Interactions (II)

Last time...

- Introduction to interactions with two continuous predictors
- Bayes galore with JJJ

Recap

We use interaction terms to test the hypothesis that the relationship between X and Y changes as a function of Z.

- social support buffers the effect of anxiety and stress

The interaction term represents how much the slope of X changes as you increase on Z, and also how much the slope of Z changes as you increase on X. Interactions are **symmetric**.

Recap: Output

```
##
## Call:
## lm(formula = mpg ~ disp * hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5153 -1.6315 -0.6346  0.9038  5.7030
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.967e+01  2.914e+00  13.614 7.18e-14 ***
## disp        -7.337e-02  1.439e-02  -5.100 2.11e-05 ***
## hp          -9.789e-02  2.474e-02  -3.956 0.000473 ***
## disp:hp      2.900e-04  8.694e-05   3.336 0.002407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.692 on 28 degrees of freedom
## Multiple R-squared:  0.8198,	Adjusted R-squared:  0.8005
## F-statistic: 42.48 on 3 and 28 DF,  p-value: 1.499e-10
```

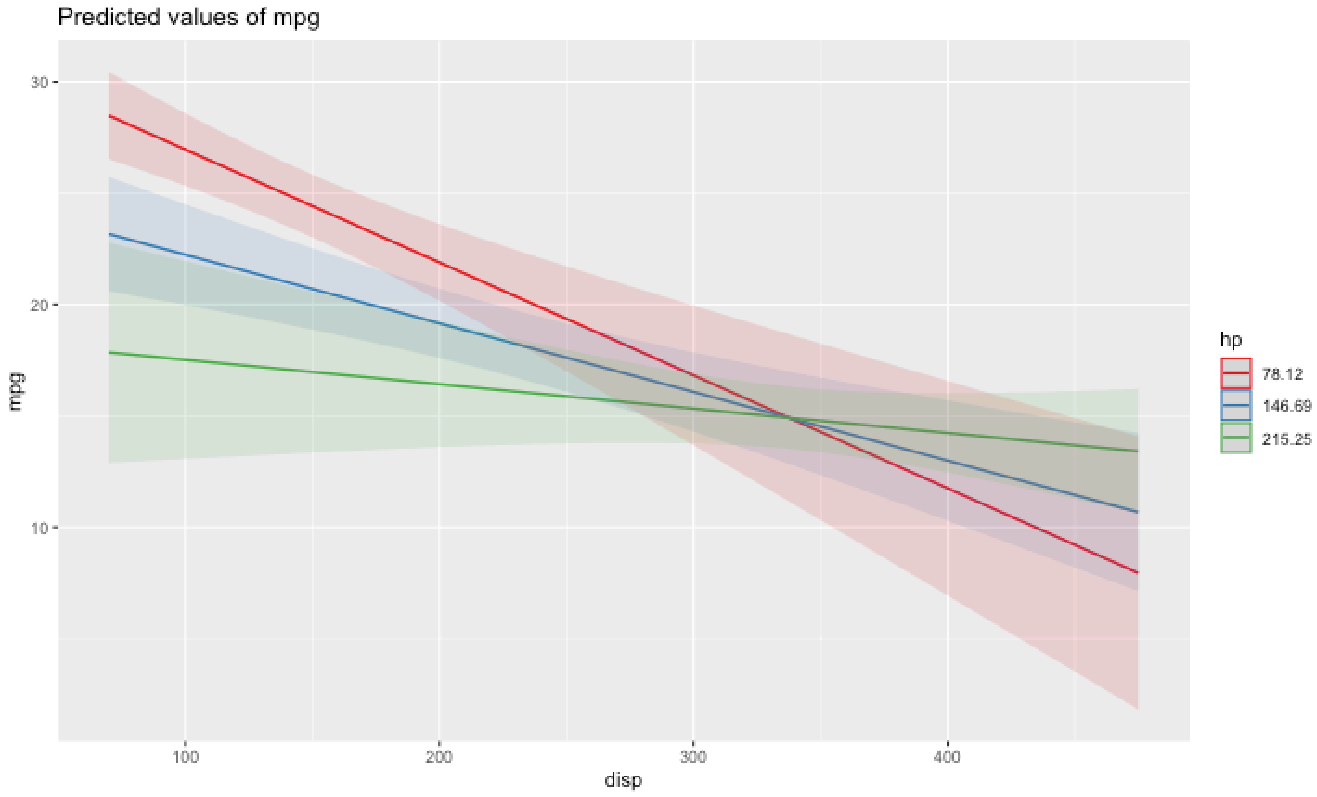
Recap: Simple slopes

```
library(reghelper)
simple_slopes(cars_model, levels = list(hp = c(78, 147, 215)))
```

```
##     disp  hp Test Estimate Std. Error t value df  Pr(>|t|) Sig.
## 1 sstest  78       -0.0507     0.0088 -5.7448 28 3.645e-06  ***
## 2 sstest 147       -0.0307     0.0064 -4.8207 28 4.528e-05  ***
## 3 sstest 215       -0.0110     0.0086 -1.2782 28    0.2117
```

Recap: Plot simple slopes

```
plot_model(cars_model, type = "int", mdrt.values = "meansd")
```



Today

Mixing categorical and continuous predictors

Two categorical predictors

Start discussing Factorial ANOVA

Mixing categorical and continuous

Consider the case where *D* is a *d*ummy coded variable representing two groups. In a univariate regression, how do we interpret the coefficient for D?

$$\hat{Y} = b_0 + b_1 D$$

b_0 is the mean of the reference group, and *D* represents the difference in means between the two groups.

Interpreting slopes

Extending this to the multivariate case, where *X* is continuous and *D* is a dummy code representing two groups.

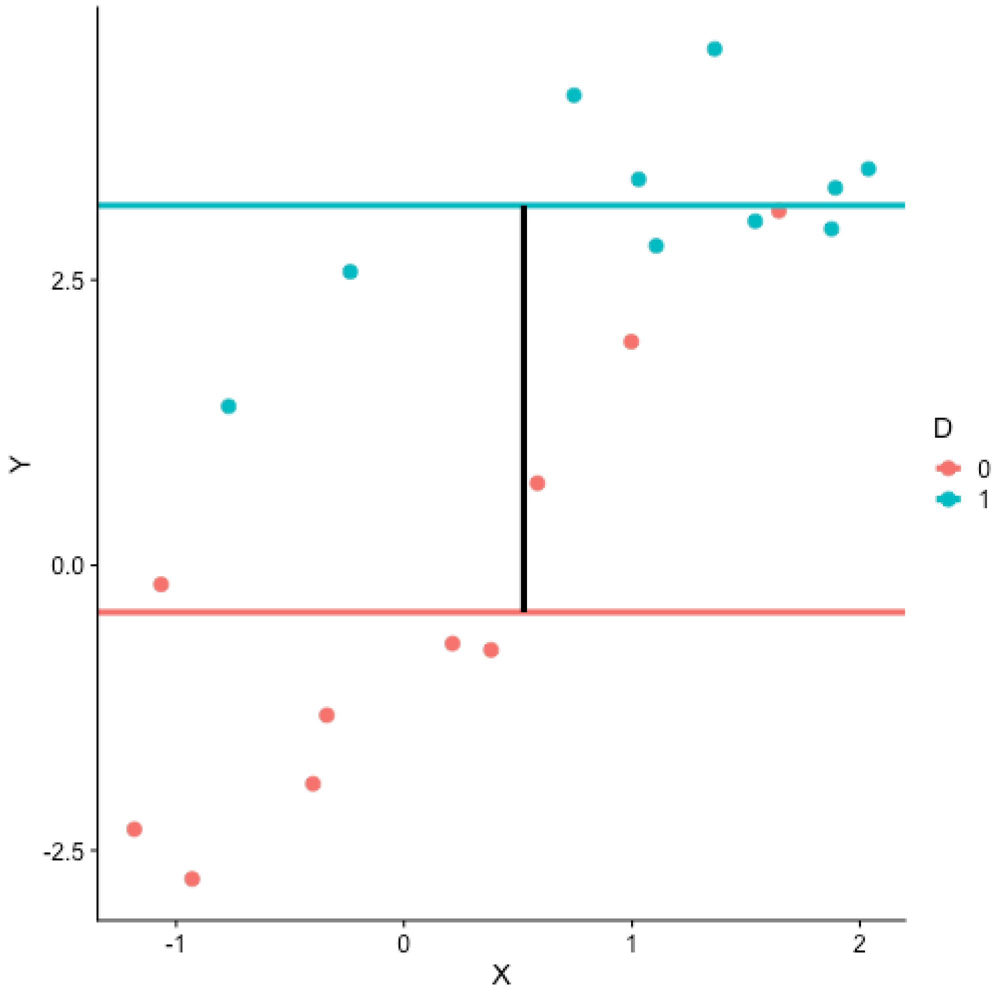
$$\hat{Y} = b_0 + b_1 D + b_2 X$$

How do we interpret b_1?

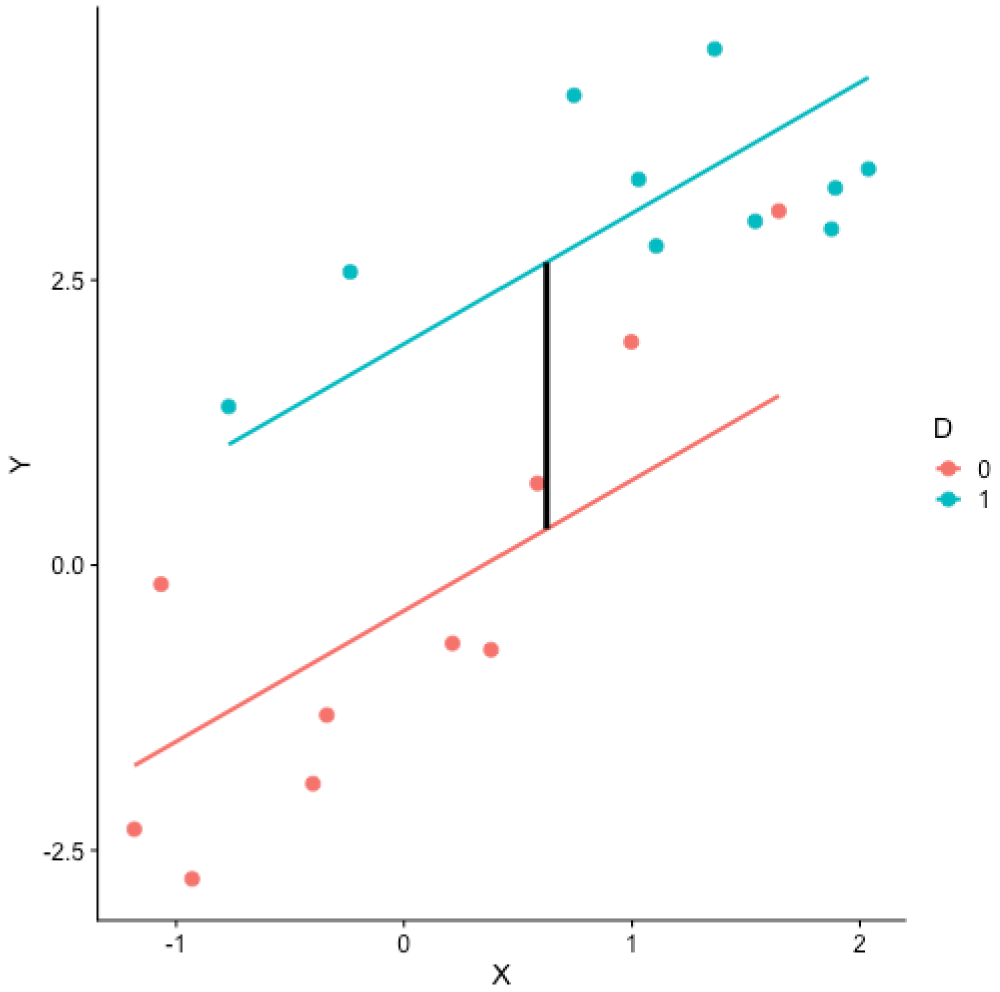
b_1 is the difference in means between the two groups *if the two groups have the same average level of X* or holding X constant.

This is called "ANCOVA". It's just regression.

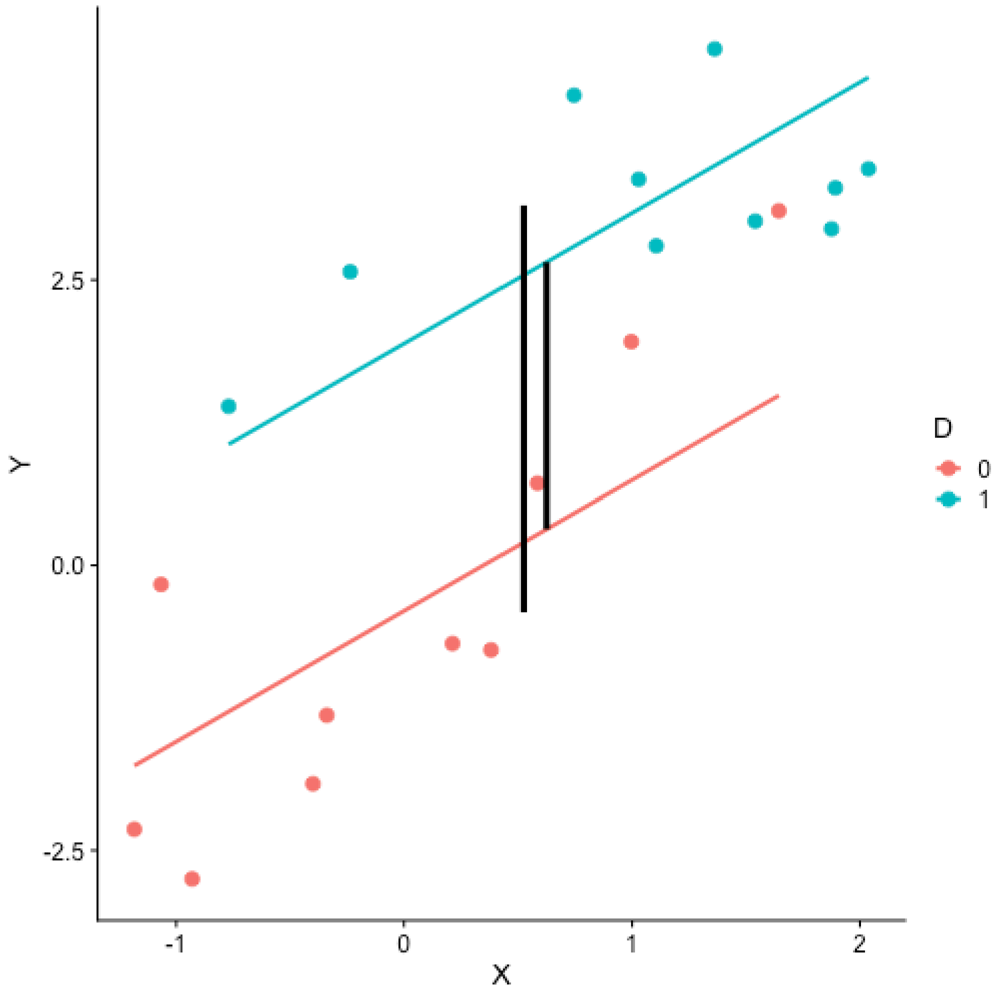
Visualizing



Visualizing



Visualizing



3 or more groups

We might be interested in the relative contributions of our two variables, but we have to remember that they're on different scales, so we cannot compare them using the unstandardized regression coefficient.

Standardized coefficients can be used if we only have two groups, but what if we have 3 or more?

Just like we use R^2 to report how much variance in Y is explained by the model, we can look at the contributions of each variable in the model, including factors with 3+ levels

```r
mod = lm(Y ~ X + D, data = df)
anova(mod)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X          1 64.045  64.045  61.489 4.788e-07 ***
## D          1 20.071  20.071  19.270 0.0003998 ***
## Residuals 17 17.707   1.042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\eta_X^2 = \frac{64.045}{64.045 + 20.071 + 17.707} = .62899$$

$$\eta_D^2 = \frac{20.071}{64.045 + 20.071 + 17.707} = .19712$$

Now extend this to include joint effects:

$$\hat{Y} = b_0 + b_1 D + b_2 X + b_3 DX$$

How do we interpret b_1?

b_1 is the difference in means between the two groups *when X is 0*.

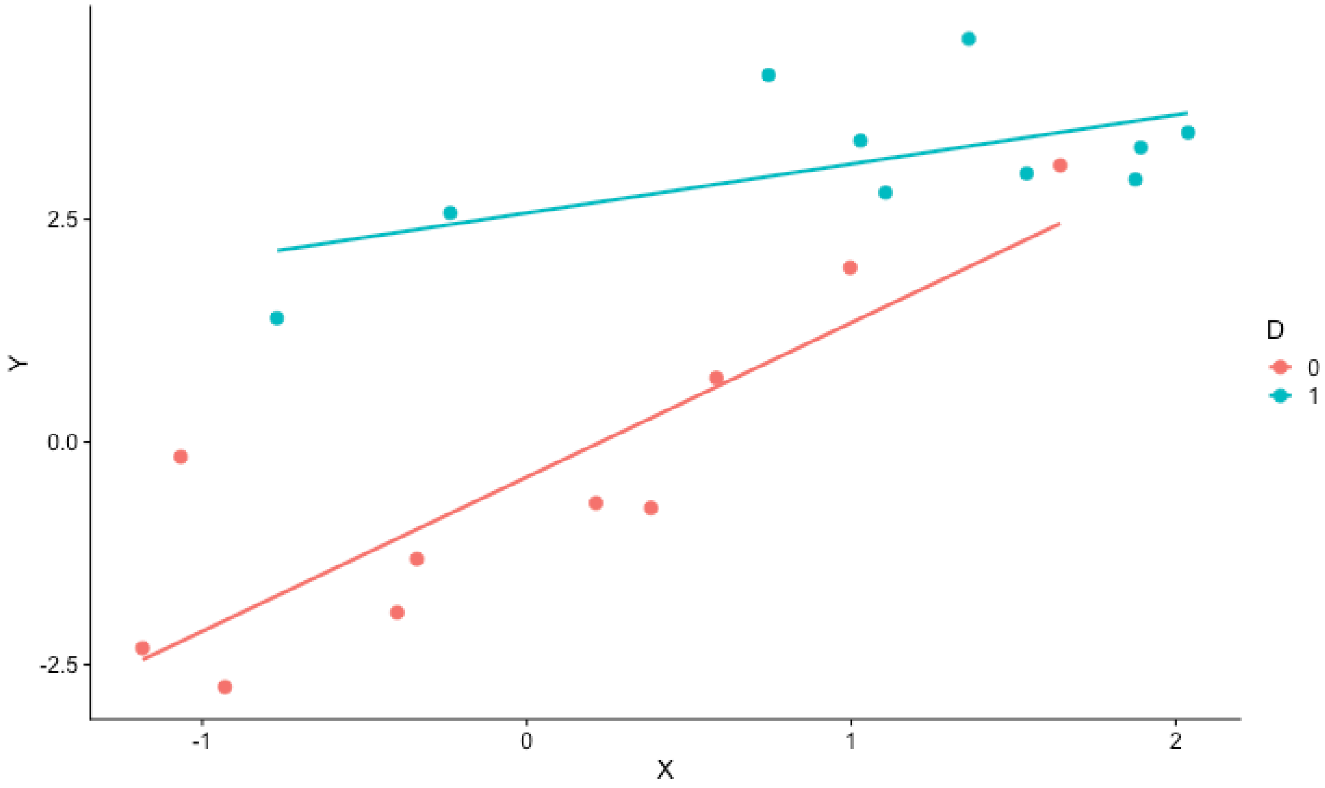
What is the interpretation of b_2?

b_2 is the slope of X among the reference group.

What is the interpretation of b_3?

b_3 is the difference in slopes between the reference group and the other group.

Visualizing



Where should we draw the segment to compare means?

Wash U contacts 150 alumni and collects their current salary (in thousands of dollars), their primary undergraduate major, and their GPA upon graduating.

```
library(psych)
table(inc_data$major)
```

```
##
##    Econ English   Psych
##      50      50      50
```

```
describe(inc_data[,c("gpa", "income")], fast = T)
```

```
##        vars   n  mean   sd   min    max  range   se
## gpa       1 150  3.36  0.4  2.44   4.19   1.74 0.03
## income    2 150 84.35 34.0 24.67 160.27 135.60 2.78
```

```r
career.mod = lm(income ~ gpa*major, data = inc_data)
summary(career.mod)
```

```
##
## Call:
## lm(formula = income ~ gpa * major, data = inc_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.625 -11.869   0.376   9.301  40.942
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -59.181     22.902  -2.584   0.0108 *
## gpa                 59.660      7.705   7.743 1.58e-12 ***
## majorEnglish       -81.747     37.149  -2.201   0.0294 *
## majorPsych        -175.314     35.462  -4.944 2.10e-06 ***
## gpa:majorEnglish    -4.562     11.089  -0.411   0.6814
## gpa:majorPsych      29.545     10.949   2.698   0.0078 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.91 on 144 degrees of freedom
## Multiple R-squared:  0.8142,    Adjusted R-squared:  0.8077
## F-statistic: 126.2 on 5 and 144 DF,  p-value: < 2.2e-16
```

Pop Quiz

For the model just defined in the previous slide...

- Write out the regression equation
- Interpret each term
- Where is Econ?

```r
inc_data$gpa_c = inc_data$gpa - mean(inc_data$gpa)
career.mod_c = lm(income ~ gpa_c*major, data = inc_data)
summary(career.mod_c)
```

```
##
## Call:
## lm(formula = income ~ gpa_c * major, data = inc_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.625 -11.869   0.376   9.301  40.942
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        141.428      3.752  37.691  < 2e-16 ***
## gpa_c               59.660      7.705   7.743 1.58e-12 ***
## majorEnglish       -97.086      4.907 -19.783  < 2e-16 ***
## majorPsych         -75.965      4.384 -17.327  < 2e-16 ***
## gpa_c:majorEnglish  -4.562     11.089  -0.411   0.6814
## gpa_c:majorPsych    29.545     10.949   2.698   0.0078 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.91 on 144 degrees of freedom
## Multiple R-squared:  0.8142,    Adjusted R-squared:  0.8077
## F-statistic: 126.2 on 5 and 144 DF,  p-value: < 2.2e-16
```

Pop Quiz 2

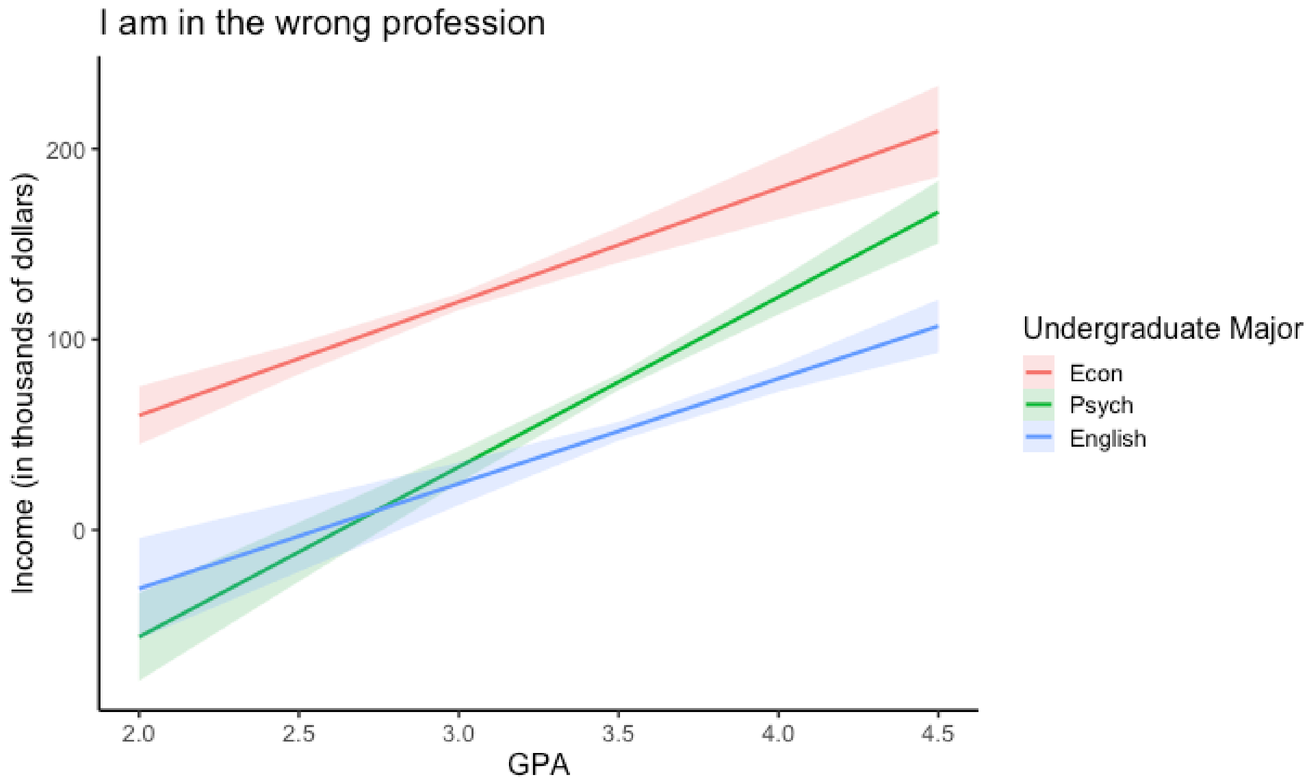
Now re-interpret each term for for the model where gpa is centered. What changes?

Plotting results

```
library(ggeffects)
predictedvals = ggpredict(model = career.mod, terms = c("gpa"

ggplot(data = predictedvals, aes(x = x, y = predicted, group
  geom_smooth(aes(ymin = conf.low,
                  ymax = conf.high,
                  color = group,
                  fill = group),
              stat = "identity",
              alpha = .2) +
  labs(x = "GPA",
       y = "Income (in thousands of dollars)",
       title = "I am in the wrong profession",
       color = "Undergraduate Major",
       fill = "Undergraduate Major",
       group = "Undergraduate Major") +
  theme_classic(base_size = 16)
```

Plotting Results



Two categorical predictors

If both X and M are categorical variables, the interpretation of coefficients is no longer the value of means and slopes, but means and differences in means.

Recall our Solomon's paradox example from a few weeks ago:

```
head(solomon[,c("PERSPECTIVE", "DISTANCE", "WISDOM")])
```

```
##   PERSPECTIVE  DISTANCE      WISDOM
## 1       other  immersed -0.27589395
## 2       other distanced  0.42949213
## 3       other distanced -0.02785874
## 4       other distanced  0.53271500
## 5        self distanced  0.62299793
## 6        self distanced -1.99578129
```

Model Means

```
solomon %>%
  group_by(DISTANCE, PERSPECTIVE) %>%
  summarize(meanWISDOM = mean(WISDOM, na.rm = TRUE))
```

```
## # A tibble: 4 × 3
## # Groups:   DISTANCE [2]
##   DISTANCE  PERSPECTIVE meanWISDOM
##   <fct>     <fct>            <dbl>
## 1 distanced other            0.334
## 2 distanced self             0.122
## 3 immersed  other            0.195
## 4 immersed  self            -0.559
```

```
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                 -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed                -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

```r
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                  -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed                 -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

```r
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                 -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed                -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

```
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed               -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed  -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

```r
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

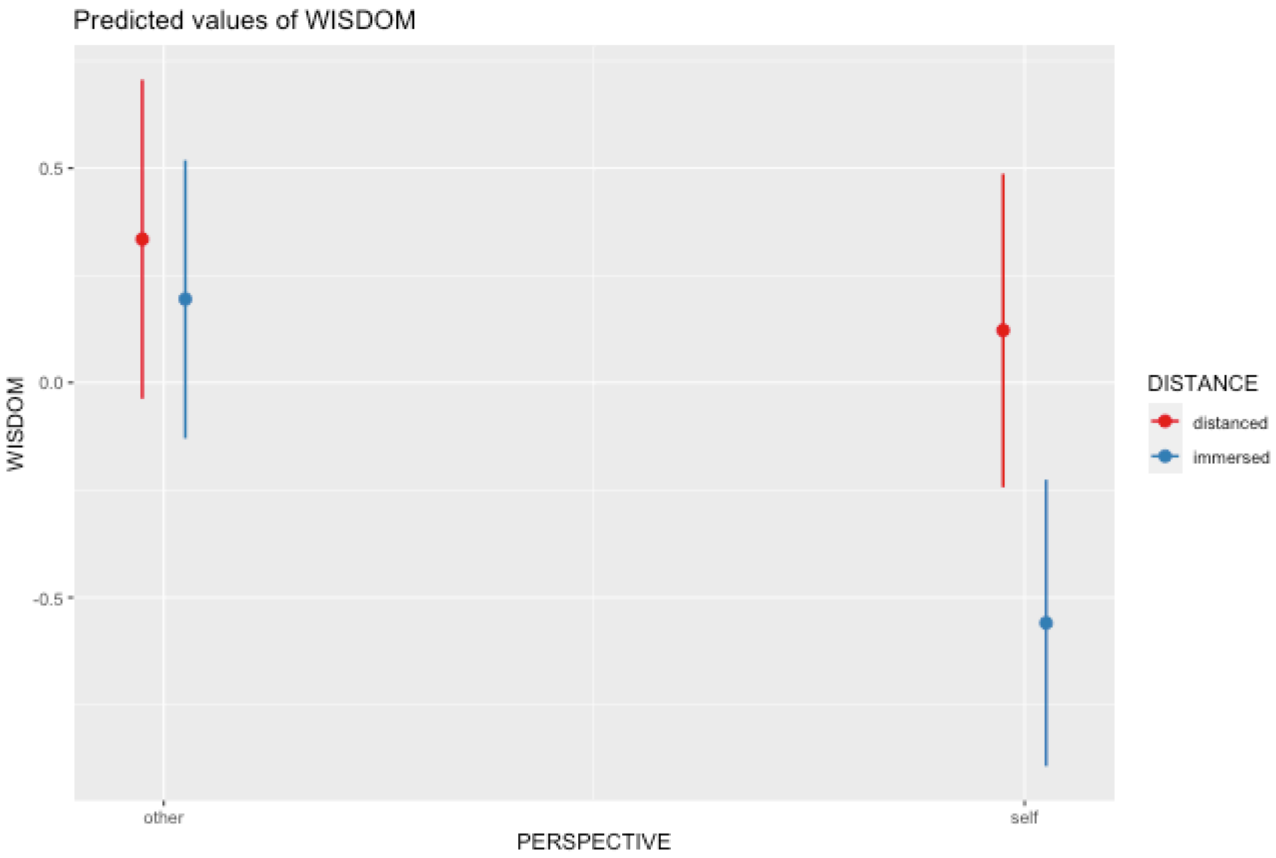
```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed               -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed  -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

```r
summary(lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomon))
```

```
##
## Call:
## lm(formula = WISDOM ~ PERSPECTIVE * DISTANCE, data = solomon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6809 -0.4209  0.0473  0.6694  2.3499
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.3345     0.1878   1.781   0.0776
## PERSPECTIVEself                 -0.2124     0.2630  -0.808   0.4210
## DISTANCEimmersed                -0.1396     0.2490  -0.561   0.5760
## PERSPECTIVEself:DISTANCEimmersed -0.5417     0.3526  -1.536   0.1273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9389 on 111 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.1262,    Adjusted R-squared:  0.1026
## F-statistic: 5.343 on 3 and 111 DF,  p-value: 0.001783
```

Plotting results

```
solomon.mod = lm(WISDOM ~ PERSPECTIVE*DISTANCE, data = solomo
plot_model(solomon.mod, type = "int")
```



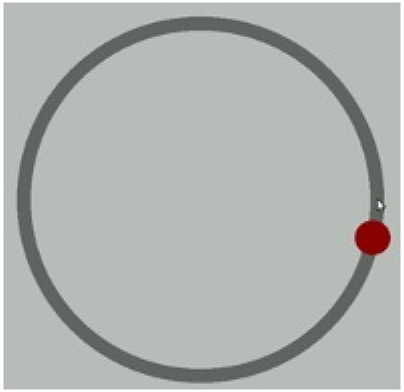
The interaction of two or more categorical variables in a general linear model is formally known as *Factorial ANOVA*.

A factorial design is used when there is an interest in how two or more variables (or factors) affect the outcome.

- Rather than conduct separate one-way ANOVAs for each factor, they are all included in one analysis.

- The unique and important advantage to a factorial ANOVA over separate one-way ANOVAs is the ability to examine interactions.

The example data are from a simulated study in which 180 participants performed an eye-hand coordination task in which they were required to keep a mouse pointer on a red dot that moved in a circular motion.



The outcome was the time of the 10th failure. The experiment used a completely crossed, 3 x 3 factorial design.

Coordination Study

One factor was dot speed: .5, 1, or 1.5 revolutions per second.

The second factor was noise condition. Some participants performed the task without any noise; others were subjected to periodic and unpredictable 3-second bursts of 85 dB white noise played over earphones. Of those subjected to noise, half could do nothing to stop the noise (uncontrollable noise); half believed they could stop the noise by pressing a button (controllable noise).

Terminology

In a **completely crossed** factorial design, each level of one factor occurs in combination with each level of the other factor.

If equal numbers of participants occur in each combination, the design is **balanced**. This has some distinct advantages (described later).

	Slow	**Medium**	**Fast**
No Noise	X	X	X
Controllable Noise	X	X	X
Uncontrollable Noise	X	X	X

Terminology

We describe the factorial ANOVA design by the number of **levels** of each **factor.**

- Factor: a variable that is being manipulated or in which there are two or more groups
- Level: the different groups within a factor

In this case, we have a 3 x 3 ANOVA ("three by three"), because our first factor (speed) has three levels (slow, medium, and fast) and our second factor (noise) also has three levels (none, controllable, and uncontrollable)

Noise	Slow	Medium	Fast	Marginal
None	630.72	525.29	329.28	495.10
Controllable	576.67	492.72	287.23	452.21
Uncontrollable	594.44	304.62	268.16	389.08
Marginal	600.61	440.88	294.89	445.46

3 Hypotheses. 2 correspond to questions that would arise in a simple one-way ANOVA:

Regardless of noise condition, does speed of the moving dot affect performance?

Regardless of dot speed, does noise condition affect performance?

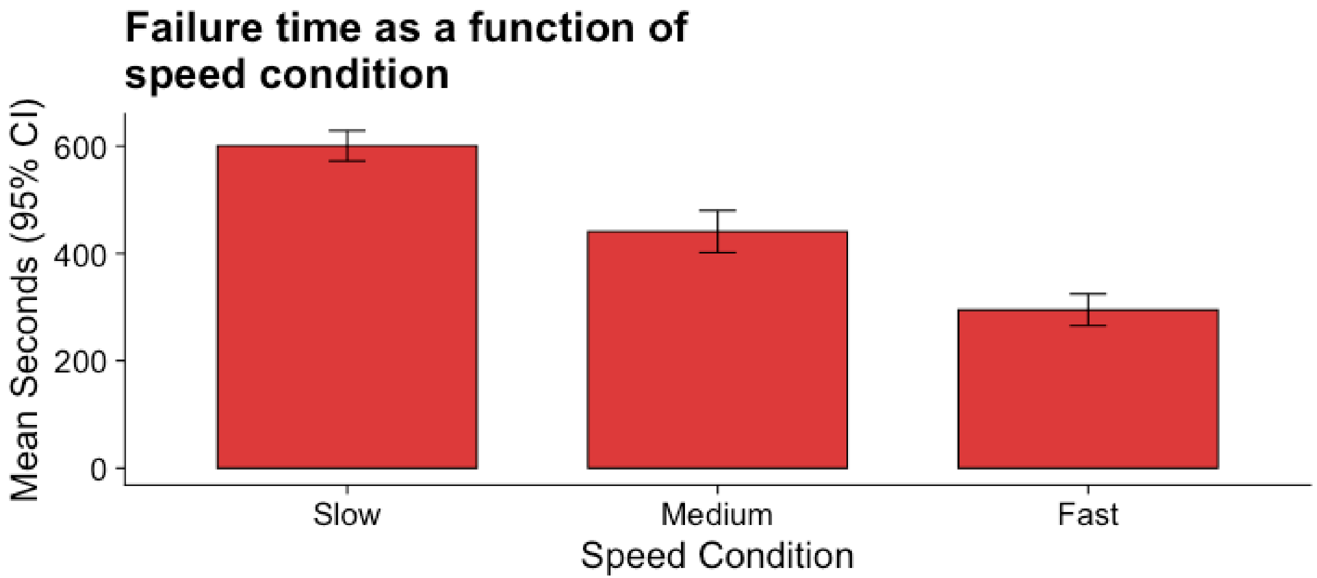
Marginal means

Noise	Slow	Medium	Fast	Marginal
None	630.72	525.29	329.28	495.10
Controllable	576.67	492.72	287.23	452.21
Uncontrollable	594.44	304.62	268.16	389.08
Marginal	**600.61**	**440.88**	**294.89**	**445.46**

We can answer those questions by examining the **marginal means**, which isolate one factor while collapsing across the other factor.

Regardless of noise condition, does speed of the moving dot affect performance?

```r
library(ggpubr)
ggbarplot(data = Data, x = "Speed", y = "Time", add = c("mean
```



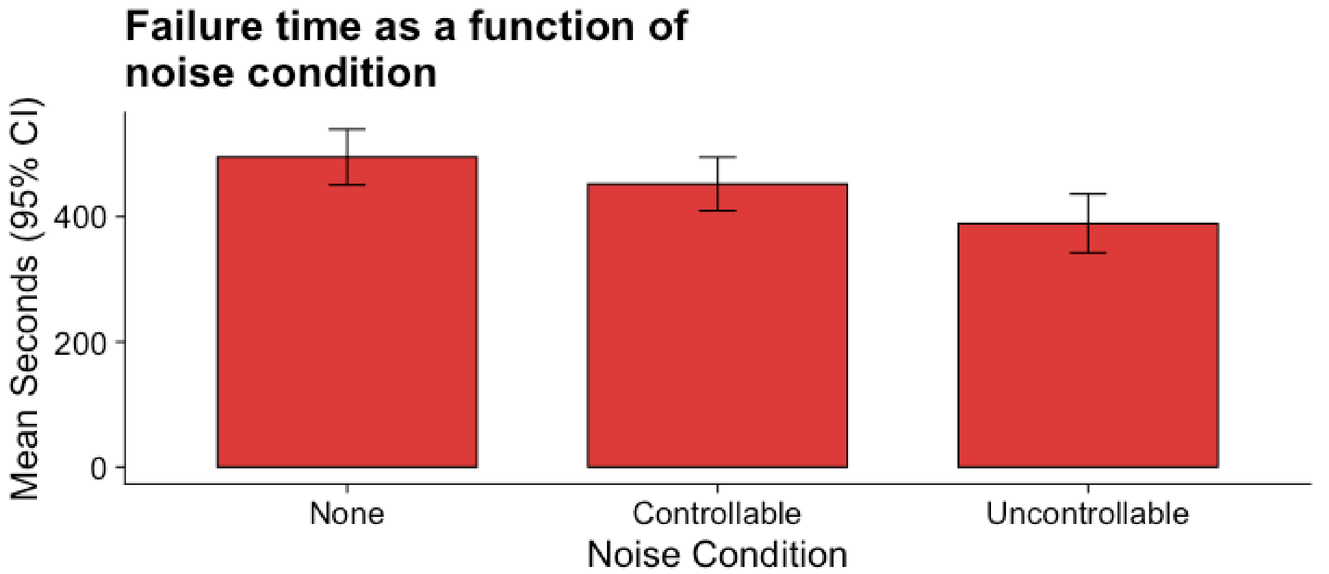
The ANOVA will be able to tell us if the means are significantly different and the magnitude of those differences in terms of variance accounted for.

Marginal means

Noise	Slow	Medium	Fast	Marginal
None	630.72	525.29	329.28	**495.10**
Controllable	576.67	492.72	287.23	**452.21**
Uncontrollable	594.44	304.62	268.16	**389.08**
Marginal	600.61	440.88	294.89	**445.46**

Regardless of dot speed, does noise condition affect performance? Performance declines in the presence of noise, especially if the noise is uncontrollable.

```r
ggbarplot(data = Data, x = "Noise", y = "Time", add = c("mean
```



The mean differences are not as apparent for this factor. The ANOVA will be particularly important for informing us about statistical significance and effect size.

Marginal means

Noise	Slow	Medium	Fast	Marginal
None	630.72	525.29	329.28	495.10
Controllable	576.67	492.72	287.23	452.21
Uncontrollable	594.44	304.62	268.16	389.08
Marginal	600.61	440.88	294.89	445.46

The **marginal mean differences** correspond to main effects. They tell us what impact a particular factor has, ignoring the impact of the other factor.

Marginal means

Noise	Slow	Medium	Fast	Marginal
None	630.72	525.29	329.28	495.10
Controllable	576.67	492.72	287.23	452.21
Uncontrollable	594.44	304.62	268.16	389.08
Marginal	600.61	440.88	294.89	445.46

The remaining effect in a factorial design, and it primary advantage over separate one-way ANOVAs, is the ability to examine **conditional mean differences**

One-way vs Factorial

Marginal Mean Differences

Results of one-way ANOVA

```
lm(y ~ GROUP)
```

$$\hat{Y} = b_0 + b_1 D$$

Conditional Mean Differences

Results of Factorial ANOVA

```
lm(y ~ GROUP*other_VARIABLE)
```

$$\hat{Y} = b_0 + b_1 D + b_2 O + b_3 DO$$

Next time

More Factorial ANOVA