Homework Notes  Chapter 6

These notes do not include full answers. Students are expected to read the full answers in StatsPortal before looking at these notes. I am preparing these notes to provide more details about what you should be thinking about and learning from each of the exercises.

The multiple choice questions just before the Exercises at the end of the chapter are not discussed here because there is information in StatsPortal beside the correct answer to help students see why that answer is better than the others. (And also because it would take me twice as long to write these if I had to address all of those questions as well!)

Chapter 6: [6.1, 6.3, 6.5, 6.7, 6.9 – 6.18], 6.19, 6.21, 6.23, 6.25, 6.27, 6.29, 6.31

6.1, 6.3: I can’t think of anything to say that wasn’t said in StatsPortal.

6.5. While the solution in StatsPortal is quite complete, I find that students often have trouble deciding which distributions to compute to answer questions like this. This is an illustration of the discussion in the second half of the last paragraph before these exercises. Here the response variable is whether the person has arthritis later in life. That’s the response variable because the exercise says that this is a study of arthritis and we are investigating whether the types of experiences playing soccer affect whether the arthritis occurs. So, we need the conditional distribution of arthritis or not for each of the three types of soccer players.

6.7. As mentioned earlier in the chapter, whenever percentages are required, the first step is to determine what number goes in the denominator. When the problem says “Find the percentage of …. who meet the criterion of ….,” then the condition in the first blank tells you what goes in the denominator.

In part a. here, when it says “of all Alaska Airlines flights” then you know that the denominator is the total number of all Alaska Airlines flights. Even though the answer in StatsPortal doesn’t show exactly what numbers were divided to get the percentages, you should show that in your work for the problem.

In part b, you must compute five percentages for the Alaska Airline flights and another five percentages for the America West flights. The first one of these is the answer to the question “Of all Alaska Airlines flights to Los Angeles, what percentage were delayed?” So the denominator is the number of all Alaska Airline flights to Los Angeles, which is 497 + 62 = 559. The numerator is the number of those that are delayed, which is 62.

Part c. Some students need to see several examples of Simpson’s Paradox and think carefully about what is happening before this makes sense to them. Notice that the answer in StatsPortal points out that one of the airlines had a lot larger portion of the “easy” cases than the other airline, and that’s what causes the reversal of the comparison.

6.19, 6.21: I can’t think of anything to say that wasn’t said in StatsPortal.

6.23. This is an open-ended question and there can be different correct answers. However, if you identified a different lurking variable, read the answer in StatsPortal to see what the author thought about and see whether you agree that is a reasonable answer too. After you have thought about that, then think about whether that lurking variable and the lurking variable you identified are probably equally important, or whether one is more important than the other. This is an important part of thinking about open-ended questions in this class.
6.25. Remember the hint at the end of the section on conditional distributions, page 157, about how to choose which percentages to compute. What is the response variable here? (Answer: whether there was a relapse or not.) So then you want to compute the distribution of that response variable for each of the values of the other variable. Look in StatsPortal to see how to do that.

b. Think about what you’d have to do to believe this was evidence of actual causation. If these subjects were allowed to just decide which treatment to use, do you think that would make the study better or worse? Why? So what’s the crucial thing about making this a study which will give evidence for (or against) causation?

6.27. Again, as in exercise 6.25, think carefully about what is the response variable here in order to decide which conditional distributions to compute. Remember the hint on page 157. (see the comments about exercise 6.25 above.)

6.29. Again, as in exercise 6.25 and 6.27, think carefully about what is the response variable here in order to decide which conditional distributions to compute. This problem is different from the others, as they point out when they say this is not a two-way table, as we have been using. Do you see why? Answer: Those two-way tables have the totals on the margins, not as part of the table. Look carefully at the data in this problem, which are displayed in a similar manner, but not a two-way table.

You will need to change it to presenting the data in a two-way table. Do you see how to do that? Look at the data for the category “Cold.” Do you see that the total number of eggs observed in this category is 27? Look at the information in the table. What would be good labels for the two categories whose labels will be on the left of the table? Answer: Maybe “Hatched” and “Not hatched.” After you can see how to change it to a two-way table, then prepare your solution as before and compare it to the solution in StatsPortal.

6.31. Data given in this manner is sometimes called a three-way table. That’s because it shows the relationships between three variables – gender, whether they were admitted, and which type of school.

a. The hardest thing for students to understand about this question is that the two-way table, which is your answer, DOES NOT contain all the information shown in the three-way table. You lose the information about what school it is, because you sum up, so you can’t see how many are from each school in the resulting two-way table. To make a two-way table from a three-way table, I think the first thing to do is to label the categories. (When you get them labeled then read the labels to see what numbers are needed in that cell.) Since the question says “two-way table of gender versus admission decision” then, on one side you have the categories of gender and the other side is the categories of admission decision.

b. Nothing to say more than what’s in StatsPortal.

c. Here, you must go back to the original three-way table to find the data to compute these percentages.

d. Remember the general idea we saw back in exercise 6.7 – look for whether two groups had different proportions of “easy” cases. What would “easy” mean here? And what are the two groups? (Answer: “Easy” means easier to get into the program. The two groups are male and female applicants.)