MSCI 9110:EXA/EXD Spring 2017

*Advanced Analytics* Des Moines

**Memo #1 (Due February 14)**

In this assignment you will conduct a statistical investigation of the salaries of tenure-track faculty at the Tippie College of Business. The Excel file *TT Faculty FY17.xls*, available on the course ICON site, contains the following variables for all full-time tenure-track faculty, except the dean of the college:

ID: Identification number for faculty member in data set

Salary: Base salary for 2016-17 academic year (does not include administrative stipends, summer salary, salary from fellowship/professorship/chair appointments, or any extra compensation)

Dept.: Home department of faculty member (Accounting, Economics, Finance, Marketing, Management & Organizations or Management Sciences)

Rank: Assistant, Associate or Full Professor

PhD: Year in which faculty member received PhD

Years: Years since receiving PhD at start of 2016-17 academic year (= 2016 - PhD)

Gender: Male (M) or Female (F)

All data is public information and can be obtained from public web pages.

1. Start by producing a cross-tabulation (pivot table of counts) using the variables Rank and Gender. How evenly is the faculty distributed within the three ranks? How even is the percentage of female faculty within the three ranks?
2. Use pivot tables/charts to generate a multiple histogram that compares the salary distributions for males and females. (Since the number of females is much less than the number of males you should display the percentage of men and percentage of women in each salary range, rather than raw counts.) Describe any differences that you see in the two distributions.
3. Use pivot tables/charts to generate a multiple histogram that compares the salary distributions for assistant, associate and full professors. (Again display the percentages within each group, not the raw counts.) How do these distributions compare?
4. Run a regression of LN(Salary) on Years, dummy variables for the departments, and a dummy variable for gender. Use Marketing for the base case department, and Female for the base case gender. Interpret the regression equation (intercept and slope coefficients) and discuss significance of variables. How do the residuals look (residual plot on Years and histogram of residuals)? Ignore all plots of residuals on dummy variables.
5. Repeat the regression in Part D, but adding dummy variables for rank (use Assistant for the base case). What effect does this have? How would you interpret this result?
6. Run separate regressions for the faculty in each of the three ranks, in each case using the remaining independent variables (Years, dummy variables for the departments and the dummy variable for gender.) Describe differences that you see between these three regressions. How consistent are the estimates for the salary differences between the departments?
7. Describe how this analysis might be improved by considering additional variables.

**Note:** In the course of your analysis you may want to sort the original data on certain variables. If you have previously run regressions, then any residual plots will be scrambled when the data is sorted. However if you subsequently put the data back in the original order then any residual plots will again be correct. For part F, it is recommended that you create separate sheets that contain the data for each of the three different ranks.