1. Review: Interpreting regression parameters

Answers to questions at end of section notes.

A. Levels, logs, rescaling, and LPM

Model	DepVar	IndepVar	How does Δy relate to Δx?	Interpretation
Linear	y	x	$\Delta y = \beta_1 \Delta x$	$\Delta y = \beta_1 \Delta x$
Logarithmic	y	$\log(x)$	$\Delta y = \beta_1 \frac{\Delta x}{x}$	$\Delta y = (\beta_1/100)\%\Delta x$
Exponential	$\log(y)$	x	$\frac{\Delta y}{y} = \beta_1 \Delta x$	$\%\Delta y = (100\beta_1)\Delta x$
Log-Log	$\log(y)$	$\log(x)$	$\frac{\Delta y}{y} = \beta_1 \frac{\Delta x}{x}$	$\%\Delta y = \beta_1 \%\Delta x$
Standardized	$\tilde{y} = \left(\frac{y-\bar{y}}{\sigma_y}\right)$	$\tilde{x} = \left(\frac{x-\bar{x}}{\sigma_x}\right)$	$\Delta \tilde{y} = \beta_1 \Delta \tilde{x}$	$\Delta s.d.(y) = \beta_1 \Delta s.d.(x)$
LPM	$y(y \in \{0,1\})$	x	$\Delta y = \beta_1 \Delta x$	$\Delta P(Y=1) = \beta \Delta x$

Questions: Consider the following regression output analyzing the relationship between sleep and work, both measured in hours per day. Interpret the coefficient on the parameter associated with work hours and test the null hypothesis that the coefficient is equal to 0, noting that there are 706 observations in the data.

1. $sleep = 8.538 - 0.151 work$; $SE(\beta_{work}) = 0.0168$

2. $sleep = 8.753 - 0.578 log(work)$; $SE(\beta_{log(work)}) = 0.0727$

3. $log(sleep) = 2.144 - 0.024 work$; $SE(\beta_{work}) = 0.0024$

4. $log(sleep) = 2.169 - 0.076 log(work)$; $SE(\beta_{log(work)}) = 0.0104$

5. We standardized the sleep and work variables (recall that standardizing means subtracting the mean and dividing by the standard deviation) and obtained
$scale(sleep) = -2.6e^{-16} - 0.321 scale(work)$; $SE(\beta_{scale(work)}) = 0.0357$

6. We created a dummy variable indicating whether an individual sleeps at least 8 hours per day ($sleep8up$). We estimate the following linear probability model
$sleep8up = 0.687 - 0.055 work$; $SE(\beta_{work}) = 0.00794$

B. Qualitative variables

In order to incorporate qualitative data, we will usually transform it so it can be represented by dummy (or indicator, or 0-1) variables. If our qualitative data can be represented by g different categories, we create $g - 1$ dummy variables. This is important: if we develop dummy variables for all values of the categorical variable and include them in our regression model, then we will have perfect collinearity—one of your dummies can be predicted perfectly with a combination of all the other category dummies.

Examples

- Data: whether or not someone finished high school $\Rightarrow$ Create a dummy variable *highschool* equal to 1 if finished high school and 0 otherwise. The left out category is implicitly *nohighschool*
- Data: highest educational degree obtained $\Rightarrow$ Create dummy variables *doctoral*, *masters*, 4*yearuni*, 2*yeartechvoc*, *highschool*, and *ged* equal to 1 if the individual has received a doctoral, masters, 4 year university, 2 year university/other technical or vocational degree, high school degree, or general educational degree, respectively, and 0 otherwise. The left out category is implicitly *nodegree*.

How do we interpret the parameters on dummy variables? It is important to know the **reference group**, i.e. the group that has been left out. In the above two examples, the reference groups are individuals who have not finished high school and individuals who have not obtained any educational degree. We therefore interpret the parameters on the dummy variables as the average difference between observations with a particular characteristic and those in the left-out/reference group, holding other variables in the regression model constant.

Note also that with dummy variables in a regression, the intercept term provides information about the reference group. In this case, β_0 gives the average outcome variable value for observations in the reference group(s) when any other continuous variables in the regression model are equal to 0. We can interpret the coefficients on the dummy variabels as intercept shifters for observations with different characteristics.

Questions: You have the following output from a regression of sleep in hours per day on hours of work per day and dummy variables for whether the individual has young children (*yngkid*) and whether the individual is Black (*black*). There are 706 observations in the data.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.55358    0.09432  90.685   <2e-16 ***
wrkhrs      -0.15117    0.01676  -9.021   <2e-16 ***
yngkid      -0.02961    0.11275  -0.263    0.793
black       -0.17398    0.17411  -0.999    0.318
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.003 on 702 degrees of freedom
Multiple R-squared:  0.1047,    Adjusted R-squared:  0.1008
F-statistic: 27.36 on 3 and 702 DF,  p-value: < 2.2e-16
```

1. Interpret the coefficient for the intercept (ignore significance).

2. Interpret the coefficient estimate for *yngkid* (including significance).

3. What is the average hours of sleep per day for individuals who work 5 hours per day, have young children, and are not Black?

4. Comparing the above output controlling for *yngkid* and *black* to the below output without, test the joint null hypothesis that neither having young children nor being Black affect sleep controlling for total hours of work per day.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.53899    0.09265  92.165   <2e-16 ***
wrkhrs      -0.15075    0.01674  -9.005   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.003 on 704 degrees of freedom
Multiple R-squared:  0.1033,    Adjusted R-squared:  0.102
F-statistic: 81.09 on 1 and 704 DF,  p-value: < 2.2e-16
```

C. Interactions and total marginal effects

What if we believe the effect of one variable depends on the value of another variable – use an interaction term.
Does the effect of education on wages vary by sex?

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

Interpreting β_0: This parameter reflects the intercept (value of wages) for males with no education.
Interpreting β_1: This parameter reflects the difference in the intercepts between women and men with no education.
Interpreting β_2: This parameter reflects the effect of an additional year of education for males.
Interpreting β_3: This parameter reflects the difference in the returns to education (differential effect of education on wage) for males and females.

Notice that individual variables appear in multiple terms in this regression. This means that interpreting the effect of one variable will require calculating a total marginal effect. We estimate total marginal effects by taking the derivative of the regression output with respect to the variable of interest.

- Total effect of $female = \beta_1 + \beta_3 \times educ$
- Total effect of $educ = \beta_2 + \beta_3 \times female$

Note that because we have an interaction term, the total marginal effect for each variable depends on the level of the other variable. For *educ*, we can calculate two different effects, one for females and one for males. For *female*, we would need to choose a level of *educ* to be able to estimate a total effect - for example, the total effect of being female on wages at the mean years of education.

The following regression output analyzes whether the effect of age (in years) on sleep hours per day varies by age and by work hours per day. Note that include a quadratic term (in this case $age2 = age^2$) is essentially including an interaction of a variable with itself, testing whether the effect of a variable on an outcome changes depending on the level of the variable.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.5059657  0.6132099  13.871   <2e-16 ***
wrkhrs      -0.1043907  0.0580052  -1.800   0.0723 .
age         -0.0118602  0.0285583  -0.415   0.6781
age2         0.0002884  0.0003228   0.894   0.3719
wrkhrs:age  -0.0010902  0.0014329  -0.761   0.4470
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.001 on 701 degrees of freedom
Multiple R-squared:  0.1108,    Adjusted R-squared:  0.1057
F-statistic: 21.84 on 4 and 701 DF,  p-value: < 2.2e-16
```

1. Interpret the coefficient for the intercept (ignore significance).

2. Interpret the coefficient estimate for *wrkhrs* (including significance).

3. What is the total marginal effect of *age* on sleep? What is the effect of *age* on sleep at age 25 for individuals who work 8 hours per day?

4. For individuals who work 8 hours per day, at what age does the sign of the relationship between age and sleep change, and how does it change?

5. Comparing the above output controlling for *age* and interactions with age and work hours to the below output without, test the joint null hypothesis that age does not affect sleep controlling for total hours of work per day.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.53899    0.09265  92.165   <2e-16 ***
wrkhrs      -0.15075    0.01674  -9.005   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.003 on 704 degrees of freedom
Multiple R-squared:  0.1033,    Adjusted R-squared:  0.102
F-statistic: 81.09 on 1 and 704 DF,  p-value: < 2.2e-16
```

6. Now consider the below regression output. What would be the null and alternative hypothesis to test whether the association between work hours and sleep does not vary by whether the individual has young children? Implement this test.

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     8.61512    0.09864  87.339   <2e-16 ***
wrkhrs         -0.16499    0.01784  -9.251   <2e-16 ***
yngkid         -0.62474    0.28252  -2.211   0.0273 *
wrkhrs:yngkid  0.11607    0.05084   2.283   0.0227 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1 on 702 degrees of freedom
Multiple R-squared:   0.11,     Adjusted R-squared:  0.1062
F-statistic: 28.92 on 3 and 702 DF,  p-value: < 2.2e-16
```

7. What would be the null and alternative hypothesis to test whether average sleep hours are the same for individuals with and without young children who work the same number of hours per week?

2. Chow Test

A Chow test is a shorter way to perform a specific type of F-test. We use a **Chow test to check if our regression parameters are different between two different groups**. In Section 7, we saw that interacting dummy variables with a) other dummies and b) continuous variables allows us to test whether different groups have different intercepts and different slopes, respectively. We may also wish to test the null that two groups follow the same regression function, against the alternative that one or more of the slope or intercept parameters differs across groups.

We will work through an example that uses sleep75 data set from the Woodridge text that we are already familiar with. Suppose we are interested in seeing whether age and total hours worked affect time slept (in minutes per week). Given this, we're interested in the following regression:

$$sleep = \beta_0 + \beta_2 age + \beta_4 totwrk + u$$

You suspect that the relationship between *sleep* and *age* and *totwrk* is different if you have young kids vs. if you do not have young kids (defined in the data as kids under the age of 3), as individuals with young kids have more time-consuming childcare responsibilities.

Question: What regression would you run as the unrestricted model to test your hypothesis that people with vs. without young kids get different amounts of sleep?

Answer: We can rewrite restricted and unrestricted regressions as:

$$\begin{aligned} \text{Unrestricted} : sleep &= \beta_0 + \beta_1 yngkids + \beta_2 age + \beta_3 yngkids * age \\ &\quad + \beta_4 totwrk + \beta_5 yngkids * totwork + e \\ \text{Restricted} : sleep &= \beta_0 + \beta_2 age + \beta_4 totwrk + e \end{aligned}$$

Before we proceed, let's practice interpreting the coefficients in the unrestricted model. Write down the interpretation of each coefficient below (answers at end):

- β_1:

- β_2:

- β_3:

- β_4:

- β_5:

Question 1: What is the average number of minutes slept per week for individuals who are 50 years old, work 40 hours per week, and do not have young kids? (answer at end) Question 2:

What is the average number of minutes slept per week for individuals who are 30 years old, work 45 hours per week, and have a young child? (answer at end) Now back to our hypothesis test!

There are two ways you could formally test this hypothesis:

A. F-test

If you suspect that this whole regression might be different if we ran it for only people with young kids, that's equivalent to saying that each of the βs is different depending on whether the respondent has young kids.

The F-test that will tell us whether there is a significant difference between these two models:

$$H_0 : \beta_1, \beta_3, \beta_5 = 0$$
$$H_1 : \beta_1 \neq 0 \text{ and/or } \beta_3 \neq 0 \text{ and/or } \beta_5 \neq 0$$

And our F stat would be:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n - k_{UR} - 1)}$$

B. Chow-Test

The Chow test is just a way to complete that F-test *without running the UR regression with all of those pesky interactions*. What Chow realized is that we can get everything we need to compute the F-stat from running the following regressions for different subsamples:

$$\begin{aligned} \text{(A) } sleep &= \beta_0 + \beta_1 age + \beta_2 totwrk + u \quad \textit{Have young kids only} \\ \text{(B) } sleep &= \beta_0 + \beta_1 age + \beta_2 totwork + u \quad \textit{Do not have young kids only} \end{aligned}$$

Chow noticed that:

1. $SSR_{UR} = SSR_A + SSR_B$
2. $q = k + 1$ the hypothesis that each beta is the same across the two groups involves $k + 1$ restrictions, since we are testing whether the intercept and each slope term may differ by group.
3. The unrestricted model, which we can think of as having a group dummy variable and k interaction terms in addition to the intercept and variables themselves, has $n - 2(k + 1)$ degrees of freedom

We can use these three facts to rewrite our F-statistic in a way so that we only need to run (1) Restricted Model, (2) Model A and (3) Model B instead of the usual restricted and unrestricted regressions:

$$F = \frac{\left(SSR_{pooled} - (SSR_A + SSR_B)\right)/q}{(SSR_A + SSR_B)/(n - 2(k+1))}$$

What this means is that we can calculate the F-statistic that tests whether or not each parameter in our original model (1) is different for individuals with vs without young kids without actually running the unrestricted model. From here, the Chow test is the same as the usual F-tests.

Bottom-line: The Chow test is just an F-test for a specific situation: when you want to see if the regression is totally different (every parameter) between different groups, but you want to avoid running the unrestricted regression. This is helpful since it's usually easier to run regressions for sub-samples that to run regressions that require creating new interaction terms.

Let's try this test with data. To compute the F/Chow-statistic, in R we need to run:

- Estimate equation (1) for all people together (call it reg_restricted):

```
Call:
lm(formula = sleep ~ age + totwrk, data = sleep75)

Residuals:
     Min       1Q   Median       3Q      Max 
-2384.63  -242.13     7.81   262.45  1302.19 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3469.20059   68.11787  50.929   <2e-16 ***
age            2.92388    1.39671   2.093   0.0367 *  
totwrk        -0.14901    0.01672  -8.912   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 420.1 on 703 degrees of freedom
Multiple R-squared:  0.1088,	Adjusted R-squared:  0.1063 
F-statistic: 42.93 on 2 and 703 DF,  p-value: < 2.2e-16
```

- Estimate equation (1) for only people with young kids (call it A_reg):

```
Call:
lm(formula = sleep ~ age + totwrk, data = sleep75[which(sleep75$yngkid ==
    1), ])

Residuals:
     Min       1Q   Median       3Q      Max
-1210.02  -258.66   -16.41   290.26  1062.36

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 3153.34833  331.43332   9.514 3.58e-15 ***
age            7.45240   11.40556   0.653    0.515
totwrk        -0.05659    0.05146  -1.100    0.274
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 442.2 on 88 degrees of freedom
Multiple R-squared:  0.01544,   Adjusted R-squared:  -0.006935
F-statistic: 0.6901 on 2 and 88 DF,  p-value: 0.5042
```

- Estimate equation (1) for only people without young kids (call it B_reg):

```
Call:
lm(formula = sleep ~ age + totwrk, data = sleep75[which(sleep75$yngkid ==
    0), ])

Residuals:
     Min       1Q   Median       3Q      Max
-2376.03  -228.19     9.76   241.80  1297.21

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3498.7306    74.1757  47.168   <2e-16 ***
age            2.8570     1.4759   1.936   0.0534 .
totwrk        -0.1628     0.0177  -9.194   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 416.1 on 612 degrees of freedom
Multiple R-squared:  0.1294,    Adjusted R-squared:  0.1266
F-statistic: 45.49 on 2 and 612 DF,  p-value: < 2.2e-16
```

To compute the Chow statistic we use:

- $SSR_{pooled} = 124084606$

```
    > sum(reg_restricted$residuals^2)
[1] 124084606
```

- $SSR_A = 17207661$

```
    > sum(A_reg$residuals^2)
[1] 17207661
```

- $SSR_B = 105982703$

```
    > sum(B_reg$residuals^2)
[1] 105982703
```

- $q = 3$ restrictions we are testing
- $k = 2$ variables in the restricted model
- $n = 706$ sample size
- $\Rightarrow F = \frac{(SSR_{pooled} - (SSR_A + SSR_B))/q}{SSR_A + SSR_B/(n-2(k+1))} = \frac{124084606 - (17207661 + 105982703)/3}{(17207661 + 105982703)/700} = 1.69378$

The critical value for $F_{3,700}$ and $\alpha = 0.05$ is about 2.63. Hence we fail to reject the critical value at the 5% significance level. This might seem surprising initially, but we notice that there are only about 90 people (out of around 700) in the dataset who have small kids. It could just be that we don't have enough observations to be picking up the effect of having small kids once we control for age and total hours worked.

3. Answers

Section 1A

1. $sleep = 8.538 - 0.151work$; $SE(\beta_{work}) = 0.0168$
   An increase of 1 hour of work per day is associated with a decrease of 0.151 hours of sleep per day.
   $t = \beta/SE(\beta) = -0.151/0.0168 = -9.005$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{work} = 0$ with a high level of confidence.

2. $sleep = 8.753 - 0.578log(work)$; $SE(\beta_{work}) = 0.0727$
   An increase of 1% in hours of work per day is associated with a decrease of 0.0058 hours (0.35 minutes) of sleep per day.
   $t = \beta/SE(\beta) = -0.578/0.0727 = -7.918$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{log(work)} = 0$ with a high level of confidence.

3. $log(sleep) = 2.144 - 0.024work$; $SE(\beta_{work}) = 0.0024$
   An increase of 1 hour of work per day is associated with a decrease of 2.4% in hours of sleep per day.
   $t = \beta/SE(\beta) = -0.024/0.0024 = -8.522$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{work} = 0$ with a high level of confidence.

4. $log(sleep) = 2.169 - 0.076log(work)$; $SE(\beta_{work}) = 0.0104$
   An increase of 1% in hours of work per day is associated with a decrease of 0.076% in hours of sleep per day.
   $t = \beta/SE(\beta) = -0.076/0.0104 = -7.302$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{log(work)} = 0$ with a high level of confidence.

5. We standardized the sleep and work variables (recall that standardizing means subtracting the mean and dividing by the standard deviation) and obtained $scale(sleep) = -2.6e^{-16} - 0.321scale(work)$; $SE(\beta_{scale(work)}) = 0.0357$
   An increase of 1 standard deviation of hours of work per day decreases hours of sleep per day by 0.321 standard deviations.
   $t = \beta/SE(\beta) = -0.321/0.0357 = -9.005$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{scale(work)} = 0$ with a high level of confidence.

6. We created a dummy variable indicating whether an individual sleeps at least 8 hours per day ($sleep8up$). We estimate the following linear probability model $sleep8up = 0.687 - 0.055work$; $SE(\beta_{work}) = 0.00794$
   An increase of 1 hour of work per day is associated with a decrease of 0.055 (or 5.5%) in the probability that an individual sleeps at least 8 hours per day.
   $t = \beta/SE(\beta) = -0.055/0.00794 = -6.979$. This is larger in absolute value than any reasonable critical value. We therefore reject the null that $\beta_{work} = 0$ with a high level of confidence.

Section 1B

1. Interpret the coefficient for the intercept (ignore significance).
   Individuals who work 0 hours per day, have no young kids, and are not Black sleep 8.55 hours per day on average.

2. Interpret the coefficient estimate for *yngkid* (including significance).
   Individuals with young children sleep 0.0296 hours (1.78 minutes) less per day on average, holding hours of work and whether they are Black constant. We have $t = -0.263$ and $p = 0.793$, indicating that we cannot reject the null that it is equal to 0 (there is no association between having a young child and sleep hours) at any reasonable level of confidence.

3. What is the average hours of sleep per day for individuals who work 5 hours per day, have young children, and are not Black?
   $8.55 + 5 * (-0.151) - 0.0296 * 1 - 0.174 * 0 = 7.765$ hours per day.

4. Test the joint null hypothesis that neither having young children nor being Black affect sleep.
   Formula: $F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1 - R^2_{UR}\right)/(n - k_{UR} - 1)} = \frac{(0.1047 - 0.1033)/2}{(1 - 0.1047)/(706 - 3 - 1)} = 0.549$
   With $q = 2$ and $dof = 702$, we can look up $c_{.05} \sim 3$. Since $0.549 < 3$, we cannot reject the null hypothesis that neither having young children nor being Black affect sleep at a 5% significance level (or any other reasonable level).

Section 1C

1. Interpret the coefficient for the intercept (ignore significance).
   Individuals who work 0 hours per day and are 0 years old sleep 8.51 hours per day on average.

2. Interpret the coefficient estimate for $wrkhrs$ (including significance).
   One additional hour of work per day is associated with 0.104 fewer hours of sleep for individuals at age 0. We observe $p = 0.0723$, suggesting we can reject the null that this effect is 0 at the 10% significance level, but not at the 5% significance level.

3. What is the total marginal effect of age on sleep? What is the effect of age on sleep at age 25 for individuals who work 8 hours per day?
   Taking the derivative with respect to age gives $-0.012 + 0.2 * 00029age - 0.0011wrkhrs$.
   One additional year of age at age 25 for individuals who work 8 hours per day will decrease sleep by $-0.012 + 0.00058 * 25 - 0.0011 * 8 = -0.0063$ hours per day.

4. At what age does the sign of the relationship between age and sleep change, and how does it change?
   Set $-0.012 + 0.00058age - .0011 * 8 = 0$ and solve for $age = 35.86$. Sleep hours decrease with age up until age 36, and increase with age afterward.

5. Test the joint null hypothesis that age does not affect sleep, controlling for total hours of work.
   The null hypothesis here is that $\beta_{age} = 0$ & $\beta_{age2} = 0$ & $\beta_{age:wrkhrs} = 0$ against the alternative that any of these coefficients are not 0.
   Formula: $F = \frac{\left(R^2_{UR} - R^2_R\right)/q}{\left(1-R^2_{UR}\right)/(n-k_{UR}-1)} = \frac{(0.1108-0.1033)/3}{(1-0.1108)/(706-4-1)} = 1.971$
   With $q = 3$ and $dof = 701$, we can look up $c_{.05} \sim 2.6$. Since $1.971 < 2.6$, we cannot reject the null hypothesis that age does not affect sleep at a 5% significance level (it would be closer at a 10% level where $c_{.05} \sim 2.08$).

6. Now consider the regression of sleep on work hours and having young children. What would be the null and alternative hypothesis to test whether the association between work hours and sleep does not vary by whether the individual has young children? Implement this test.
   The null hypothesis is that $\beta_{wrkhrs:yngkid} = 0$, and the alternative is that $\beta_{wrkhrs:yngkid} \neq 0$.
   We can test this hypothesis by looking at the t and p-values for this coefficient. With $p = 0.0227$ we can reject the null at the 5% significance level, but not at the 1% significance level.

7. What would be the null and alternative hypothesis to test whether average sleep hours are the same for individuals with and without young children who work the same number of hours per week?

$$H_0 : \beta_{yngkid} = 0 \;\&\; \beta_{wrkhrs:yngkid} = 0 \;\; vs. \;\; H_1 : \beta_{yngkid} \neq 0 \;\&/or\; \beta_{wrkhrs:yngkid} \neq 0$$

Section 2

We rewrite the unrestricted regressions as:

$$\begin{aligned}\text{Unrestricted}: sleep = {} & \beta_0 + \beta_1 yngkids + \beta_2 age + \beta_3 yngkids * age \\ & + \beta_4 totwrk + \beta_5 yngkids * totwork + e\end{aligned}$$

- β_1: the average difference in minutes of sleep for people with young kids relative to those without young kids, holding constant age and total hours worked.[1]
- β_2: being one year older is associated with a β_2 change in minutes of sleep, holding constant total hours worked and whether the respondent has young kids.
- β_3: the difference in the change in minutes slept associated with being one year older for people who have young kids vs those who don't, holding total hours worked constant.
- β_4: working one more hour per week is associated with a β_4 change in minutes of sleep per week, holding constant age and whether the respondent has young kids.
- β_5: the difference in the change in minutes slept associated with working one more hour per week for people who have young kids relative to those who don't, holding age constant.

Question 1: What is the average number of minutes slept per week for individuals who are 20 years old, work 20 hours per week, and do not have young kids?
Answer 1: $\beta_0 + \beta_2(20) + \beta_4(20)$. For these individuals, $yngkids = 0$.

Question 2: What is the average number of minutes slept per week for individuals who are 40 years old, work 40 hours per week, and have a young child?
Answer 2: $\beta_0 + \beta_1 + \beta_2(40) + \beta_3(40) + \beta_4(40) + \beta_5(40)$.

[1] Any time we include a variable in an interaction in our model, we also want to include it on its own. This is important to be able to correctly interpret the coefficient on the interaction term.