STA 250F — Assignment #2 — Due in lecture on December 8 at 11:10am

Late assignments will be accepted only with a valid medical or other excuse.
This assignment is to be done by each student individually.

In this assignment you will look at the same data as you looked at for the first assignment. See the first assignment for how to get your data, and the general description of the problem. Note that you **must** analyse your own data, not someone else's. The conclusions you come to may be different from those of some other student, since your data, and also the underlying situation, will be different from theirs.

Your first assignment involved exploratory data analysis, to find what relationships seem to exist. In this assignment, you will do formal hypothesis tests and construct confidence intervals. This will help in deciding what relationships that you saw are most likely real, and which might just be due to chance. However, you should discuss whether the assumptions behind these tests and confidence interval procedures are close enough to being true that you can trust the results.

As in the first assignment, your aim should be to analyse the data as if this were a real problem. You should do whatever is necessary (within the limits of what we've covered in the course so far) in order to get good answers to whatever questions would be of interest if this were a real problem. Certain specific hypothesis tests and other tasks are mentioned below, which you should certainly do, but you may think of other things to do as well.

Cleaning up the data

In the first assignment, you were supposed to look for any unusual data values that might be erroneous, and replace these values with "*", indicating that the correct value is unknown. Doing this will be essential for this assignment, and you should begin by making sure that you have done it correctly.

To make sure everyone can do this, a list of the actual errors that exist in each data set is available on the course web page. These are the errors that were deliberately put into your data by my simulation program. The list gives the number of errors involving food, involving weight, and involving lifespan. A "food" error is where food-amount is greater than food-allowed, which shouldn't be possible. A "weight" error is where weight200 is much less than weight100, so much less that it can't be right, or if it is right, indicates a really strange mouse. A "lifespan" error is where the lifespan recorded is less than 100 days, which isn't possible because the experiment started with mice that had already lived for 100 days.

For the first assignment, some people decided that some other data points also were errors, or were due to extraordinary events that should be ignored. This was not necessarily wrong (though it sometimes was), but for this assignment, you should change only the data points on the list of deliberate errors to "*". All the other data points should be left as they are in the original data (except that you can remove them temporarily to see what the effect is, if you think that is an interesting thing to look at). This is so that the marker will be able to compare the results of your tests to results that they have for your data set with just those erroneous points omitted.

Looking at the mice that were allowed exercise and lots of food

As for the first assignment, you should begin by looking at only the 20 mice who were allowed to exercise and to eat up to 10 grams of food per day. (The `Manip > Subset Worksheet` option is probably the most convenient way of isolating just these 20 mice, and is useful for other purposes as well.)

For these 20 mice, you should first perform regressions of lifespan on food-amount and of lifespan on ex-amount, and discuss the meaning of the results. Your discussion should include an interpretation of the p-value for each regression, and of the apparent strength of the relationship, when the p-value is such that you think the relationship seen is not just due to chance.

You should then perform a regression of lifespan on **both** food-amount and ex-amount, and discuss the results. You should go on to perform whatever other regressions or other procedures you think will help you understand what is going on.

Looking at the full experimental data set

You should then investigate the full data set of 200 mice, starting with looking at how lifespan varied among the ten experimental groups.

You should create two worksheets using `Manip > Subset Worksheet`, one that contains just the 100 mice allowed exercise, and one that contans just the 100 not allowed exercise. For each of these worksheets, you should perform a one-way ANOVA test to see whether lifespan is related to food-allowed. You should also perform a regression of lifespan on food-allowed. You should discuss the results of these tests (for the mice allowed exercise, and for those not allowed exercise), and what they say about the effect of food on lifespan.

You should also perform a two-sample t test comparing lifespan in the group fed 4 grams of food and allowed exercise to lifespan in the group fed 4 grams of food and not allowed exercise. Discuss what the results of this test mean. Similarly, you should perform a two-sample t test comparing lifespan in the group fed 10 grams of food and allowed exercise to lifespan in the group fed 10 grams of food and not allowed exercise. (You needn't do t tests for the groups with other amounts of food allowed, unless you think that these tests would provide substantial additional information.) For these t tests, you should decide whether or not to use a pooled variance estimate (ie, assume the variances are equal), and you should justify your decision. You should also look at the confidence intervals for the differences in lifespan that you obtain along with these tests. You should discuss the meaning of your results in terms of how exercise affects lifespan.

Finally, you should do any other regressions, hypothesis tests, or confidence interval calculations that you think will help you to understand what is going on, and why.

Organization of your report

You should hand in a report that describes how you went about your analysis, and what your conclusions were. These conclusions should be supported by a moderate amount of relevant MINITAB output.

Your report should be organized into three sections, as follows:

1. Analysis of observations on mice allowed exercise and lots of food.
2. Analysis of the full experimental data set.
3. Conclusions.

You conclusions should include a comparison of the results from the full experimental data to what you might have concluded from just the observational data.