1. The .csv file `http://scofield.site/teaching/data/csv/aspirinAndStrokes.csv` contains raw data from a blind experiment to look at the effect of aspirin on instances of stroke. Here are some of the cases:

	progress	treatment
139	unfavorable	placebo
49	favorable	aspirin
106	favorable	placebo
9	favorable	aspirin
127	unfavorable	placebo
149	unfavorable	placebo
92	favorable	placebo
50	favorable	aspirin

   (a) Summarize the data in this file.
   (b) State a research question that may have prompted the collection of this data. (I suggest you access one appropriately-chosen app from the "Descriptive Statistics and Graphs" section of StatKey, import the data, and work from there to obtain the desired value.)
   (c) Write hypotheses ($\mathbf{H_0}$ and $\mathbf{H_a}$) appropriate for your stated research question.
   (d) With regard to the hypotheses you wrote, what is the *test statistic* (write it symbolically, and compute its value) produced from the data.
   (e) In StatKey, produce a randomization distribution appropriate for an hypothesis test from this data. Then compute the corresponding P-value, and state a conclusion at the $\alpha = 5\%$ level.
   (f) The original data had two variables, "progress" and "treatment". The Lock text describes how to produce a randomization statistic in two-variable settings, *without* using software:

      quantitative data for samples taken from 2 groups: p. 273, Example 4.32
      binary categorical data for samples taken from 2 groups: p. 268, Example 4.28
      bivariate quantitative data: p. 242, the first three paragraphs

      Tailor one of these descriptions to the data at hand, and give a thorough set of instructions (like the instructions you might read for setting up and playing a game like Uno, Risk, or Chess) for how you could produce a

      - randomization sample,
      - randomization statistic, and
      - randomization distribution.

   (g) What steps in your prior instructions would be different if you needed to produce a

      - bootstrap sample,
      - bootstrap statistic, and
      - bootstrap distribution?

2. Suppose you are interested in the variable *height* (measured in inches) for the population "current female students at Calvin". Suppose a sample of size n is taken from this population motivated by the question, "Is the average height of a female Calvin student above 68 inches?"

   (a) If $\overline{X}$ denotes the mean of a random sample, where would the given distribution be centered? (Answer using an appropriate name or symbol.)

   - the sampling distribution of $\overline{X}$
   - a bootstrap distribution
   - a randomization distribution

   (b) Tailor the description of a bootstrap sample in Example 3.20 (specifically, part (b) of that example) to the current setting, and describe how you would produce a single bootstrap statistic.

   (c) Assuming that you have already found

   - the mean height $\overline{x}$ for a *sample* of female Calvin students, and
   - a bootstrap distribution

   describe **two** different ways you might construct an 80% confidence interval for μ, the mean height of the population.

   (d) Tailor the description given on p. 270, Example 4.30 so as to describe how you would produce a single randomization statistic.

3. Let p represent the proportion of people who voted in the 2020 Presidential election who would, if given the opportunity to vote again today, cast that vote for a different candidate. If a 90% confidence interval for p were $(0.06, 0.31)$, then what could you say about the P-value associated with the hypotheses

   (a) $\mathbf{H}_0 : p = 0.25$ vs. $\mathbf{H}_a : p \neq 0.25$?
   (b) $\mathbf{H}_0 : p = 1/3$ vs. $\mathbf{H}_a : p \neq 1/3$?

4. True or False. Place a "T" by those statements that are true without reservation. Place an "F" by those statements which are not unequivocally true.

   (a) ______ A Type I error occurs when we reject a null hypothesis which is true.
   (b) ______ Using significance level $\alpha = 0.05$, the chance of committing a Type II error is 5%.
   (c) ______ The P-value one obtains in an hypothesis test is the probability that the null hypothesis is true.
   (d) ______ We reject the null hypothesis in precisely those instances where an effect is *significant*.
   (e) ______ When $P = 0.97$, we would reject the null hypothesis at the 5% level.
   (f) ______ If a P-value is significant at the 5% level, then it is also significant at the 1% level.

5. Why don't we ever construct 100% confidence intervals?