More Multiple Linear Regression

For those interested in delving a bit deeper into the world of linear regression, a few additional examples are included below. In the first example, you can work through a multiple linear regression model with one binary covariate and one continuous covariate. In the second example, we add an interaction between these covariates to examine effect modification/interaction between covariates in the context of linear regression. It is important to think about how the interpretation of the regression coefficients changes in the presence of an interaction term.

Example 1:

Fit a linear regression model with `recommendyes` as the outcome and `highnurse` and `quietalways` as the covariates.

1. Make a scatterplot with quietalways on the x-axis and recommendyes on the y-axis. Stratify by highnurse when you are plotting, so that you can distinguish between hospitals with highnurse $= 0$ and highnurse $= 1$. Overlay a linear prediction line for highnurse $= 0$ and highnurse $= 1$.

   Via the dropdown menus, go to Graphics $\rightarrow$ Two-way graph. Within the two-way window, you will need to create four different plots: two scatter plots (go to Basic plots $\rightarrow$ Scatter and then fill in Y and X variables) and two linear prediction lines (go to Fit plots $\rightarrow$ Linear Prediction and then fill in Y and X variables). Or, via command line:

   ```
   twoway (scatter recommendyes quietalways if highnurse==1) ///
   (scatter recommendyes quietalways if highnurse==0) ///
   (lfit recommendyes quietalways if highnurse == 1)///
   (lfit recommendyes quietalways if highnurse==0)
   ```

2. State your model.

   $Yi =$ percent of patients who recommend the hospital always
   $X_{1i} =$ percent of patients in a hospital who say that the hospital is always quiet
   $D_i = 1$ if at least 75% of patients at the hospital report that nurses communicate well, and is 0 otherwise

   $$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 D_{2i} + \epsilon_i$$

   where $\epsilon_i \sim N(0, \sigma^2)$.

3. Fit the model.

```
regress recommendyes highnurse quietalways
```

4. Evaluate the model assumptions.

   Suggestions: Check the residual plots to look for outliers and heteroskedasticity. Do the residuals look approximately normal? Patterns in the residual plot could suggest that your model for the mean of the outcome is misspecified (linearity is violated).

5. Interpret the coefficients.

   - α - the average percent of patients who always recommend a hospital when highnurse is 0 and quietalways is 0. α does not have a meaningful interpretation for this study since quietalways never drops to 0.
   - β_1 - the average increase in the percent of patients who always recommend a hospital for a one percent increase in quietalways, for a given value of highnurse.
   - β_2 - the average increase in the percent of patients who always recommend a hospital for hospitals with highnurse $= 1$ compared to hospitals with high nurse $= 0$, fixing quietalways.
   - $\alpha + 80\beta_1 + \beta_2$ - the average percent of patients who recommend a hospital with highnurse $= 1$ and quietalways $= 80$.
   - $\alpha + 80\beta_1$ - the average percent of patients who recommend a hospital with highnurse $= 0$ and quietalways $= 80$.

Example 2 - Multiple Linear Regression with an Interaction

Now, we examine whether there is an interaction between `highnurse` and `quietalways` on `recommendyes`. Equivalently, we look for evidence of effect modification of the relationship between quietalways and recommendyes by highnurse.

1. Check out the scatter plot from the previous example. Does the plot suggest that an interaction term might improve the model?

   Yes. We can look for evidence of effect modification by comparing the slopes of the overlayed lines in the scatter plot. Because the slopes appear to differ by highnurse, there is evidence of effect modification.

2. State your model.

   $Yi =$ percent of patients who recommend the hospital always
   $X_{1i} =$ percent of patients in a hospital who say that the hospital is always quiet
   $D_i = 1$ if at least 75% of patients at the hospital report that nurses communicate well, and is 0 otherwise

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 D_i + \beta_3 X_{1i} D_i + \epsilon_i$$

   where $\epsilon_i \sim N(0, \sigma^2)$.

3. Fit the model.

```
. xi: regress recommendyes i.highnurse*quietalways
```

4. Evaluate the model assumptions.

   You can use the same approach as the previous question.

5. Interpret the coefficients.

   - α - the average percent of patients who always recommend a hospital when highnurse is 0 and quietalways is 0. α does not have a meaningful interpretation for this study since quietalways never drops to 0.
   - β_1 - the average increase in the percent of patients who always recommend a hospital for a one percent increase in quietalways, when highnurse $= 0$.

- $\beta_1 + \beta_3$ - the average increase in the percent of patients who always recommend a hospital for a one percent increase in quietalways, when highnurse $= 1$.
- β_2 - the average increase in the percent of patients who always recommend a hospital for hospitals with highnurse $= 1$ compared to hospitals with high nurse $= 0$, when quietalways equals 0. β_2 does not have a meaningful interpretation in this analysis. Note that we could have centered the covariate quietalways around its mean, so that the covariate would be more interpretable.
- $\beta_2 + 70\beta_3$ - the average increase in the percent of patients who always recommend a hospital for hospitals with highnurse $= 1$ compared to hospitals with high nurse $= 0$, when quietalways equals 70.
- $\alpha + 80\beta_1 + \beta_2 + 80\beta_3$ - the average percent of patients who recommend a hospital with highnurse $= 1$ and quietalways $= 80$.
- $\alpha + 80\beta_1$ - the average percent of patients who recommend a hospital with highnurse $= 0$ and quietalways $= 80$.