Homework # 1—Due Mon 6 Feb

1. Briefly, neatly and concisely (1/2 to 1 legible page) describe a particular designed experiment that you might conduct. Be specific about what questions you are asking and how you would conduct the experiment, with attention to components discussed in class. Keep it simple; you may consider an experiment with only one factor.
(a) Identify in your own words:

key questions
experimental units
design structure
treatment structure
method of randomization
method of replication

(b) State the four assumptions important for analysis (it may help to write down a model).
(c) Examine potential problems with assumptions that may arise during the course of your experiment.

2. The data for this problem are abstracted from a research project by Dr. Michel Wattiaux, a graduate of the Department of Dairy Science. The experiment involved 60 cows initially, although some were lost during the study. The cows were randomly assigned to one of 6 diets and followed for a number of weeks. Diets were begun after the 3rd week, allowing the animals some initial time to adjust to their new environment. Interest focuses on the effect of diet on the average Dry Matter Intake (DMI), the amount of food eaten by each cow. The data you have are a baseline value (`covar` – DMI for week 3), average DMI during subsequent weeks (`dmi`), the number of subsequent weeks (`wt`). In addition there are cow and block identifiers.

For now, we treat this experiment as if it were a **completely randomized design**. In fact, the cows were blocked by time, the first 6 cows randomly assigned among the 6 diets, and so on. In addition, "proper" analysis should take account of the initial capacity of each animal (its "covariate" of DMI at 3 weeks, `covar`) and should be weighted by the number of weeks the cow was on trial.

The data can be copied from the course area:

```
% cp ~st850-1/data/hwk1.dat .
```

There are some SAS suggestions for this problem in `hwk1.sas` for your information. The command (on `atlas`)

```
% sas hwk1.sas
```

produces the files `hwk1.log` and `hwk1.lst`, while the command

```
% saspr hwk1.lst > hwk1.prt
```

produces a more condensed output `hwk1.prt`, which could then be further edited. The first two procedures, `proc anova` and `proc glm`, consider the experiment as a completely randomized design. Interpret these for this assignment. Notice that `proc glm` is a more general procedure which we will use throughout the semester. There are copies of `hwk1.log, hwk1.lst` and `hwk1.prt` in the course data directory if you are interested.

The second `proc glm` includes blocking, covariate and weighting, as well as estimates of specific contrasts of interest. You may want to study this on your own.

There are also ideas for S in `hwk1.s`, with accompanying data in `hwk1.sdat`. [The reason there are two copies of the data is that SAS and S record missing data differently: SAS uses a period "." while S uses "`NA`".] Copy the two files and then perform the commands one by one. You will want to learn about the S commands `sink()` and `postscript()` for saving output in files for later viewing or printing; see file `hwk1.sink`, which was created by `sink("hwk1.sink")` right after entering S. There is also a UNIX command `script` which saves all printing into a file `typescript` – but be sure to `exit` from it! (See `hwk1.script` for an example.)

You may use a pencil and paper, or any package you like for this problem. However, it is very important that you annotate any package output and include it as **part** of your "report". Keep it neat and concise.
(a) Plot a histogram (or stemleaf diagram) of all the data. Comment on the pattern.
(b) Find means and standard deviations for each diet group.
(c) Estimate a common SD. Does this seem appropriate for this problem?
(d) Along a line next to the histogram of (a), identify the means for the treatment groups. Add a bar for the estimate of common SD to indicate precision.
(e) Produce a scatter plot of group mean against individual values, identifying groups by symbols or with some annotation.
(f) Comment on the comparison of group means, with reference to the contrasts specified by the scientist (see the second `proc glm`).