Statistics done "by hand"

1. The main thing these formulas do for us is provide, when certain assumptions (or "rules of thumb") are in place, a formula for the approximate standard error. Standard error gets used both when building confidence intervals and when calculating an approximate P-value. Before Chapter 6, we used either bootstrapping or a randomization distribution to find an approximate standard error. These methods generally work quite well, but it took the invention of fast modern computers before it became practical carry out these methods from Chapters 3 and 4.

   That comes pretty close to summing up the entirety of Chapter 6: formulas for SE instead of the repetetive simulations of Chapters 3 and 4. We did not learn how to address any new research questions through material of Chapter 6 (unless you consider the idea of "paired data" in Section 6.13 new).

2. There are many correct answers one might pose here. Some examples:

   - **For a t-distribution** "Do students who study all seven days of the week have higher gpas, on average, than students who build a 'day of rest' into their study schedule?"
     This one involves a comparison of two groups. It is also possible to write one involving just one group, as in
     "What is the average gpa of a Calvin student?"
   - **For a χ^2-distribution** You might write a question that considers whether two categorical variables have an association (like in Section 7.2):
     "Are hair color and eye color associated?"
     Or, you might compare a sample of values for a single categorical variable against a supposed standard (like in Section 7.1): "When The College Board puts together its AP tests with multiple choice answers A, B, C, D, and E, does it make each of these options equally-likely (i.e., is the population proportion for each letter $p = 0.2$)?"
   - **For an F-distribution** Your research question could simply repeat the one I gave for two groups and a t-distribution above; that is, two-sample t tests of hypothesis can be carried out (arriving at the same P-value) using one-way ANOVA instead. However, ANOVA affords the ability of comparing *more* than two groups, as in my research question below: Is there a difference in average SAT score between students from the four regions we commonly call the East, South, Midwest and West?

3. For the question, "What is the average gpa of a Calvin student?", if you only have a sample of size $n = 10$ (something less than 30), that may call the reliability of results using a t-distribution into question. In such cases, it isn't only about sample size, but also the shape of the distribution of the underlying population. That is, if the distribution of gpas of Calvin students were normal already, then $n = 10$ would make the t-distribution results entirely reliable. (So, in fact, would $n = 2$.)

   For the question involving equal-likelihood of answers for SAT questions, use of a χ^2-distribution would not be entirely reliable if we sampled only a total of $n = 20$ questions.

For in that case, each of the expected counts would be $np = (20)(0.2) = 4$, less than 5.

For the question about SAT scores from four regions of the country, results from an F-distribution might be less reliable if my sample of SAT scores from the Midwest had a sample standard deviation of 47.2 but my sample of SAT scores from the South had a sample standard deviation of 12.3, thereby violating the 2:1 ratio rule of thumb.

4. Instead of providing an example here, I'll refer you to the app found at this website
`https://people.richland.edu/james/ictcm/2004/anovagen.php`
where you can try your hand at finishing a partial ANOVA table, even setting it to easy or harder levels of difficulty.