STA 290 STATISTICAL LABORATORY **HOMEWORK 1**

Fall Semester 2002

Solutions should be written up using Latex. Include graphs only where required or (sparingly) where they will help with your discussion of solutions. You may include graphs in text or as an appendix.

Exercise (1)

(a) In the binomial sampling model, for $n = 100$ and for **each** of $\theta = 0.2$, 0.4, and 0.9:

(i) Graph the sampling density $p(t|\theta)$, where $t = y/n$,

(ii) Draw a random sample of size $K = 5,000$ from the binomial, and explore histograms of the resulting values of $t = y/n$. Try different histogram bin-widths to see if they make a difference.

(b) Briefly describe the ways in which the distribution $p(t|\theta)$ changes its shape and spread for different values of θ.

(c) For $n = 10$, 100, 1000, and 10000, use S/R to compute the (theoretical) 0.005 and 0.995 quantiles of $p(t|\theta)$ when $\theta = 0.4$ and graph these against n. Interpret this graph.

Exercise (2)

Show that if $X \sim Bin(n, \theta)$, then $Var(X/n) = \theta(1 - \theta)/n$. (This should be straightforward — derive the variance of a Bernoulli and note that a binomial is the sum of independent Bernoullis. The point of this problem is to practice typesetting math in $\LaTeX$.)

Exercise (3)

For the 1992 and 1993 VA datasets:

(a) Create a vector of the difference in failure rate between 1993 and 1992 (i.e., `f93-f92`). Compute summary statistics for the differences.

(b) Plot the differences by hospital against total number of patients in 1992. Comment on any interesting aspects of this plot and suggest explanations for them if possible.

(c) Plot the differences against the failure proportion in 1992 and explain why this plot looks this way.

(d) Assume the data are generated independently each year from a binomial with $p = 0.41$ and with n as observed for each hospital each year. Simulate a vector of differences by sampling from a random binomial for each hospital for each year. Compute summary statistics for the simulated differences and compare to the observed values from part (a). Do four more simulations (for a total of five) and compare the summary statistics. What have you learned from this simulation?

Exercise (4)

[From GCSR Ex 1.3; Lindley 1965.] Suppose that in each individual of a large population there is a pair of genes, each of which can be x or X that controls eye color. Those with xx have blue eyes, while heterozygotes (those with xX or Xx) and those with XX have brown eyes. The proportion of blue-eyed individuals is p^2 and the proportion of heterozygotes is $2p(1 - p)$, where $0 < p < 1$. Each parent transmits one of its own genes to the child; if a parent is a heteorzygote, the probability that it transmits the gene of type X is $1/2$. Assuming random mating, show that among brown-eyed children of brown eyed parents the expected proportion of heterozygotes is $2p/(1+p)$. Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children all brown-eyed. Find the posterior probability tha Judy is a heterozygote.
