**STAT 6740**

**Homework 5**

**Due: Friday, November 20**

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. Also add the tertiles for average faculty salary and total undergraduate enrollment.
3. For each university type, create a separate set of side-by-side boxplots that show the distribution of average assistant professor salary by region.
   1. Connect the mean values across regions
   2. Create titles that include the university type and footnotes that indicate the sources of the data
4. For each combination of the salary and enrollment tertiles, create a scatterplot of average combined SAT scores versus student-faculty ratio.
   1. Use a different symbol for each university type;
   2. Put reference lines in for SAT scores at 1200 and student-faculty ratio of 15;
   3. Create titles that contains both tertile values.
5. Create a single bar chart that shows the total number of professors of all rank by region and university type.
   1. Create an appropriate title.
6. 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). Calculate the difference between AC and BC percent micellization (HW 4).
3. Create a bubbleplot where the axes are AC and BC percent micellization and the bubble size is lutein percent micellization.
   1. Use AXIS statements to format the axes (label, values);
   2. Make the bubbles green;
   3. Add an appropriate title and footnote (data source).
4. Use PROC GPLOT to fit a linear regression line to predict BC percent micellization as a function of AC percent micellization.
   1. Plot the data in black and the regression line in red;
   2. Include a 95% confidence band for the regression line;
   3. Add an appropriate title.
5. Create a block chart that shows the average lutein percent micellization for every combination of fiber and fiber level.
   1. Add an appropriate title;
   2. Use PATTERN statements to define different solid colors for each fiber level.
6. Submit your SAS program, SAS log, and SAS output.
7. Recall the Ohio Department of Health data from Homework Assignments 2 - 4.
8. Import and merge the four SAS datasets and create the four high-risk indicators from Homeworks 3 and 4. Format the risk indicators to note “High Risk” and “Low Risk.”
   1. For risk factors associated with “House Age” and “Female Head of Household,” create a single variable with 4 levels (Low-Low, Low-High, High-Low, High-High).
   2. Calculate the number of children age 5 or less in each census tract using the equation pct\_age\_5\*num\_pop/100.
9. Create a chart that shows the total number of children across all census tracts for each of the combination of the risk factors. (Hint: use the double-risk factor variable defined in Question 1a as the subgroup variable; the other two risk factors as the chart variable and the group variable, and use TYPE=SUM).
   1. Create different patterns for each subgroup value, but make sure that the Highest Risk (multiple high-risk variables) children stand out in the chart.
   2. Add an appropriate title.
10. Using a format procedure, divide the population into 10 groups based on the percent of renters (0-10, 10-20, … 90-100). Create a pie chart that shows the number of census tracts in each of the 10 groups.
    1. Add an appropriate title.
    2. Place the group labels in a legend.
    3. Move the frequency counts inside the pie.
    4. Use a GOPTIONS state to specify a color list for the pie chart.
11. Create a 3-dimensional scatterplot that shows the relationship between percent of households with a married couple (pct\_married), percent of people with a high school degree (pct\_hs\_ed), and percent of households that immigrated to the U.S. since 2000 (pct\_fr\_00). Make percent HS educated the z axis.
    1. Label each of the variables appropriately.
    2. Modify the plot so that the plotting symbol is a square, the color is green, the rotation is 30 degrees, and angle is 45 degrees.
    3. Add a grid to the plot.
    4. Add an appropriate title.
    5. Add a right-justified footnote that indicates the data come from the U.S. Census Bureau.

Submit the SAS program, SAS log, and SAS output