FIN 373 Homework 6

due: **10/5/21**

Instructions: Please submit solutions on canvas. Only a knitted pdf of an `Rmarkdown` file will be accepted.

Problem 1: In class and our weekly readings, we encountered an important programming method called the *loop*. In this exercise, we practice using loops with data on the ideological positions of United States Supreme Court Justices. Just like legislators, justices make voting decisions that we can use to estimate their ideological positions.[1]

The file `justices.csv` contains the following variables:

Variable	*Description*
`term`	Supreme Court term (a year)
`justice`	Justice's name
`idealpt`	Justice's estimated ideal point in that term
`pparty`	Political party of the president in that term
`pres`	President's name

The ideal points of the justices are negative to indicate liberal preferences and positive to indicate conservative preferences.

a. We wish to know the median ideal point for the Court during each term included in the dataset. First, calculate the median ideal point for each term of the Court. Next, generate a plot with term on the horizontal axis and ideal point on the vertical axis. Include a dashed horizontal line at zero to indicate a "neutral" ideal point. Be sure to include informative axis labels and a plot title.

b. Next, we wish to identify the name of the justice with the median ideal point *for each term*. Which justice had the median ideal point in the most (potentially nonconsecutive) terms? How long did this justice serve on the Court overall? What was this justice's average ideal point over his/her entire tenure on the Court?

c. We now turn to the relationship between Supreme Court ideology and the president. Specifically, we want to see how the ideology of the Supreme Court changes over the course of each president's time in office. Begin by creating two empty "container" vectors: one to hold Democratic presidents, and another for Republican presidents. Label each vector with the presidents' names.

d. Next, for each Democratic president, calculate the shift in Supreme Court ideology by subtracting the Court's median ideal point in the president's first term from its median ideal point in the president's last term. Use a loop to store these values in your Democratic container vector. Repeat the same process for Republican presidents.

e. What was the mean and standard deviation of the Supreme Court ideology shifts you just calculated when looking only at the Democratic presidencies? What about the Republican presidencies? Which Republican president's tenure had the largest conservative (positive) shift on the Court? Which Democratic president's tenure had the largest liberal (negative) shift?

[1] This exercise is based in part on Andrew Martin and Kevin Quinn (2002) "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis*, 10:2, pp.134-154.

f. Create a plot that shows the median Supreme Court ideal point over time. Then, add lines for the ideal points of each unique justice to the same plot. The color of each line should be red if the justice was appointed by a Republican and blue if he or she was appointed by a Democrat. (You can assume that when a Justice first appears in the data, they were appointed by the president sitting during that term.) Label each line with the justice's last name. Briefly comment on the resulting plot.

Problem 2: Across industrialized countries, it is a well-studied phenomenon that childless women are paid more on average than mothers. In this exercise, we use survey data to investigate how the structural aspects of jobs affect the wages of mothers relative to the wages of childless women.[2]

In this paper, the authors examine the association between the so-called *mother wage penalty* (i.e., the pay gap between mothers and non-mothers) and occupational characteristics. Three prominent explanations for the motherhood wage penalty–"stressing work-family conflict and job performance," "compensating differentials," and "employer discrimination"–provide hypotheses about the relationship between penalty changes and occupational characteristics. The authors use data from 16 waves of the National Longitudinal Survey of Youth to estimate the effects of five occupational characteristics on the mother wage penalty and to test these hypotheses.

This paper uses a type of data known as "panel data." Panel data consist of observations on the same people over time. In this example, we are going to analyze the same women over multiple years. When analyzing panel data, each time period is referred to as a *wave*, so here each year is a wave. **The most general form of model for working with panel data is the *two-way fixed effects model*, in which there is a fixed effect for each woman and for each wave.**

The data file is `yu2017sample.csv`. The names and descriptions of variables are:

Variable	*Description*
`PUBID`	ID of woman
`year`	Year of observation
`wage`	Hourly wage, in cents
`numChildren`	Number of children that the woman has (in this wave)
`age`	Age in years
`region`	Name of region (North East = 1, North Central = 2, South = 3, West = 4)
`urban`	Geographical classification (urban = 1, otherwise = 0)
`marstat`	Marital status
`educ`	Level of education
`school`	School enrollment (enrolled = `TRUE`, otherwise = `FALSE`)
`experience`	Experience since 14 years old, in days
`tenure`	Current job tenure, in years
`tenure2`	Current job tenure in years, squared

[2]The exercise is based on: Wei-hsin Yu and J anet Chen-Lan Kuo (2017) "The Motherhood Wage Penalty by Work Conditions: How Do Occupational Characteristics Hinder or Empower Mothers?" *American Sociological Review* 82(4): 744-769.

Variable	*Description*
`fullTime`	Employment status (employed full-time = `TRUE`, otherwise = `FALSE`)
`firmSize`	Size of the firm
`multipleLocations`	Multiple locations indicator (firm with multiple locations = 1, otherwise = 0)
`unionized`	Job unionization status (job is unionized = 1, otherwise = 1)
`industry`	Job's industry type
`hazardous`	Hazard measure for the job (between 1 and 5)
`regularity`	Regularity measure for the job (between 1 and 5)

a. How many different women are in the data? How many observations per year? We will refer to each row as a "person-year observation" since the row contains data on a given person in a particular year. In a few sentences, describe one advantage and one disadvantage of using a contemporary cohort of women rather than an older cohort in estimating the predictors of the mother wage gap.

b. `numChildren` is the variable representing how many children the woman had at the time of an observation. Please provide a table that shows the proportion of observations by the number of children. Provide a brief substantive interpretation of the results.

c. Create a new indicator variable `isMother` that takes a value of 1 if the woman has at least one child in that year and a value of 0 otherwise. Tabulate the new variable. Briefly comment on the results.

d. Create a new variable called `logwage` that is the log of `wage`. Make two boxplots, one for `wage` and the other for `logwage`, as a function of educational level (`educ`). Compare the two boxplots and discuss the purpose of the log transformation.

e. In the same graph, plot the mean `logwage` against year for mothers, then for non-mothers in a different color or line type. Include a legend and a proper title. Make sure the figure and axes are clearly labeled. Give a brief interpretation of the results.

f. Run a regression using fixed effects for both *woman* and *year*. You should be sure to include variables for number of children (`numChildren`) and job characteristics (`fullTime`, `firmSize`, `multipleLocations`, `unionized`, `industry`). Note: that you should *not* use the `isMother` variable you created earlier in this model. Report the coefficient of `numChildren`. Provide a brief substantive interpretation of this coefficient and the coefficients for any two other variables. (*Hint*: fixed effects means including the relevant factor variables in the regression model – see the bolded statement in the problem introduction).

g. Add interactions between `numChildren` and `regularity` and between `numChildren` and `hazardous` to the model in the previous question. Report the five coefficients involving these variables. Interpret the interaction term for `numChildren` and `hazardous`. Can we interpret the effect of occupation characteristics on motherhood wage penalty as causal? Why or why not?