HW 1: Solution

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?

**A p-value is the probability of observing an observation as or more extreme than what was observed given that the null hypothesis is true and the assumptions are correct.**

**In this example, if µ = 3 (and if the assumptions of the t-test are true (normality, etc.)) then the probability of observing something as extreme or more extreme (in either direction) than what was observed is only .04. For my purposes, this is smaller than my threshold 0.01 and hence we will reject Ho and conclude that the evidence supports the alternative that the true mean is not equal to 3.**

1. Speed of Evolution. How fast can evolution occur in nature? Are evolutionary trajectories

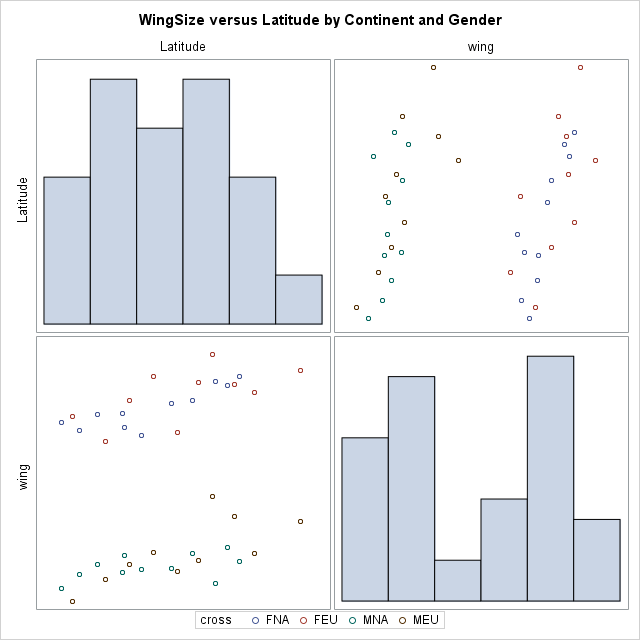
predictable or idiosyncratic? To answer these questions R. B. Huey et al. (“Rapid Evolution of a

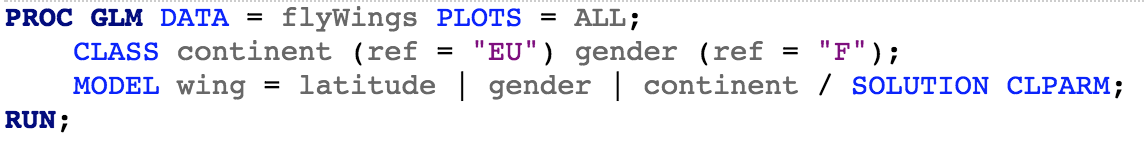
Geographic Cline in Size in an Introduced Fly,” Science 287 (2000): 308–9) studied the develop-

ment 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.

1. 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?

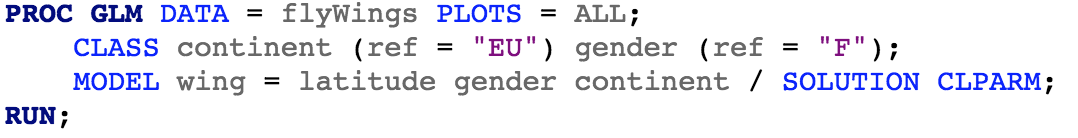
**For flies in EU, wing size appear to be positively associated with latitude for both males and females. For flies in NA, there is not an obvious linear association between wing size and latitude.**



1. Is wing size related to continent after taking into account gender and latitude?
   1. The fully interactive model takes the pairwise products of all three of the explanatory variables plus the product of all three explanatory variables:
2. 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 “Male” that equals 0 if “F” and 1 if “M”).
3. What would be the form of for a fly at a latitude=lat, gender = M, and continent = EU?
4. What would be the form of for a fly at a latitude=lat, gender = F, and continent = EU?
5. Write down estimated model (using the usual notation/parameters and rounding to the nearest 0.1).
6. What would be the form of for a fly at a latitude=lat, gender = M, and continent = EU?
7. What would be the form of for a fly at a latitude=lat, gender = F, and continent = EU?
   1. Test for whether this fully interactive model should include the term latitude\*gender\*continent

at significance level 0.05.

P-value is 0.1117 > 0.05, so fail to reject . Under the fully interactive model specification, there is little evidence suggesting the parameter for three-way interaction term is nonzero. Therefore, the term latitude\*gender\*continent should be removed.

* 1. Test for whether there should be any interactions at all relative to the “main effects” model:

Using extra sum of square F-test (subscripts R for “reduced model”, F for “full model”),

p-value = 0.60. Do not reject H0.

There is little evidence that any interaction term is significant. A main effect model would be sufficient.

* 1. 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)

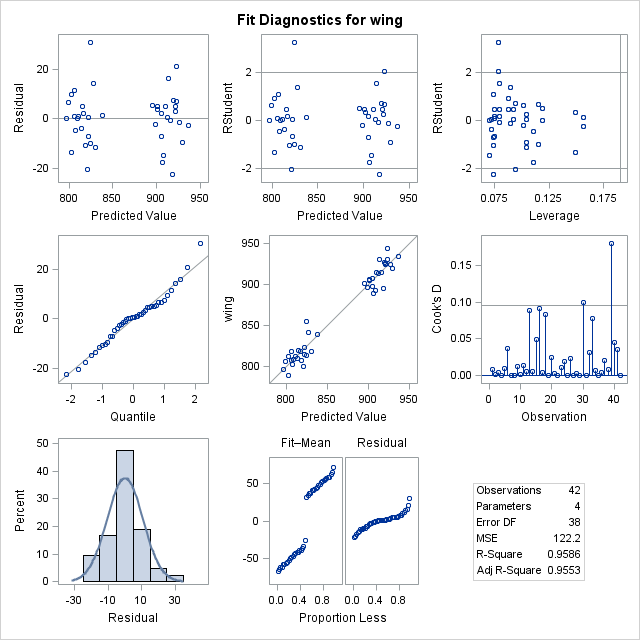
: For fixed levels of gender and continent, each one degree increase in latitude is associated with 1.8 units increase in the fly’s wing size on average.

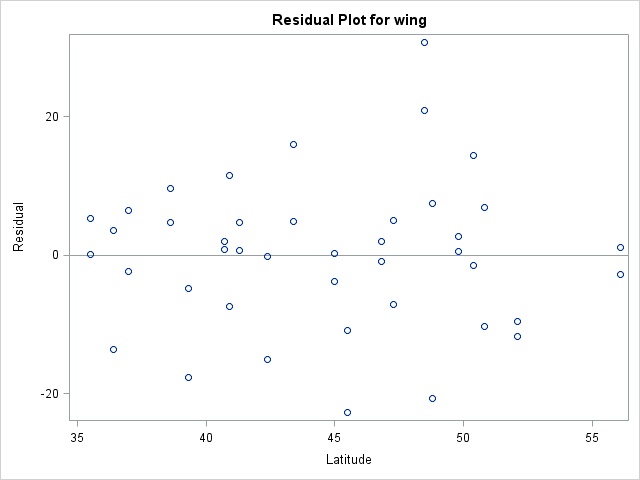
* 1. Once you have settled on a model, answer the above scientific question. A complete answer would examine whether there is evidence that continent would interact with latitude or if gender interacts with continent. Be thorough in your answer. You need to verify the assumptions are met and then use your estimated model to answer the question, including a statistical conclusion and scope of inference (see last semester’s lecture “\_scopeOfInferenceStatisticalConclusionDarren” for examples).

As suggested by the F-test in part c, there is little to no evidence that continent would interact with latitude or that gender interacts with continent. The model selected only contains main effects. Continents and gender are kept in the model, because they are of research interest.

**Checking Assumptions**

The residual vs. predicted value plot shows residuals scatter about 0 with no obvious pattern. There are no observations with large residual and high leverage. Residuals appear to be approximately normally distributed based on qq-plot.





**Conclusion**

The research question of interest is whether there is a detectable “Cline effect” in NA like there appears to be in EU. There is no evidence for an interaction term between continent and latitude, so if there is an effect, we have evidence that it would be the same for both continents (F-test p-value = 0.6). Same discussion goes for gender. For the fly, Drosophila subobscura, the estimated effect for a 1 degree difference in latitude is that it is associated with a 1.8 units increase in wing size for fixed gender and continent. At 95% confidence level, the positive increment is between 1.15 and 2.43. There is no evidence suggesting cline effect depends on continent.

**Scope**

This is an observational study, because continents and latitude that the flies live in are not randomly assigned. Therefore, only association between continent and wing size can be concluded, not causation.

Sampling method does not mention random selection of flies, so the conclusion only applies to flies in selected localities, not the entire continents of interest. However, there seems to be some attempt to do a “stratified sample” by selecting flies at specific locations on the continents.

(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)