Lab 9 Activity

We will look at some data from the U.S. Department of Commerce Census of 1977.

Variable	Description
Population	population estimate as of July 1, 1975
Income	per capita income (1974)
Illiteracy	illiteracy (1970, percent of population)
Life Exp	life expectancy in years (1969–71)
Murder	murder and non-negligent manslaughter rate per 100,000 population (1976)
HS Grad	percent high-school graduates (1970)
Frost	mean number of days with minimum temperature below freezing (1931–1960) in capital or large city
Area	land area in square miles

Run the following code to name the data you will be using as `dat`:

```
dat <- data.frame(state.x77)

# Rename 2 columns that are named in slightly unusual way
colnames(dat)[c(4, 6)] <- c("Life_Exp", "HS_Grad")
```

1. Create two new columns that represent the mean-centered versions of the `HS_Grad` variable and the `Income` variable. For the remainder of the activity, use these centered variables as the predictors and treat `HS_Grad` as the moderator.

2. Run a regression predicting `Life_Exp` with an interaction between the mean-centered `HS_Grad` variable and the `Income`. Is the interaction term significantly different from 0? Should you interpret the other regression coefficients?

3. What is the expected value of the slope of `Income` when `HS_grand` is equal to 20?

4. Calculate the simple slopes of `Income` at some values of `HS_grand` that you believe should be informative. What is the meaning of the simple slopes? Focus on the sign of the slopes, the value will be really small due to the scale of the outcome.

5. Create both a simple slopes plot and a johnson-neyman plot to visualize the interaction between the two predictors. Outside what interval is the slope of `Income` significantly different from 0?