NOTES FOR M358K INSTRUCTORS USING DEVEAUX, VELLEMAN AND BOCK, *STATS: DATA AND MODELS*, 3RD EDITION

NOTES ON PART I (Chapters 2 – 6)

Chapter 2:

- Exercises 5 and 6 look good in concept, but (in the printing I have) are poorly worded – the phrase "answer the same questions …" appears to be left over from a previous edition.
- I suggest assigning some exercises from each category (1-4, 5-12, 13 – 26)

Chapter 3:

- You may want to add descriptions (or ask the students to do the translation, or guide them in the translation) of some of the topics in this chapter in the language of the probability prerequisite to the course. For example:
    - On p. 20 (last sentence of first paragraph), you may want to point out how the word "distribution" can be interpreted as the probability distribution (probability mass function) of a discrete random variable, with universe/population all people on the ship.
    - In Figure 3.4, if the percentages are converted to decimals, the first bar's height shows the probability that a person on the ship was in first class – i.e., the probability that a randomly selected person on the ship was in first class, or the probability of being in first class, when the universe is all people on the ship.
    - In Table 3.5, converting the percentages to decimals, the various entries describe conditional and joint probabilities such as P(survived| third class), P(third class|survived), and P(third class and survived)
- The "percent of what" box on p 23 focuses on something a lot of students have trouble with.
- I have a quibble with the sentence, "We can look at this question in two ways," on p. 24. The question is "… whether the chance of surviving the *Titanic* sinking *depended* on ticket class." I would interpret this phrasing as asking about the conditional distribution of survival given ticket class, but not asking about the conditional distribution of ticket class given survival. (Of course, there turns out to be a connection, but it is not obvious.)
- Also, it may be considered less than best practice to use a pie chart with an ordered categorical variable (ticket class in the Titanic example; amount of fish in diet on p. 30).
- Exercises (pp. 37 – 43)
    - Assigning one of #1- 4 (or the student's choice of one of the four) for class discussion might be good.
    - Other exercises that look particularly good: #14-16, 19 (especially part d), 20, 23
    - I suggest assigning one of 31e or 33, and at least one of 38 – 40.

Chapter 4:

- It would be good to illustrate in class the point in the last sentence in the box on p.45 (how bin width may affect the appearance of a histogram). The teaching applet Learning about Histograms on the ActivStats disk shows this, but it is probably easier for class demonstration to use the applet at http://www.amstat.org/publications/jse/v6n3/applets/histogram.html .
- The Just Checking features on pp. 51 and 61 are good.
- Re the caption for Figure 4.13 (p. 58): The sentence, "Because the distribution is skewed to the right, the mean (2.28%) is higher than the median," is not quite correct. As a general rule of thumb, most distributions that are skewed to the right have mean higher than median, but this is not true for all distributions that are skewed to the right. For examples, see Hippel, Mean, Median and Skew: Correcting a Textbook Rule, *Journal of Statistics Education* Volume 13, Number 2 (2005), www.amstat.org/publications/jse/v13n2/vonhippel.html, and the letter linked from the Addendum of that paper.
- Re p. 59: Be sure to point out that students may have seen a definition of variance with n in the denominator in M362K. However, it's probably best to delay an explanation of the n-1 used here until Chapter 18. A supplement "Why Does the Sample Variance Have n-1 in the Denominator?" is provided for this.
- Some specific suggestions will be given under that chapter in Part IV.)
- The For Example on p.61 is nice.
- Exercises: (pp. 72 – 78)
    - Assigning one (or their choice) of #1- 4 for class discussion might be good.
    - Other exercises that look good: #5- 12; at least one of 13-16; 18, 26, 33 or 34; 35 or 36; 41 or 42; 44; 45 and/or 46; 50
    - You might also consider assigning the Exercise from the supplement "Why Does the Sample Variance Have n-1 in the Denominator?" now.

Chapter 5:

- This chapter looks very good.
- The Just Checking questions on p. 86 look good.
- Be sure to include the starred section on Smoothing Timeplots (pp. 89 – 90) to the extent in the textbook, but resist the temptation to go further than that, since other topics will need the time. (In case you don't know: "lowess" stands for "locally weighted scatterplot smoother.)
- Box on p. 92: If you have time or if someone asks why it doesn't make a difference which base of logs you use for transforming, show (or assign) that the base of the log doesn't make a difference in the shape of the transformed distribution: $\log_a y = c\log_b y$, where $c = \log_a b$
- p. 92 is worthwhile.
- Exercises (pp. 99 – 107): There are lots of good exercises here. Some of my favorites:
    - Any (or a choice of) #1-4 would be good for class discussion.
    - 5-14, 16, 18, 20 – 22 (especially 21), 28 (also ask them what the ogive was called in M362K), 35, 37-40 (especially the combination 37 plus 39)

- Another good exercise: Give the information in #3 on p. 135 and have them make a box plot.

Chapter 6:

- The answer (p. 140) to part (b) in the Just Checking box on p. 113 seems unclear to me – it seems to depend on how the teacher computes the average: using the raw scores or the curved scores.
- p. 114: Do the math (or have them do it) to verify that the mean is shifted and the standard deviation stays the same.
- p. 117: A nice introduction to the idea of statistical models. The disk applet Learn about Density Curves might be helpful for students who need supplementation on the concept of probability densities or comparing and contrasting with histograms.
- p. 118: The disk applet Work With the Normal Model might be of some use to show how different samples from the normal distribution might appear, but you may prefer other ways of doing this.
- p. 119: Just Checking question 5 is good. Possible homework or extra credit problem: Do the math to show that the inflection points of the normal curve are one standard deviation from the mean.
- p. 121: Nice discussion of Chebychev's inequality.
- Note re pp. 122 – 128: Students should have done normal model problems in the prerequisite course M 362K, so this part can be regarded as review or independent reading rather than covering in class.
- pp. 129-130: See the Math Supplement on Normal Probability Plots. Also, point out that some software calls the normal scores just "z-scores," which can cause confusion.
- Exercises (pp. 135 – 140) Some calculational exercises are in order, but bear in mind that students should have seen normal calculations in the prerequisite M 362K, but probably have not dealt with calculations involving IQR. Suggestions for exercises to consider: # 1 – 5, 11, 13 – 14, 16, 23, 31 – 36, 47, 48. See also the suggestion under p. 119, above.

Review Exercises (pp. 141 – 149):

- The following look good: #2-4, 6, 7, 9-12, 14, 16-24, 26 (especially part c), 28, 30-33, 35-38.
- Caution: The book's wording on #20 is unclear.
- A nice addition to #24: Make a time plot of the data for each player, using each player's first full season as the starting point, and discuss what these show.