

## 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. Do not make any changes to the data sets in Excel; any data manipulation must take place within your SAS programs.

1. Read the Excel worksheet Q1 into WORK data set Q1; it lists weekday commuting times over a 3-week period (1=Monday, 2=Tuesday, etc)
  - (a) Using IF-ELSE and OUTPUT statements, separate the file into two data sets based on the value of the variable Commute.
  - (b) Transpose the original data set so that, when printed, it has the layout shown below:

Week	Day	Home to Work	Work to Home
1	Monday	18	17
1	Tuesday	17	18
1	Wednesday	15	17
		...	
3	Friday	14	18

2. Use the growth curve data on yellowfin tuna in Excel worksheet Q2 to re-create Plot 1 on the website using PROC SGPLOT. In addition to using PROC SGPLOT options, you may find it useful to modify the data in previous DATA steps. *Grad students only:* Replicate Plot 2 as well.
3. The data in Excel worksheet Q3 contains bacteria readings from two stations (representing two different streams) over the course of a year.
  - (a) Write SAS code to find the monthly maximum and minimum readings for E Coli for each of the stations. Print the monthly summary data for each station (don't worry about the format).
  - (b) Use PROC PRINT to output the monthly minimum and maximum readings exclusively for station CW-208. The title should be "Monthly E Coli Readings for CW-208". The only columns printed should be labeled "Month", "Minimum" and "Maximum", and Month should be displayed as January, February, etc. E Coli readings should be rounded to the nearest integer. *Grad students:* The numbers for the minimum should be green, while the headers for all three variables should be blue.
  - (c) Use a many-to-one match merge by station and month to attach the monthly maximum and minimum to the original data set. Print the first 20 observations from the data set.
4. Refer to the same data set used in Q3
  - (a) Use %LET so that the user can select only the records for station CW-088, and then print those records with the title "Bacterial WQ indicators for Station CW-088"

- (b) Write code to find and store the monthly mean for each combination of station and month, much as you did in 3(a) for the monthly minimum and maximum. Use multiple CALL SYMPUT statements to identify the station and month with the highest mean. Print all records for the resulting station-month combination (there should be 4-5 records total) with the title: "Station xxxx had the highest monthly mean E Coli readings".
- (c) *Grad students only.* Create a macro to print all records for both stations for a given month, with the title, e.g., "Monthly records for August". Ideally, the macro would have the single numeric argument, month, but it may be preferable to create a character version of month in the data set prior to writing the macro and use a character argument instead (printing the recommended title is difficult otherwise). Demonstrate the macro for the month of August.