MODERATION

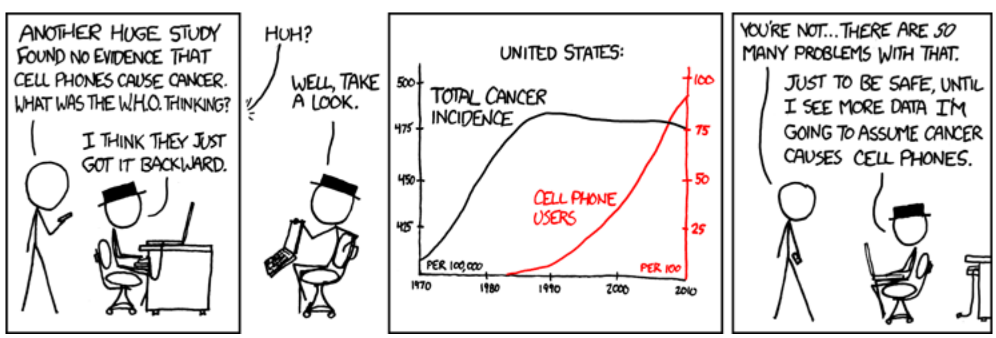
Research Methods in Psychology I & II ▪ Department of Psychology ▪ Colorado State University

BY THE END OF THIS UNIT YOU WILL:

1. Understand what it means for two variables to interact to predict an outcome.
2. Know how to test for an interaction between combinations of continuous and categorical variables.
3. Know how to plot and describe the results of a model that includes an interaction.

What is a moderator?

A moderator is a variable that affects (i.e., moderates) the relationship between two (or more) other variables. If the effect of x on y depends on z, then z is a moderator of the effect of x on y. For example, we can determine if the effect of a training program (x) on knowledge gained (y) is more beneficial for men compared to women (i.e., gender, z, moderates the effect of the training program).



© Kimberly L. Henry, Ph.D. ▪ kim.henry@colostate.edu

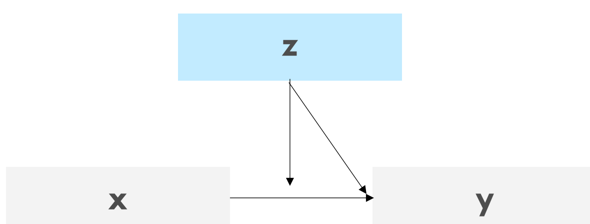
Overview

What is moderation in the context of regression analysis?

Moderation models are used to determine if the magnitude and/or direction of a certain regression slope varies as a function of some third variable. In a MLR model, we test for moderation by including interaction terms as predictors in our model. For example, if we hypothesize that the effect of variable x on outcome y varies as a function of z, then we test for this by regressing y on x, z and x•z (i.e., a new variable formed by multiplying x and z). If the regression slope for the interaction term (i.e., x•z) is statistically significant, then this provides evidence for moderation. Two variables interact to predict an outcome when above and beyond the additive component of each (i.e., y regressed on x and z), they have a joint or synergistic effect as well.

Some examples of moderation research questions in Psychology

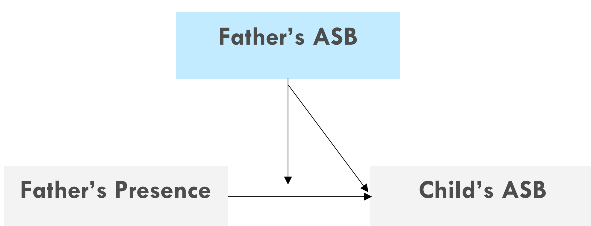
- Counseling Psychology: Is a new counseling therapy more effective for certain personality types?
- I-O Psychology: Does the effect of work autonomy on job satisfaction depend on the length of time an individual has been on the job?
- Social and Health Psychology: Does a strong bond between parent and child buffer the harmful effects of delinquent peer affiliation on adolescent substance use?
- Cognitive Neuroscience: Does cognitive reframing reduce the harmful effects of trauma history on depression?
- Cognitive Psychology: Does the effect of age on memory vary as a function of physical fitness?

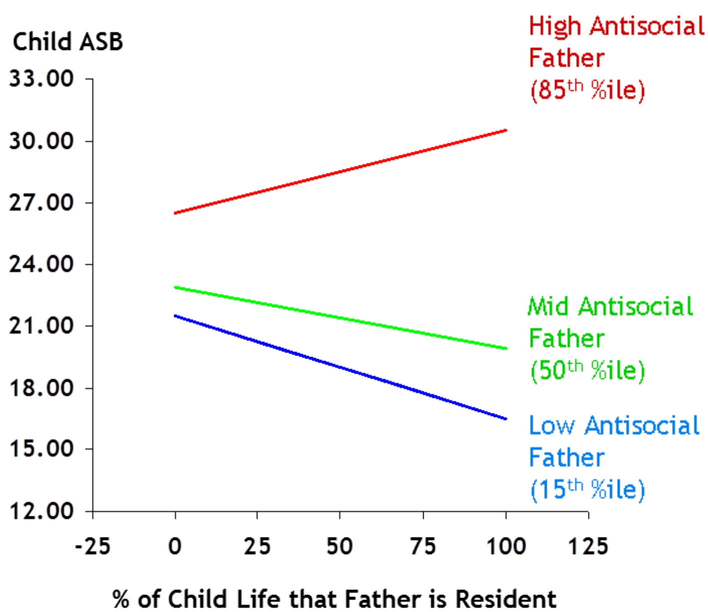


An Example

Jaffee, S.R., Moffitt, T.E., Caspi, A., & Taylor, A. (2003). Life with (or without) father: The benefits of living with two biological parents depend on the father's antisocial behavior. *Child Development, 74(1),* 109-126.

"The salutary effects of being raised by two married, biological parents depend on the quality of care parents can provide. Using data from an epidemiological sample of 1,116 5–year–old twin pairs and their parents, this study found that the less time fathers lived with their children, the more conduct problems their children had, but only if the fathers engaged in low levels of antisocial behavior. In contrast, when fathers engaged in high levels of antisocial behavior, the more time they lived with their children, the more conduct problems their children had. Behavioral genetic analyses showed that children who resided with antisocial fathers received a "double whammy" of genetic and environmental risk for conduct problems. Marriage may not be the answer to the problems faced by some children living in single–parent families unless their fathers can become reliable sources of emotional and economic support."





Steps to Test Moderation

Step 1: Center x and z to ensure that each has a meaningful 0 point (i.e., create 2 new variables xcen and zcen).

Step 2: Create an interaction term—multiply xcen and zcen.

Step 3: Regress y on xcen, zcen and xcen*zcen.

= x - 30 **= z - 45** **= xcen • zcen**

y	x	z	xcen	zcen	xcen*zcen
25	5	25	-25	-20	500
28	30	24	0	-21	0
32	60	29	30	-16	-480
17	7	45	-23	0	0
22	31	40	1	-5	-5
30	65	39	35	-6	-210
10	3	60	-27	15	-405
20	29	59	-1	14	-14
28	55	65	25	20	500

$$\hat{y}_i = b_0 + b_1 xcen_i + b_2 zcen_i + b_3 xcen_i \cdot zcen_i$$

b_0 = predicted value of y when all predictor variables are 0

b_1 = predicted change in y for a one unit increase in xcen **when zcen=0**

b_2 = predicted change in y for a one unit increase in zcen **when xcen=0**

b_3 = predicted difference in the effect of xcen on y (b_1) for a one unit increase in zcen

Therefore:

b_1 is the effect of xcen on y when zcen = 0

$b_1 + (b_3 \bullet 1)$ is the effect of xcen on y when zcen = +1

$b_1 + (b_3 \bullet -1)$ is the effect of xcen on y when zcen = -1

When we calculate the effect of x on y at a certain level of the moderator, we call these effects "simple slopes."

Example 1: Interaction of one categorical variable and one continuous variable.

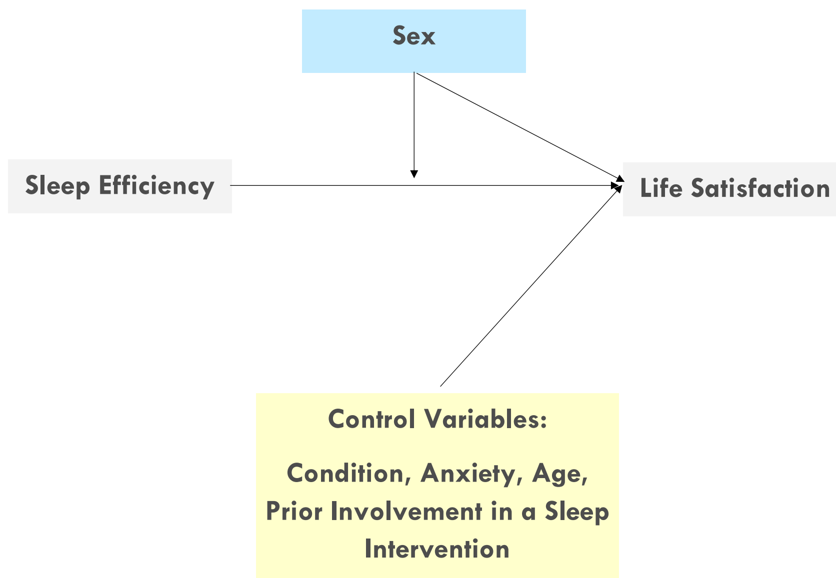
AN EXAMPLE

Sleep Intervention

A team of sleep researchers sought to study the effects of a 6-week sleep intervention aimed to improve participant's sleep hygiene. Sleep hygiene encompasses a variety of practices and habits that are necessary to have good nighttime sleep quality and full daytime alertness. The team formulated three different versions of the intervention. The first version (Condition 1) provided participants with a self-help book on the topic of sleep hygiene. The second version (Condition 2) brought participants together once per week in groups of 10-12 to teach the principles of sleep hygiene in a classroom setting. The final version (Condition 3) also used the group-based classroom setting of Condition 2, but in addition, each participant's partner was invited to also take part in the group sessions. Six-hundred male and female adults living with an intimate partner and suffering from a sleep disorder were recruited to take part in the study. The participants were randomly assigned to one of the three conditions.

DATASET: slpdata.csv

Research Question: Does the effect of sleep efficiency on life satisfaction differ for males compared to females?



Model 1: Regress life satisfaction on sex and sleep efficiency.

Model 2: Add the interaction term for sex and sleep efficiency to Model 1.

Model 3: Add the control variables to Model 2.

Import Data and Create Needed Dummy and Factor Variables

```
slp <- read_csv("slpdata.csv")

# Create a dummy coded variable for sex and dummy coded set for condition
slp <- mutate(slp,
        female = ifelse(sex == 1, 0, 1),
        female.f = factor(female, levels = c(0,1), labels = c("male", "female")),
        cond2 = ifelse(cond == 2, 1, 0),
        cond3 = ifelse(cond == 3, 1, 0),
        cond.f = factor(cond, levels = c(1,2,3), labels = c("self help", "group-based", "group + partner")))
```

Describe the Data

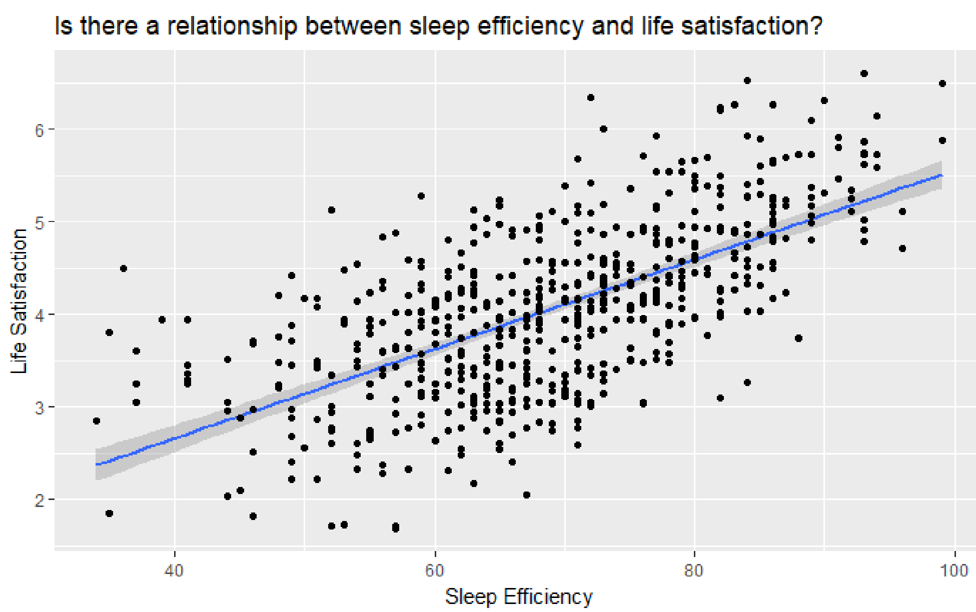
```
describe(slp)
```

```
        vars   n   mean     sd median trimmed    mad   min    max  range  skew kurtosis   se
cond       1 600   2.00   0.82   2.00    2.00   1.48  1.00   3.00   2.00  0.00    -1.50 0.03
prior      2 600   0.72   0.45   1.00    0.78   0.00  0.00   1.00   1.00 -1.00    -1.01 0.02
age        3 600  44.94  12.87  45.20   45.12  16.46 20.00  67.80  47.80 -0.10    -1.14 0.53
anxiety    4 600   3.88   0.90   3.86    3.89   0.93  1.05   6.84   5.79 -0.07    -0.06 0.04
hygiene    5 600   5.99   1.57   6.05    6.04   1.57  1.68   9.74   8.06 -0.23    -0.29 0.06
support    6 600   3.04   0.68   2.96    3.02   0.73  1.09   4.91   3.82  0.21    -0.51 0.03
sleep      7 600  68.88  12.14  69.00   69.09  11.86 34.00  99.00  65.00 -0.16    -0.17 0.50
lifesat    8 600   4.06   0.92   4.05    4.04   0.96  1.68   6.61   4.93  0.13    -0.23 0.04
sex        9 600   1.41   0.49   1.00    1.39   0.00  1.00   2.00   1.00  0.36    -1.87 0.02
id        10 600 300.50 173.35 300.50  300.50 222.39  1.00 600.00 599.00  0.00    -1.21 7.08
```

Make a Plot of Our Key Variables

Let's begin by making a scatterplot of the relationship between sleep efficiency and life satisfaction.

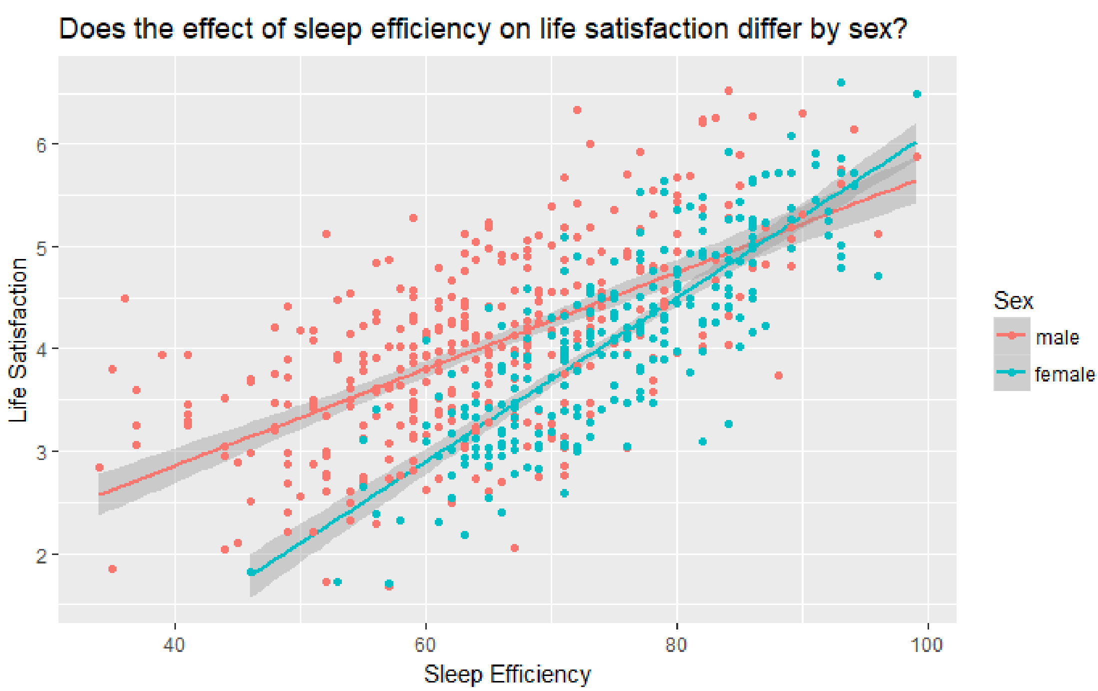
```
ggplot(slp, aes(x = sleep, y = lifesat)) +
  geom_smooth(method = lm, se = TRUE) +
  geom_point() +
  labs(title = "Is there a relationship between sleep efficiency and life satisfaction?",
      x = "Sleep Efficiency", y = "Life Satisfaction")
```



Layer Sex onto the Plot—Does the Relationship Differ for Males & Females?

Let's add sex as a grouping variable to out plot.

```
ggplot(slp, aes(x = sleep, y = lifesat, group = female.f, color = female.f)) +
  geom_smooth(method = lm, se = TRUE) +
  geom_point() +
  guides(color = guide_legend("Sex")) +
  labs(title = "Does the effect of sleep efficiency on life satisfaction differ by sex?",
     x = "Sleep Efficiency", y = "Life Satisfaction")
```



Model 1: Main Effects of Sleep Efficiency and Sex, No Interaction

In the first model, we will examine the main effects of our key predictor (sleep efficiency) and our potential moderator (sex) on our outcome (life satisfaction), without the inclusion of the interaction.

We will create a new dataset for this model, called slp_mix, and within it create a mean centered version of sleep (sleep_m). Then we will regress life satisfaction on female and sleep_m.

```
# center sleep at the mean
slp_mix <- mutate(slp, sleep_m = sleep - mean(sleep))

mix1 <- lm(data=slp_mix, lifesat ~ female + sleep_m)
ols_regress(mix1)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.688       RMSE                0.666
R-Squared               0.473       Coef. Var          16.418
Adj. R-Squared          0.471       MSE                 0.444
Pred R-Squared          0.467       MAE                 0.533
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
---------------------------------------------------------------------
              Sum of
             Squares         DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression   237.344          2        118.672    267.574    0.0000
Residual     264.776        597          0.444
Total        502.119        599
---------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
-----------------------------------------------------------------------------------------
(Intercept)     4.268         0.037                 115.202    0.000     4.195     4.341
     female    -0.514         0.061       -0.276     -8.407    0.000    -0.634    -0.394
    sleep_m     0.057         0.002        0.759     23.076    0.000     0.052     0.062
-----------------------------------------------------------------------------------------
```

Plot the Results of Model 1

Let's use the estimates from Model 1 to plot the fitted model.

```
predgrid_mix1 <- slp_mix %>%
  group_by(female.f) %>%
  data_grid(sleep_m = seq_range(sleep_m, 10),
           female) %>%
  ungroup() %>%
  add_predictions(mix1) %>%
  mutate(sleep = sleep_m + mean(slp_mix$sleep))

ggplot(predgrid_mix1, aes(x = sleep, y = pred, group = female.f, color = female.f)) +
  geom_line(size = 1) +
  guides(color = guide_legend("Sex")) +
  labs(title = "Results of Model 1 - No interaction",
     x = "Sleep Efficiency", y = "Life Satisfaction")
```

This plot does not do a very good job of reproducing what we observed in the raw data. This is because the model we fit forces the effect of sleep efficiency on life satisfaction to be the same for males and females.

Model 2: Add an Interaction Between Sleep Efficiency and Sex

```
mix2 <- lm(data=slp_mix, lifesat ~ female + sleep_m + female*sleep_m)
ols_regress(mix2)
```

Model Summary

R	0.711	RMSE	0.646
R-Squared	0.505	Coef. Var	15.918
Adj. R-Squared	0.503	MSE	0.417
Pred R-Squared	0.499	MAE	0.510

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	253.625	3	84.542	202.769	0.0000
Residual	248.494	596	0.417		
Total	502.119	599			

The R^2 is .51, indicating that we are explaining about 51% of the variability in life satisfaction with this model. The F* for the model is 202.77, and the p-value is less than alpha, indicating that we are explaining a significant portion of the variability.

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.225	0.037		115.515	0.000	4.153	4.297
female	-0.612	0.061	[illegible]329	-9.986	0.000	-0.733	-0.492
sleep_m	0.047	0.003	0.[illegible]28	16.431	0.000	0.042	0.053
female:sleep_m	0.033	0.005	0.2[illegible]	6.249	0.000	0.022	0.043

Do not rely on the default standardized betas when you have an interaction term — we'll come back to this later.

Interpretation of parameter estimates:

Intercept: The predicted life satisfaction when all predictor variables are 0. That is, we predict males with an average level of sleep efficiency to have a life satisfaction score of 4.23.

female: This variable is involved in an interaction, so this is a simple slope. Specifically, it is the effect of female when sleep_m = 0. Therefore, it is the predicted difference in life satisfaction between females and males, **among people with an average level of sleep efficiency**. The slope is negative, meaning that among people with an average sleep efficiency, we predict that females will have a life satisfaction score that is .61 units lower than males. This is a statistically significant difference (p-value is less than alpha).

sleep_m: This variable is also involved in an interaction, so it too is a simple slope. That is, the effect of sleep_m when female = 0. Therefore, it is the predicted change in life satisfaction for a one-unit increase in sleep efficiency **among males**. Among males, we predict that each one-unit increase in sleep efficiency will be associated with a .05 unit increase in life satisfaction. This is a statistically significant effect (p-value is less than alpha).

female:sleep_m: This is the interaction term between female and sleep_m. Recall that an interaction term is interpreted as the predicted difference in the effect of x on y for a one unit increase in z. It captures how the simple slope of x on y changes as z increases. It is the predicted **difference in the effect** of sleep_m on life satisfaction for females compared to males— each one unit increase in sleep has a larger effect on life satisfaction for females as compared to males. It is statistically significant (p-value is less than alpha). To obtain the effect of sleep efficiency on life satisfaction for females, we add this slope (.033) to the slope for males (.047): .047 + .033 = .08. Among females, we predict that each one-unit increase in sleep efficiency will be associated with a .08 unit increase in life satisfaction.

Let's rewrite the equation to specify the relationship between sleep_m and lifesat for males and females separately:

$$\hat{Y}_i = 4.23 + (-.61 \cdot female_i) + (.05 \cdot sleep\_m_i) + (.03 \cdot female_i \cdot sleep\_m_i)$$

$$\hat{Y}_i \text{ for males: } 4.23 + (-.61 \cdot 0) + (.05 \cdot sleep\_m_i) + (.03 \cdot 0 \cdot sleep\_m_i) = 4.23 + (.05 \cdot sleep\_m_i)$$

$$\hat{Y}_i \text{ for females: } 4.23 + (-.61 \cdot 1) + (.05 \cdot sleep\_m_i) + (.03 \cdot 1 \cdot sleep\_m_i) = 3.61 + (.08 \cdot sleep\_m_i)$$

Modify Model 2 by Switching Reference Group for Sex

Let's create a new dummy with female as the reference group and refit the model.

```
slp_mix <- mutate(slp_mix, male = ifelse(sex == 1, 1, 0))

mix2_switch <- lm(data=slp_mix, lifesat ~ male + sleep_m + male*sleep_m)
ols_regress(mix2_switch)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.711       RMSE                0.646
R-Squared               0.505       Coef. Var          15.918
Adj. R-Squared          0.503       MSE                 0.417
Pred R-Squared          0.499       MAE                 0.510
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
---------------------------------------------------------------------
               Sum of
              Squares        DF    Mean Square       F          Sig.
---------------------------------------------------------------------
Regression    253.625         3         84.542    202.769     0.0000
Residual      248.494       596          0.417
Total         502.119       599
---------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
-----------------------------------------------------------------------------------------
 (Intercept)     3.612         0.049                 73.375    0.000      3.516     3.709
        male     0.612         0.061        0.329     9.986    0.000      0.492     0.733
     sleep_m     0.080         0.004        1.061    18.316    0.000      0.071     0.089
male:sleep_m    -0.033         0.005       -0.336    -6.249    0.000     -0.043    -0.022
-----------------------------------------------------------------------------------------
```

Notice that now the intercept and slope for sleep_m represent the corresponding effects for females, since they are the group coded 0 for the dummy variable called male. In the previous model (with the female dummy code) we were able to use the estimates of the model to obtain the intercept and slope for females; however, we didn't get a significance test for them. Now, with this refit model, we obtain the estimate and significance test for females. For example, the effect of sleep_m for females (.08) is statistically significant. Across the two models, we now know that there is a positive and significant effect of sleep efficiency on life satisfaction for males and females, but the beneficial effect is larger for females.

Y-hat$_i$ for males = 4.23 + (.05 ▪ sleep_m$_i$)

Y-hat$_i$ for females: = 3.61 + (.08 ▪ sleep_m$_i$)

Obtain the Same Result with a Linear Hypothesis Test

We can also test whether the effect of sleep_m on lifesat is significant for females by using the linearHypothesis function with the original model. Recall from Unit 6 that the null hypothesis is that the calculated value is 0. Here we specify the slope for females (the male slope + the interaction), and the function tells us whether this calculated value is significantly different from 0. The F* is 335.49, and the corresponding p-value is less than alpha, indicating that the null hypothesis can be rejected, this calculated slope (.08) is significantly different that 0. Notice that the t* for sleep_m in the model above (which is the same slope we're testing here) is 18.316, this t* squared equals the F* below (335.49). As usual, this is because both tests are assessing the same quantity in the same manner.

```
linearHypothesis(mix2, "sleep_m + female:sleep_m")
```

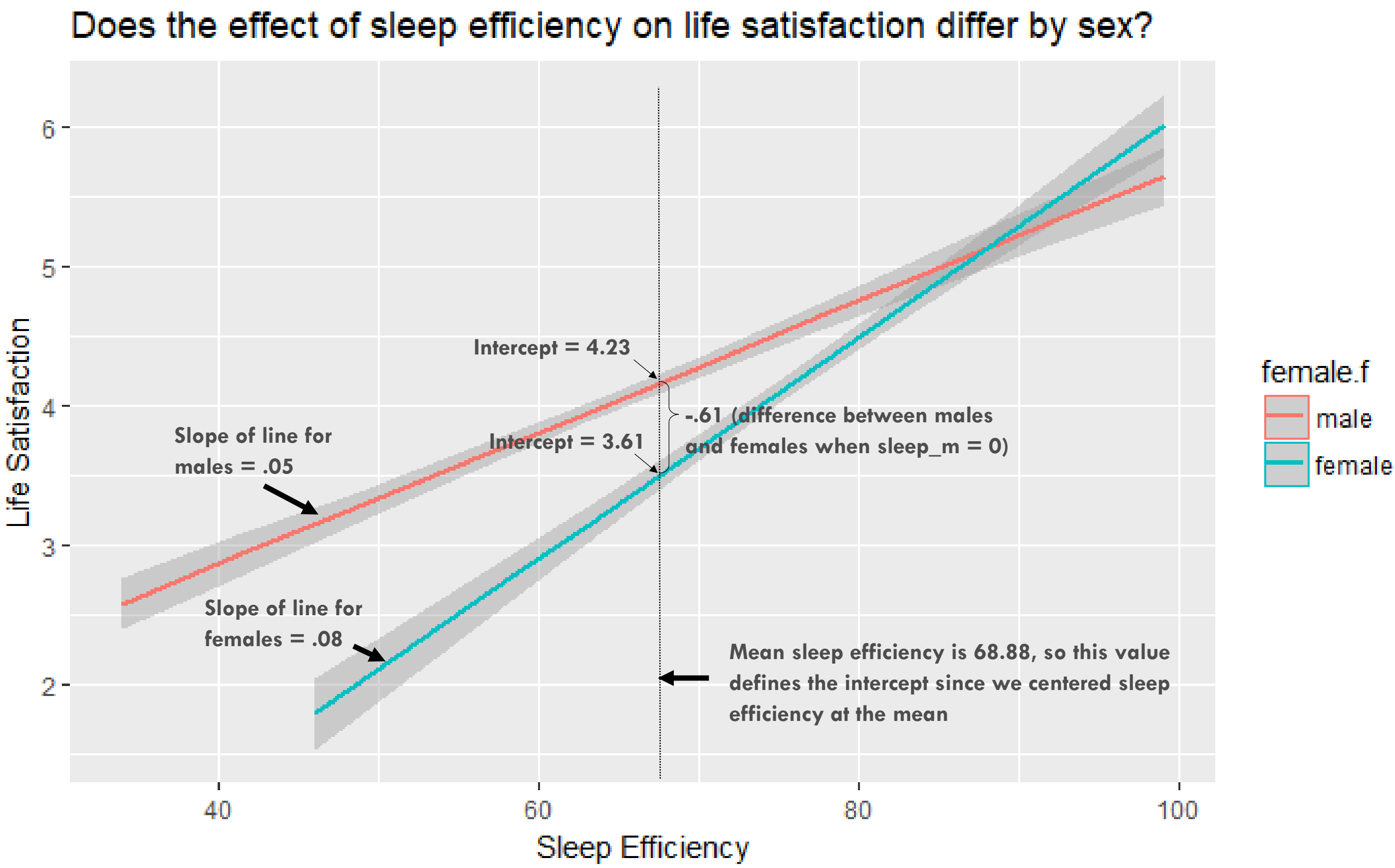
```
Linear hypothesis test

Hypothesis:
sleep_m  + female:sleep_m = 0

Model 1: restricted model
Model 2: lifesat ~ female + sleep_m + female * sleep_m

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    597 388.37
2    596 248.49  1    139.88 335.49 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot the Results of Model 2 and Add Confidence Bands

```
predgrid_mix2 <- slp_mix %>%
  group_by(female.f) %>%
  data_grid(sleep_m = seq_range(sleep_m, 10),
            female) %>%
  ungroup()

predictions_mix2 <- predict(mix2, predgrid_mix2, interval = "confidence") %>%
  as_data_frame()

adjeff_mix2 <- cbind(predgrid_mix2, predictions_mix2) %>%
  mutate(sleep = sleep_m + mean(slp_mix$sleep))

ggplot(adjeff_mix2, aes(x = sleep, y = fit, group = female.f, color = female.f)) +
  geom_ribbon(aes(ymin = lwr, ymax = upr, color = NULL), alpha = .4, fill = "grey60") +
  geom_line(size = 1) +
  guides(color = guide_legend("Sex")) +
  labs(title = "Does the effect of sleep efficiency on life satisfaction differ by sex?",
     x = "Sleep Efficiency", y = "Life Satisfaction")
```

Now that we have accounted for the interaction in our model, the fitted model is a good representation of the observed data.

Map Estimates onto Graph



Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.225	0.037		115.515	0.000	4.153	4.297
female	-0.612	0.061	-0.329	-9.986	0.000	-0.733	-0.492
sleep_m	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep_m	0.033	0.005	0.242	6.249	0.000	0.022	0.043

Y-hat for males: 4.23 + (-.61 ▪ 0) + (.05 ▪ sleep_m$_i$) + (.03 ▪ 0 ▪ sleep_m$_i$) = 4.23 + (.05 ▪ sleep_m$_i$)

Y-hat for females: 4.23 + (-.61 ▪ 1) + (.05 ▪ sleep_m$_i$) + (.03 ▪ 1 ▪ sleep_m$_i$) = 3.61 + (.08 ▪ sleep_m$_i$)

Probe the Effect of Sex on Life Satisfaction at Different Levels of Sleep Efficiency

We can think about our interaction effect in an alternative way by considering the differential effect of sex on life satisfaction at different levels of sleep efficiency. In our plot, we can see that when sleep efficiency is low, males tend to have higher life satisfaction, but that difference dissipates as sleep efficiency increases. Let's quantify the sex difference at a few prototypical values of sleep efficiency (sleep efficiency = 60, 75, and 95). This technique is called probing the interaction.

```
slp_mix <- mutate(slp_mix,
            sleep60 = sleep - 60,
            sleep75 = sleep - 75,
            sleep95 = sleep - 95)

mix2_60 <- lm(data=slp_mix, lifesat ~ female + sleep60 + female*sleep60)
ols_regress(mix2_60)

mix2_75 <- lm(data=slp_mix, lifesat ~ female + sleep75 + female*sleep75)
ols_regress(mix2_75)

mix2_95 <- lm(data=slp_mix, lifesat ~ female + sleep95 + female*sleep95)
ols_regress(mix2_95)
```

First, we create several new versions of sleep centered around our prototypical values of interest.

Then we estimate the regression model with each centered version of sleep — in this way, we will obtain the simple slope for the effect of female (i.e., comparison of females to males) when sleep efficiency = 60, when sleep efficiency = 75, and when sleep efficiency = 95. The different slopes for female are called simple slopes, because they represent the effect of the variable at a particular level of the other variable.

When sleep = 60, sleep60=0. Therefore, the slope for female, is the predicted difference in life satisfaction between females and males when sleep efficiency = 60. Here, we see a large, negative and statistically significant difference.

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	3.805	0.037		103.493	0.000	3.733	3.877
female	-0.903	0.086	-0.486	-10.505	0.000	-1.071	-0.734
sleep60	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep60	0.033	0.005	0.342	6.249	0.000	0.022	0.043

In this model, the slope for female, is the predicted difference in life satisfaction between females and males when sleep efficiency = 75. Here, we see a moderate, negative and statistically significant difference.

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.515	0.046		98.751	0.000	4.425	4.604
female	-0.412	0.061	-0.222	-6.706	0.000	-0.533	-0.291
sleep75	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep75	0.033	0.005	0.216	6.249	0.000	0.022	0.043

In this model, the slope for female, is the predicted difference in life satisfaction between females and males when sleep efficiency = 95. Here, we see a small, positive, but not significant difference.

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	5.461	0.094		57.947	0.000	5.276	5.646
female	0.242	0.135	0.130	1.796	0.073	-0.023	0.506
sleep95	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep95	0.033	0.005	0.411	6.249	0.000	0.022	0.043

Map Estimates onto Graph

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	3.805	0.037		103.493	0.000	3.733	3.877
female	-0.903	0.086	-0.486	-10.505	0.000	-1.071	-0.734
sleep60	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep60	0.033	0.005	0.342	6.249	0.000	0.022	0.043

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.515	0.046		98.751	0.000	4.425	4.604
female	-0.412	0.061	-0.222	-6.706	0.000	-0.533	-0.291
sleep75	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep75	0.033	0.005	0.216	6.249	0.000	0.022	0.043

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	5.461	0.094		57.947	0.000	5.276	5.646
female	0.242	0.135	0.130	1.796	0.073	-0.023	0.506
sleep95	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep95	0.033	0.005	0.411	6.249	0.000	0.022	0.043

Use the Initial Equation to Obtain the Simple Slopes

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.225	0.037		115.515	0.000	4.153	4.297
female	-0.612	0.061	-0.329	-9.986	0.000	-0.733	-0.492
sleep_m	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep_m	0.033	0.005	0.242	6.249	0.000	0.022	0.043

Y-hat = 4.23 + (-.61 ▪ $female_i$) + (.05 ▪ $sleep\_m_i$) + (.03 ▪ $female_i$ ▪ $sleep\_m_i$)

Y-hat_i for sleep efficiency = 60 (sleep_m = -8.88): 4.23 + (-.61 ▪ $female_i$) + (.05 ▪ -8.88) + (.03 ▪ $female_i$ ▪ -8.88) = 3.81 + (-.90 ▪ $female_i$)

Y-hat_i for sleep efficiency = 75 (sleep_m = 6.12): 4.23 + (-.61 ▪ $female_i$) + (.05 ▪ 6.12) + (.03 ▪ $female_i$ ▪ 6.12) = 4.52 + (-.41 ▪ $female_i$)

Y-hat_i for sleep efficiency = 95 (sleep_m = 26.12): 4.23 + (-.61 ▪ $female_i$) + (.05 ▪ 26.12) + (.03 ▪ $female_i$ ▪ 26.12) = 5.46 + (.24 ▪ $female_i$)

You must be sure to plug in the prototypical value that is associated with the version of the variable that you included in the model. We included sleep_m—so we have to transform our desired prototypical value of interest before plugging it into the equation. For example, a sleep score of 60 corresponds to a sleep_m score of –8.88 (60 - 68.88 = -8.88)

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	3.805	0.037		103.493	0.000	3.733	3.877
female	-0.903	0.086	-0.486	-10.505	0.000	-1.071	-0.734
sleep60	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep60	0.033	0.005	0.342	6.249	0.000	0.022	0.043

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	4.515	0.046		98.751	0.000	4.425	4.604
female	-0.412	0.061	-0.222	-6.706	0.000	-0.533	-0.291
sleep75	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep75	0.033	0.005	0.216	6.249	0.000	0.022	0.043

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	5.461	0.094		57.947	0.000	5.276	5.646
female	0.242	0.135	0.130	1.796	0.073	-0.023	0.506
sleep95	0.047	0.003	0.628	16.431	0.000	0.042	0.053
female:sleep95	0.033	0.005	0.411	6.249	0.000	0.022	0.043

Regions of Significance

Rather than considering the effect of some variable (e.g., female) at just a few prototypical values of the moderator (e.g., sleep efficiency), we can create a Johnson-Neyman graph to present the regions of the moderator for which the effect of the predictor on the outcome is statistically significant.

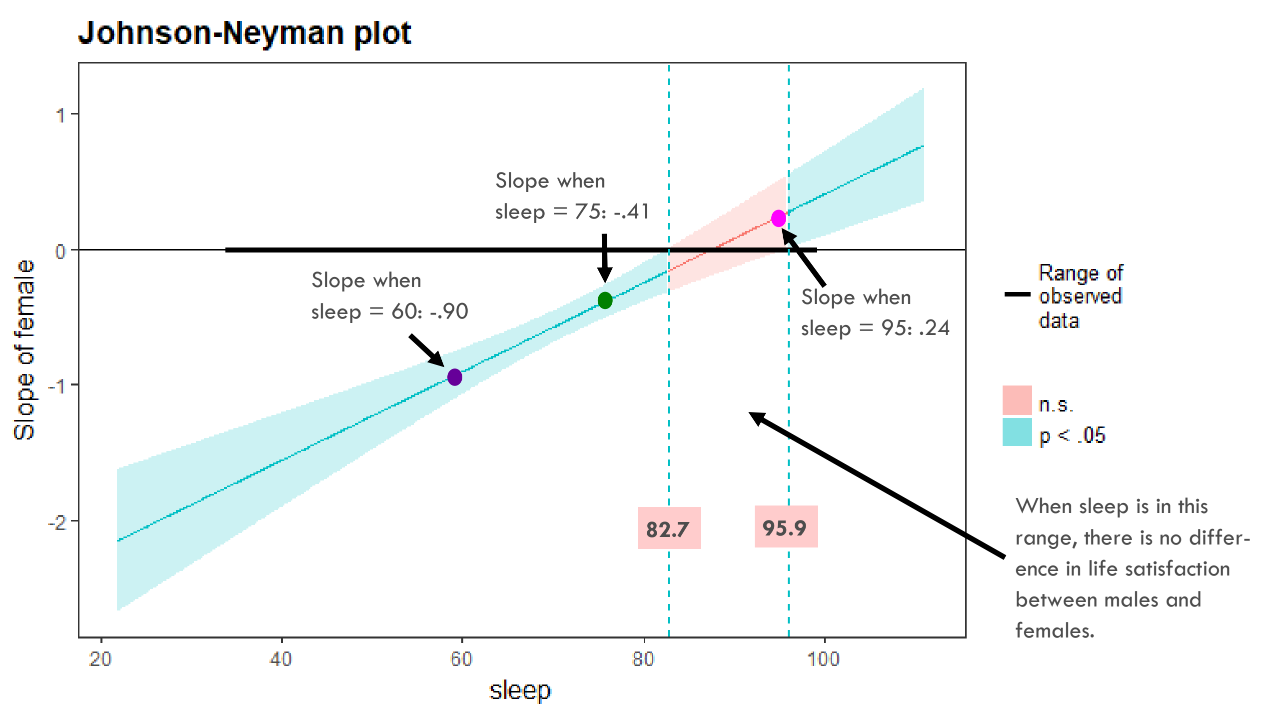
Johnson, P. O. & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs, 1*, 57–93.

Long, J.A. (2017). jtools: Analysis and Presentation of Social Scientific Data. R package version 0.8.0, https://cran.r-project.org/package=jtools.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437-448.

```
mix2_jplot1 <- lm(data = slp_mix, lifesat ~ female + sleep + female*sleep)

probe_interaction(mix2_jplot1, pred = female, modx = sleep,
          jnplot = TRUE,
          x.label = "Sex",
          y.label = "Life Satisfaction",
          main.title = "Differential effect of sex on life satisfaction by sleep efficiency",
          legend.main = "Sleep Efficiency")
```

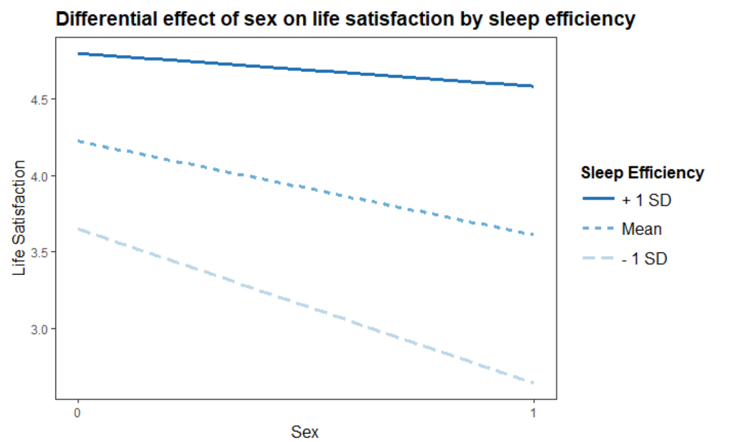
For the jtools package, it is better to feed in a model without any centering. Therefore, I create a new model that doesn't include the centered version of sleep and feed those model results to the probe_interaction function.



Rather than assessing the effect of sex on life satisfaction at just a few values of the moderator, we now get to see the effect of sex on life satisfaction across the whole range of sleep efficiency.

Other Good Things from the jtools Package

```
JOHNSON-NEYMAN INTERVAL

The slope of female is p < .05 when sleep is OUTSIDE this interval:
[82.6941, 95.9424]
Note: The range of observed values of sleep is [34, 99]

SIMPLE SLOPES ANALYSIS

Slope of female when sleep = 81.016 (+ 1 SD):
  Est.   S.E.      p
-0.215  0.076  0.005

Slope of female when sleep = 68.877 (Mean):
  Est.   S.E.      p
-0.612  0.061  0.000

Slope of female when sleep = 56.737 (- 1 SD):
  Est.   S.E.      p
-1.009  0.099  0.000
```



The package will calculate the simple slopes of the key predictor on the outcome at different levels of the moderator: by default you will get 1 SD below the mean, at the mean, and 1 SD above the mean. Notice these line up with the in class activity that you completed.

The package will also plot the effects of the key predictor on the outcome at different levels of the moderator, this is the same plot that you created in the in class activity. Check out this page for even more options: https://cran.r-project.org/web/packages/jtools/vignettes/interactions.html

Let's take a look at the plot when we put sleep on the x axis (our original specification).

```
mix2_jplot2 <- lm(data = slp_mix, lifesat ~ female.f + sleep + female.f*sleep)

probe_interaction(mix2_jplot2, pred = sleep, modx = female.f,
             jnplot = FALSE,
             x.label = "Sleep Efficiency",
             y.label = "Life Satisfaction",
             main.title = "Differential effect of sleep efficiency on life satisfaction by sex",
             legend.main = "Female")
```

For a categorical moderator, the Johnson-Neyman plot isn't appropriate, so I add jnplot = FALSE. Refitting the model with female.f rather than female will allow the legend to be automatically labeled.

Ignore this part since we have a categorical moderator.

```
JOHNSON-NEYMAN INTERVAL

The slope of sleep is p < .05 when female.f is OUTSIDE this interval:
[-2.2655, -1.0107]
Note: The range of observed values of female.f is [0, 1]

SIMPLE SLOPES ANALYSIS

Slope of sleep when female.f = 1 (female):
 Est.  S.E.     p
0.080 0.004 0.000

Slope of sleep when female.f = 0 (male):
 Est.  S.E.     p
0.047 0.003 0.000
```

These are the slopes for the effect of sleep efficiency on life satisfaction for females and males that we calculated earlier.

Model 3: Add the Control Variables

Finally, we will add the control variables to the model. We will mean center the two predictors that don't have a meaningful 0 point first.

```
# center control variables
slp_mix <- mutate(slp_mix,
                  age30 = age-30,
                  anxiety_m = anxiety - mean(anxiety))

mix3 <- lm(data=slp_mix, lifesat ~ cond2 + cond3 + prior + age30 + anxiety_m + female + sleep_m + female*sleep_m)
ols_regress(mix3)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.796       RMSE                0.558
R-Squared               0.633       Coef. Var          13.762
Adj. R-Squared          0.628       MSE                 0.312
Pred R-Squared          0.622       MAE                 0.439
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
---------------------------------------------------------------------
                 Sum of
                Squares         DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression      317.959          8         39.745    127.547    0.0000
Residual        184.161        591          0.312
Total           502.119        599
---------------------------------------------------------------------

                                    Parameter Estimates
-------------------------------------------------------------------------------------------
        model        Beta    Std. Error    Std. Beta        t        Sig      lower     upper
-------------------------------------------------------------------------------------------
   (Intercept)      3.699         0.076                48.458      0.000      3.549     3.848
         cond2      0.116         0.063        0.060    1.843      0.066     -0.008     0.239
         cond3      0.374         0.072        0.193    5.207      0.000      0.233     0.515
         prior     -0.056         0.051       -0.028   -1.101      0.271     -0.157     0.044
         age30      0.021         0.002        0.289   11.328      0.000      0.017     0.024
     anxiety_m     -0.250         0.027       -0.247   -9.320      0.000     -0.302    -0.197
        female     -0.348         0.060       -0.187   -5.806      0.000     -0.465    -0.230
       sleep_m      0.034         0.003        0.444   11.294      0.000      0.028     0.039
female:sleep_m      0.028         0.005        0.206    6.104      0.000      0.019     0.037
-------------------------------------------------------------------------------------------
```

R^2 is .63, indicating that about 63% of the variability in life satisfaction is predicted by the variables in this model. The F* is statistically significant, indicating that this amount is significantly greater than 0.

Interpretation of Parameter Estimates from Model 3

Interpretation of parameter estimates:

Intercept: The predicted life satisfaction when all predictor variables are 0. That is, a 30 year old male in Condition 1 who has no prior sleep intervention history and who has an average level of anxiety and sleep efficiency.

cond2: The predicted difference in life satisfaction for people in Condition 2 compared to Condition 1, **holding constant all other predictors**. This is not a significant difference.

cond3: The predicted difference in life satisfaction for people in Condition 3 compared to Condition 1, **holding constant all other predictors**. This is a significant difference, people in Condition 3 tend to have better life satisfaction than people in Condition 1.

prior: The predicted difference in life satisfaction between those who have and haven't participated in a sleep intervention in the past, **holding constant all other predictors**. This is not a significant difference.

age30: The predicted change in life satisfaction for each 1 year increase in age, **holding constant all other predictors**. This is a positive (i.e., older people have more life satisfaction) and statistically significant effect.

anxiety_m: The predicted change in life satisfaction for each 1 unit increase in anxiety, **holding constant all other predictors**. This is a negative (i.e., increased anxiety is associated with less life satisfaction) and statistically significant effect.

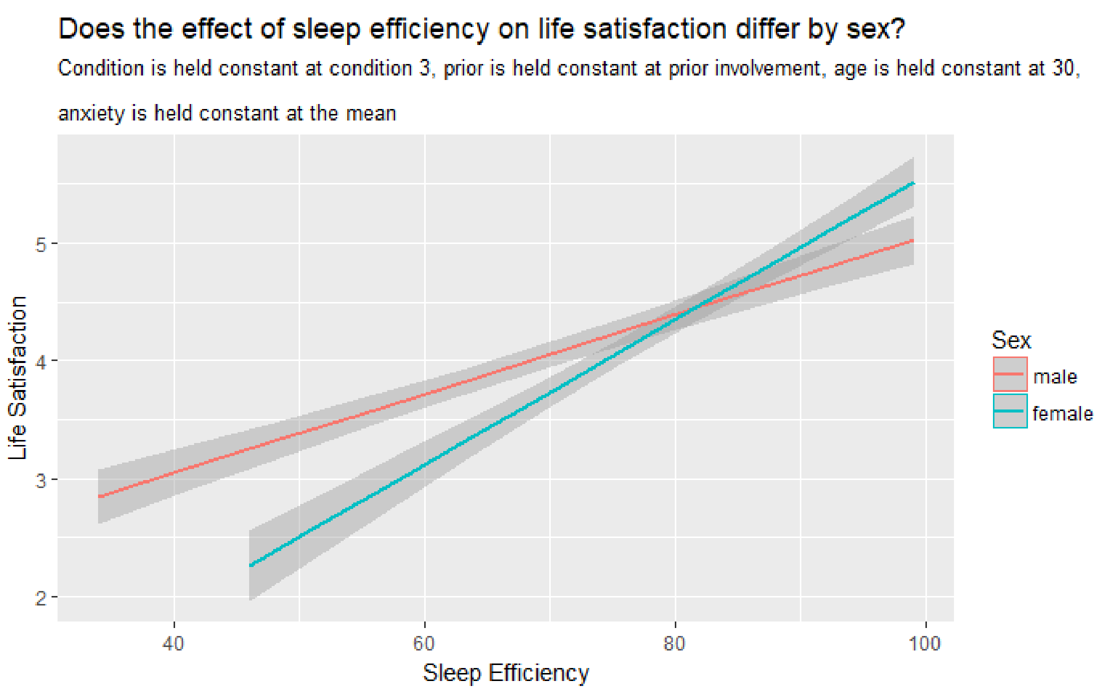
female: The predicted difference in life satisfaction between females and males among people with an average level of sleep efficiency, and **holding constant all other predictors**. The slope is negative, meaning that among people with an average sleep efficiency, we predict that females will have a life satisfaction score that is .35 units lower than males. This is a statistically significant difference (p-value is less than alpha).

sleep_m: The predicted change in life satisfaction for a one-unit increase in sleep efficiency among males, **holding constant all other predictors**. Among males, we predict that each one-unit increase in sleep efficiency will be associated with a .03 unit increase in life satisfaction. This is a statistically significant effect (p-value is less than alpha).

female:sleep_m: This is the interaction term between female and sleep_m. It is the predicted **difference** in the effect of sleep_m on life satisfaction for females compared to males, **holding constant all other predictors**. Each one unit increase in sleep has a larger effect on life satisfaction for females as compared to males. It is statistically significant (p-value is less than alpha). The slope for females is .034 + .028 = .06.

Plot the Results of Model 3

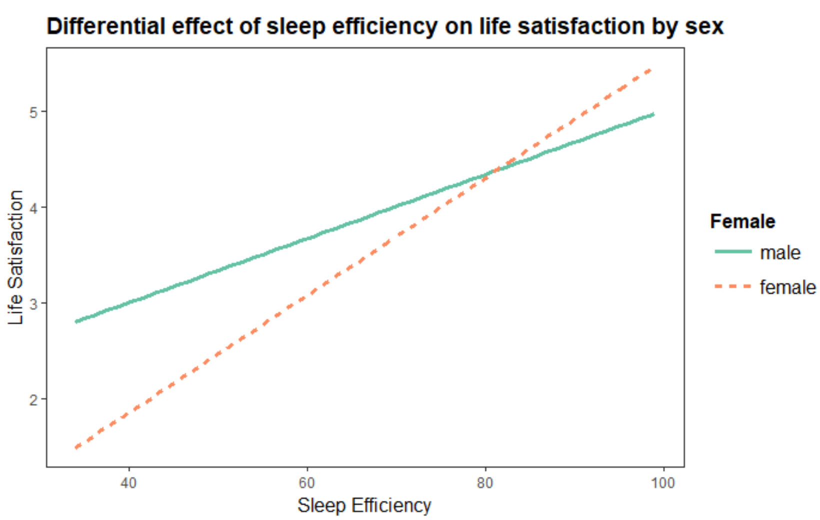
```
predgrid_mix3 <- slp_mix %>%
  group_by(female.f) %>%
  data_grid(sleep_m = seq_range(sleep_m, 10),
            female,
            cond2 = 0,
            cond3 = 1,
            prior = 1,
            age30 = 0,
            anxiety_m = 0) %>%
  ungroup()

predictions_mix3 <- predict(mix3, predgrid_mix3, interval = "confidence") %>%
  as_data_frame()

adjmeans_mix3 <- cbind(predgrid_mix3, predictions_mix3) %>%
  mutate(sleep = sleep_m + mean(slp_mix$sleep))

ggplot(adjmeans_mix3, aes(x = sleep, y = fit, group = female.f, color = female.f)) +
  geom_ribbon(aes(ymin = lwr, ymax = upr, color = NULL), alpha = .4, fill = "grey60") +
  geom_line(size = 1) +
  guides(color = guide_legend("Sex")) +
  labs(title = "Does the effect of sleep efficiency on life satisfaction differ by sex?",
       subtitle = "Condition is held constant at Condition 3,
       prior is held constant at prior involvement, age is held constant at 30,
       \nanxiety is held constant at the mean",
       x = "Sleep Efficiency", y = "Life Satisfaction")
```



Make the Same Plot with jtools

The jtools plotting function automatically centers the variables for you so you can enter the non-centered variables. Also, by providing the ".f" (factor) version of categorical variables it will dummy code for you as well.

```
mix3_jplot <- lm(data = slp_mix, lifesat ~ cond2 + cond3 + prior + age + anxiety + female.f + sleep + female.f*sleep)

probe_interaction(mix3_jplot, pred = sleep, modx = female.f,
            jnplot = FALSE,
            x.label = "Sleep Efficiency",
            y.label = "Life Satisfaction",
            main.title = "Differential effect of sleep efficiency on life satisfaction by sex",
            legend.main = "Female")
```




```
SIMPLE SLOPES ANALYSIS

Slope of sleep when female.f = 1 (female):
 Est.  S.E.     p
0.061 0.004 0.000

Slope of sleep when female.f = 0 (male):
 Est.  S.E.     p
0.034 0.003 0.000
```

These simple slopes are calculated holding all other variables at the mean.

Example Write-up of Results

The differential effect of sleep efficiency on life satisfaction between males and females was examined among 600 participants who took part in a sleep intervention. A moderation model was estimated, life satisfaction was regressed on a binary indicator of sex, sleep efficiency, the interaction between the two, and a set of relevant control variables. The results of the model are presented in Table 1. The interaction term is statistically significant, indicating that, holding constant all control variables, the effect of sleep efficiency on life satisfaction is larger for females (*b* = .061, 95% CI .053, .070) than for males (*b* = .034, 95% CI .028, .039). The model fitted slopes for females and males are presented in Figure 1. We further probed the significant interaction using the techniques outlined by Preacher, Curran & Bauer (2006). Figure 2 presents a Johnson-Neyman regions of significance graph that depicts the nature of the difference in life satisfaction between females and males as sleep efficiency varies from the observed minimum to the maximum value. Based on the fitted model, when sleep efficiency is below 77, males tend to have a higher life satisfaction. Between a sleep efficiency of 77 and 88, there is no difference in life satisfaction between males and females. Above a sleep efficiency of 88, females tend to have a higher life satisfaction.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437-448.

Table 1

Results of multiple linear regression model to predict life satisfaction.

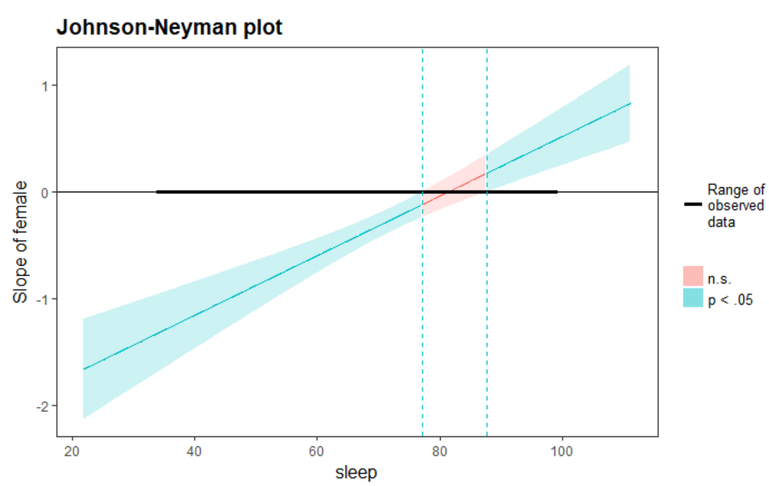
	Life Satisfaction
Intercept	3.699 ***
	(0.076)
Condition 2	0.116
	(0.063)
Condition 3	0.374 ***
	(0.072)
Prior Intervention	-0.056
	(0.051)
Age	0.021 ***
	(0.002)
Anxiety	-0.250 ***
	(0.027)
Female	-0.348 ***
	(0.060)
Sleep Efficiency	0.034 ***
	(0.003)
Female*Sleep Efficiency	0.028 ***
	(0.005)
N	600
R-Squared	0.633
F statistic	127.547
P value	0.000

Age was centered at 30, anxiety was centered at the mean, sleep efficiency was centered at the mean. Tabled values are unstandardized regression coefficients and (standard errors). *** p < 0.001; ** p < 0.01; * p < 0.05.

Figure 1

Figure 2

Differential effect of sex on life satisfaction as a function of sleep efficiency.



Complete R Syntax for Write-up

```
library(huxtable)
library(jtools)

# estimates for table
mix3 <- lm(data=slp_mix, lifesat ~ cond2 + cond3 + prior + age30 + anxiety_m + female + sleep_m + female*sleep_m)
ols_regress(mix3)

# model to get slope of sleep for females (switch reference group)
mix3_switch <- lm(data=slp_mix, lifesat ~ cond2 + cond3 + prior + age30 + anxiety_m + male + sleep_m + male*sleep_m)
ols_regress(mix3_switch)

# use huxtable to get a good looking table 1 — if you knit to html, then the table looks like the one in the handout
huxreg("Life Satisfaction" = mix3, error_format = "({std.error})",
      coefs = c("Intercept" = "(Intercept)","Condition 2" = "cond2", "Condition 3" = "cond3", "Prior Intervention" = "prior",
                "Age" = "age30", "Anxiety" = "anxiety_m", "Female" = "female", "Sleep Efficiency" = "sleep_m",
                "Female*Sleep Efficiency" = "female:sleep_m"),
      note = "Age was centered at 30, anxiety was centered at the mean, sleep efficiency was centered at the mean.
              Tabled values are unstandardized regression coefficient and (standard errors). {stars}.",
      statistics = c("N" = "nobs", "R-Squared" = "r.squared", "F statistic" = "statistic", "P value" = "p.value"))

# produce figure 1
mix3_jplot1 <- lm(data = slp_mix, lifesat ~ cond.f + prior + age + anxiety + female.f + sleep + female.f*sleep)

probe_interaction(mix3_jplot1, pred = sleep, modx = female.f,
          x.label = "Sleep Efficiency",
          y.label = "Life Satisfaction",
          main.title = "Differential effect of sleep efficiency on life satisfaction by sex",
          legend.main = "Female")

# produce figure 2
mix3_jplot2 <- lm(data = slp_mix, lifesat ~ cond.f + prior + age + anxiety + female + sleep + female*sleep)

# this produces only the jnplot
johnson_neyman(mix3_jplot2, pred = female, modx = sleep, alpha = .05)
```

Your Own Johnson-Neyman Graph

If you want to code up your own JN graph rather than use jtools — here is an example.

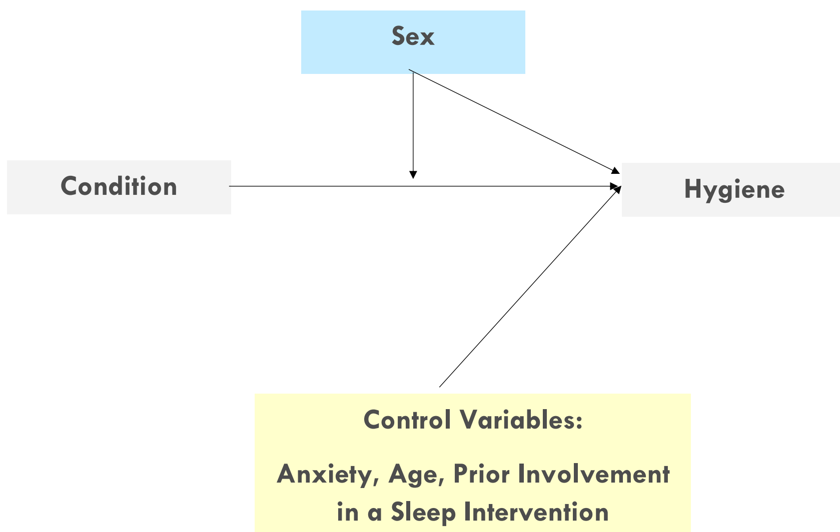
```
# Rename your dataset, and your X and Z variables
my.data <- slp_mix
my.data$x <- my.data$female
my.data$z <- my.data$sleep

# Supply information for plot
ct=1.963986 # define critical t based on error/residual df from MLR
z1=34  # supply lowest score for z (variable that will be on the x axis)
z2=99    # supply highest score for z

# MLR for plot - no need to center anything - note, don't change x and z
my.output <- lm(lifesat ~ cond2 + cond3 + prior + age + anxiety + x + z + x:z, data=my.data)

# No need to change anything here
coefficients(my.output)
b0 <- my.output$coefficients[["(Intercept)"]]
bx <- my.output$coefficients[["x"]]
bz <- my.output$coefficients[["z"]]
bxz <- my.output$coefficients[["x:z"]]
varmat <- as.matrix(vcov(my.output))
vb0 <- varmat["(Intercept)","(Intercept)"]
vbx <- varmat["x","x"]
vbz <- varmat["z","z"]
vbxz <- varmat["x:z","x:z"]
cvb0bz <- varmat["(Intercept)","z"]
cvbxbxz <- varmat["x","x:z"]

z <- seq(z1,z2,length=1000)
fz <- c(z,z)
y1 <- (bx+bxz*z)+(ct*sqrt(vbx+(2*z*cvbxbxz)+((z^2)*vbxz)))
y2 <- (bx+bxz*z)-(ct*sqrt(vbx+(2*z*cvbxbxz)+((z^2)*vbxz)))
fy <- c(y1,y2)
ssz <- (bx+bxz*z)
se_ssz <- sqrt(vbx+(2*z*cvbxbxz)+((z^2)*vbxz))
t_ssz <- ssz/se_ssz
ranget <- data.frame(cbind(z, ssz, se_ssz, t_ssz))
options(max.print=10000)
ranget

ggplot(data = ranget, aes(x = z, y = ssz)) +
  geom_ribbon(aes(ymin = (ssz - se_ssz*ct), ymax = (ssz + se_ssz*ct), color = NULL), alpha = .4, fill = "grey60") +
  geom_line(size = 1) +
  geom_hline(yintercept=0, linetype=2) +
  scale_y_continuous(breaks=seq(-2,2,.5)) + # this will need to be changed to correspond with simple slope range
  labs(title = "Does the simple slope for sex predicting life satisfaction differ by sleep efficiency?",
       subtitle = "Johnson-Neyman Graph",
       x = "Value of the moderator (Sleep Efficiency)", y = "Simple Slope for Effect of Female",
       caption = "Simple slope is significant in regions where the confidence bounds do not cross the dashed line")
```

The Result


Does the simple slope of sex predicting life satisfaction differ by sleep efficiency?
Johnson Neyman Graph
Simple Slope for Effect of Female
0.5
0.0
-0.5
-1.0
-1.5
40
60
80
100
Value of the moderator (Sleep Efficiency)
Simple slope is significant in regions where the confidence bounds do not cross the dashed line

EXAMPLE 2: Interaction of Two Categorical Variables

Research Question: Do the benefits of the group and group + partner conditions, as compared to the self-help condition, differ as a function of participant sex?



Model 1: Regress hygiene on female, cond2, cond3, female*cond2, female*cond3.

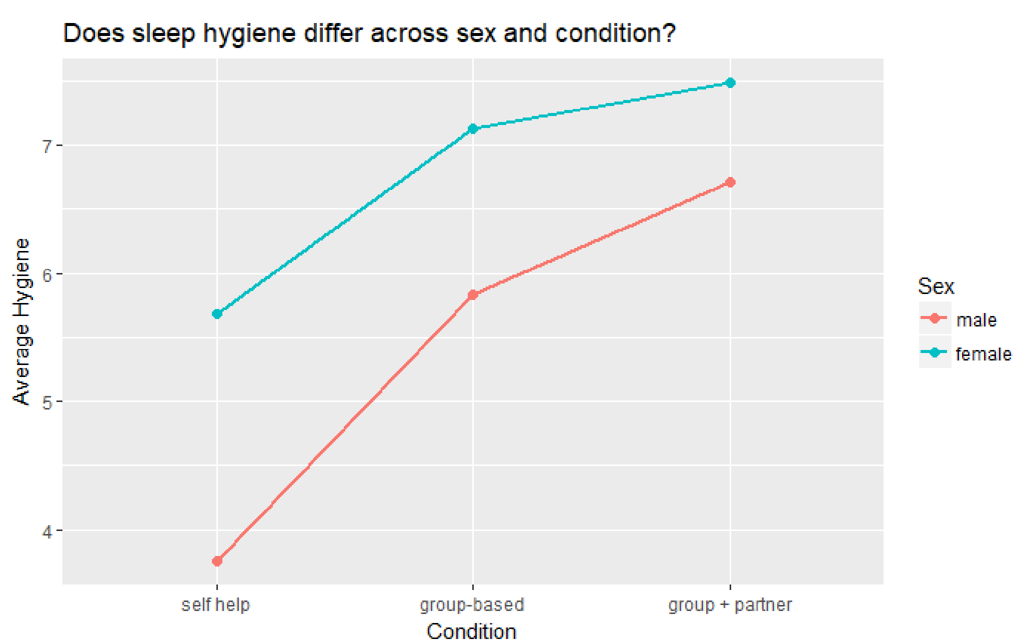
Model 2: Add the control variables to Model 1.

Describe the Key Variables

Let's begin by taking a look at average hygiene as a function of sex and condition.

```
slp_cat_means <- slp %>%
  group_by(female.f, cond.f) %>%
  summarize(mean_hyg = mean(hygiene)) %>%
  ungroup()

slp_cat_means

ggplot(slp_cat_means, aes(x = cond.f, y = mean_hyg, group = female.f, color = female.f)) +
  geom_line(lwd = 1) +
  geom_point(size = 2) +
  guides(color=guide_legend("Sex")) +
  labs(title = "Does sleep hygiene differ across sex and condition?", x = "Condition", y = "Average Hygiene")
```

female.f <fctr>	cond.f <fctr>	mean_hyg <dbl>
male	self help	3.750943
male	group-based	5.835000
male	group + partner	6.712821
female	self help	5.686064
female	group-based	7.127286
female	group + partner	7.488313



Model 1: Fit MLR with Condition and Sex Interaction

To determine if the effect of condition on hygiene differs as a function of sex, we will regress hygiene on the two dummy codes for condition, the dummy code for sex, and the interaction between each condition dummy code and the sex dummy code.

```
cat1 <- lm(data = slp, hygiene ~ female + cond2 + cond3 + female*cond2 + female*cond3)
ols_regress(cat1)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.771       RMSE                1.007
R-Squared               0.594       Coef. Var          16.801
Adj. R-Squared          0.591       MSE                 1.014
Pred R-Squared          0.585       MAE                 0.799
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
----------------------------------------------------------------------
                Sum of
               Squares         DF    Mean Square       F         Sig.
----------------------------------------------------------------------
Regression     881.213          5        176.243    173.768    0.0000
Residual       602.459        594          1.014
Total         1483.671        599
----------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
-----------------------------------------------------------------------------------------
 (Intercept)     3.751         0.098                 38.346    0.000      3.559     3.943
      female     1.935         0.143        0.606    13.562    0.000      1.655     2.215
       cond2     2.084         0.132        0.625    15.813    0.000      1.825     2.343
       cond3     2.962         0.135        0.888    21.933    0.000      2.697     3.227
female:cond2    -0.643         0.207       -0.131    -3.113    0.002     -1.048    -0.237
female:cond3    -1.160         0.203       -0.255    -5.710    0.000     -1.558    -0.761
-----------------------------------------------------------------------------------------
```

Interpretation of parameter estimates:

Intercept: The predicted sleep hygiene score when all x variables are zero, so males in Condition 1.

female: This variable is involved in an interaction, so it's a simple slope. Specifically, it is the effect of female when both cond2 = 0 and cond3 = 0, so people in Condition 1. Therefore, it is the predicted difference in sleep hygiene between females and males in Condition 1. The slope is positive, meaning that females in Condition 1 tend to have better sleep hygiene than males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

cond2: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 2 (compared to Condition 1) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 2 compared to males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

cond3: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 3 (compared to Condition 1) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 3 compared to males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

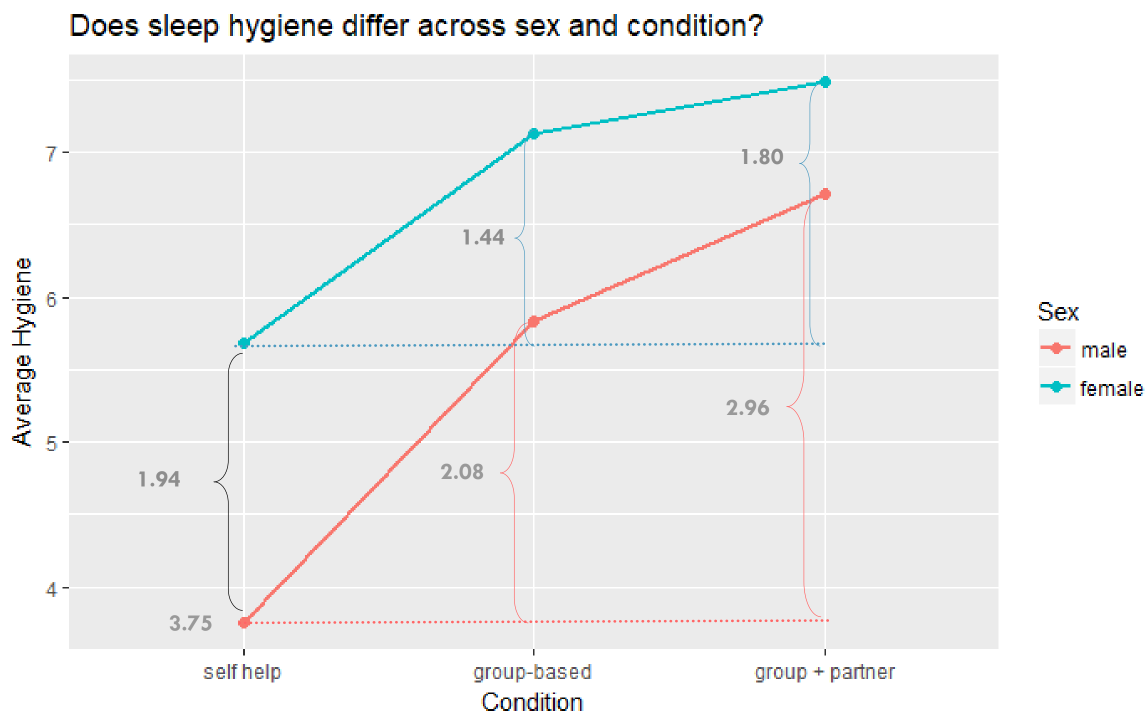
female:cond2: The predicted differential effect of being in Condition 2 compared to Condition 1 for females compared to males. This is a statistically significant difference (p-value is less than alpha). The coefficient for cond2 presents the effect (i.e., benefit) of being in Condition 2 (compared to 1) for males. To get the effect for females, we take the effect for males (2.084) and add the female:cond2 interaction term (-.643). Therefore, the effect (i.e., benefit) of being in Condition 2 (compared to Condition 1) for females is 1.441.

female:cond3: The predicted differential effect of being in Condition 3 compared to Condition 1 for females compared to males. This is a statistically significant difference (p-value is less than alpha). The coefficient for cond3 presents the effect (i.e., benefit) of being in Condition 3 (compared to 1) for males. To get the effect for females, we take the effect for males (2.962) and add the female:cond3 interaction term (-1.160). Therefore, the effect (i.e., benefit) of being in Condition 3 (compared to Condition 1) for females is 1.802.

Map Results onto Graph

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	3.751	0.098		38.346	0.000	3.559	3.943
female	1.935	0.143	0.606	13.562	0.000	1.655	2.215
cond2	2.084	0.132	0.625	15.813	0.000	1.825	2.343
cond3	2.962	0.135	0.888	21.933	0.000	2.697	3.227
female:cond2	-0.643	0.207	-0.131	-3.113	0.002	-1.048	-0.237
female:cond3	-1.160	0.203	-0.255	-5.710	0.000	-1.558	-0.761



2.08 - 1.44 = -.64

2.96 - 1.80 = -1.16

Determine if the Treatment Effects are Significant for Females

In our initial model, the regression estimates for cond2 and cond3 represented the treatment effect for males. If we want to determine if the treatment effect is statistically significant for females, we can refit the model specifying females as the reference group.

```
slp_cat <- mutate(slp, male = ifelse(sex == 1, 1, 0))

cat1_switch <- lm(data = slp_cat, hygiene ~ male + cond2 + cond3 + male*cond2 + male*cond3)
ols_regress(cat1_switch)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.771       RMSE                1.007
R-Squared               0.594       Coef. Var          16.801
Adj. R-Squared          0.591       MSE                 1.014
Pred R-Squared          0.585       MAE                 0.799
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
----------------------------------------------------------------------
                Sum of
               Squares         DF    Mean Square       F          Sig.
----------------------------------------------------------------------
Regression     881.213          5        176.243    173.768     0.0000
Residual       602.459        594          1.014
Total         1483.671        599
----------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta         t        Sig      lower     upper
-----------------------------------------------------------------------------------------
(Intercept)       5.686         0.104                  54.740     0.000      5.482     5.890
       male      -1.935         0.143       -0.606    -13.562     0.000     -2.215    -1.655
      cond2       1.441         0.159        0.432      9.065     0.000      1.129     1.753
      cond3       1.802         0.152        0.540     11.881     0.000      1.504     2.100
 male:cond2       0.643         0.207        0.168      3.113     0.002      0.237     1.048
 male:cond3       1.160         0.203        0.292      5.710     0.000      0.761     1.558
-----------------------------------------------------------------------------------------
```

The results indicate that both Conditions 2 and 3 produced significantly better sleep hygiene, as compared to Condition 1, for females as well.

Obtain the Predicted Values from Model 1

```
predgrid_cat <- slp %>%
  group_by(cond.f, female.f) %>%
  data_grid(cond2, cond3, female) %>%
  ungroup() %>%
  add_predictions(cat1)

predgrid_cat
```

cond.f <fctr>	female.f <fctr>	cond2 <dbl>	cond3 <dbl>	female <dbl>	pred <dbl>
self help	male	0	0	0	3.750943
self help	female	0	0	1	5.686064
group-based	male	1	0	0	5.835000
group-based	female	1	0	1	7.127286
group + partner	male	0	1	0	6.712821
group + partner	female	0	1	1	7.488313

Let's compare these to the simple means we obtained earlier—**they are the same.** We've just simply reproduced the means with this model.

female.f <fctr>	cond.f <fctr>	mean_hyg <dbl>
male	self help	3.750943
male	group-based	5.835000
male	group + partner	6.712821
female	self help	5.686064
female	group-based	7.127286
female	group + partner	7.488313

Add the Control Variables to Model 1

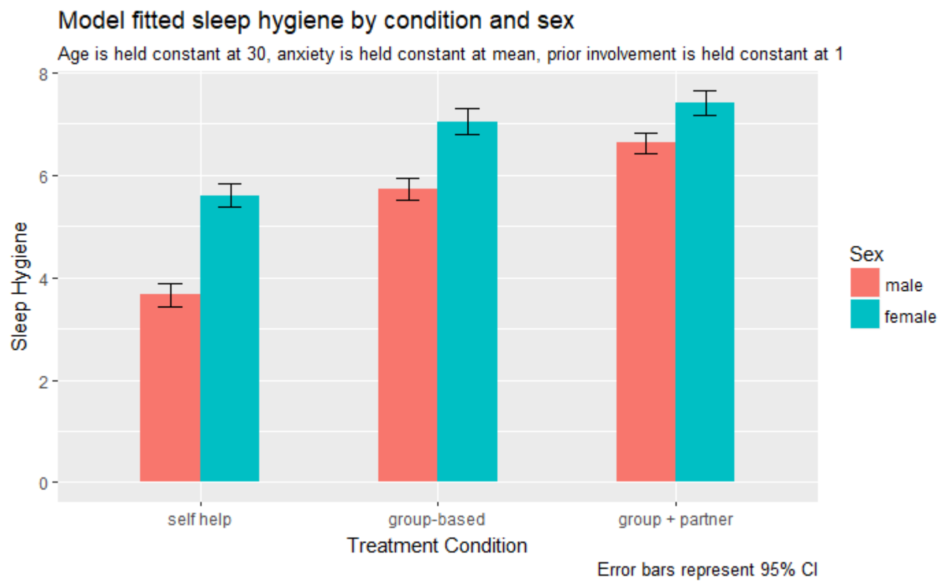
```
# center predictors and fit model
slp_cat <- mutate(slp_cat,
                  age30 = age-30,
                  anxiety_m = anxiety - mean(anxiety))

cat2 <- lm(data = slp_cat, hygiene ~ female + cond2 + cond3 + female*cond2 + female*cond3 +
age30 + prior + anxiety_m)
ols_regress(cat2)
```

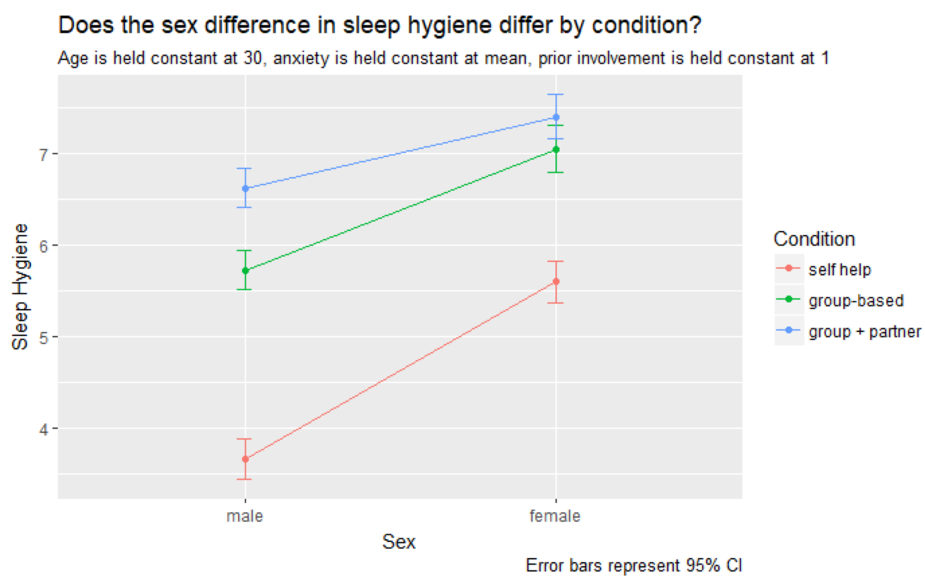
```
                        Model Summary
--------------------------------------------------------------
R                       0.772       RMSE                1.007
R-Squared               0.596       Coef. Var          16.794
Adj. R-Squared          0.591       MSE                 1.013
Pred R-Squared          0.583       MAE                 0.798
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
----------------------------------------------------------------------
                Sum of
               Squares         DF    Mean Square       F         Sig.
----------------------------------------------------------------------
Regression     884.784          8        110.598    109.142     0.0000
Residual       598.887        591          1.013
Total         1483.671        599
----------------------------------------------------------------------

                                   Parameter Estimates
-----------------------------------------------------------------------------------------
       model        Beta    Std. Error    Std. Beta       t        Sig      lower    upper
-----------------------------------------------------------------------------------------
 (Intercept)       3.665         0.127                  28.871    0.000     3.416    3.914
      female       1.939         0.143        0.607     13.570    0.000     1.658    2.220
       cond2       2.069         0.132        0.620     15.652    0.000     1.809    2.328
       cond3       2.962         0.135        0.888     21.910    0.000     2.697    3.228
       age30       0.006         0.003        0.049      1.876    0.061     0.000    0.012
       prior      -0.004         0.092       -0.001     -0.040    0.968    -0.185    0.177
   anxiety_m      -0.003         0.046       -0.002     -0.075    0.940    -0.094    0.087
female:cond2      -0.624         0.207       -0.127     -3.016    0.003    -1.030   -0.218
female:cond3      -1.161         0.203       -0.255     -5.712    0.000    -1.560   -0.762
-----------------------------------------------------------------------------------------
```

Create Plots of the Fitted Model

```
# bar graph
predgrid_cat <- slp %>%
  group_by(cond.f, female.f) %>%
  data_grid(cond2, cond3, female,
            prior = 1,
            age30 = 0,
            anxiety_m = 0) %>%
  ungroup()

predictions_cat <- predict(cat2, predgrid_cat, interval = "confidence") %>%
  as_data_frame()

adjmeans_cat <- cbind(predgrid_cat, predictions_cat)

ggplot(adjmeans_cat, aes(x = cond.f, y = fit, ymin = lwr, ymax = upr, group = female.f, fill = female.f)) +
  geom_bar(position="dodge", stat = "identity", width = .5) +
  geom_errorbar( position = position_dodge(.5), colour="black", width = .2) +
  guides(fill = guide_legend("Sex")) +
  labs(title = "Model fitted sleep hygiene by condition and sex",
       subtitle = "Age is held constant at 30, anxiety is held constant at mean, prior involvement is held constant at 1",
       x = "Treatment Condition", y = "Sleep Hygiene",
       caption = "Error bars represent 95% CI")


# line graph
ggplot(adjmeans_cat, aes(x = female.f, y = fit, ymin = lwr, ymax = upr, group = cond.f, color = cond.f)) +
  geom_line() +
  geom_errorbar(width = .05) +
  geom_point() +
  guides(color = guide_legend("Condition")) +
  labs(title = "Does the sex difference in sleep hygiene differ by condition?",
       subtitle = "Age is held constant at 30, anxiety is held constant at mean, prior involvement is held constant at 1",
       x = "Sex", y = "Sleep Hygiene",
       caption = "Error bars represent 95% CI")
```

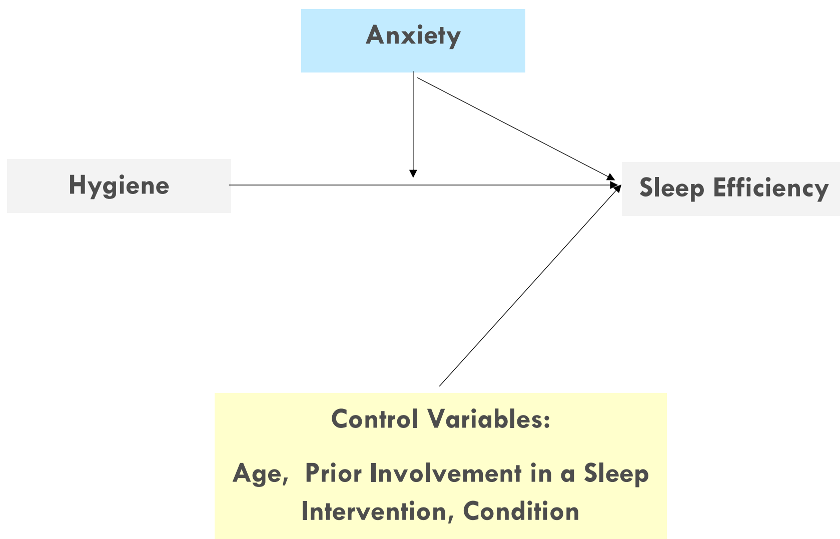
The Plots


Model fitted sleep hygiene by condition and sex
Age is held constant at 30, anxiety is held constant at mean, prior involvement is held constant at 1
Sleep Hygiene
0
2
4
6
8
self help
group-based
group + partner
Treatment Condition
Sex
male
female
Error bars represent 95% CI


Does the sex difference in sleep hygiene differ by condition?
Age is held constant at 30, anxiety is held constant at mean, prior involvement is held constant at 1
Sleep Hygiene
4
5
6
7
male
female
Sex
Condition
self help
group-based
group + partner
Error bars represent 95% CI

Example 3: Interaction of Two Continuous Variables

In this example we will consider only the males in the sample. We seek to determine if the benefit of practicing better sleep hygiene on improved sleep efficiency is attenuated if participants suffer from anxiety. Our hypothesis is that as anxiety increases the beneficial effect of good sleep hygiene on better sleep outcomes is attenuated.

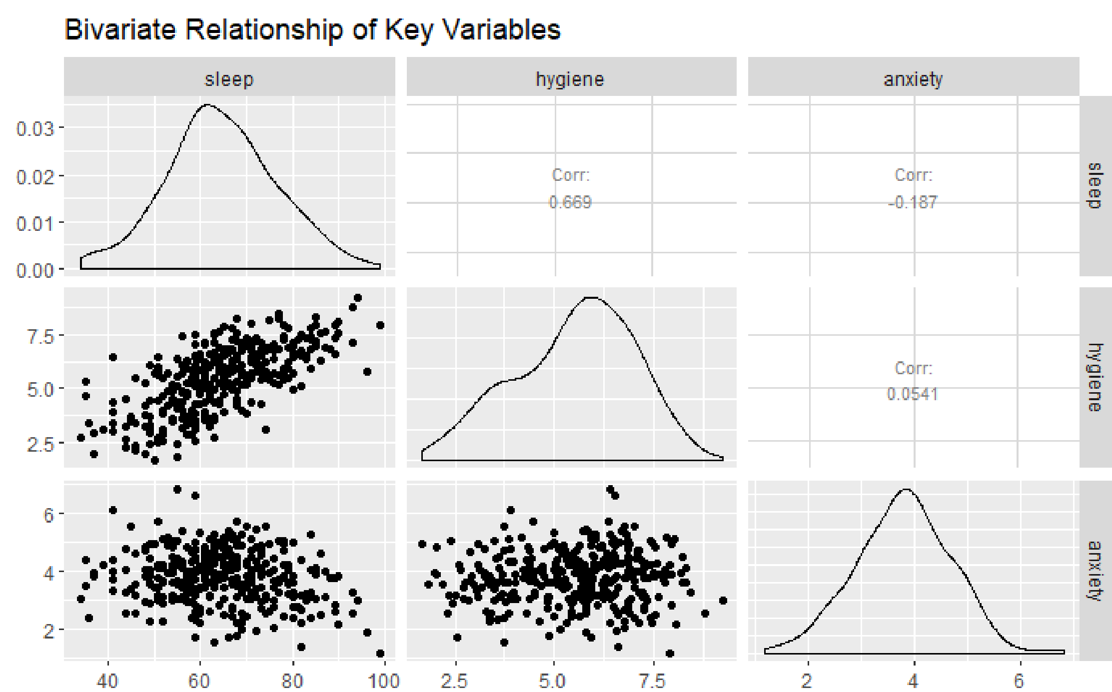


Get Descriptive Statistics for the Subsample and Plot the Data

```
# subset to obtain males only
slp_cont <- filter(slp, female == 0)

describe(slp_cont)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
cond	1	353	2.03	0.80	2.00	2.04	1.48	1.00	3.00	2.00	-0.06	-1.42	0.04
prior	2	353	0.71	0.45	1.00	0.77	0.00	0.00	1.00	1.00	-0.94	-1.11	0.02
age	3	353	45.58	11.96	46.10	45.78	15.72	23.90	64.70	40.80	-0.12	-1.23	0.64
anxiety	4	353	3.80	0.89	3.81	3.81	0.87	1.18	6.84	5.66	-0.04	0.16	0.05
hygiene	5	353	5.50	1.50	5.69	5.55	1.57	1.68	9.21	7.53	-0.32	-0.52	0.08
support	6	353	3.00	0.70	2.86	2.97	0.71	1.46	4.76	3.30	0.41	-0.75	0.04
sleep	7	353	64.53	11.95	64.00	64.44	10.38	34.00	99.00	65.00	0.08	-0.05	0.64
lifesat	8	353	4.02	0.91	4.02	4.00	0.89	1.68	6.53	4.85	0.20	-0.11	0.05

```
scatterplot <- ggpairs(slp_cont, columns = c("sleep", "hygiene", "anxiety"),
        upper = list(continuous = wrap("cor", size=3)),
        title = "Bivariate Relationship of Key Variables")
print(scatterplot, progress=FALSE )
```



Fit the Regression Model

```
# center the predictors
slp_cont <- mutate(slp_cont,
              age30 = age - 30,
              hygiene_m = hygiene - mean(hygiene),
              anxiety_m = anxiety - mean(anxiety))

cont1 <- lm(data=slp_cont, sleep ~ age30 + prior + cond2 + cond3 + hygiene_m + anxiety_m + hygiene_m*anxiety_m)
ols_regress(cont1)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.736       RMSE                8.176
R-Squared               0.541       Coef. Var          12.669
Adj. R-Squared          0.532       MSE                66.842
Pred R-Squared          0.521       MAE                 6.416
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
----------------------------------------------------------------------
                Sum of
               Squares         DF    Mean Square       F         Sig.
----------------------------------------------------------------------
Regression    27195.497         7       3885.071    58.123     0.0000
Residual      23060.378       345         66.842
Total         50255.875       352
----------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------------
              model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
-----------------------------------------------------------------------------------------------
        (Intercept)    60.194         1.510                39.856    0.000     57.223   63.164
              age30     0.122         0.037        0.122    3.290    0.001      0.049    0.194
              prior    -0.162         0.966       -0.006   -0.167    0.867     -2.062    1.739
              cond2     1.656         1.480        0.067    1.119    0.264     -1.255    4.566
              cond3     6.082         1.823        0.240    3.336    0.001      2.496    9.668
          hygiene_m     4.076         0.493        0.512    8.261    0.000      3.106    5.046
          anxiety_m    -2.967         0.493       -0.222   -6.020    0.000     -3.936   -1.998
hygiene_m:anxiety_m    -0.918         0.333       -0.102   -2.755    0.006     -1.574   -0.263
-----------------------------------------------------------------------------------------------
```

Interpret the Parameter Estimates

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	60.194	1.510		39.856	0.000	57.223	63.164
age30	0.122	0.037	0.122	3.290	0.001	0.049	0.194
prior	-0.162	0.966	-0.006	-0.167	0.867	-2.062	1.739
cond2	1.656	1.480	0.067	1.119	0.264	-1.255	4.566
cond3	6.082	1.823	0.240	3.336	0.001	2.496	9.668
hygiene_m	4.076	0.493	0.512	8.261	0.000	3.106	5.046
anxiety_m	-2.967	0.493	-0.222	-6.020	0.000	-3.936	-1.998
hygiene_m:anxiety_m	-0.918	0.333	-0.102	-2.755	0.006	-1.574	-0.263

Interpretation of parameter estimates:

Intercept: The predicted sleep efficiency score when all x variables are zero, so 30 year old males in Condition 1 with no prior intervention history, an average level of anxiety, and average sleep hygiene.

age30: The predicted change in sleep efficiency for a 1 year increase in age, holding all other variables constant. This is a statistically significant effect (p-value is less than alpha).

prior: The predicted difference in sleep efficiency for participants with and without prior intervention experience, holding constant all other variables. This is not a statistically significant difference.

cond2: The predicted difference in sleep efficiency for participants in Condition 2 compared to Condition 1, holding constant all other variables. This is not a statistically significant difference.

cond3: The predicted difference in sleep efficiency for participants in Condition 3 compared to Condition 1, holding constant all other variables. This is a statistically significant difference (p-value is less than alpha).

hygiene_m: Holding constant age, prior experience, and condition, this is the predicted change in sleep efficiency for a one unit increase in hygiene, **among participants with an average level of anxiety (i.e., anxiety_m = 0)**. Among participants with an average level of anxiety, better hygiene is associated with better sleep efficiency. This is a statistically significant effect (p-value is less than alpha).

anxiety_m: Holding constant age, prior experience, and condition, this is the predicted change in sleep efficiency for a one unit increase in anxiety, **among participants with an average level of hygiene (i.e., hygiene_m = 0)**. Among participants with an average level of hygiene, higher anxiety is associated with worse sleep efficiency. This is a statistically significant effect (p-value is less than alpha).

hygiene_m*anxiety_m: Holding constant age, prior experience, and condition, this is the predicted difference in the effect of hygiene on sleep efficiency for a one unit increase in anxiety. So, while the effect of hygiene is 4.08 when anxiety is at the mean, for each one unit increase in anxiety, we expect the beneficial effect of sleep hygiene to decrease by .92 units.

From the descriptive statistics table, we see that the average anxiety for males is 3.80. So, holding constant age, prior intervention experience and condition, we know that:

When anxiety = 3.80, the simple slope for the effect of hygiene on sleep efficiency is 4.08.

When anxiety = 4.80 (one unit higher than the mean), the simple slope for the effect of hygiene on sleep efficiency is 4.08 + (1 ▪ -.92) = 3.16

When anxiety = 2.80 (one unit lower than the mean), the simple slope for the effect of hygiene on sleep efficiency is 4.08 + (-1 ▪ -.92) = 5.00

Writing Reduced Forms of the Equation for Prototypical Groups

The equation from our fitted model is:

$\text{y-hat}_i = 60.19 + (.12 \cdot \text{age30}_i) + (-.16 \cdot \text{prior}_i) + (1.66 \cdot \text{cond2}_i) + (6.08 \cdot \text{cond3}_i) + (4.08 \cdot \text{hygiene\_m}_i) + (-2.97 \cdot \text{anxiety\_m}_i) + (-.92 \cdot \text{hygiene\_m}_i \cdot \text{anxiety\_m}_i)$

Let's write the equation that relates sleep hygiene to sleep efficiency for a 30 year old participant with no prior intervention history in Condition 1 who has an average level of anxiety.

$\text{y-hat}_i = 60.19 + (.12 \cdot 0) + (-.16 \cdot 0) + (1.66 \cdot 0) + (6.08 \cdot 0) + (4.08 \cdot \text{hygiene\_m}_i) + (-2.97 \cdot 0) + (-.92 \cdot \text{hygiene\_m}_i \cdot 0)$

$\text{y-hat}_i = 60.19 + (4.08 \cdot \text{hygiene\_m}_i)$

Let's write the equation that relates sleep hygiene to sleep efficiency for a 30 year old participant with no prior intervention history in Condition 1 who has an anxiety score that is one unit above the mean.

$\text{y-hat}_i = 60.19 + (.12 \cdot 0) + (-.16 \cdot 0) + (1.66 \cdot 0) + (6.08 \cdot 0) + (4.08 \cdot \text{hygiene\_m}_i) + (-2.97 \cdot 1) + (-.92 \cdot \text{hygiene\_m}_i \cdot 1)$

$\text{y-hat}_i = 57.22 + (3.16 \cdot \text{hygiene\_m}_i)$

Let's write the equation that relates sleep hygiene to sleep efficiency for a 30 year old participant with no prior intervention history in Condition 1 who has an anxiety score that is one unit below the mean.

$\text{y-hat}_i = 60.19 + (.12 \cdot 0) + (-.16 \cdot 0) + (1.66 \cdot 0) + (6.08 \cdot 0) + (4.08 \cdot \text{hygiene\_m}_i) + (-2.97 \cdot -1) + (-.92 \cdot \text{hygiene\_m}_i \cdot -1)$

$\text{y-hat}_i = 63.16 + (5.00 \cdot \text{hygiene\_m}_i)$

```
predgrid_cont <- slp_cont %>%
  data_grid(hygiene_m = c(-1, 0, 1),
            anxiety_m = c(-1, 0, 1),
            age30 = 0,
            prior = 0,
            cond2 = 0,
            cond3 = 0) %>%
  add_predictions(cont1)
```

hygiene_m	anxiety_m	age30	prior	cond2	cond3	pred
-1	-1	0	0	0	0	58.16649
-1	0	0	0	0	0	56.11761
-1	1	0	0	0	0	54.06873
0	-1	0	0	0	0	63.16066
0	0	0	0	0	0	60.19361
0	1	0	0	0	0	57.22655
1	-1	0	0	0	0	68.15483
1	0	0	0	0	0	64.26960
1	1	0	0	0	0	60.38437

Use the table above to solve for the following.

1. What is the difference in y-hat (pred) for a person with 0 compared to 1 for hygiene_m if anxiety_m is held constant at 0?
2. What is the difference in y-hat for a person with 0 compared to 1 for hygiene_m if anxiety_m is held constant at 1?
3. What is the difference in the values that you obtained for step 1 and step 2?
4. Reconcile each of these values with the regression model output from our fitted model.

More Reduced Equations

$\text{y-hat}_i = 60.19 + (.12 \cdot \text{age30}_i) + (-.16 \cdot \text{prior}_i) + (1.66 \cdot \text{cond2}_i) + (6.08 \cdot \text{cond3}_i) + (4.08 \cdot \text{hygiene\_m}_i) + (-2.97 \cdot \text{anxiety\_m}_i) + (-.92 \cdot \text{hygiene\_m}_i \cdot \text{anxiety\_m}_i)$

Using the same technique, write a reduced form of the equation to represent the relationship between hygiene_m and sleep efficiency when anxiety_m is one standard deviation above the mean in the subsample of males, and when anxiety_m is one standard deviation below the mean in the subsample of males.

Probing the Interaction

```
slp_cont <- mutate(slp,
                   anxiety_1above = anxiety - (mean(anxiety) + sd(anxiety)),
                   anxiety_1below = anxiety - (mean(anxiety) - sd(anxiety)))

cont1_above <- lm(data=slp_cont, sleep ~ age30 + prior + cond2 + cond3 + hygiene_m + anxiety_1above + hygiene_m*anxiety_1above)
ols_regress(cont1_above)

cont1_below <- lm(data=slp_cont, sleep ~ age30 + prior + cond2 + cond3 + hygiene_m + anxiety_1below + hygiene_m*anxiety_1below)
ols_regress(cont1_below)
```

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	57.542	1.599		35.975	0.000	54.396	60.687
age30	0.122	0.037	0.122	3.290	0.001	0.049	0.194
prior	-0.162	0.966	-0.006	-0.167	0.867	-2.062	1.739
cond2	1.656	1.480	0.067	1.119	0.264	-1.255	4.566
cond3	6.082	1.823	0.240	3.336	0.001	2.496	9.668
hygiene_m	3.255	0.603	0.409	5.400	0.000	2.070	4.441
anxiety_1above	-2.967	0.493	-0.222	-6.020	0.000	-3.936	-1.998
hygiene_m:anxiety_1above	-0.918	0.333	-0.153	-2.755	0.006	-1.574	-0.263

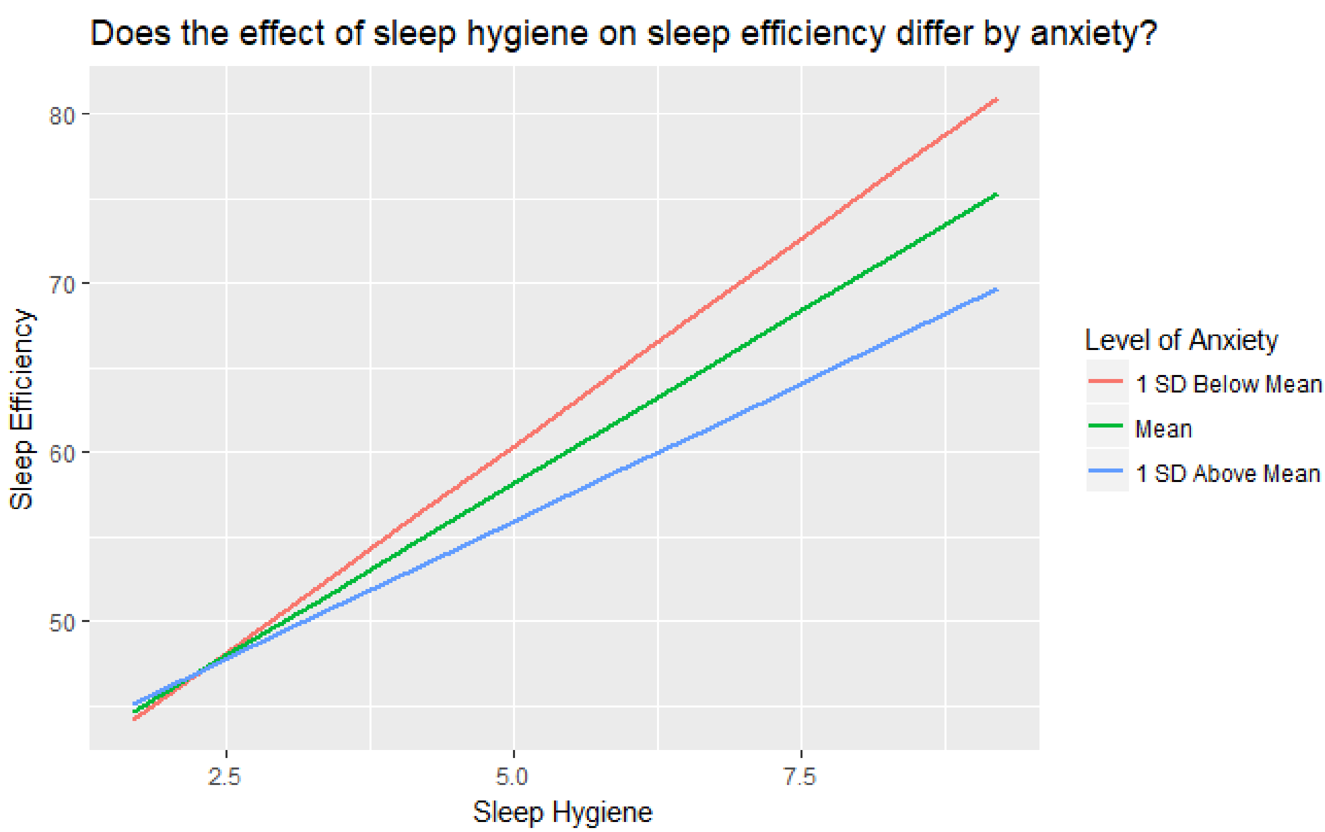
Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	62.846	1.547		40.636	0.000	59.804	65.888
age30	0.122	0.037	0.122	3.290	0.001	0.049	0.194
prior	-0.162	0.966	-0.006	-0.167	0.867	-2.062	1.739
cond2	1.656	1.480	0.067	1.119	0.264	-1.255	4.566
cond3	6.082	1.823	0.240	3.336	0.001	2.496	9.668
hygiene_m	4.897	0.549	0.615	8.925	0.000	3.818	5.976
anxiety_1below	-2.967	0.493	-0.222	-6.020	0.000	-3.936	-1.998
hygiene_m:anxiety_1below	-0.918	0.333	-0.137	-2.755	0.006	-1.574	-0.263

While we were able to calculate the simple slope for hygiene at one standard deviation above and below the mean of anxiety using our equation, by centering hygiene at one standard deviation above and one standard deviation below the mean and then refitting the model, we can determine if the effect of hygiene is significant at each of these levels of anxiety. Here, we see that hygiene remains a significant, positive predictor at each level.

Plot the Simple Slopes

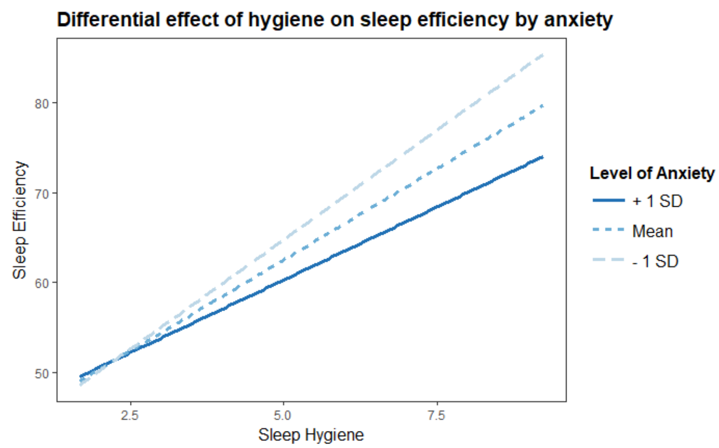
```
predgrid_cont <- slp_cont %>%
  data_grid(hygiene_m = seq_range(hygiene_m, 10),
            age30 = 0,
            prior = 0,
            cond2 = 0,
            cond3 = 0,
            anxiety_m = c(-.89, 0, .89)) %>%
  add_predictions(cont1) %>%
  mutate(hygiene = hygiene_m + mean(slp_cont$hygiene),
         anxiety.f = factor(anxiety_m, levels = c(-.89, 0, .89), labels = c("1 SD Below Mean", "Mean", "1 SD Above Mean")))


ggplot(predgrid_cont, aes(x = hygiene, y = pred, group = anxiety.f, color = anxiety.f)) +
  geom_line(size = 1) +
  guides(color = guide_legend("Level of Anxiety")) +
  labs(title = "Does the effect of sleep hygiene on sleep efficiency differ by anxiety?",
     x = "Sleep Hygiene", y = "Sleep Efficiency")
```

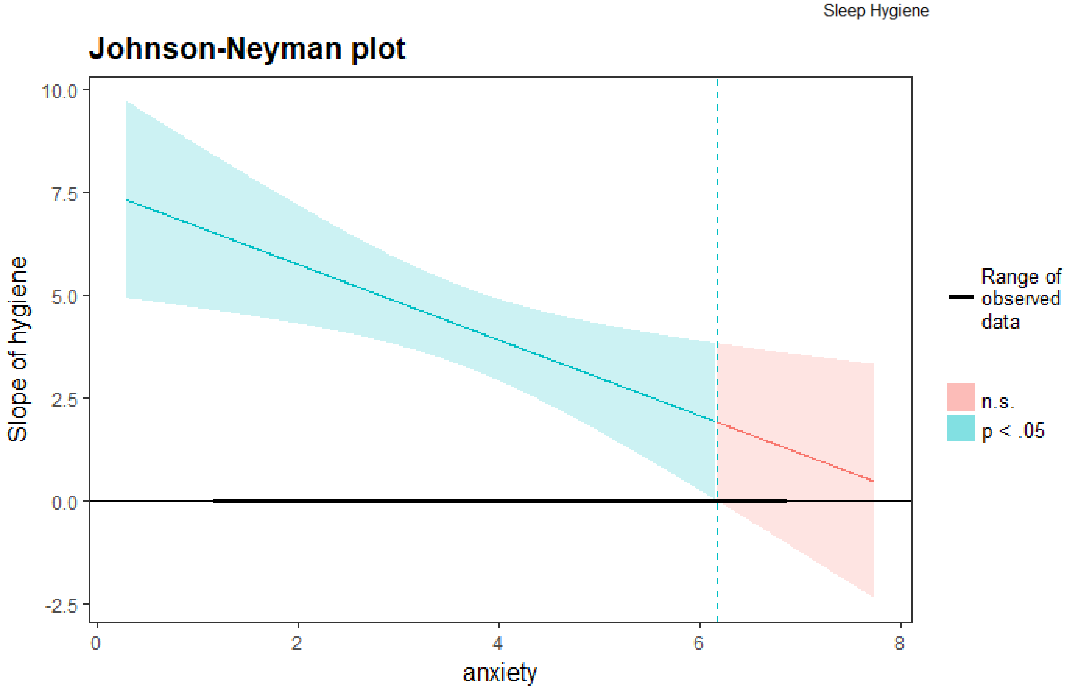


Use the jtools Package to Get Plots

```
cont1.jn <- lm(data=slp_cont, sleep ~ age30 + prior + cond2 + cond3 + hygiene + anxiety + hygiene*anxiety)

probe_interaction(cont1.jn, pred = hygiene, modx = anxiety,
            jnplot = TRUE,
            x.label = "Sleep Hygiene",
            y.label = "Sleep Efficiency",
            main.title = "Differential effect of hygiene on sleep efficiency by anxiety",
            legend.main = "Level of Anxiety")
```

```
JOHNSON-NEYMAN INTERVAL

The slope of hygiene is p < .05 when anxiety is OUTSIDE this interval
[6.167, 19.7434]
Note: The range of observed values of anxiety is [1.18, 6.84]

SIMPLE SLOPES ANALYSIS

Slope of hygiene when anxiety = 4.698 (+ 1 SD):
 Est.  S.E.    p
3.255 0.603 0.000

Slope of hygiene when anxiety = 3.804 (Mean):
 Est.  S.E.    p
4.076 0.493 0.000

Slope of hygiene when anxiety = 2.91 (- 1 SD):
 Est.  S.E.    p
4.897 0.549 0.000
```





This graph suggests that at the very highest levels of anxiety, the benefits of sleep hygiene on sleep efficiency are no longer realized (i.e., the effect is no longer significantly different from zero).

Method to Obtain the Correct Standardized Coefficients

In order to get the correct standardized estimates for variables in an interaction, you should z-score the variables (y and any x variables for which you want the standardized coefficient) and then refit the model.

```
slp_z <- mutate(slp,
                anxiety_z = zscore(anxiety),
                hygiene_z = zscore(hygiene),
                sleep_z = zscore(sleep))

z <- lm(data = slp_z, sleep_z ~ hygiene_z + anxiety_z + hygiene_z*anxiety_z)
ols_regress(z)
```

```
                           Model Summary
--------------------------------------------------------------------
R                        0.761       RMSE                    0.650
R-Squared                0.579       Coef. Var        5.086487e+17
Adj. R-Squared           0.577       MSE                     0.423
Pred R-Squared           0.574       MAE                     0.513
--------------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
---------------------------------------------------------------------
              Sum of
             Squares         DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression   347.064          3        115.688    273.681     0.0000
Residual     251.936        596          0.423
Total        599.000        599
---------------------------------------------------------------------

                                   Parameter Estimates
------------------------------------------------------------------------------------------------
              model      Beta    Std. Error    Std. Beta       t        Sig      lower    upper
------------------------------------------------------------------------------------------------
        (Intercept)     0.005         0.027                  0.183    0.855    -0.047    0.057
          hygiene_z     0.748         0.027        0.748    28.019    0.000     0.696    0.801
          anxiety_z    -0.191         0.027       -0.191    -7.125    0.000    -0.244   -0.139
hygiene_z:anxiety_z    -0.055         0.026       -0.056    -2.081    0.038    -0.107   -0.003
------------------------------------------------------------------------------------------------
```

For a nice discussion on why you need to z-score variables prior to entering them into a linear regression model to obtain standardized coefficients see: https://www.ssc.wisc.edu/~hemken/Stataworkshops/stdBeta/Getting%20Standardized%20Coefficients%20Right.pdf

And, remember, don't standardize categorical variables.