Public Policy 529
Fall 2023: Problem Set #8

Due on Monday, November 20th

1. In a 2014 survey of Americans, respondents were asked "How important is it for an American to be able to speak English?" They were also asked about their political party identification. The following table shows a raw frequency distribution of these two variables:

| | Party Identification | | | |
Speak English	Democratic	Independent	Republican	Total
Very Important	401	172	312	885
Fairly Important	134	57	66	257
Not Important	51	12	13	76
Total	586	241	391	1,218

   (a) Calculate f_e, the expected frequency in each cell under the scenario that the variables are independent. Do you have sufficient sample size to perform a χ^2 test?

   (b) Use the χ^2 table from Canvas or lecture slides to look up the critical value of χ^2 that would be necessary to reject the null hypothesis that the variables are independent (α=.05). You will first have to find the correct degrees of freedom.

   (c) Calculate the χ^2 statistic from the data presented above. Can you reject the null hypothesis? What can you say about the p-value?

   (d) You are informed by a statistician that the gamma statistic for this relationship is -0.21 with an ASE of 0.05. The Kendall's tau-b statistic is -0.11 with an ASE of .026. Explain the meaning of these findings.

2. Use the dataset PEW_HigherEd_subset for this question. The variable JobSatis asks how satisfied a person is with their job (very dissatisfied, somewhat dissatisfied, somewhat satisfied, very satisfied). The variable OverQual asks if the person feels overqualified for their job (not overqualified, overqualified).

   (a) What are the measurement levels of these variables?

   (b) Make a table showing the joint frequency distribution of these two variables, with JobSatis forming the rows. Have your software report a χ^2 test for this table. Interpret the reported statistics. See lecture slides for code.

(c) What is the critical value of the χ^2 statistic in the test that was just performed?

3. Use the dataset anes2016subset for this question. The variable ScientistsThem contains the respondent's feeling thermometer score for scientists. The variable PartyID measures whether the respondent identifies as a Democrat, Republican, Independent, or Other. The variable ReligImpt indicates whether respondents say religion is an important part of their life (yes, no).

   (a) Use the tabulate command with the summarize option to obtain the mean values of the variable ScientistsTherm across the categories of PartyID:

      Stata: tabulate PartyID, summarize(ScientistsTherm)

      R: aggregate(ScientistsTherm ~ PartyID, data = anes2016, FUN = mean)

      Interpret the findings from this table.

   (b) Use ANOVA to test whether these means are different from each other. Interpret Stata output to perform the significance tests. The correct command is:

      Stata: anova ScientistsTherm PartyID

      R: summary(aov(ScientistsTherm ~ PartyID, data = anes2016))

   (c) Now add ReligImpt to the ANOVA analysis, using a twoway ANOVA to test whether the category means of each variable are different from each other when controlling for the effect of the other variable.

      Stata: anova ScientistsTherm PartyID ReligImpt

      R: summary(aov(ScientistsTherm ~ PartyID + ReligImpt, data = anes2016))