**STAT 6740**

**Homework 4**

**Due: Wednesday, October 28**

1. Recall the colleges and university data project from previous homework assignments. Do the following:
2. Read the AAUP and US News data, combine the two datasets, and add the Department of Education region to the entire combined dataset. For the DOE region, assign the region a number but use formatting to print the values using Roman Numerals.
3. Create a summary table using PROC TABULATE that presents summary statistics for each combination of university type and region for student-to-faculty ratio and average salary for full professors. Your summary statistics should include sample size, minimum, maximum, median, mean, and standard deviation and use some type of formatting for the statistics aside from the keyword. Add variable labels for all variables included in the table. Add an appropriate title and a footnote to indicate the sources of the data.
4. Within each state, create 2 new variables that each take the value “1” if the value is in the lower third of the data, “2” if the value is in the middle third of the data, and “3” if the value is in the upper third of the data based on average faculty salary (all ranks) and total undergraduate enrollment. To do this:
   1. Use PROC UNIVARIATE to calculate and save the 33rd and 67th percentiles for each variable (using PCTLPTS and PCTLPRE options in OUTPUT statement);
   2. Merge the percentiles with the original data; and
   3. Use IF-THEN statements to assign each observation to the appropriate two categories.
5. Format the two variables so that each number is associated with an appropriate description (upper third, middle third, lower third).
6. Using PROC TABULATE, create a table that has the minimum, first quartile, median, third quartile, and maximum of the average combined SAT score for each combination of the two categorical variables created in Step 3. Make sure variables are labeled and do NOT include any rows and columns with sub-totals. Print the statistics with a single decimal point. Add an appropriate title and a footnote to indicate the sources of the data.
7. Submit the SAS program, SAS log, and SAS output.

1. Recall Ashley Hart’s data from previous homework assignments.
2. Read the data into SAS and change the value of the food variable for observations where it is blank to “No food”. Also, create new observations for each replicate and new variables for each carotenoid (from HW 3).
3. Add appropriate labels to all variables.
4. Create a summary tables using PROC TABULATE that contains summary statistics for each combination of fiber, level, enzyme strength, and carotenoid. Your summary statistics should include sample size, mean, standard deviation, and 90th percentile. Add labels for the statistics. Add an appropriate title and a footnote to indicate the source of the data.
5. Create a new dataset that calculates the difference between AC and BC percent micellization. Use PROC TABULATE to create a table that shows the mean differences and the results of a t-test to determine if the differences are significant (t-test) for each fiber, level, food, and enzyme level. Include the test statistic and p-value in the table and label the columns appropriately. Add an appropriate title and a footnote to indicate the source of the data.
6. Submit your SAS program, SAS log, and SAS output.
7. Recall the Ohio Department of Health data from previous homework assignments.
8. Import and merge the four SAS datasets and create the four high-risk indicators from Homework 3. Format the risk indicator values as “High Risk” and “Low Risk.”
9. Using PROC TABULATE, create a frequency table that shows the number of census tracts (and overall percent) within each combination of the four risk indicators. Use two risk factors for rows and two risk factors for columns, and make sure to include marginal totals (i.e., ALL) for each risk factor. Add variable labels. Add an appropriate title and a footnote to indicate the sources of the data (American Factfinder / U.S. Census Bureau).
10. Using PROC TABULATE, create a table that shows the total number of people and the total number of households within each combination of the four risk factor groups (again, include all marginal totals). Change the text colors for the table so that the cell contents, the risk factor labels, and the variable labels have different colors. Add an appropriate title and a footnote to indicate the sources of the data.
11. Submit the SAS program, SAS log, and SAS output.