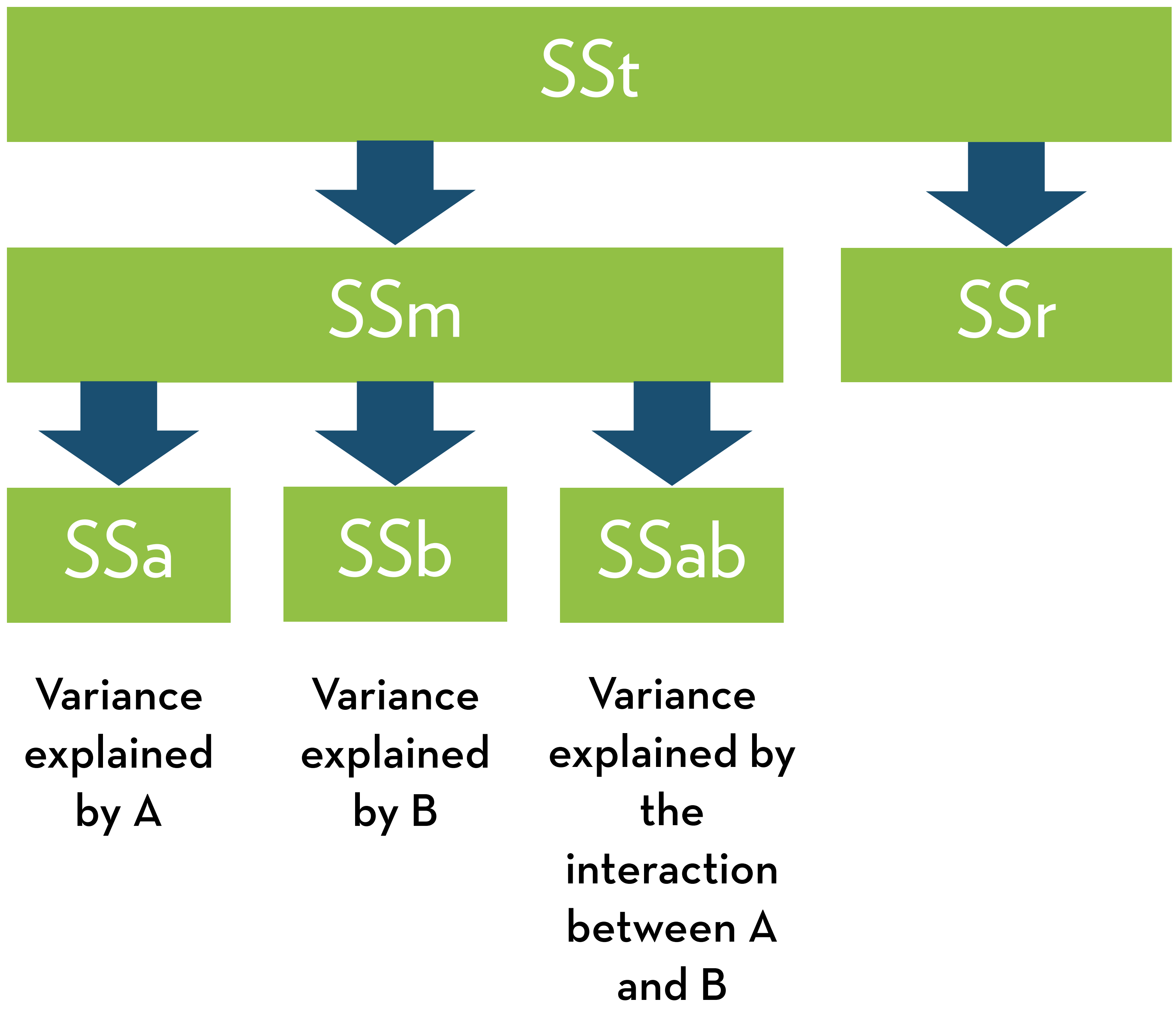
Small, blue: $48.75+17.5{*}-0.5+27.5{*}-0.5+15{*}-0.5{*}-0.5 = 30$

Large, blue: $48.75+17.5{*}0.5+27.5{*}-0.5+15{*}0.5{*}-0.5 = 40$

Small, white: $48.75+17.5{*}-0.5+27.5{*}0.5+15{*}-0.5{*}0.5 = 50$

Large, white: $48.75+17.5{*}0.5+27.5{*}0.5+15{*}0.5{*}0.5 = 75$

Sum of Squares



AB Sum of Squares

Formulas for A (r groups) and B (s groups):

SSt: same as for regular ANOVA:

$SSt = s^2(N-1)$

SSr: also the same, k is all r*s combinations of A and B:

$\sum s_k^2(N_k-1)$, with n–k df

SSm: also the same; sum of squares over r*s group means:

$\sum n_k(mean_k - \text{grand mean})^2$, with k–1 df

AB

Sum of Squares

SSa: sum of squares over r group means:

$\sum n_r(\text{mean}_r - \text{grand mean})^2$, with r–1 df

SSb: sum of squares over s group means:

$\sum n_s(\text{mean}_s - \text{grand mean})^2$, with s–1 df

SSab: what is left over:

SSm–SSa–SSb, with (r–1)(s–1) df

AB Mean Squares and F

Mean squares:

$MSa = SSa/df_a$

$MSb = SSb/df_b$

$MSab = SSab/df_{ab}$

$MSr = SSr/df_r$

F ratios:

$Fa = MSa/MSr$ (with df_a, df_r degrees of freedom)

$Fb = MSb/MSr$ (with df_b, df_r degrees of freedom)

$Fab = MSab/MSr$ (with df_{ab}, df_r degrees of freedom)

AB Lessons learned

A factorial ANOVA is a regular ANOVA, but with the SSm divided into each factor and their interaction(s)

- 2 variables: SSm = SSa + SSb + SSab
- 3 variables: SSm = SSa + SSb + SSc + SSab + SSac + SSbc + SSabc

AB Lessons learned

A factorial ANOVA is a regression model with interaction term(s)

e.g. 2x2: X_1 represents A, X_2 represents B:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{1i}X_{2i} + e_i$$

e.g. 3x2: X_1 and X_2 represent A, X_3 represents B:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{1i}X_{3i} + b_5X_{2i}X_{3i} + e_i$$

AB Lessons learned

e.g. 3x3: X_1 and X_2 represent A, X_3 and X_4 represent B:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5X_{1i}X_{3i} + b_6X_{2i}X_{3i} + b_7X_{1i}X_{4i} + b_8X_{2i}X_{4i} + e_i$$

e.g. 2x2x2: X_1 represents A, X_2 represents B, X_3 represents C:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{1i}X_{2i} + b_5X_{1i}X_{3i} + b_6X_{2i}X_{3i} + b_7X_{1i}X_{2i}X_{3i} + e_i$$

AB

Factorial ANOVA in R

because too many $X_{1i}X_{2i}X_{3i}X_{4i}X_{5i}X_{6i}$...

Factorial in R

Dataset "goggles.csv"

- Effect of beer consumption on mate attractiveness

Variables:

- gender: gender of the participant
- alcohol: amount of alcohol consumed
- attractiveness: attractiveness of the person they want to go home with at the end of the night (%)

Plotting

Relevel the alcohol variable to make "None" the baseline:

```
goggles$alcohol <- relevel(goggles$alcohol, ref="None")
```

Line plot with bootstrapped CIs:

```
ggplot(goggles, aes(alcohol, attractiveness, color =
gender)) + stat_summary(fun.y = mean, geom = "line",
aes(group = gender)) + stat_summary(fun.y = mean, geom
= "point", aes(group = gender), size = 3) +
stat_summary(fun.data = mean_cl_boot, geom =
"errorbar", width = 0.2) + ylim(0, 100)
```

Plotting

Box plots per group:

```
ggplot(goggles,aes(alcohol,attractiveness))
+geom_boxplot()+facet_wrap(~gender)
```

Normality

Stat.desc():

```
stat.desc(goggles$attractiveness, desc=F, norm=T)
```

By gender (2 groups):

```
by(goggles$attractiveness, goggles$gender, stat.desc,
desc=F, norm=T)
```

By alcohol (3 groups):

```
by(goggles$attractiveness, goggles$alcohol, stat.desc,
desc=F, norm=T)
```

AB Normality

For each of the 6 groups:

```
by(goggles$attractiveness, list(goggles$alcohol, goggles$gender), stat.desc, desc=F, norm=T)
```

Verdict:

- Overall some skewness, and failed normal test
- Failed normal test for females
- No problems in all 6 groups

AB Homoscedasticity

By gender (2 groups):

```
leveneTest(attractiveness~gender, data=goggles)
```

By alcohol (3 groups):

```
leveneTest(attractiveness~alcohol, data=goggles)
```

For each of the 6 groups:

```
leveneTest(attractiveness~alcohol*gender, data=goggles)
```

Verdict:

Heteroscedasticity by gender, but not for the interaction

Contrasts

Alcohol has 3 levels, so we should define 2 contrasts:

```
contrasts(goggles$alcohol)<-cbind(c(-2/3, 1/3, 1/3),
c(0, -1/2, 1/2)
```

Gender has 2 levels, so only one contrast is needed:

```
contrasts(goggles$gender)<-c(-1/2, 1/2)
```

(Why bother with contrasts here? — We'll get to that in a minute!)

AB Run the ANOVA

To run with both main effects of A and B, and the interaction effect AB, you can simply specify A*B

- R automatically includes the main effects

Run the model:

```
g1 <- aov(attractiveness ~ alcohol*gender, data = goggles)
Anova(g1, type=3)
```

```
               Sum Sq Df   F value    Pr(>F)    
(Intercept)    163333  1 1967.0251 < 2.2e-16 ***
gender            169  1    2.0323    0.1614    
alcohol          3332  2   20.0654 7.649e-07 ***
gender:alcohol   1978  2   11.9113 7.987e-05 ***
Residuals        3487 42                        
```

Run the ANOVA

Get the plots to test for homoscedasticity and normality:

```
plot(g1)
```

Why type=3?

You can run a factorial ANOVA in three ways, with three types of Sum of Squares:

AB Type I

In **Type I** Sum of Squares variables are added to the model one by one (this is what AOV does)

Let's say you test Y~A*B

- The first F-test is the effect of A alone
- The second F-test is the effect of B, given A
- The third F-test is the effect of AB, given A and B

The **order** in which you list your variables makes a difference!

- If you specify Y ~ B*A, you get different results!

Type II

In **Type II** Sum of Squares, main effects are added first, interaction(s) later (this is what Anova does by default)

Let's say you test Y~A*B

- The first F-test is the effect of A, given B
- The second F-test is the effect of B, given A
- The third F-test is the effect of AB, given A and B

The main effects are meaningless when there is an interaction effect (but accurate if not)

AB Type III

In **Type III** Sum of Squares everything is added to the model at the same time

Let's say you test Y~A*B

- The first F-test is the effect of A, given B and AB
- The second F-test is the effect of B, given A and AB
- The third F-test is the effect of AB, given A and B

The main effects are meaningful, but not very useful

- Because the effect of B now depends on A and vice versa

AB Type I, II or III?

Tips:

- Don't use type I
- Use type II if you expect no interaction effects at all (slightly more powerful) or if you want to use non-orthogonal contrasts
- Type II doesn't work when group sizes are very unequal
- Use type III if you do expect an interaction effect, or when group sizes are unequal
- For type III to make sense, contrasts **must** be orthogonal

AB Interpretation

```
                Sum Sq Df   F value    Pr(>F)    
(Intercept)     163333  1 1967.0251 < 2.2e-16 ***
gender             169  1    2.0323    0.1614    
alcohol           3332  2   20.0654 7.649e-07 ***
gender:alcohol    1978  2   11.9113 7.987e-05 ***
Residuals         3487 42                        
```

There is no significant main effect of gender (plot it!)

There is a significant main effect of alcohol (plot it!)

There is a significant interaction effect (see our first plot!)

- The effect of alcohol differs per gender, and vice versa
- The other two effects are therefore uninterpretable!

AB Interpret contrasts

Run summary.lm(g1):

```
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         58.333      1.315  44.351  < 2e-16 ***
gender1             -3.750      2.631  -1.426 0.161382    
alcohol1            -8.125      2.790  -2.912 0.005727 ** 
alcohol2           -18.125      3.222  -5.626 1.37e-06 ***
gender1:alcohol1   -15.000      5.580  -2.688 0.010258 *  
gender1:alcohol2   -26.250      6.443  -4.074 0.000201 ***
```

Gender1: the contrast of gender

- Since we coded the model orthogonal, this is the overall difference between males and females (which differs per alcohol level, and is therefore not very interesting)

AB Interpret contrasts

```
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)          58.333      1.315  44.351  < 2e-16 ***
gender1              -3.750      2.631  -1.426 0.161382    
alcohol1             -8.125      2.790  -2.912 0.005727 ** 
alcohol2            -18.125      3.222  -5.626 1.37e-06 ***
gender1:alcohol1    -15.000      5.580  -2.688 0.010258 *  
gender1:alcohol2    -26.250      6.443  -4.074 0.000201 ***
```

alcohol1: comparing no alcohol to the two alcohol groups

- Again, there is an overall difference, but not interesting because it differs for males and females

alcohol2: comparing 2 versus 4 pints of beer

- Same thing

AB Interpret contrasts

```
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)        58.333      1.315  44.351  < 2e-16 ***
gender1            -3.750      2.631  -1.426 0.161382    
alcohol1           -8.125      2.790  -2.912 0.005727 ** 
alcohol2          -18.125      3.222  -5.626 1.37e-06 ***
gender1:alcohol1  -15.000      5.580  -2.688 0.010258 *  
gender1:alcohol2  -26.250      6.443  -4.074 0.000201 ***
```

gender1:alcohol1: tests whether the effect of no alcohol vs. the two alcohol groups differs for males and females

gender1:alcohol2: tests whether the effect of 2 vs. 4 pints differs for males and females

The answer is yes for both!

Interpret contrasts

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	58.333	1.315	44.351	< 2e-16	***
gender1	-3.750	2.631	-1.426	0.161382	
alcohol1	-8.125	2.790	-2.912	0.005727	**
alcohol2	-18.125	3.222	-5.626	1.37e-06	***
gender1:alcohol1	-15.000	5.580	-2.688	0.010258	*
gender1:alcohol2	-26.250	6.443	-4.074	0.000201	***

Effect for females, no alcohol: $58.333 + -1/2 * -3.75 + -2/3 * -8.125 + -1/2 * -2/3 * -15 =$ **60.625**

Effect for females, 2 pints: $58.333 + -1/2 * -3.75 + 1/3 * -8.125 + -1/2 * -18.125 + -1/2 * 1/3 * -15 + -1/2 * -1/2 * -26.25 =$ **62.5**

Effect for females, 4 pints: $58.333 + -1/2 * -3.75 + 1/3 * -8.125 + 1/2 * -18.125 + -1/2 * 1/3 * -15 + -1/2 * 1/2 * -26.25 =$ **57.5**

AB Interpret contrasts

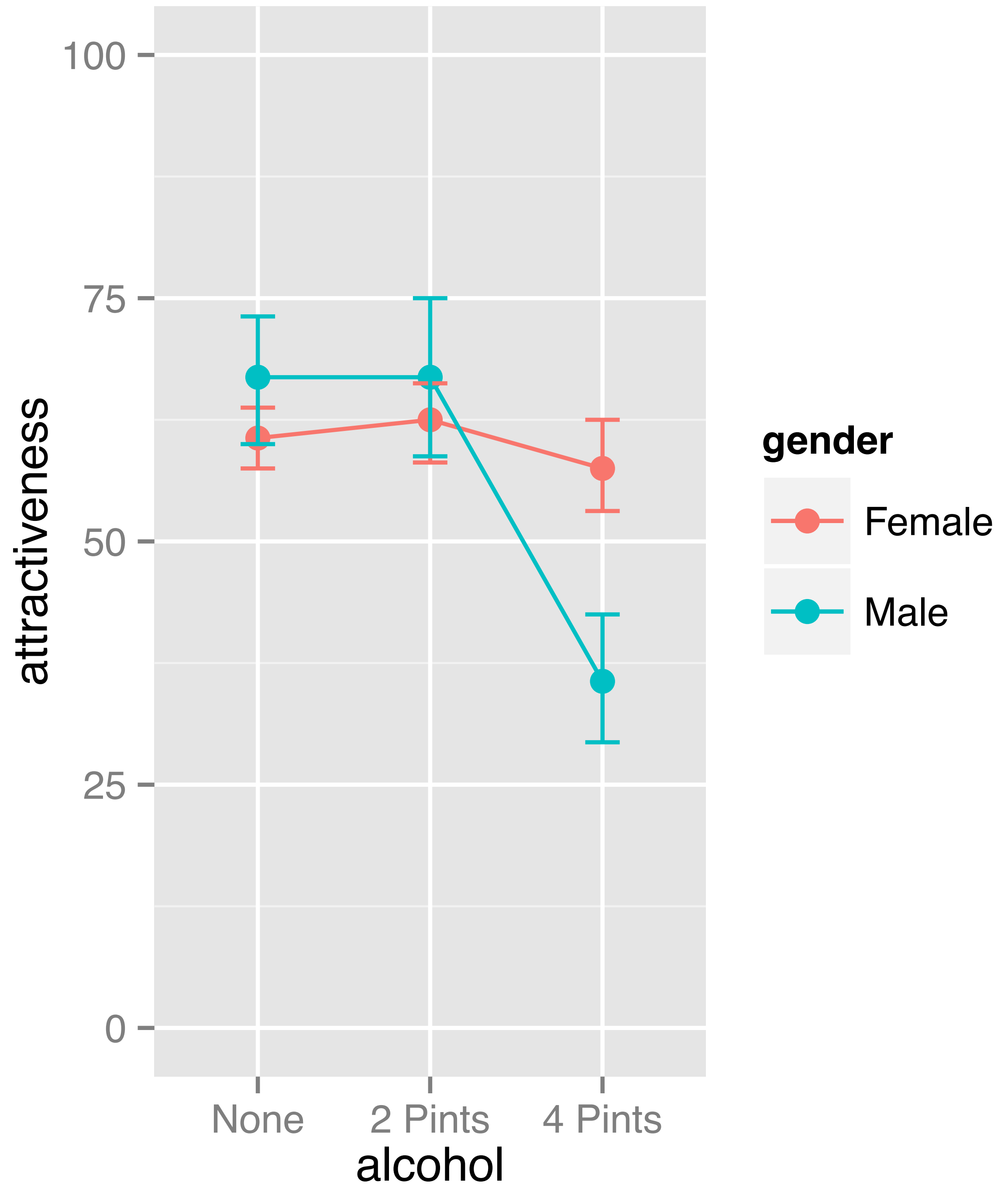
```
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         58.333      1.315  44.351  < 2e-16 ***
gender1             -3.750      2.631  -1.426 0.161382    
alcohol1            -8.125      2.790  -2.912 0.005727 ** 
alcohol2           -18.125      3.222  -5.626 1.37e-06 ***
gender1:alcohol1   -15.000      5.580  -2.688 0.010258 *  
gender1:alcohol2   -26.250      6.443  -4.074 0.000201 ***
```

Effect for males, no alcohol: 58.333 + 1/2*-3.75 + -2/3*-8.125 + 1/2*-2/3*-15 = **66.875**

Effect for males, 2 pints: 58.333 + 1/2*-3.75 + 1/3*-8.125 + -1/2*-18.125 + 1/2*1/3*-15 + 1/2*-1/2*-26.25 = **66.875**

Effect for males, 4 pints: 58.333 + 1/2*-3.75 + 1/3*-8.125 + 1/2*-18.125 + 1/2*1/3*-15 + 1/2*1/2*-26.25 = **35.625**

Interpret contrasts

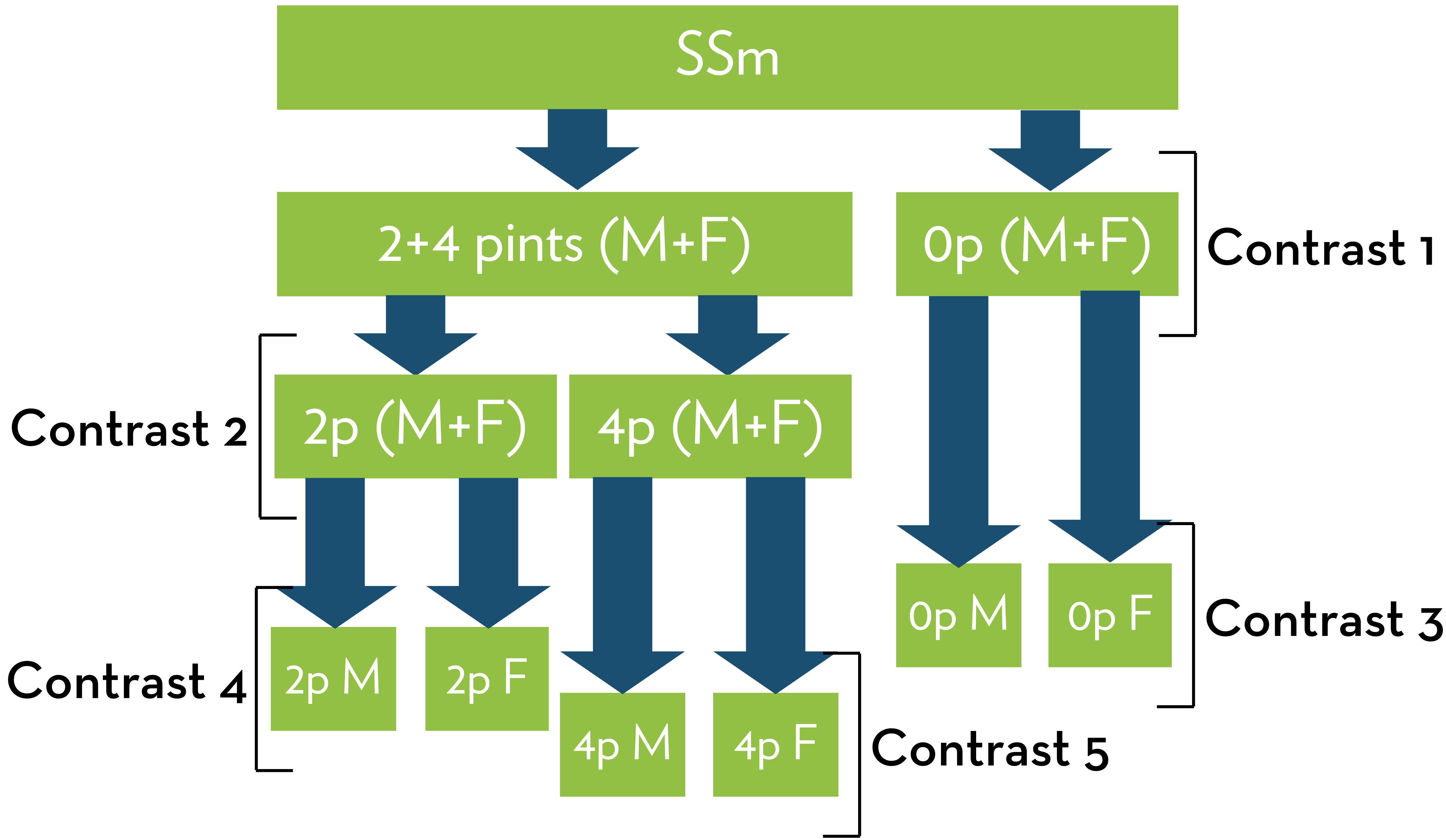


Simple effects

Test the effect of one variable for different levels of the other variable

- E.g., a kind of t-test for gender at each level of alcohol
- Or, a kind of ANOVA for alcohol separately for males and females

Simple effects



AB Simple effects

First, create a variable with all groups:

```
goggles$simple <- interaction(goggles$alcohol, goggles$gender)
```

Create dummies for the contrasts:

```
alcohol1 <- c(-2/3,1/3,1/3,-2/3,1/3,1/3)
alcohol2 <- c(0, -1/2, 1/2, 0, -1/2, 1/2)
gender_none <- c(-1/2, 0, 0, 1/2, 0, 0)
gender_2p <- c(0, -1/2, 0, 0, 1/2, 0)
gender_4p <- c(0, 0, -1/2, 0, 0, 1/2)
```

Simple effects

Load the contrasts:

```
contasts(goggles$simple) <- cbind(alcohol1, alcohol2, gender_none, gender_2p, gender_4p)
```

Run the ANOVA and get the lm summary:

```
simpleg <- aov(attractiveness ~ simple, data = goggles)
summary.lm(simpleg)
```

AB Simple effects

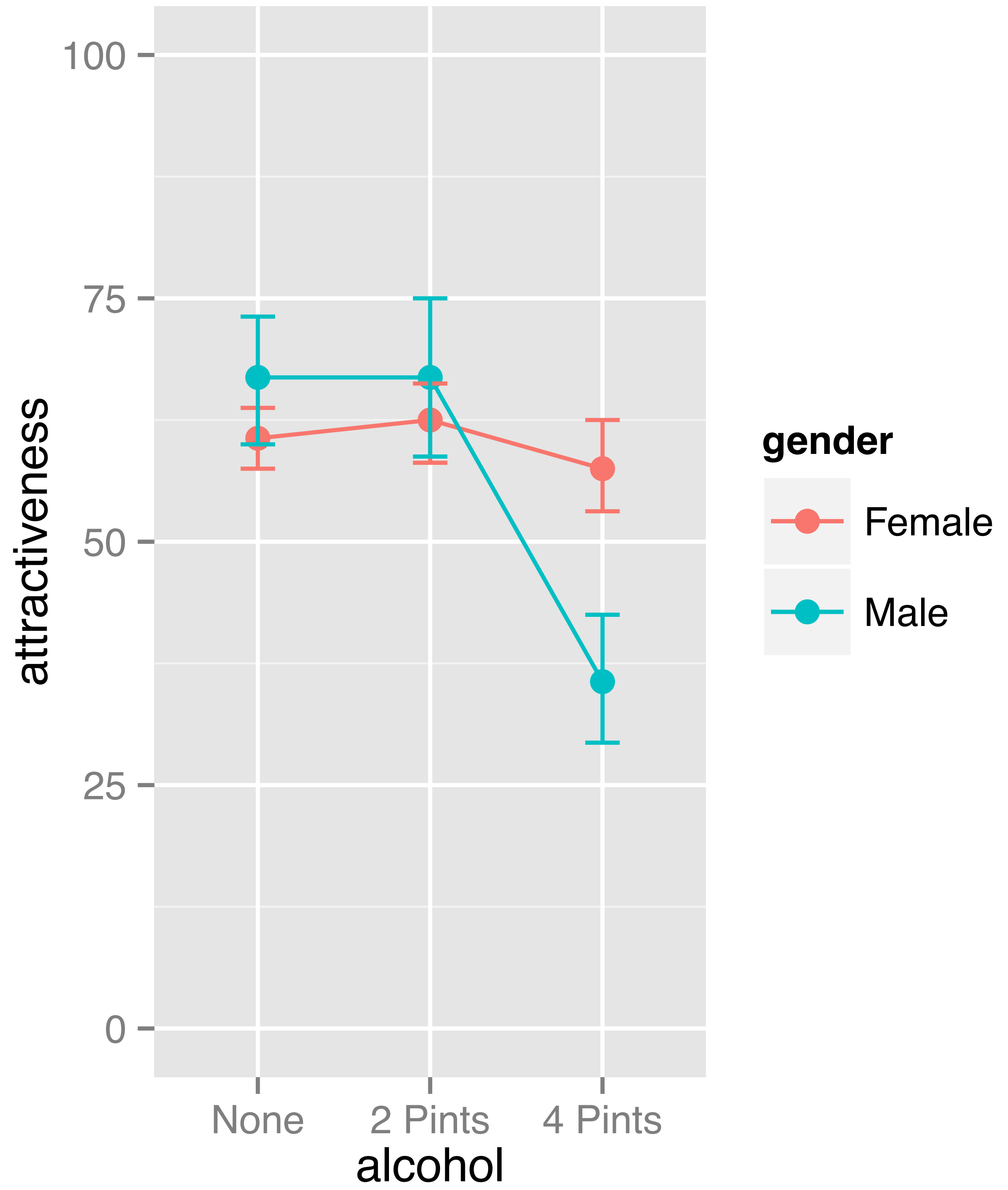
```
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          58.333      1.315  44.351  < 2e-16 ***
simplealcohol1       -8.125      2.790  -2.912  0.00573 **
simplealcohol2      -18.125      3.222  -5.626 1.37e-06 ***
simplegender_none     6.250      4.556   1.372  0.17742
simplegender_2p       4.375      4.556   0.960  0.34243
simplegender_4p     -21.875      4.556  -4.801 2.02e-05 ***
```

The first part we already knew from earlier. For the rest:

- Simplegender_none: the effect of gender with no alcohol
- Simplegender_2p: the effect of gender with 2 pints
- Simplegender_4p: the effect of gender with 4 pints

Only the last simple effect is significant!

Simple effects



Post-hoc tests?

Same as for regular ANOVA!

- You can only do post-hoc tests on main effects!

If you want to do post-hoc tests on your simple effects (e.g. on alcohol for different genders):

- Use non-orthogonal contrasts, and apply the appropriate correction (see slides on Bonferroni, Holm, and Benjamini-Hochberg corrections)

Robust methods

Use the WRS2 package!

Two-way ANOVA on 10% trimmed means:

t2way(attractiveness~alcohol*gender, data=goggles, tr=0.1)

Two way ANOVA with M-measures and a bootstrap:

pbad2way(attractiveness~alcohol*gender, data=goggles, est="median", nboot = 2000)

use est="mom" to use an automatically trimmed mean rather than the median

Robust rest

Robust contrasts? See regular ANOVA!

- You can run robust t-tests on the contrast dummies

Robust post-hoc tests? Same as for regular ANOVA!

- You can only do post-hoc tests on main effects!

If you want to do robust post-hoc tests on your simple effects (e.g. on alcohol for different genders):

- Use non-orthogonal contrasts, and apply the appropriate correction, run robust t-tests on the contrast dummies

Effect sizes

Overall R^2: from summary.lm

Omega-squared per effect: use my “omega_aov” function on the next slide

Effect sizes

```
omega_aov <- function(model){
    MS<-summary(model)[[1]]$'Mean Sq'; #get the mean squares
    df<-summary(model)[[1]]$Df; #get the Dfs
    MSr<-MS[length(MS)]; #get MSr (the last one)
    N<-sum(df)+1; #get N (sum of df+1)
    MS<-MS[-c(length(MS))]; #remove MSr from MS
    df<-df[-c(length(df))]; #remove dfr from df
    var<-df*(MS-MSr)/N; #get the variances
    varTotal<-sum(var)+MSr; #get the total variance
    omega.squared<-var/varTotal; #get the omega-squareds
    omega<-sqrt(omega.squared); #get the omegas
    labels<-attr(model$terms,"term.labels"); #get labels
    return(cbind(labels,omega,omega.squared))
}
```

Effect sizes

Cohen's *d* of specific comparisons (e.g. the simple effects): same as ANOVA

Get means, sds, and ns from stat.desc:

```
desc <- by(goggles$attractiveness, list(goggles$gender,
goggles$alcohol), stat.desc)
```

Plug values into mes, e.g.:

```
mes(desc[[“Male”, “None”]][“mean”], desc[[“Female”,
“None”]][“mean”], desc[[“Male”, “None”]][“std.dev”],
desc[[“Female”, “None”]][“std.dev”], 8, 8)
```

AB Reporting

See Field section 12.9

Since effects may be complicated, always report a graph

"It is the mark of a truly intelligent person to be moved by statistics."

T H A N K S !

George Bernard Shaw