

Using R to Investigate the Binomial Distribution
Stat 341 - Fall 2008

This file contains R code to assist you in investigating the properties of the binomial distribution for certain values of n and p . In this investigation, we will use data simulated from particular binomial distributions. We will always look at four different values of n for a given value of p . To simulate the four data sets to study, the R commands are

```
sample1<- rbinom(10000,n1,p)
sample2<- rbinom(10000,n2,p)
sample3<- rbinom(10000,n3,p)
sample4<- rbinom(10000,n4,p)
```

(Hint: to make this easier to use in R, you should copy and paste the four lines above to a text window (ex. NotePad). Then you can change the parameter values and paste into R.)

Three important aspects to any data distribution are the histogram, the mean and the variance. In comparing the four histograms, it is helpful to put them on the same graph in R. The code for producing the four histograms in R is

```
par(mfrow = c(2,2))
hist(sample1, breaks = (min(sample1) - 1):max(sample1) + 0.5)
hist(sample2, breaks = (min(sample2) - 1):max(sample2) + 0.5)
hist(sample3, breaks = (min(sample3) - 1):max(sample3) + 0.5)
hist(sample4, breaks = (min(sample4) - 1):max(sample4) + 0.5)
```

(Note: In the code above, the breaks for each histogram are set according to the minimum and maximum observed values in the simulated data.)