Homework Notes  Chapter 2

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!)

I am not promising to do this for all chapters. I continued it for Chapter 2 because there is so much interpretation and some students are not familiar with interpretation in math classes.

HW Chapter 2: [2.1, 2.3, 2.5, 2.7, 2.9(T), 2.12, 2.13-2.22], 2.23, 2.29, 2.31, 2.35, 2.37, 2.41(T), 2.43, 2.45(T)

2.1  I can’t think of anything to say more than is said in StatsPortal.

2.3. Students ask “Why should I know that housing prices are right-skewed?” Think about how expensive typical houses are. Maybe $150,000 or maybe $200,000. Lots of houses sell for around those prices. Some are less expensive – maybe as little as $75,000 or even $40,000. But then, have you heard of some houses selling for $1,000,000 or even $2,000,000. (Not one that you and I will live in!) But there are some quite expensive houses. Think of how these observations would all be graphed on a histogram. Do you see that the $1,000,000 and $2,000,000 prices a very long way from most of the prices of $200,000 or less? And the prices less than $200,000 are not nearly as far to the left as the $1,000,000 price is to the right. Thus, the distribution of housing prices must be right-skewed. (Most distributions of income are right-skewed, and that is true of any other variable that is quite dependent on income, such as house prices.)

2.5. I can’t think of anything to say more than is said in StatsPortal.

Note about the section on suspected outliers.
I think this is a very useful section. After students hear about the concept of outliers, some students tend to identify the largest observation in almost any distribution as an outlier, and the smallest observation in almost any distribution as an outlier. That is not correct. So it is useful to have at least some criterion to identify when an observation is far enough outside the rest of the distribution to be suspected of being an outlier.

How to use this criterion.
1. Arrange the data in order from the smallest observation to the largest observation.
2. Find the median and quartiles of the distribution.
3. Find the IQR of the distribution. (It is one number.)
4. Multiply the IQR times 1.5 to get a number for the distance.
5. Subtract that distance from Q1 to see how small an observation would have to be to be suspected of being an outlier on the low end. To be suspected of being an outlier it must be further than this.
6. Add that distance to Q3 to see how large an observation would have to be to be suspected of being an outlier on the high end. To be suspected of being an outlier it must be further than this.
7. Look at the data in order from step 1 above and determine whether any of the data points on the ends fall outside the values you found in steps 5 and 6.

2.7 I can’t think of anything to say more than is said in StatsPortal.

2.9 Please recall what your instructor said about what technology you are expected to use to find standard deviations and how you will/won’t be tested on this.

2.12 Be sure to identify the four steps and write words for each one as well as the appropriate calculations and summaries.

2.23. The only thing I can think of in addition to what is said in StatsPortal is the same thing said for exercise 2.3.

2.29 If the discussion in the StatsPortal answer doesn’t make it clear to you how to do this, it will definitely be worth your time to actually list the 74 scores in this distribution. Start with fifteen zeros. Write them down – all fifteen of them. Follow that with eleven ones. Follow that with fifteen twos. I’ll bet that you don’t want to write them down now – it looks tedious. I promise you that it is worth it. And you can do it in under two minutes. Write out all 74 observations in this way. Then find the median and quartiles just as you learned in the examples in this chapter.

After you have done that, then look again at the histogram in Figure 1.13 and see whether you can now “see” all the essential things about the list of observations, in order, without having to write them down. If not, then you should do something like this for another histogram or two. This is an important exercise to help you fully understand what histograms are telling you about the distribution.

2.31.b. It is possible that you can just visualize the general idea of what removing these will change about the mean and standard deviation. (It will make the mean smaller because you are removing a value larger than the mean. It will make the standard deviation smaller because you are removing a point that is a long way from the mean.)

It is very tedious to compute these means and standard deviations on your calculator since you have to punch all of them into your calculator. It is easier to find these means and standard deviations using software. The dataset provided in the textbook doesn’t have these two sets of data in different columns, so that is a bit confusing. The easiest way to handle that is to use the option in the dialog box for using the different groups. Following (next page) are illustrations of the dialog box and choices needed here.

When you need to delete an observation, just make the cell with that observation blank. Do not put zero in it – that is not at all the same thing as blank.
2.35  I can’t think of anything to say more than is said in StatsPortal.

2.37  If you are using the data icons in StatsPortal, you’ll need to use the icon for 2.38. They have these two icons switched from where they should be. The data is all in one column, so to use the software to find summary statistics, use the group option in the software, as discussed in the notes for exercise 2.31 b here.
As far as the solution to the exercise, I have nothing to add to the solution in StatsPortal.

2.41  I tried this on MINITAB 14 Student Edition, CrunchIt, and Excel 2003. MINITAB computed the standard deviation correctly for all numbers it would let me enter. It would not let me enter numbers larger than a number like this with 17 zeroes in the middle. CrunchIt displayed the numbers in a strange (almost unreadable way) when I put more than six zeros in the middle, but it did correctly compute the standard deviation of them with as many as 14 zeros in the middle. However I am unwilling to use CrunchIt to analyze numbers that it displays in a way that is not easily readable.
Excel let me enter numbers as large as I wanted, but after fourteen zeros in the middle it simply rounded the number, so it did not compute the standard deviation correctly.

2.43  The explanation in StatsPortal is correct, but more algebraic than I’d prefer. Here you just need a distribution that is really strongly skewed to the right. So put all your data points except the largest fairly close together and the largest data point a long way away.
Here’s an example: 3, 4, 5, 6, 1000. Note that Q3 = 6 and the mean is 1018/5 = 203.6

2.45  I can’t think of anything to say more than is said in StatsPortal.