# STAT 540: Final Exam

All data sets you need for the exam can be found as separate worksheets in the Excel workbook; import them as needed. Any data manipulation must take place within your SAS programs. You may consult online documentation, but you cannot seek external help, other than myself.

1. Read the Excel worksheet Q1 into SAS. It lists multiple height measurements for individual trees; some trees were measured twice, some measured three times. Tree species, scientific name, individual tree ID, date of measurement and height (in feet) are recorded.

   (a) Use PROC TRANSPOSE to transform the data so that individual trees have a single record recording their multiple height measurements. The output data set should have the following appearance:

   | Tree ID | Latin Name | Height_1 | Height_2 | Height_3 |
   |---------|------------|----------|----------|----------|
   | Pt1 | Pinus taeda | 169 | 164 | . |
   | Pt2 | Pinus taeda | 168.7 | 167 | 168.7 |
   | ... | ... | ... | ... | ... |
   | Qp1 | Quercus pagoda | 141 | 162 | 141 |

   (b) Change your code so that any missing values in the output table are replaced with "No measurement".

   (c) Graduate students should generate a table similar to the table in part (a) that also includes the variables Date 1, Date 2, and Date 3, corresponding to the measurement dates of Height 1, Height 2, and Height 3 in the table above.

2. Read the Excel worksheet Q2 into SAS. It lists the average monthly high temperature in degrees Fahrenheit, average monthly low temperature in degrees Fahrenheit, and average monthly precipitation (in inches) for three US cities.

   (a) Provide commands to generate a plot similar to Plot 1 on the website. Provide ODS commands to save the graph as a JPG file on a directory of your choice.

   (b) Now provide commands to generate a plot similar to Plot 2 on the website.

   (c) Graduate students should provide commands to generate a plot similar to Plot 3 on the website.

3. Read the Excel worksheet Q3 into SAS. It lists four variables: state, county, percent of persons below poverty rate (2009-2013 average), and population density (per square mile).

   (a) Name the variables State, County, PovertyRate, PopDensity, and label the last two as "Poverty Rate" and "Population Density".

(b) Compute the product of the poverty rate (as a decimal, not a percent) and population per square mile and label this variable Poverty Density. For full credit, these statements should be added to the previous data step, rather than a second data step.

(c) Find the median Poverty Density for each state and save the medians in an output data set containing only the median and state variables. Use a one-to-many match merge to add Median Poverty Density as an additional variable to your original file.

(d) Create a character variable, PovStatus, with levels "Above State Median", "Below State Median", "State Median" based on whether a county's poverty density is above, below, or equal to the state median. Print the merged data set with attractive labels and significant digits consistent with the original data set.

4. Use the Excel worksheet Q3 from 3(a) for the following.

(a) Use %LET to select a given state and print a percent histogram of Poverty Rate for counties in that state with an appropriate title; demonstrate your code with the state of Arizona. Graduate students should also write ODS code to output the histogram as a PDF file named Arizona.pdf.

(b) Write a short macro with two arguments, State and County, that will print the data for a state/county combination. Test your macro for Covington County in Alabama.

(c) Change your macro with %IF, %THEN, %ELSE, %DO and %END statements so that it will print all the data for a state when only state is specified, but county is left blank. Test your macro for the state of Arkansas.

(d) Use FIRST./LAST. statements to find and print the county in each state with the lowest population density.