**CPH 576C, group work**

**L2: Data cleaning and exploratory analysis**

0. Fill in the following table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type of data | Descriptive statistic | Graphical displays | Test comparing two independent groups | Test comparing two dependent groups | Test comparing multiple groups |
| Binary or Categorical |  |  |  |  |  |
| Continuous  (symmetric) |  |  |  |  |  |
| Continuous  (skewed or not symmetric) |  |  |  |  |  |

1. Import the HELP data into the statistical software of your choice. Please sit next to someone who uses the same software so that you can help each other.

2. Get a quick description of the contents of the dataset

3. Consider the variables female, substance, cesd and i1. What type of variables are these, and what do they represent? Are there any values out of range?

4. What type of graph would you use to explore the association of

a. cesd and substance?

b. cesd and i1?

Create these graphs

5. Use Prog Sgscatter; matrix or

graph matrix to explore relationships between variables. Use the help files.

6. What are the variables cesd, cesd1, cesd2, cesd3 and cesd4?

7. Use the variables f11-f1t to create a CESD (Center for Epidemiologic Studies Depression Scale) score. The CESD is a commonly used measure of depression, with higher scores indicating greater depression. Be sure to reverse code items where appropriate. How does your value compare to the already created one? Why is there a difference?

8. The missing item rule for the CESD is that if 4 or less of the questions are unanswered, impute the missing values with the mean of the other values. Do this.