# **BS806 --- Homework 4**

**Reading**. Chapters 5 and 6 of (SW) Sanford Weisberg. Applied Linear Regression, 4th ed. Wiley

**(100 points)**. The data for this assignment comes from Wave I of the National Longitudinal Study of Adolescent to Adult Health and is stored in a csv file called **depress2.csv**. We are interested in the association between parental receipt of public assistance and depressive symptom index (from the Center for Epidemiologic Studies Depression Scale or CES-D). The variables are described below. We would like to perform a multiple linear regression predicting depressive symptom index from public assistance, adjusting for age.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Description** | **Coding/Unit** | **Type** |
| AID | Subject ID |  | Numeric |
| age | Subject’s age | Years | Numeric |
| cesd | Depressive symptom index (CES-D) | Symptom count | Numeric |
| cesd\_16 | Depressive symptom index category (>=16 vs. <16) | 0=<16 symptoms (low)  1=>=16 symptoms (high) | Numeric |
| publicassist | Parental receipt of public assistance | 0=No  1=Yes | Numeric |

**Preparation**: Summary data and visually examination on the relationship.

1. **[5 points]** Create appropriate summary statistics/sample characteristics to summarize variables (age and cesd only) in this dataset stratified by public assistance status. Then comment on your observations.
2. **[5 points]** Generate appropriate plot to show the relationship between depressive symptom index and public assistance and then comment on your observations based on the plot.

**Hypothesis Testing:** raw comparison and model building

1. **[10 points]** Before going forward with multiple linear regression, let’s perform the two-sample t-test first to determine whether depressive symptom index are the same based on public assistance and then comment on your results.
2. Conduct a multiple linear regression predicting depressive symptom index from public assistance, adjusting for age.
   1. **[10 points]** Write a complete report for the **global hypothesis test** for the built model including R2.
   2. **[10 points]** Write a complete report for the effect of public assistance, adjusting for age.
   3. **[10 points]** Write a complete report for the effect of age, adjusting for public assistance.
   4. **[10 points]** What is the predicted value of CESD for a 14 year old subject whose parent receives public assistance?
3. **[10 points]** Calculate the age-adjusted least square means for CESD in both ways: manually and with R
4. Conduct a multiple linear regression predicting depressive symptom index from public assistance, age and interaction between public assistance and age.
   1. **[10 points]** Based on the model, write down the fitted regression line for the people with public assistance and without public assistance, separately.
   2. **[10 points]** Perform the hypothesis test to examine whether the interaction should be included in the model.

**Result Reporting**

1. **[10 points]** Assuming you were writing a paper on this issue, write one short paragraph in plain English to summarize essential findings to address the question of interest.