STA 250F — Assignment #1 — Due in **lecture**, October 21, 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.

For both this assignment and the next you will look at data for the problem described below. This data was produced by a fairly realistic computer simulation of the situation described. Each of you has *different* data, for a different situation. 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 aim in this assignment 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. To help you in this, certain specific questions are mentioned below, which you should certainly address, but other questions might occur to you as well. You might also consider all the various techniques we have covered in the course so far, and decide which of them might be useful for this problem.

You should hand in a description of how you went about your analysis, along with a moderate amount of Minitab output with plots and statistics that support your conclusions. Note that you can expect the data to show some variation due to chance. For this assignment, you will have to judge what is just chance informally, since we haven't covered how to do it formally yet. The final part of the write-up for your assignment should be a section entitled "Conclusions", in which you explain the results of your analysis in non-technical language.

The problem: Analysis of an experiment on feeding grain to cattle

An agricultural research station has performed an experiment to determine the effect of feeding grain to year-old cattle, as a supplement to the grass they eat from a pasture. Of particular interest is the effect of feeding grain on the weight of the cattle at the end of the experimental period, since the value of cattle when sold to a slaughterhouse is approximately proportional to their weight. On the other hand, feeding cattle grain costs more money than just letting them eat grass. The health of the cattle is also of interest, both because unhealthy cattle are likely to gain less weight, and because dead cattle are worth nothing, regardless of how much they weigh.

The experiment began at the start of summer and finished 100 days later, near the end of summer. It involved 40 animals — 20 males (steers) and 20 females (heifers). All animals were approximately one year old at the start of the experiment. The animals were all of the same breed, and all had been fed and otherwise treated in approximately the same way before the start of the experiment.

The experimenters randomly divided the animals into five groups of 8 animals each, which were fed different amounts of grain each day — 0 kg, 0.5 kg, 1.0 kg, 1.5 kg, or 2.0 kg. No attempt was made to to equalize the number of males and females in each group, or to make the groups uniform in other respects. This was just left to chance.

All the animals were put out to pasture during the day, where they had access to as much grass as they wished to eat. At the end of each day, the animals were placed in individual pens, where they were fed the allotted amount of grain, and also examined by a veterinarian to determine their health.

The animals were weighed at the start of the experiment and at the end of the experiment. The age of the animal at the start of the experiment was also recorded. For each animal, a

record was kept of how many days the animal was ill, as judged by the veterinarian, and of whether the animal lived until the end of the experiment.

The data format

If you are using the new Minitab, you can read in the data with the command

```
MTB> read '\\server\radford\adata\your-cquest-account' c1-c8
```

Or you can do it via menus with File > Other Files > Import Special Text. If you are using the old Minitab, you can read the data with the command

```
MTB> read '/u/radford/adata/your-cquest-account' c1-c8
```

If you are doing the assignment from home, you can download your data from the course web page. *Note: The data will not be available until October 4.*

The data file consists of 40 lines, one for each animal, with each line containing seven numbers, as follows:

1	sex	The animal's sex: 0=male, 1=female
2	age	The animal's age at the start of the experiment, in days
3	swt	The animal's weight at the start of the experiment, in kilograms
4	group	The group the animal was assigned to (1, 2, 3, 4, or 5)
5	grain	The amount of grain fed to the animal each day, in kilograms
6	ill	The number of days the animal was ill
7	alive	Whether the animal was alive at the end: 0=dead, 1=alive
8	ewt	The animal's weight at the end of the experiment, in kilograms

The amount of grain is determined by the group, but both are included for convenience. If the animal died during the experiment, the 'ill' variable is the number of days of illness before death (which could be zero, if the animal dies suddenly). For animals that died, their weight at the end of the experiment is recorded as zero.

The names for the variables above ('sex', 'age', 'swt', etc.) do not appear in the data file, but you should give the variables these names in Minitab.

Questions to address

The primary objective is to determine what effect feeding grain has on the final weight of the animals. The most obvious way to investigate this relationship is to make a scatterplot of these two variables, and to perform a regression of final weight against amount of grain. You should do this, and interpret the results. You may need to produce other plots to do this properly, such as a plot of residuals for the regression.

It may be that you can get a more precise picture of the effect of feeding grain by looking at the *change* in weight over the experiment — ie, by looking at the difference between the 'ewt' and the 'swt' variables. This may remove some of the random variation in final weights of the animals that is just the result of random variation in the starting weights of the animals. With this random variation removed, the relationship with the amount of grain may be clearer. Would looking at the change in weight rather than the actual final weight have the same significance as far as advice to the farmer goes? How might you be able to determine whether looking at the change in weight is more informative than just looking at the final weight?

You should also try to find out as much as you can about *why* the relationship of final weight to amount of grain is the way it is. This will require looking at other scatterplots, so that you can see what other relationships there are among the variables. For instance, you should look at how the number of days of illness relates to the amount of grain and to the final weight. You should also consider the possibility that the relationship is different for males and females, or for other sub-groups.

Finally, you should be on the lookout for anything else that might be of interest to the experimenters, and ultimately to the farmers for whose benefit the research is being done. It's possible that the data will show something interesting that wasn't anticipated beforehand.

Some potential problems

You should look for data points that are extreme, or that don't follow the relationship that the other points seem to follow. These might be data recording errors, or they might be due to exceptional events. You should try to determine whether such values are really erroneous, and if you decide that they are in error, you should omit that animal from any analyses where the error would have an effect. If you can't determine that an extreme value is the result of an error, then you shouldn't ignore it completely, but it might be advisable to see what the effect of removing that item is, and to qualify your conclusions if necessary.

If any of the animals died during the experiment, you will need to decide what to do when looking at the 'ewt' and 'ill' variables, which won't mean the same for the dead animals as for the living ones. You might decide to omit the dead animals from your analysis, or it might be better to include them. You will have to judge yourself what to do, and then state your conclusions in a way that is appropriate for what you did.

Minitab commands

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

You will need to use the graphing commands, such as `hist`, or their menu equivalents in order to check for outliers and such. Normal quantile plots might also be useful.

You will need to make scatterplots, and do regressions, with the `regress` command, or using the menus in the new Minitab. You may also want to look at two-way tables, with the `table` command, or the menu equivalent.

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 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 having with the amounts of grain fed and final weights of the male animals, and another two with the amounts of grain fed and final weights of the female animals. This would let you analyse the males and females separately.

You can delete rows of certain columns with the `delete` command, or by editing the rows directly in the new Minitab. As an alternative, you can set a value to be "*" to indicate it is unknown.