Ch 6 Summary

- Introduced the concept of the random variable.
- random variable is a number that records the outcome of some random occurrence.
- random variables have some properties
  1. there is a probability assigned to each value the variable takes
  2. each probability is between 0 and 1
  3. all probabilities sum to 1
- these properties make up the probability distribution for the random variable.
- there are two main types of r.v.
  - continuous : take on any value in a given interval e.g. (0,1)
  - discrete : take on separate and distinct values in an interval e.g. $1, 2, 3, \ldots, n$.
- two of the most common/important

Normal Distribution

- the normal distribution is the most widely used and important distribution in statistics.
- has some important properties :
  - symmetric and unimodal
  - characterized by two parameters (only), μ (the mean) and σ (the standard deviation).
- many types of real-world data have been shown to have the normal distributing.
- the normal distribution is the theoretical basis for the empirical rule which can give us a good feeling for the spread of data.
- we have also seen how to use a table of cumulative probabilities to find probabilities associated w/ the normal distribution.
- (perhaps go over the density function)
- tables used by relating a normal random variable to a standard normal random variable so that $N(\mu, \sigma) \rightarrow N(0, 1)$. This is done by calculating the z score using the parameter values from the normal random variable so that

$$z = \frac{x - \mu}{\sigma}$$

Binomial

- represents the sum of independent and identically distributed binary random variables. more specifically, the following three characteristics have to be satisfied

  1. trials are independent
  2. success for each trial is the same
  3. each trial has only two outcomes : success or failure

- related to these characteristics are the two parameters : n (number of trials) and p (probability of success).

- the probability mass function defines the probability for each value the binomial can take and is given by

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

- the binomial can be approximated with the normal distribution also (makes it easier to work with). in this case we have

$$\text{Binom}(n, p) \approx N(np, \sqrt{np(1-p)})$$

What good are probability distribution?

- we use these distributions to answer questions about populations from a sample of collected data.

- a population is assumed to have some distribution w/ parameters either known or unknown.

- we make an assumption that there is a distribution that the data comes from.

- we use statistical methods to answer questions about the population based on the sample.

- this is called inference : we make general conclusions about the population from the specific examples taken from the sample.

6.4

- sampling distribution : the distribution of a statistic

- ex : sample proportion (pizza)

  let $X \sim \text{bin}(n, 0.5) = \#$ of people who prefer pizza from A over D.

  1. what is the distribution of X/n?

- look at graphs of X and X/n.

- note the mean and standard deviation of X/n
- often we use X/n because it is easier to interpret
- ex : voters, assume that 50% voted for recall
    1. give the # voting for recall
    2. give the proportion voting for recall