STA 250F — Assignment #2 — Due in **lecture**, December 10, 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.
There are three pages to this assignment.

This assignment is a continuation of assignment #1. Refer to the first assignment sheet for a description of the data and the problem. In that assignment, you looked at your data in an informal way, to see what effects seemed to be present. In this assignment, you will construct formal confidence intervals and do formal hypothesis tests. Formal tests will help in deciding whether effects that seem to be present are real, or might instead just be the result of chance variation. Confidence intervals will quantify the precision to which the magnitude of an effect has been estimated.

As was the case for the first assignment, you **must** analyse your own data, not someone else's. If you made mistakes in assignment #1, you should be sure that you now know what you should do, so that you don't make the same mistakes again. Talk to the instructor or a TA if you are in doubt about such things as which data points should be regarded as errors, and if so, what should be done about them.

The main question to be addressed in this assignment is how various factors affect the amount of weight gained by cattle (of the age and kind used in this experiment). The effect of the amount of grain fed is of particular interest, but other factors are of some interest as well, particularly if they interact with the amount of grain fed (eg, the effect of grain might be different for males and females).

The overall causal effect of feeding grain (averaged over other factors) could be found in an unbiasing way without considering these other factors, because the amount of grain was varied in a randomized experiment, which eliminates confounding effects (eg, a common cause for weight gain and amount of grain fed). However, no blocking was done in this experiment, so it is possible that these other factors may introduce a lot of variation in the results. By doing a multiple regression analysis that looks at several factors in addition to grain, it may be possible to improve the precision with which the effect of grain on weight can be estimated.

You should also investigate the secondary question of whether the amount of grain fed affects the health of the cattle.

You should hand in a description of how you went about your analysis, along with the Minitab output for your tests and regression analyses. For this assignment, you can write your report in statistical language — ie, you don't need to translate your conclusions to language that a farmer would understand.

Effect of grain on illness

You should start by considering whether or not the amount of grain fed has an effect on the health of the cattle, because this is the simpler part of the assignment. In particular, you should consider two questions:

1) Does feeding *any* grain to the cattle have an effect on their health, compared to feeding no grain?

2) Does feeding *lots* of grain to the cattle have an effect on their health, compared to feeding a smaller amount of grain?

Question (1) would be of interest if we suspected that there was something wrong with the grain — contamination with some poisonous substance, for instance. Question (2) would be of interest if we suspected that eating too much grain is bad, even though eating a smaller amount is fine.

To address Question (1), you should do a two-sample t-test of the hypothesis that the mean days of illness in cattle fed no grain is the same as the mean days of illness in cattle fed 0.5kg of grain. Discuss the results, and in particular comment on whether or not this test will be reliable.

To address Question (2), do a two-sample t-test to test the hypothesis that the mean days of illness in the cattle fed 0.5kg or 1.0kg of grain is the same as for the cattle fed 1.5kg or 2.0kg of grain. Again, comment on the reliability of this test, and discuss the meaning of whatever results you obtain.

For both questions, you will have to decide what to do if some of the cattle in question died. Remember that the number of days of illness recorded will be only up to the day of death. Unfortunately, nobody thought to record what day these animals died. Discuss why you did what you did to handle this, and whether or not it solved the problem.

Effect of grain on weight gain

In the first assignment, you should have done a least squares regression of the animals' final weight (ewt) on the amount of grain they were fed. You should also have done a regression of the change in weight (ewt-swt) on the amount of grain fed. You should now do these regressions again, looking not only at the estimates for the regression coefficients, but also at the standard errors for these estimates. Compute a 95% confidence interval for the effect of grain on weight for each of these regressions (separately). If there were no random variation, would you expect the two regression coefficients to be the same? What do the results say about which regression is a better way to estimate the magnitude of the effect of grain on weight?

When doing the regressions above (and also those below), be sure to eliminate any data values that you have decided are erroneous. Eliminate animals that died as well. (However, your final conclusions may need to take account of these deaths, if it seems that they weren't just exceptional occurrences.)

One might suspect, either from the beginning, or after looking at the data, that the effect of grain on weight is different for the males and for the females. Separate the data for the males and females, and do regressions of change in weight on amount of grain separately for each sex. Discuss the results.

A different way of seeing whether sex influences how weight gain depends on the amount of grain fed is by doing a multiple regression for the change in weight, with the following explanatory variables:

sex	The animal's sex: 0=male, 1=female
grain	The amount of grain fed
fgrain	For females: the amount of grain fed, for males: 0 (ie, this is the product of the 'grain' and 'sex' variables)

The regression coefficients for 'grain' will indicate how much effect grain has on weight gain for males — ie, how much more male animals will weigh, on average, if they are fed one more kilogram of grain — since 'fgrain' is zero for males, eliminating that term from the regression equation. The sum of the regression coefficients for 'grain' and 'fgrain' will indicate how much effect feeding grain has on weight gain in females – ie, how much more female animals will weigh, on average, if they are fed one more kilogram of grain — since 'fgrain' and 'grain' are the same

for females. The regression coefficient for 'sex' will be the difference in average change in weight for males and females that are fed no grain.

You should compare the regression coefficients found this way to those found with the two separate regressions. Is the result what you would expect? An advantage of the multiple regression approach is that you can test the null hypothesis that the regression coefficient for 'fgrain' is zero — ie, that the effect of feeding grain is the same for males and females. You should do this, and discuss the result.

Finally, do a multiple regression for the final weight ('ewt') in which you include as explanatory variables 'sex', 'grain', 'fgrain', 'swt', and 'age'. These are all the variables that the farmer would know about before the beginning of the feeding period. Interpret the results of this regression. Qualify your conclusions if necessary to take account of the fact that some animals died.

Minitab commands

Here are some of the Minitab commands you may need to use in order to analyse the data. Some of these you may have used for the first assignment. For details, see the Minitab Manual for the text, or the Minitab on-line help.

The `name` command is used to give names to variables. In the new Minitab, you can also do this by just clicking on the box for the name at the top of the column and then typing a name in that box.

You can set a value to be "*" to indicate it is unknown. This is the preferred way to deal with values that you decide are erroneous. You might also change the final weight of the dead cattle to "*", when you don't want these included in a regression.

You can compute new variables (ie, columns) from existing variables using the `let` command. For instance, you could create a new variable whose values are the differences of two existing variables.

The `unstack` command can be used to create new columns that contain data for just some of the animals. For instance, you could create four new columns, two of them containing the amounts of grain fed and final weights of the male animals, and another two containing the amounts of grain fed and final weights of the female animals. This would let you analyse the males and females separately. Similarly, you could separate the days of illness according to group.

To do two-sample t tests, you can use either the `twosample` command or the `twot` command, or their menu equivalents. The `twot` command, in which the data is all in one column, is probably more convenient. You need another column in which you say which items are in which group — eg, "0" for one group and "1" for the other group. You can set this column to "*" for items that shouldn't be in either group.

The `regress` command does single and multiple regressions. You can also do regressions using the menus in the new Minitab.