HW 1

1. Let Ho: μ = 3 Ha: μ≠3 and let p-value = .0004. Define a p-value and interpret it in this problem. How would you explain what this p-value means to a client with little knowledge of statistics?
2. Speed of Evolution. How fast can evolution occur in nature? Are evolutionary trajectories predictable or idiosyncratic? To answer these questions R. B. Huey et al. ("Rapid Evolution of a Geographic Cline in Size in an Introduced Fly," Science 287 (2000): 308–9) studied the development of a fly—Drosophila subobscura—that had accidentally been introduced from the Old World into North America (NA) around 1980. In Europe (EU), characteristics of the flies' wings follow a "cline"—a steady change with latitude. One decade after introduction, the NA population had spread throughout the continent, but no such cline could be found. After two decades, Huey and his team collected flies from 11 locations in western NA and native flies from 10 locations in EU at latitudes ranging from 35–55 degrees N. They maintained all samples in uniform conditions through several generations to isolate genetic differences from environmental differences.
   (a) Construct a scatterplot of average wing size against latitude. Using different colors and/or plotting characters, indicate on your plot the continent and gender of the fly. What does this plot indicate to you about the relationship between these variables?
   (b) Is wing size related to continent after taking into account gender and latitude?
      a. The fully interactive model takes the pairwise products of all three of the explanatory variables plus the product of all three explanatory variables:

```
PROC GLM DATA = flyWings PLOTS = ALL;
    CLASS continent (ref = "EU") gender (ref = "F");
    MODEL wing = latitude | gender | continent / SOLUTION CLPARM;
RUN;
```

         i. Write down the general model (using the usual notation/parameters) that would match the MODEL statement in this code (for example, use an indicator variable called "Female" that equals 1 if "F" and 0 if "M").
         ii. What would be the form of $\mu\{Y|X\}$ for a fly at a latitude=lat, gender = M, and continent = EU?
         iii. What would be the form of $\mu\{Y|X\}$ for a fly at a latitude=lat, gender = F, and continent = EU?
         iv. Write down estimated model (using the usual notation/parameters and rounding to the nearest 0.1).
         v. What would be the form of $\hat{\mu}\{Y|X\}$ for a fly at a latitude=lat, gender = M, and continent = EU?
         vi. What would be the form of $\hat{\mu}\{Y|X\}$ for a fly at a latitude=lat, gender = F, and continent = EU?
      b. Test for whether this fully interactive model should include the term latitude*gender*continent
      c. Test for whether there should be any interactions at all relative to the "main effects" model:

```
PROC GLM DATA = flyWings PLOTS = ALL;
    CLASS continent (ref = "EU") gender (ref = "F");
    MODEL wing = latitude gender continent / SOLUTION CLPARM;
RUN;
```

d. In the “main effects” model, interpret the estimated coefficient for latitude (do this regardless of whether you find evidence that this model is sufficient for explaining this data)
e. Once you have settled on a model, answer the above scientific question. A complete answer would examine whether there is evidence that continent would interact with latitude or if gender interacts with continent. Be thorough in your answer. You need to verify the assumptions are met and then use your estimated model to answer the question, including a statistical conclusion and scope of inference (see last semester’s lecture “_scopeOfInferenceStatisticalConclusionDarren” for examples).

(Note: when including a single numeric explanatory variable (latitude) and then categorical explanatory variables (gender and continent), the multiple regression is sometimes called an ANCOVA. As it is a straight-forward special case of multiple regression, we won’t address the topic explicitly)