Statistics done "by hand"

1. Prior to Chapter 6, there we had encountered few formulas. Chapter 6 changed that abruptly. What role do these formulas play? Were there any research questions we learned to address in Chapter 6 that we had been entirely unable to address beforehand?

2. Also prior to Chapter 6, the only theoretical distributions we had encountered were from the *normal* family of distributions. Since then, we have encountered the Student t-distribution, the chi-square distributions, and the F-distributions. Write three research questions, each of which might call upon one of these distributions.

   In the case of *normal* distributions (not one of the three families requested here of you), here is an example: "Is a certain coin *fair*?". For this question, it is natural to focus on the proportion p of flips which are "Heads", and we have used a *normal* approximation to the sampling distribution of $\widehat{p}$ to obtain an approximate standard error $\mathrm{SE}_{\widehat{p}}$, used both in constructing a confidence interval for p, and in determining the P-value in a test of hypotheses.

3. In the last question, you wrote three research questions. My intention was that one of them would be addressable using a t-distribution, one using a chi-square distribution, and one using an F-distribution. Nevertheless, even if you have written appropriate questions, it is still the case that, for some populations and data sets, use of the specified distribution isn't always appropriate. For each of your three questions, contrive a data set for which use of the given distribution could lead to unreliable results.

   Example: Piggy-backing on my normal-distribution example of the last problem, suppose my sample consisted only of $n = 14$ coin flips. Then, given my null hypothesis with *null value* $p_0 = 1/2$, we would have

   $$np_0 = (14)(1/2) = 7 \qquad \text{and} \qquad n(1-p_0) = (14)(1-1/2) = 7.$$

   We should generally not rely on a normal distribution to determine the P-value for the corresponding test of hypothesis if either one of these numbers is less than 10.

   Look at the contrived data sets from other group members and write a respond by describing how one might modify their data sets so as to make the targeted theoretical distribution family relevant again. Say, explicitly, which *member* from the distribution family (i.e., what number of *degrees of freedom*) would be relevant after your suggested modifications.

4. (Do this only after working through Module 8.) Write a skeleton ANOVA table for regression—that is, provide just enough numbers in your table that the missing ones can be found.

   Look at the ANOVA tables submitted by fellow group members. See if you can both fill in the missing numbers, then find and interpret the **coefficient of determination**.