Where to find material in Lock5 textbook:

- Lock5 Chapter 11: the entire chapter, but focus will be on Sections 11.3-11.4.
- Lock5 Section 3.1
- Lock5 Section 5.1
- Central limit theorem and related results: some at this link and some in Section 6.1 (when applied to $\hat{p}$)

Terminology of Data

- population vs. sample
- sampling
  - SRS (sampling without replacement) vs. i.i.d. (sampling with replacement)
  - sample statistics and their sampling distributions
    - When can they be well approximated using a normal distribution model?
    - What is meant by $SE_{\bar{x}}$? by $SE_{\hat{p}}$?

Probability (general)

- terminology: random process, outcomes, sample space, event, complement, independence, mutual excusivity (a.k.a. "disjointness")
- conditional probability, independent events
- rules
  - axioms
  - addition/multiplication rules
  - Bayes' rule
- (discrete) random variables (r.v.s)
  - the role of the probability (mass) function $f(x) = \Pr(X = x)$ and cumulative distribution function giving $\Pr(X \leq x)$
  - computing the mean of an r.v. X from the formula $\mu_X = \sum_x xf(x)$
  - computing the variance from the formula $\sigma_X^2 = \sum_x (x - \mu_X)^2 f(x)$
  - formulas: $\mu_{X+Y} = \mu_X + \mu_Y$ and $\mu_{X-Y} = \mu_X - \mu_Y$ when r.v.s are combined through addition/subtraction
  - formulas: $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ and $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$ when independent r.v.s are combined through addition/subtraction

Binomial Random Variables

- recognizing a binomial scenario, identifying parameters n and p
- mean and standard deviation, when r.v. $X \sim \mathsf{Binom}(n, p)$
- the probability function to go with $X \sim \mathsf{Binom}(n, p)$
  - how to read it from a table and understand what it provides
  - the command, `dbinom()`, in R that provides its values
  - how to graph it
  - how to simulate it
- role of the cdf, and the R command, `pbinom()`, used to evaluate it

Normal distributions

- Empirical (68-95-99.7%) Rule
- Z-scores (or standardized scores), and using the standard normal distribution
- normal distributions in play if you are taking the
  - sum of $n \geq 30$ iid variables from a population with mean μ, standard deviation σ: $\mathsf{Norm}(n\mu, \sigma\sqrt{n})$
  - sample mean of $n \geq 30$ iid variables from a population with mean μ, standard deviation σ: $\mathsf{Norm}(\mu, \sigma/\sqrt{n})$
  - sum of two independent variables from normal populations, $X \sim \mathsf{Norm}(\mu_1, \sigma_1)$ and $Y \sim \mathsf{Norm}(\mu_2, \sigma_2)$: $(X + Y) \sim \mathsf{Norm}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
  - difference of two independent variables from normal populations, $X \sim \mathsf{Norm}(\mu_1, \sigma_1)$ and $Y \sim \mathsf{Norm}(\mu_2, \sigma_2)$: $(X - Y) \sim \mathsf{Norm}(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
- R commands
  - Evaluating probabilities of normal variable $X \sim \mathsf{Norm}(\mu, \sigma)$ using `pnorm()`
  - Finding quantiles of a normal variable $X \sim \mathsf{Norm}(\mu, \sigma)$ using `qnorm()`