A quick trip through the text book shows that during our work in chapters 1, 3 and 4 that we have discussed the following:

- types of variables: categorical, quantitative;

- center, spread, shape

- stemplot

- histogram

- boxplot

- density plot

- modes, symmetry, skew

- outliers (suspected outliers 1.5IQR)

- mean, median

- quartiles, IQR, standard deviation

- five-number summary

- effect of linear transformation

- normal: rules of thumb, z-scores, standard normal, inverse CDF, qqplot

- lurking variables, confounding variables,

- Causation; common response, confounding (Ch. 2, p. 174)

- types of experiments

- comparative experiment

- placebo effect

- bias: lack of realism

- randomization

- randomized comparative experiment

- double blind, completely randomized

- matched pairs design

- block design

- population; sample

- statistic; parameter

- samples, SRS, stratified, multistage

- bias: response bias, wording of questions, lack of realism, ...

- sampling variability (Margin of Error)

- sampling distriubtion: center, spread, shape

- sample space, outcome, event, random variable

- rules of probability

- equally likely probabilities

- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

- independence

- conditional probability: $P(A \mid B) = P(A \text{ and } B)/P(B)$

- distribution of a discrete random variable X: $P(X = k) = ...$

- mean of a discrete random variable. Expectation, $E(X) = \mu = \Sigma k P(X = k)$

- variance and standard deviation of a discrete random variable. $\sigma^2 = \Sigma(k - \mu)^2 P(X = k)$.

Here are some sample problems to work on during class. These are not meant to exhaust the full range of questions I may ask.

1. A 6-sided die has faces 1,1,2,3,5,8. Let X be the amount on one roll and Y the amount on 2 rolls or the die.

    (a) Write down the probability distribution for X

    (b) What is the expected value of X?

    (c) What is the expected value of Y?

    (d) The variance of X is **56/9**. What is the variance of Y?

2. A teacher has 4 items from the snack machine: a candy bar, a pack of gum, a bag of chips and a pack of licorice. She hands out 2 of them.

    (a) What is the probability the second one handed out is the pack of licorice?

    (b) What is the probability the last one handed out is the pack of licorice?

    (c) What is the probability the second one handed out is the pack of licorice, given that the first was a bag of chips?

3. A randomly chosen student has a 30% chance of being enrolled in a math class and a 35% chance of being enrolled in an English class. Suppose that the chance they are enrolled in both a math class *and* an English clash is only 10%. What is the probability that a randomly chosen student is enrolled in either a math class *or* an English class?

4. If one tosses a fair coin 5 times a sequence results, say x_1, x_2, x_3, x_4, x_5, where each x is either *H* or *T*.

   (a) Find the probability of a sequence that produces 3 heads and 2 tails? (Does it matter which order the heads or tails are?)

   (b) There are 10 different sequences that will produce 3 heads and 2 tails. Each of these outcomes is disjoint from the others. Use the last question to find the probability one will get 3 heads in 5 coin tosses.

   (c) Explain, without doing any calculations, why this should be the same probability as getting 2 heads in 5 coin tosses *if* the coin is fair.

5. Recalling the rules of thumb for a normal, if math SAT scores for CSI students are normally distributed with a mean of 500 and standard deviation of 100, find the probability a randomly chosen student has an SAT score exceeding 600? Between 500 and 700?

6. Imagine that when 100 dice are rolled, the sum has a normal distribution with a mean of 350 and standard deviation of 17.

   (a) Find the probability that in a given roll of 100 dice the sum is more than 370?

   (b) Find a range of values centered around 350 for which 90% of the rolls (on average) will be in this interval. (You need to find te 5th and 95th percentiles(

7. A data set consists of values 1,1,2,3,4,5,8,15.

   (a) What is n

   (b) What is $\bar{x}$?

   (c) What is s

   (d) What is the median?

   (e) What is the IQR

8. A boxplot of a data set is shown. Based on this identify

   (a) the median

   (b) The IQR

   (c) The max.

   (d) Will the mean or median be greater? Why

Benford's law?

sample(1:9, 100, replace = T, prob = p)
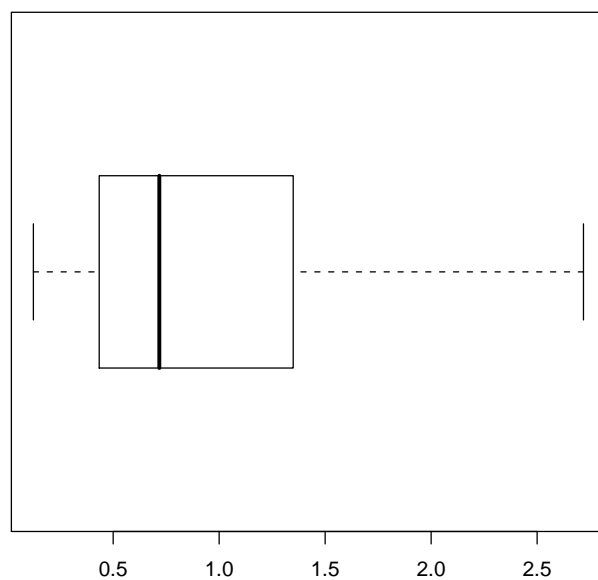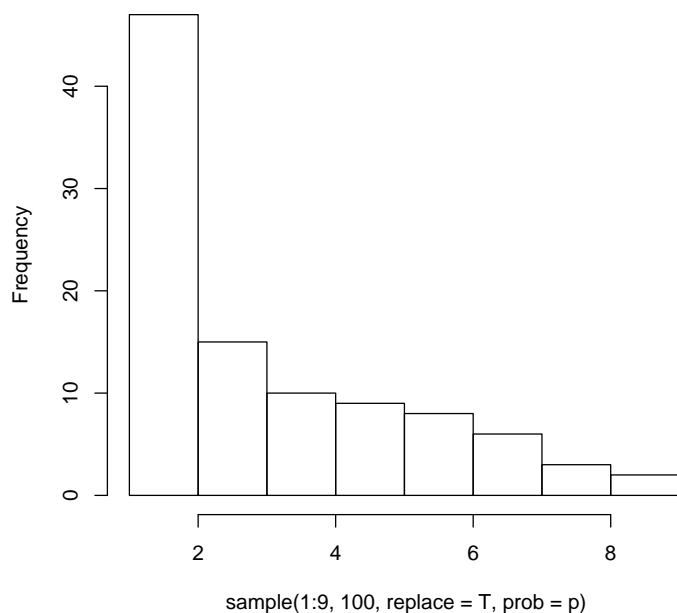
9.

A statistics teacher scrupolously records the first digit of each of receipts they receive. A histogrram is shown.

(a) Is this data set skewed? symmetric? neither?

(b) Is this data set unimodal, bimodal, multimodal?

(c) Estimate the mean value for this data set.

(d) there is something funny about this graphic in terms of representing the data. Can you tell what it is?

10. Which of these data sets appears to be normally distributed?