9.07
Introduction to Statistical Methods
Homework 3

Name: _____

1. **Reaction time data.** Download the data files RTdiscrim.txt and RTweb.txt.
RTdiscrim.txt has data from a reaction time experiment in which observers were asked to
discriminate two words as quickly as they could. RTweb.txt has data from a task in
which observers did a web search to find side effects of the drug halcion. Plot histograms
of the two sets of data, and normal quantile plots for the two datasets (using qqplot).
Compare the shapes of the two distributions. Does one look closer to normal than the
other? Based on what you know about the central limit theorem, does this make sense?
Why or why not?

2. **More use of the z-table.** What is the z value corresponding to the 90th percentile point,
i.e. the z value such that 90% of the area under the normal curve lies between –infinity
and z? The 80th percentile point? If x is normal with mean 5 and standard deviation 2,
what x value corresponds to the 80th and 90th percentile points?

9.07
Introduction to Statistical Methods
Homework 3

Name: _____

3. **Plot your own normal-quantile plot.** Though we have access to qqplot, it's good to do at least one normal-quantile plot "by-hand" (you may use MATLAB, but not qqplot), to make sure you've got the idea. Here's the data:

0, 9, 2, 10, 8, 3, 5, 1, 7, 4

(The normal-quantile plot is not going to look like much with so few data points, but this is just for practice.) Here are the steps, from Lecture 5: Sort the data. For each data point, compute its corresponding percentile using the equation perc = (i-0.5)/N*100, where i=1 for the smallest data point, i=2 for the next smallest, and so on. For each percentile, look up in your z-tables the z value corresponding to the perc percentile, as in Problem 2. Plot the sorted data vs. corresponding z-values. That's your normal-quantile plot. (You might try qqplot just to confirm, but you don't need to hand it in.)

4. Do problem 7, p. 95.

5. Do problem 6, p. 261. What would the answer be if in addition to picking the right number of people for the committee, you also picked who was head of the committee. I.E. a committee with Beth as head and Ann as member is a different choice than a committee with Ann as head and Beth as member. Note that we are not asking you to compute the number of ways of picking the committees and their chairs, just whether there are more possible ways when the committees have 6 members than when they have 2 members.

9.07
Introduction to Statistical Methods
Homework 3

Name: _____

6. Do part of problem 11, p. 262-3. Do part (a), and part (d) (considering only the results of your work in part (a)). That is, pretend the parts of the question having to do with lung disease and coronary heart disease are not there.

7. Do problem 8, p. 265.

8. Do problem 9, p. 265.

9.07
Introduction to Statistical Methods
Homework 3

Name: _____

9. **The range over which the normal approximation to the binomial distribution is a good one.** In this problem, you'll be generating coin flip data and counting the number of heads, just as in the first HW assignment. We are interested in estimating, for a given value of p (the probability of seeing a head on one coin flip), how many coin flips we need for the normal distribution to be a good approximation to the binomial distribution. Use qqplot to judge how close the distribution of the number of heads is to a normal distribution (don't worry about the granularity of the data – just whether it falls on a straight line otherwise). You will just be "eyeballing" whether the qqplot says the data is approximately normal or not. Try to be consistent, but use whatever criterion you like. (Note that for binomial data, the line that qqplot draws is sometimes kind of bogus – nowhere near the data at all. You may want to draw your own line over the plot, to see if it looks like the data points fall on a line.)

(a) p=0.5. What is N, the number of coin flips, which makes the binomial distribution approximately normal?

(b) p=0.4. Now what N is required, for an approximately normal distribution.

In both cases, show a qqplot for the N you believe is required, and a qqplot for approximately N/2.

9.07
Introduction to Statistical Methods
Homework 3

Name: _____

10. **Percentiles for non-normal data.** For the data in RTdiscrim.txt, what is the 10th percentile point? What is the difference in reaction time between the 30th and 20th percentile points? Between the 90th and 80th percentile points?