Small Assignment 2 Solutions

Due September 24th by 11:59pm PT. Download your notebook with your code and responses as a PDF via HTML and submit to Gradescope.

The following model is a simplified version of the multiple regression model used by Biddle and Hammermesh (1990) to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{yrsmarr} + \beta_3 \text{yngkid} + u$$

where sleep and total work are measured in minutes per week, yrsmarr is the number of years married, and yngkid is a dummy variable (taking a value of 0 or 1) for whether the person has a young child.

1. If adults tradeoff sleep for work, what is the sign of β_1?

(Note: typing the words "beta 1" is fine. If you want to have it as math notation, type: $ \beta_1 $. You can double click this cell and copy the code.)

If adults trade off sleep for work, we would expect a negative β_1, as the more work adults do, the less sleep they will have.

2. What signs to you think β_2 and β_3 will have?

Example Answer: I expect β_2 to be positive – holding constant total work time and the presence of a young child, I expect being married longer to be associated with increases in sleep because people who have been married for longer are also more likely to be older, which may mean that they have a more consistent sleep schedule and take more naps (you can also argue for the opposite sign).

I expect β_3 to be negative – holding constant years married and total work, I expect having a young child to be associated with decreases in sleep because having a young child likely increases stress and makes it harder to sleep, particularly if the child has trouble sleeping. (Again, you might also argue for the opposite sign).

Regardless of the sign you chose for each, remember that because you are dealing with multiple regression, you need to think of the sign of the beta you're asked about **holding the other variables in the model constant!**

3. Load in the dataset: sleep75.dta and estimate the above model using `lm()`. Display the results with the `summary()` command.

Hint: *Remember that this is a .dta file, meaning that the `haven` package and `read_dta()` commands will be useful. You may also look at the Jupyter Notebook for section 3, which uses these data, for some example code.*

```
In [1]:  library(haven)
         sleepdata<-read_dta("sleep75.dta")

         sleepreg<-lm(sleep~totwrk+yrsmarr+yngkid, data=sleepdata)
         summary(sleepreg)
```

```
Call:
lm(formula = sleep ~ totwrk + yrsmarr + yngkid, data = sleepdata)

Residuals:
     Min       1Q   Median       3Q      Max
-2415.52  -247.80     2.68   253.82  1355.80

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 3568.03648   44.46777  80.239   <2e-16 ***
totwrk        -0.14927    0.01682  -8.876   <2e-16 ***
yrsmarr        1.35323    1.39702   0.969    0.333
yngkid        -5.51006   48.10672  -0.115    0.909
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 421.4 on 702 degrees of freedom
Multiple R-squared:  0.1046,    Adjusted R-squared:  0.1008
F-statistic: 27.33 on 3 and 702 DF,  p-value: < 2.2e-16
```

4. Interpret the effect $\hat{\beta}_1$ (for now, don't worry about significance).

An increase of 1 minute per week of work is associated with a decrease of 0.1546 minutes per week of sleep, holding constant years married and having a young child.

5. Discuss whether assumptions MLR1-MLR4 are reasonable or may be violated in these data.

MLR1–It seems plausible that the model linear in its parameters (no reason to think otherwise).

MLR2–Having a random sample is contingent on the sampling method, so it may be violated if the authors were very selective about how they chose their sample (for example, do they only include workers?). We would need to refer to the paper which describes the data to understand this better.

MLR3–None of the independent variables in the model are likely to be perfectly collinear with one another. It is reasonable to expect that years married, having a young child, and total work do not perfectly explain one another. We can confirm this is true because we were able to estimate the model without any variable being dropped from the estimation.

MLR4–It is likely the case that this assumption is not satisfied. We are perhaps not including something that matters within our model, such as the total number of people in the household (where you might imagine that the larger the household, the more noise people will make, and the less sleep one is likely able to have) that is also correlated with the included variables.

```
In [ ]:
```