# Assignment 3 — Due September 26, 2003

1. This problem is intended as an exercise in using the R language; thus, you will need to have access to a computer. Please refer to the handout "R Software Introduction for Stat 571". A data file contains 83 observations on a pair of nesting cranes. An observation is made when the nest (containing 2 eggs) was not covered by the body of one of the parents. The length of time that the eggs were left uncovered (time off eggs) was recorded along with some other information. (An indicator of time of day and the air temperature are included in this data file.) The eggs required about 30 days to hatch. The data given here come from the middle third of this 30-day interval. One objective of the study was to determine the distribution of the times off eggs data. — Your job is to use R to perform some exploratory data analysis of the time off eggs values. In particular, you are to conclude if the data are approximately bell-shaped and, if not, whether some transformation of the data can make them approximately bell-shaped. In the scale in which the data are most bell-shaped, you are to find some summary statistics.

   The data are stored in the file `crane.dat`, which is available on the course website. The times off eggs are in column 1, the time of day indicator in column 2, and the temperature in column 3; the units for times off eggs are minutes. You can look at the data directly on the Web to see what they look like. In order to read the data into R, you will need to create a file containing these data. There are several ways to do this; the easiest might be to select `Save As` from the `File` menu and save as a text file (e.g. `crane.txt`). The file can be saved in several places; I might suggest the desktop.

   Now access R as described in the handout on R. Then after the prompt, type:

   ```
   crane = read.table(file.choose())
   ```

   This allows you to choose the file using the WINDOWS menu system. Select your file and click `Open`. This creates an object named `crane` (since that appears to the left of the '=' in the above command) that contains the data from the file.

   After you have read in the data into your object, `crane`, it is wise to look at the data to make sure that they contain what you expect. This is accomplished by typing `crane` after the prompt. Note that the times of eggs data are in the first column. Although it is not essential, it probably makes sense to create a new variable composed of only the first column of the data in order to simplify certain manipulations. We can create a new variable, `toe`, composed of the times off eggs data in column one by typing:

   ```
   toe = crane$V1
   ```

   This takes the data in the first column of `crane (V1)` and places them in the new variable `toe`. Again, it makes sense to examine the new variable. You can view a stem and leaf display of `toe` by typing `stem(toe)` after the prompt. You can also find various summary statistics, for example `mean(toe)` , `sd(toe)` , `summary(toe)` . You can easily transform the data. To find the square root of `toe` , and to store the result as a new variable, `sqtoe` , type `sqtoe = sqrt(toe)` after the prompt. You can now make displays of `sqtoe` . To find the log of toe, and to store the result as a new variable, `logtoe` , type `logtoe = log(toe)` .

   The goal of this problem is to determine which of the scales – the original data, the square root transform, or the log transform – results in data that are closest to bell-shaped. You are to hand in the stem-leaf display on the original scale plus the stem-leaf display (if you choose a different scale) and some key summary statistics for the chosen scale. Summarize in a concise paragraph what you have learned about the time off eggs data from your exploratory analysis; write this paragraph as if you were addressing biologists.

   (One of the major goals of this problem is for you to gain some familiarity with R. Feel free to experiment as much as possible.)

IF YOU ARE WORKING ON THIS IN THE CALS COMPUTER LAB, DO NOT HESITATE TO ASK THE STAFF IN THE COMPUTING AREA FOR HELP IF YOU ARE HAVING DIFFICULTIES IN USING R OR GAINING ACCESS TO THE DATA.

2. Horses suffering from a particular contagious disease require special isolation rooms for treatment. Each room has space for only one horse. It is known that 20% of all horses suffer from this particular disease.

   (a) A veterinary laboratory has exactly one special isolation room. Suppose 6 randomly selected horses are brought to the veterinary laboratory. Find the probability that, of the 6 horses brought in, exactly 1 of them has the disease and will need to be put into the isolation room.

   (b) A second laboratory has 2 isolation rooms. Suppose that 8 randomly selected horses are brought into the laboratory. Find the probability that all the diseased horses can be housed in the available special isolation rooms.

3. This problem is designed to show the effect of assumption violations for a "binomial-type" situation. Consider flipping a coin. For the first flip there is a probability .5 of heads and .5 of tails. The probabilities for each flip after the first depend on the result of the immediately preceding flip as now described. If the immediately preceding flip results in heads, the probability of heads on the coming flip is .6 (prob of tails is .4). If the immediately preceding flip results in tails, the probability of tails on the coming flip is .6 (prob of heads is .4). The coin is flipped a total of 3 times.

   (a) Let $X = \#$ of heads. Find the probability distribution of $X$. (Hint — Construct the sample space and find the probability for each elementary outcome. You may find it useful to make use of calculations like the following: $P(\text{H on first flip AND H on second flip}) = P(\text{H on first flip}) \times P(\text{H on second flip}|\text{H on first flip})$. )

   (b) Find the probability that the first flip results in heads. Find the probability that the second flip results in heads. Find the probability that the third flip results in heads. (Hint — Again start with the sample space.)

   (c) Repeat (a) and (b) using a probability of .9 for heads following heads and for tails following tails instead of .6 (and .1 for tails following heads and for heads following tails instead of .4).

   (d) Summarize your conclusions for this problem. Which, if any, of the 3 binomial assumptions are violated? What is the effect?

4. Let $Z \sim N(0, 1)$ (i.e. $Z$ is a standard normal random variable.)

   (a) Find the following probabilities using tables.

      i. $P(Z \leq 0.65)$
      ii. $P(Z \geq -1.32)$
      iii. $P(-1.25 \leq Z \leq 0.58)$
      iv. $P(1.4 \leq Z \leq 2.4)$
      v. $P(-1.5 \leq Z \leq -1.2)$

   (b) With the aid of R, find the following probabilities.

      i. $P(-0.639 \leq Z)$
      ii. $P(0.427 \leq Z \leq 1.295)$

5. (a) In the following table assume $X$ is normal with mean $\mu$ and standard deviation $\sigma$. With the aid of tables, fill in the missing values treating each row as a separate problem.

| | $\mu$ | $\sigma$ | $a$ | $P(X \leq a)$ | $b$ | $P(X \geq b)$ |
|---|---|---|---|---|---|---|
| (i) | 0 | 1 | ?? | .30 | ?? | .80 |
| (ii) | 10 | 6 | ?? | .75 | ?? | .15 |
| (iii) | -3 | 2 | 0 | ?? | ?? | .010 |
| (iv) | 33 | 2.5 | ?? | .15 | 31.81 | ?? |
| (v) | -11 | ?? | -5 | .9 | -17 | .9 |

   (b) Using R, find $P(X \geq -5.7)$ if $X \sim N(-4, 10)$. (Hint: The second number in $N(-4, 10)$ is the variance, $\sigma^2$, not the standard deviation, $\sigma$.)

   (c) Using R, find $P(-2.0 \leq X \leq 8.6)$ if $X \sim N(6, 36)$.

   (d) Using R, find $x_*$ so that $P(X \geq x_*) = 0.70$ if $X \sim N(104, 1.44)$.

6. The total nitrogen concentration of the blood plasma of healthy 40 day old albino rats is known to be approximately normally distributed with mean 0.80 g/100ml and variance 0.0040 (g/100ml)$^2$.

   (a) If a healthy 40 day old albino rat is chosen at random, what is the probability that the total nitrogen concentration of its blood plasma will be between 0.65 g/100ml and 0.90 g/100ml?

   (b) What is the value of nitrogen concentration such that 60% of 40 day old albino rats have a nitrogen concentration higher than that value?

   (c) Find symmetric limits around the mean such that 95% of all 40 day old albino rats will have nitrogen concentrations between those limits.

Readings:

Week 3:

- Course Notes: Chapter 4