

## STAT 541: Final Exam

- Hourly temperature and windspeed data were collected from two different weather stations on January 31, 2011 (January 31 2011 Weather worksheet). A separate chart contains wind chill conversions (Wind Chill worksheet) for a variety of temperatures and windspeed.
  - Use informat statements to create versions of the windspeed and temperature data that are converted to the nearest multiple of 5 (We will need this in order to be able to use the windchill chart).
  - Use the windchill chart as a lookup table to add windchill to the weather data.
  - Graduate students should create a PICTURE format that prints both the windchill as a signed 2-digit number and a message on frostbite danger (use the following chart as a reference):

Wind Chill	Message
-30 to -40	Frostbite in ten minutes
-20 to -30	Frostbite in thrity minutes
-10 to -20	Frostbite in one hour
> -10	Lessened danger of frostbite

- Refer to the same data set as for Problem 1. Note that the times for the two cities vary slightly. How could you convert these values to a common variable hour?
- Great Egret nestlings (a species of large white wading bird) were captured and tagged on two different islands in Lake Huron during the summer of 2011. A large circular wing patch was attached to each wing, with a 3-character ID that can be seen with a spotting scope, and weight, gender and location were recorded (Egret Database worksheet).
  - Update the database with the Egret Sighting worksheet using UPDATE and BY rather than an INDEX without necessarily trying to resolve multiple/ambiguous sightings. How did this approach handle any problematic features with the sighting data set?
  - Use a Full Join in PROC SQL to combine the two data sets. Clean up any problems with labeling, etc. in the output table.
- The tab-delimited text files **April09.txt**, **May09.txt**, and **June09.txt** contain fields Record, Station, Station Description, Collection Date and Fecal Coliform from the EColi data set. Use INFILE and INPUT statement (Chapter 14) to successively add each of the data sets in the file into a new SAS data set called FC.Q2. Do not modify the data sets in order to simplify input.
- Consider typical code we learned for sampling with replacement:

```
data subset (drop=i samplesize);  
samplesize=20;
```

```
do i=1 to samplesize;
  obsnum=ceil(ranuni(0)*totobs);
  set original point=obsnum nobs=totobs;
  output;
end;
stop;
run;
```

- (a) Create a macro that allows you to specify the proportion of the data set to be sampled and placed in SUBSET.
- (b) Test your macro by taking a 1% subsample ( $p=.01$ ) of the EColi data set. How many observations and which observations were sampled? Double-check your macro by taking a 5% subsample.
- (c) Graduate students should add a second statement that specifies the library in which to place SUBSET. Demonstrate the macro by placing the sample in STAT541.SUBSET (STAT541 can be created in a directory of your choice).