

STAT 530 - Fall 2003 - Homework 2

Due: Monday, September 22nd

1) When looking at a data set with a large number of variables there are a variety of descriptive methods we could use, including:

- Chernoff Faces or Star Plots
- The estimated Equation from performing a Multiple Regression
- The Residual vs. Predicted plot from performing a Multiple Regression
- Scatter Plot Matrix (a.k.a. Draftsman's Display)
- Side-by-side Boxplots

Consider the following four settings. Choose one method that would be appropriate **and** briefly say how you would use it. You may use the same technique for more than one question, you do not need to use all four techniques listed above, and you can choose a descriptive method not listed above if desired.

- a) We want to know how each of the variable relate to each of the other variables.
- b) We want to know how a single variable relates to the other variables.
- c) We want to examine the observations to find a pattern to them or to find clusters of observations.
- d) We want to search for an observation that was "unusual".

2) The web page <http://www.stat.sc.edu/~habing/courses/data/bears.txt> contains a subset of a data set described in *Reader's Digest* (April, 1979) and *Sports Afield*, (September, 1981).

The data set consists of several measurements for bears that were captured, measured, and released. (The full data set actually caught several of the bears multiple times over a period of years.) The variables in the data set are:

- | | | |
|------------------------------|--------------------------------|---------------------------------|
| 1. estimated age in months | 4. width of head in inches | 7. girth of the chest in inches |
| 2. gender (1=male, 2=female) | 5. girth of the neck in inches | 8. weight in pounds |
| 3. length of head in inches | 6. body length in inches | 9. name |

The observations are currently ordered by name.

a) Examine each of the eight Chernoff-like face methods developed in class by plotting out the faces for the first sixteen bears on variables 3-7. Comment briefly on each of the face methods (e.g. "Method qwerty: The size of the face and eyes make those variables easy to use, but the thickness of the eyebrows isn't very helpful.") and select the method you think is best overall. You don't need to print anything out for this one.

It is desired to do an exploratory analysis of this population of bears using a faces plot on variables 3-7.

- b) What variable would it make more sense to order the bears by than name? Why?
- c) Construct the display of all the bears on variables 3-7 using the method chosen in a and the order chosen in b.
- d) Using the faces plot in c, choose three groups of at least three bears each that seem to be similar to the others in there group but distinct from those in other groups. Briefly describe in terms of the variables, not the faces, what distinguishes these groups from each other.

3) Test whether the population means of male bears and female bears are different from each other on variables 3-7 by using two-sample t-tests.

- a) Report your conclusion at an overall α level of 0.05.
- b) Do the assumptions for performing these tests seem to be met? Why or why not?