MATH 1342 Textbook Homework

There are five levels of working on problems in this course. If you try to skip one or more of these levels, it is likely to cause you to be confused and waste much time at the next level or two. Homework is levels 2 and 3 in this list.

Level 1. Examples in the text and other resources

Level 2. Odd-numbered exercises scattered through the chapter at the end of each section and focusing on the idea(s) of that particular section. (Fairly complete solutions are available.) READ THESE and READ SOLUTIONS. Ask questions as needed.

Level 3. Odd-numbered exercises at the end of the chapter which usually bring together two or more ideas in the chapter. (Fairly complete solutions are available.) WORK THESE. COMPARE YOUR WORK TO SOLUTIONS. Ask questions as needed.

Level 4. Quiz, Part 2 problems (For which you do not have solutions before you submit them.)

Level 5. Test problems. (You demonstrate that you have learned to confidently do problems without have solutions or hints to help.)

Technology:

Many of the problems require you to use technology to do some parts. In the first few chapters we cover, our homework list identifies some problems where you are definitely expected to use some technology. Problems with (A) require an applet. In addition to the applets in LaunchPad, we have the set of Visualize applets from http://visualize.tlok.org/elem-stat .

Problems with (M) have some part for which you are expected to use MINITAB, specifically. Most of these also require work by hand as well – interpretations and other questions. You may use MINITAB on additional exercises if you wish. CrunchIt in LaunchPad has menus very similar to Minitab and is available to you wherever you have web access (while Minitab is just at school.) Use CrunchIt when you need to use technology at home and then come to school to re-do the problems you need to turn in in Minitab.

Find the data files for the textbook and some brief suggestions about how to use both Minitab and Crunch-It from http://www.austincc.edu/mparker/software/. This website is used to get the data into both Minitab and into Crunch-It.

On the problems requiring technology, as on all homework, you should spend as much time and thought answering questions about what the results mean as you spend calculating. Your answers should reflect this. During the tests, you will not have computer access and may not even be allowed to use a graphing calculator. Test questions will be adjusted to reflect the tools you have available at that time. Ask your instructor in advance what you will be allowed to use on each test.

Chapter 0:

What is statistics? How does statistics help us solve problems in today's world?

0.1, 0.3

Part II Introduction:

No homework assigned.

Chapters 8 and 9:

Did you read and think about this page in the Prerequisite Material I gave you?
https://sites.google.com/site/mpstat22/home/produc
Review it now. The key to understanding Chs. 8 and 9 is to NOT think of the ideas as all new, but to think of them as giving structure and words to ideas you already had about sampling and designing experiments. This is especially important in Ch. 9, where many "jargon" terms are introduced. It is important to think about the ideas mainly and then learn to use the jargon terms as the language to communicate about those ideas. This is a good time to look at the reviews of vocabulary in the Flashcards in the additional resources in LaunchPad.

Chapter 8:

What are good types of sampling and why? What are poor types of sampling and how does each of them cause difficulties?
There are various ways of sampling that involve randomness and you should recognize those described in the text. The formulas in the rest of the course will assume the random sampling was simple random sampling – if it was a different type of random sampling, then you would need to use the more complicated formulas that are appropriate. That's for software or for a later course.
You are expected to learn how to do the method of simple random sampling described in the text using the table. It's not the only way to obtain a SRS or even the only way to use the table to produce a SRS, but when we all use the same method, we can have a solution key, and that's convenient.
In your solutions, describe how you are doing the simple random sample when you are asked to do it. (It is the process that makes it random, not the result.) Say how you assigned IDs to those in the population. When you look at a line in the random digits table, list all the ID numbers in that line and then circle those that correspond to the individuals in your population and then identify the sample you chose, based on those. (About the list: if the list from the table gets too long, after you list all of them for awhile, you may just say "etc." and then keep looking in the same way and just tell me the resulting numbers that you found for your sample.)

[8.1, 8.3, 8.5, 8.7, 8.9, 8.11, 8.13, 8.15] **8.27, 8.31, 8.33, 8.35, 8.37, 8.39, 8.43, 8.45, 8.47**

Chapter 9:

Again, please look at the Prerequisite Review material before you begin this chapter and really think through how you would describe your idea of a good experiment. Then read the material in the text and think about what words they are using to say what you were thinking. And what ideas they bring up that you hadn't thought about.
Learn to describe (and distinguish between) a completely randomized design, a block design, and a matched pairs design. For the first two, learn to use a diagram for the description.
In the three principles of experimental design, students often misunderstand "control" and define it too narrowly. It includes ALL the ways that we control the effects of outside variables – including

using randomization to assign the subjects to treatments so that the effects of the outside variables that we aren't thinking about are reasonably evenly divided among the treatment groups.

[9.1, 9.3, 9.5, 9.7, 9.9, 9.11, 9.13, 9.15, 9.17] **9.29, 9.31, 9.33, 9.37, 9.39, 9.43, 9.45, 9.49, 9.51**

Chapter 10. Data Ethics

This is a very important chapter, which should be read and considered along with the other chapters all the way through the semester. As you learn more about statistics, you'll come up with additional ideas about what seems to be ethical and what doesn't. And you can critique the various studies we see in later chapters. Most of the important parts of this won't be on tests explicitly, but will be some of the main things that you remember from this course for years.
In previous semesters I have found that students find this the most engaging material in the course.

10.1, 10.3, 10.11, 10.13 (These are to begin. During the rest of the semester pick at least three more questions from this chapter to discuss on the Disc Bd Forum on Data Ethics. Be sure to read some other students' submissions and react to at least some of them. This is a much richer subject with actual discussion.)

Chapter 11 Review.

You have learned many different techniques by now. Use these problems especially to review how to tell from the statement of the problem itself what you need to do to solve it.

[11.1 – 11.17 (odd-numbers)] **11.19, 11.21, 11.23, 11.25**

Part I Introduction:

No homework assigned.

Chapter 1:

Pay the most attention to how to interpret the graphs and the similarities (like dotplots and histograms) and differences (like bar graphs and histograms require different types of data.)

For categorical variables, do not spend time making graphs with software. Notice how you would make bar graphs by hand, including what choices you need to make (labels, order of the bars.) Practice making a bar graph by hand by hand once.

For quantitative variables, do not take the time now to make histograms or stemplots by hand. If you need to make a frequency graph of a quantitative variable by hand, make a dotplot. At this point, mainly focus on understanding how to read histograms. Recognize that dotplots and stemplots are giving the same information as histograms. Use the Visualize applets to explore the choices that can be made in making dotplots, histograms, and stemplots as well as using them to simply make the required plots.

Notice that all of these graphs for quantitative variables are interpreted in terms of center, shape, and spread of the data. Notice that bar graphs cannot correctly be interpreted in terms of shape or spread, and usually not even in terms of center.

[1.1, 1.3, 1.5, 1.7*(A), 1.9, 1.11(M, NOT by hand), 1.12] **1.23, 1.27*, 1.29*, 1.31, 1.33, 1.37(M** for graph. Interpretations are most important**), 1.41, 1.45(M), 1.46*(A)**

*1.27 and 1.29 It isn't necessary here to actually draw the bar graphs. Just notice how you could. Then discuss the answers to part b in each.

*1.45 In a timeplot, the pattern is the important thing. It is sometimes time-consuming to get the labels you want on the horizontal axis. Don't spend time on that now. Instead, focus on what different information you obtain from the two different graphs. What does this say to you about how the question asked helps determine what kind of graphical summary is useful?

*1.7(A) and 1.46(A) are assigned to help you explore how stemplots and histograms work. There are errors in the statement of these problems. The LaunchPad applet does not have the appropriate data in it. Instead, use the Visualize applets.

1.7 Open a separate window from LaunchPad. In it, open http://visualize.tlok.org/elem-stat/ and choose to make a histogram from your own data. (This is the applet that gives you a choice of the number of bins.) Delete the values in the box, leaving it empty. To obtain the data from problem 1.7, click on the Data icon in the e-book and choose PC Text (or Mac Text.) That will open a new window with the data. Copy it and then paste it into the box in the Visualize applet. To answer the questions in the problem, change the number of bins by typing in the number you want or by using the arrows to increase or decrease the number. Then answer the questions in the problem.

1.46 Since there is no data icon by this problem, you must find the data in Exercise 2.28 in Chapter 2. Follow the same instructions given for 1.7 above.

Chapter 2:

Pay the most attention to how what each statistic tells you about the center, shape, or spread of the data, and then how to choose which statistic to use for a particular set of data. Pay enough attention to how to compute them for the meaning and use of the statistic to make sense to you.

Given a histogram or dotplot of data, be able to estimate each of these statistics. Given comparative histograms of data on two groups be able to compare them in terms of these statistics (e.g. Histogram A has a larger IQR than Histogram B)

Be able to compute mean and median "by hand" (with a non-graphing calculator for the mean) as well as with software. Use CrunchIt or Minitab or a Visualize applet to compute the standard deviation. Do not compute the standard deviation by hand, even with a calculator. Know that the standard deviation is "the amount a typical score differs from the mean" and use that idea to make a rough estimate of it from a frequency graph.
For fairly small data sets, compute median and quartiles by hand. Note that software uses a slightly different method of finding quartiles, so the results may differ slightly from what you find by hand, and that's OK.

The trickiest part of this chapter is using software to compare data from two or more groups. Your comparative graphs must be on the same scale. Thus you need to instruct the software to make both graphs at the same time. That is a different command than making a single graph. The data for such comparative graphs is usually in "stacked" form. Be sure to obtain help on using that well before your quiz/homework is due. Using software to do comparisons like this is very important at several times in the course. Learn to do it now!

[2.1, 2.3, 2.5, 2.7(M), 2.9, 2.11*(M), 2.13*] **2.25, 2.30a*, 2.31, 2.33, 2.35(A), 2.37(M), 2.39, 2.47(M), 2.51**

* 2.11 Use Minitab to compute the numbers and to make the graph

* 2.13 Give comparative graphs and comparative numerical summaries as part of "Solve." Be sure to "Conclude."

* 2.30a is assigned to help you better understand how to read histograms. If you don't see how to answer the questions, start making a list of all the 74 observations. You probably won't need to completely write out that list to understand how to "make the list mentally" to answer this question.

Chapters 4 and 5 are about the same topic – relationships between quantitative variables.

Chapter 4. Scatterplots and Correlation Coefficients

What does a scatterplot tell you about the relationship between the two variables?

> What are typical ways to describe a scatterplot?
> Use a scatterplot with different types of dots to compare the relationships of the two variables for different groups of people.
>
> If there is a "response variable" which axis does it go on?
>
> Make scatterplots by hand. Make a scatterplot with Minitab.

What does the correlation coefficient tell (and NOT tell you) about the relationship between the two variables?

> Estimate the correlation coefficient "by eye" to within about 0.2. (I don't intend to give you tricky problems on this – just enough to make sure you have the idea.) Compute the correlation coefficient with Minitab – never by hand.

[4.1, 4.3, 4.5(M), 4.7, 4.9*, 4.11, 4.13(M)] **4.25(M), 4.27(M), 4.31*(M), 4.35*a (by hand. Don't try to get software to give you a specific scale), 4.35b (M), 4.37, 4.39, 4.41*(A), 4.42*(A), 4.43(M), 4.45(M)**

4.9* Since this is a small dataset, it's easier to make the scatterplot by hand so that you can just write the two separate symbols. Minitab will do it, but you aren't required to learn that at this time.

4.35* Don't try to change the scale of a graph in Minitab. It takes up time and isn't essential to learning the concepts of the course. But, for part b, of course you need Minitab to compute correlation coefficients.

4.41* and 4.42* "Playing" with this applet and watching what happens is an excellent way to get a feeling for how the correlation coefficient works.

Chapter 5. Regression

If the relationship can usefully be summarized by a straight line, what is the equation of the usual line?

> And why is the usual line called the "least squares regression line?" List all the other names you have seen for it. ("The line" in an elementary or intermediate statistics class is always this same one, even though various names are used.)
>
> These two mean the same thing – "deviations from the line" and "residuals." Do small or large values for these indicate the line is a good approximation to the data?
>
> What does each of the coefficients tell you about the relationship, **in context**? ("Interpret" them.)

How do we measure how well the line fits the data?

> What does the coefficient of determination, r-squared, tell you? ("Interpret" r-squared.)
>
> Consider a residual plot of the residuals versus the explanatory variable. Sketch a plot that indicates the data DO NOT have a linear relationship. Sketch another plot that indicates the data can be well-modeled by a linear relationship.

If a point is "influential" then it affects the result considerably more than other points. So we want to look very carefully at it – were there any errors in measuring it or is it from an individual that really shouldn't be considered a part of that population? We won't actually make those judgments in this class, but you need to recognize influential points so you will know which ones to pay attention to in this way. (And, are the ideas of influential for correlation and influential for the regression line the same? Or can points be one of these without being the other?)

How do you predict y from x using the line?

Why do we **not** use the line to predict x from y? (How is that related to the "least squares" idea?) Calculate the equation of the line and predictions both by hand and with Minitab.

Sketch a picture to clarify why a strong linear relationship doesn't necessarily mean that x has a large effect on y. Thus, a question about the nature of a relationship needs an answer for both strength and size of the effect. What numbers summarize each?

When there is a linear relationship between two variables – even a strong relationship - then x and y are associated / correlated, so x is useful for predicting y, but this does NOT mean that x causes y. Why not? How do "lurking variables" come into this conversation? How do we establish causation?

Why are ecological correlations stronger than correlations on individuals?

What is extrapolation and why should we be careful about it? Do you suppose anyone ever does it and actually finds it useful? What circumstances might lead to that?

[5.1, 5.3, 5.5(M), 5.7(M), 5.9, 5.11*(A), 5.13*(M), **5.14*,** 5.17, 5.19]
5.31, 5.33, 5.37(M for correlation and equation, the rest by hand), 5.39(M), 5.41*(M), 5.42*(M), 5.45*(M and then draw the lines on the graph by hand), 5.47
5.51(A), 5.53(M), 5.55(M), 5.59*(M), 5.61(M)

5.14* is included as required because it is the only problem about ecological correlation, which you are supposed to learn to recognize.
5.11*(A), 5.13*(M), 5.41*(M),and 5.45*(M) are assigned to help you develop your intuition about which points are influential in terms of correlation and in terms of regression.
5.42*(M) illustrates a very important idea
5.59*(M) It is important to practice the four-step process, even though the solutions don't include that. It is also important in this problem to think about what different things the correlation coefficient and the slope coefficient tell us. (Hint: What if the correlation coefficient was very close to 1, but the slope coefficient was very close to zero? What would that tell you about the answer to this question?

Chapter 6. Relationships between two categorical variables.

In the Prerequisite Review, you saw a question where college students were asked their sex and their major, and the results were compiled into a table. You noticed there that the difference in answering the these two questions came from how you choose the denominator: "What percentage of the women were accounting majors?" and "What percentage of the accounting majors were women?"

In almost every problem in this chapter, one of the first, and most crucial questions is- what is the denominator?

There are some "jargon words" in this chapter. Learn about the differences between the various questions first, and after those are clear to you, then learn the words we use to identify those.

When you are learning about Simpson's Paradox, notice that the reason the paradox occurs (when it does) is that there is some imbalance in how many of what type of people are in the two categories. So when you are asked to explain why it happened, you are expected to identify that imbalance. However, in these problems, if you can correctly make the required tables and compute the required percentages / proportions, you will earn most of the credit – even if you do not correctly identify that imbalance.

[6.1, 6.3, 6.5, 6.7] **6.19, 6.20*, 6.21, 6.22*, 6.23, 6.25, 6.27, 6.33**
6.20* and 6.22*. One of the differences between statistics and algebra courses is that sometimes we have rules in statistics that we have to learn to read "flexibly" because the rules presume exact numbers (like those we learned to use in algebra) and in statistics we are almost always using rounded (approximate) numbers. If you are working hard to get your percentages to add up to exactly 100%, then you are not spending your time appropriately. Spend your time understanding the meaning of these various distributions, not in trying to find the best rounding rule in this problem to make them sum to 100%.

Chapter 7: Review. You have learned many different techniques by now. Use these problems especially to review how to tell from the statement of the problem itself what you need to do to solve it.

[7.1 – 7.35 (odd-numbers)] **7.37, 7.45, 7.53**

Chapter 3:

Here we are discussing distributions of populations. So the datasets that produced these graphs have a very large number of points in them. Use the Visualize applet to see how, for a very large dataset, a histogram can have so many bins that it seems silly to continue to draw it with all the separate bars – it makes more sense to sketch a smooth curve that describes the shape of the distribution.

We move beyond talking about a distribution in terms of just center, shape, and spread, to more details. For example, what proportion of the distribution is higher than the value 37.2? Or between 29.1 and 37.2. Or what score has 5% of the scores above it?
For each question of this type, first draw a picture of the distribution and shade in the appropriate area. Use that to estimate the answer. Then use the computational techniques discussed in this chapter to give a more precise answer.

It is important to note that not all distributions are normal. Exercise 3.2 is about a different shape - the uniform distribution. Be sure to read the problems carefully to see what distribution they are asking about. Then sketch the picture. Use that to answer the questions.

There are many possible distributions. In this chapter, we will only consider answering these questions for a uniform distribution and a normal distribution. NOT NOW, but in last few weeks of the course, we will answer questions like these for a t-distribution, a chi-squared distribution and an F distribution.

We do not use software in this chapter to find areas in a normal distribution because that might lead you to not draw/look at the appropriate picture to understand what you are doing. Instead use the

Normal table**,** after drawing appropriate pictures to illustrate how you are using that table. **Solutions on the test must include appropriate pictures to be fully correct.**
Some problems involving normal calculations can be solved either by the 68-95-99.7 rule or by using the table. In that case, either way is OK. The difference between the answer you get from each of those methods is an illustration of something you will see often in this course – there is not just one correct way of expressing an answer. Begin to develop your intuition about how much rounding is acceptable.

[3.1, 3.2*, 3.3, 3.5, 3.7, 3.9, 3.11, 3.13] **3.25, 3.27 3.29, 3.31*, 3.33, 3.35, 3.37, 3.39, 3.43, 3.45, 3.47*, 3.49*(M), 3.53***

* 3.2 Answers are (b) 2*(1/5)=0.40 (c) (4/5)*(1/5)=0.16 (d) 3*(1/5)=0.60

* 3.31 – 3.43. Practice as many of these as you need to reliably answer both types of questions ((1) find the proportion or (2) use a given proportion to find a value.) Draw pictures for every problem you solve. If you think "I can solve these, but the pictures confuse me" then there will be later material in the course that will be difficult for you. Understand them now so you don't get behind.

* 3.45 Part b is confusing – too hard for a test in this class. This is included in the homework because it is such an obviously useful result for anyone thinking about health of a friend who is an older woman.

* 3.47 Assigned because it emphasizes two important subtle points that students often miss.
(1) There is a difference between the distribution of actual data and a theoretical normal distribution. You must be able to hold both those ideas in your mind as separate, but somewhat related, ideas.
(2) In a theoretical distribution where we are computing proportions from areas under the curve, including or excluding an endpoint does not affect the area.

*3.49 Assigned to emphasize the difference between a distribution of actual data and a theoretical normal distribution. The new idea illustrated here is that we claim that some distributions of actual data are normal because their shape is so close to normal that approximating the data distribution by a normal distribution gives answers that are accurate enough. We will do MUCH more with this later in the course. You could look now at a Visualize applet to show the shapes of different samples of size 20, for instance, from a normal distribution.

* 3.53 Do not use the applet, as the problem says. Use the Normal table. This is assigned because the result is a useful result to know – not to memorize, to improve your understanding of the picture of a normal distribution.

Part III Introduction

No homework assigned.

Chapter 12:
In this chapter we extend the ideas we learned in Ch. 3. In Ch. 3, we talked about finding areas, and now we learn that those areas are called "probabilities." And we also learn to find probabilities in other situations. Students find the notation challenging in this chapter. Pay careful attention to that.

[12.1, 12.3, 12.5, 12.7, 12.9, 12.11, 12.13, 12.15, 12.17, 12.19, 12.21] **12.33, 12.35, 12.37 12.41, 12.43, 12.47, 12.51, 12.52*, 12.53, 12.54*, 12.59(A)**

12.52* and 12.54* are included to help you review what you learned in Chapter 3 about how to distinguish between the uniform distribution and the normal distribution from the statement of the problem, and how to calculate probabilities in them.
12.57 and 12.59 use the applet to help you develop your intuition about probability as a long-term relative frequency. That's a very important idea.