**Exam 1 – Fall 2022**

**Problem #1:**

We have data for 44 physicians working in emergency medicine at a major hospital system (file; “Emergency-Service-E-22”, attached). The concern is about the number of complaints each received the previous year (Complaint). Why is it different from physician to physician? Could it be explained using other factors? Data available consist of the number of patient visits for each clinician (N-Visits; it’s the volume of work resulting in the number of complaints (Complaint) - and four covariates (X1= Revenue, in dollars per hour; X2= Hours, work load at the emergency service; X3= Gender, Female/Male, and X4= Residency training (fellowship) in emergency medicine, No/Yes). It has been known in the medical community about the heavy work load (say, about 50-60 hours a week), so X2 is the primary focus. In addition, Emergency Medicine was a relatively new specialty at the time of data collection, so X4 is also at some interest. Use Y = (Complaint/N-Visits)x(1000), the number of complaints per 1000 visits, as the dependent variable and fit the “Normal Error Regression Model” (NERM) with 4 covariates X1-X4.

1. State the model (including assumptions) and provide summary results (ANOVA Table – including estimate of the variance of residuals, regression coefficients, p-values).
2. Draw inferences about the effects of X2 (interpretation of the regression coefficient, standard errors, level of significance, 95% CI of regression coefficient).
3. Create a graph to investigate if the variance, stipulated by the NERM, is constant across fitted values and draw your conclusion.
4. Create a graph to investigate whether the relationship between Y1 and X2 is linear (in the presence of other three covariates), explain the method and draw your conclusion.

**Problem #2:**

The onset of COVID-19 pandemic in spring 2020 was notable for the number of deaths happening in more densely-populated areas, with New York one of the worst-hit cities in the first few months of the virus' arrival. But the introduction of vaccines has subsequently marked a shift nationwide, with deaths increasingly coming from more rural areas where there is greater skepticism of the vaccines, and where COVID mitigation measures like mask wearing and social distancing are less likely to be observed. That shift can be seen in Minnesota since the onset of the highly contagious delta COVID variant, which has caused cases to surge since it became the dominant strain of the virus over the summer of 2021.

We have a data set for the state of Minnesota, obtained in mid-November 2021 (file; “Covid-Vaccine-E-22). Data available for each county consist of number of Covid-19 deaths from June 3 to November 12 (Delta Deaths), Total Covid-19 deaths since the beginning of the pandemic (Total Deaths), county population size (in thousands, Size), Vaccination Rate (% as of end of November 2021, X1), Region (X2 = 1 for seven county metro, 0 for outstate).

1. Compare the mean vaccination rates, Metro versus Outstate, and draw your conclusion. Would you prefer to use the two-sample t-test or the Wilcoxon test? Explain the reason for your preference or choice.
2. Focus on the Covid-19 deaths from June 3 to November 12; form the death rate Y = Delta Deaths/Size (number of deaths per 1000 population, then compare the mean death rates, Metro versus Outstate, and draw your conclusion.
3. Use Y as the dependent variable as the dependent variable and fit the “Normal Error Regression Model” (NERM) with 2 covariates X1and X2. Draw your inference about the difference between Metro and Outstate; compare the result to the result of (b) and draw your conclusion.
4. Refit the model in (c) using weighted Least Squares with the Size being the weight. Focus on X2; Focus on the difference Metro versus Outstate, compare the result to the result in (c) and draw your conclusion.