STAA552 Homework 1

Due Friday, October 28, 2016.

1. Read Sections 1.1, 1.2, 1.3, 1.4 (you can skip 1.4.4 if you like), 1.5.1–1.5.5 and 2.1 of Agresti.

2. This is a HUGE EDIT.

3. Complete exercises 1.2, 2.1, 2.2a–b, and 2.3 of Agresti.

4. One version of the *prosecutor's fallacy* occurs when the probability of some evidence given that a defendant is not guilty is presented as if it were the probability that the defendant is not guilty given the evidence. For example, let M be the event that DNA known to have been left by the perpetrator at a crime scene matches the DNA of the defendant. Let G denote the event that the defendant is guilty and $\bar{G}$ the event that the defendant is innocent. Suppose that $\mathrm{P}(M \mid \bar{G}) = 1 \times 10^{-6})$ (that is, the DNA could match by chance even if the defendant is not guilty). Because the defendant's DNA matches the crime scene DNA, the prosecutor claims that there is a one in a million chance that the defendant is not guilty. Suppose that the weight of all other evidence leads to $\mathrm{P}(G) = 1 \times 10^{-4}$ (for example, other evidence narrows the set of suspects to 10,000 individuals, one of whom is the defendant). Use the provided information to compute $\mathrm{P}(\bar{G} \mid M)$. Comment on your result and the prosecutor's claim.

5. Use the data of Example 1.5.4 in Agresti. Assume that the count of yellow seeds among the $n = 8023$ hybrid seeds in Mendel's experiment follows a Binomial(n, π) distribution.

   (a) Plot the log-likelihood function of the unknown parameter π given the data, for a fine grid of values in the interval $(0, 1)$.

   (b) Zoom in by plotting the log-likelihood function of the unknown parameter π given the data, for a fine grid of values in the interval $(0.65, 0.85)$. Large-sample normal approximations rely on the approximate quadratic shape of the log-likelihood in a neighborhood of the true value. Does this quadratic approximation appear plausible here?

(c) Add a vertical line to your plot to indicate the location of $\widehat{\pi}$, the maximum likelihood estimator of π.

(d) Use equation (1.11) of Agresti to compute the score statistic for testing the null hypothesis $H_0 : \pi = 0.75$ versus the alternative $H_a : \pi \neq 0.75$. Compute the p-value of your test statistic using the normal approximation, and compute the p-value of your *squared* test statistic using the χ^2_1 approximation. Interpret your results.

(e) Use equation (1.16) of Agresti to compute Pearson's chi-squared statistic X^2 for Mendel's data. Compute the p-value of the test statistic using the χ^2_1 approximation, and compare to the results of the previous problem (4d).

6. Complete exercise 1.10 of Agresti. Use five categories: 0 deaths, 1 death, 2 deaths, 3 deaths, or $\geq$ 4 deaths, and note that you must estimate one parameter (see Agresti §1.5.5).