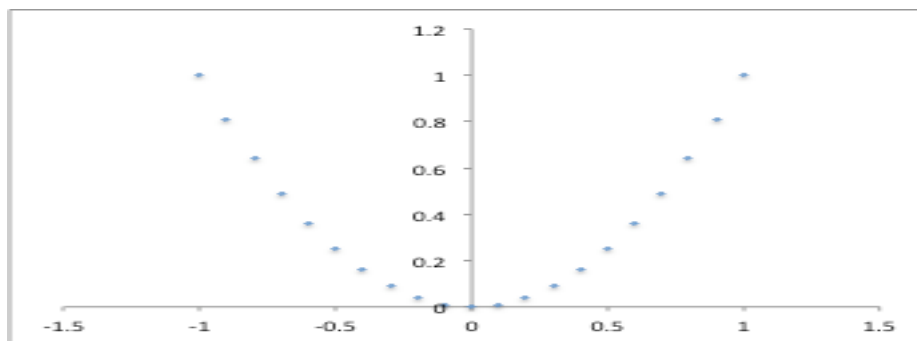


INSTRUCTOR NOTES FOR M358K FOR PART II  
(EXPLORING RELATIONSHIPS BETWEEN VARIABLES: CHAPTERS 7 – 10) OF  
DEVEAUX, VELLEMAN AND BOCK, *STATS: DATA AND MODELS*, 3<sup>RD</sup> EDITION

*Important Note on Sequencing:* If you are having students design and carry out data collection for projects, you may want to do Chapters 11 – 13 after 7, 8, or 9, then come back to the rest of Part II. (Chapters 7 and 8 are referred to on p. 285 of Chapter 12 in explaining parameters vs statistics; Chapters 7 and 9 are referred to on p. 320 in Chapter 13 in discussing lurking and confounding variables.)

Chapter 7: This is an unusually well done exposition of correlation.

- A possible exercise (perhaps in class or perhaps before class) before starting this chapter: Have students draw freehand what they think a plot of weight vs height for a class of about sixty undergraduates might look like. Then compare with the graph of weight vs height on p. 153.
- p. 151 The top blue box makes a point that is important for some students.
- p. 155 You may want to mention that the reason for the  $n-1$  in the denominator will be discussed more later.
- p. 158 The “For Example” is important for stressing that the conditions are important to check. The “Just Checking” is especially good.
- p. 160 An example to illustrate the last bullet: The data plotted below lie exactly on a parabola, hence have a strong (but nonlinear) relationship, but their correlation is 0. (The data have  $x$ -values from  $-1$  to  $1$  in increments of  $0.1$ , and  $y = x^2$ . Note that taking asymmetric endpoints for  $x$  would give different correlations.)



- p. 162 The warnings about correlation tables are definitely worthwhile! You may want to mention that a better alternative (at least, if there aren't too many variables) might be to show a scatterplot matrix. (There's an example on p. 796)
- pp. 162 – 164 Omit Kendall's Tau and Spearman's Rho (but they would be nice to come back to if you have time at the end of the semester – and be sure to refer students to these sections if their projects turn out to need them.)
- pp. 164- 165

- Include the section Straightening Scatterplots (and note the footnote on p. 165), but bear in mind that this is just an introduction to using transformations – Chapter 10 goes into more detail.
- A math (calculus) supplement that could be assigned as an exercise now or when you cover Chapter 10: Explain why statisticians often refer to the logarithm as the 0<sup>th</sup> power. (The idea: Consider the limit of  $x^\lambda$  as  $\lambda \rightarrow 0$ , and use l'Hopital's Rule.)
- p. 166 The example in Figure 7.10 is particularly nice – especially if you make the point that the real question of interest might be to find the optimal baking temperature.
- Exercises (pp. 171 – 177) Try to assign at least one exercise from each of the following groups (or other exercises covering the points made in each group): 1-4, 5-6, 9-10, 11-12, 13-14, 15-16, 19-20, 23-24, 25-26, 27-28, 29-30.

#### Chapter 8:

- p. 179
  - The quote from George Box is a gem I hadn't seen before.
  - The ActivStats activity Learn About the Least Squares Equation shows how changing the slope and intercept changes the sum of squared residuals. However, I think the web demo at [http://www.dynamicgeometry.com/JavaSketchpad/Gallery/Other\\_Explorations\\_and\\_Amusements/Least\\_Squares.html](http://www.dynamicgeometry.com/JavaSketchpad/Gallery/Other_Explorations_and_Amusements/Least_Squares.html) does this more effectively.
- pp. 180 – 181 Good discussion of interpreting slope and intercept.
  - Note especially the common misleading misinterpretation pointed out in the blue box on p. 222.
- Note the comments on units (bottom of p. 181 – top of p. 182).
- pp. 181 – 191 There are some really good points here, but for this class, you should probably be more rigorous on some of the points mentioned, in particular:
  - Why the point  $(\bar{x}, \bar{y})$  lies on the least square line (p. 182).
  - Derivations of the equations for the coefficients of the least squares line (pp. 186 – 187).
  - Why the mean of the residuals is zero (p. 189)
  - Why  $R^2$  is the fraction of the variation accounted for by regression (p. 191).

See [Supplementary Exercises for Chapter 8: Least Squares Equations and Related Identities](#)

(<http://www.ma.utexas.edu/users/mks/M358KInstr/Ch8Exercise.pdf>) and

[Supplement for Chapter 8: Why  \$R^2\$  is the Fraction of Variation Accounted for by Regression](#)

(<http://www.ma.utexas.edu/users/mks/M358KInstr/CH8Supplement.pdf>) for handouts doing this.

- pp. 192 – 193 are important and well done.
- pp. 196 – 197 Supplement with [Supplement for Chapter 8: Another “What Can Go Wrong”](#) (<http://www.ma.utexas.edu/users/mks/M358KInstr/AnotherCh8WCGW.pdf>)

- p. 200 Have students read the first box of “Regression on the Computer.” This explains typical conventions in software output for regression. Point out that there are variations. For example, some software may say “Constant” instead of “Intercept,” as in Exercises 29 and 30 in this chapter. Some software may give output in a table with rows and columns switched from what is here.
- Exercises (pp. 202 – 211): (Be sure students read the top of p. 200, as suggested in previous comment, before doing exercises.) Assigning the odds from 1 – 7, 11 – 17, and 21 – 27 would be good for a self-check (i.e., not to be handed in) of several points made in the chapter. The Supplementary Exercises mentioned above are also recommended. And, as usual, there are lots of interesting multi-part exercises to choose from.

#### Chapter 9:

- This is an important chapter, with well-chosen examples to make important points.
- p. 220 The web demo at [http://www.barrodale.com/java\\_demo/javademo.htm](http://www.barrodale.com/java_demo/javademo.htm) may be better than the activity Learning about Leverage on the ActivStats disk. The paragraph starting “Unusual points ...” is especially important.
- p. 221 The “Warning” and “Just Checking” boxes are especially good.
- p. 222 The blue box is especially important.
- pp. 223-225 make an important point which is often overlooked.
- pp. 225 – 226 give a good summary – urge students to read and refer to these as needed.
- p. 226 The Connections box is worth emphasizing.
- Exercises (pp. 229 – 336) Perhaps assign the odds from 1 – 12 for self-check, then a selection of exercises from 15 on.

#### Chapter 10:

- This has a lot of well-chosen, well-discussed examples.
- p. 237 – 239 A nice beginning!
- p. 238 The comment (middle of page) that “extrapolation can go dangerously wrong when your model is wrong” is worth stressing.
- p. 242
  - See notes for Chapter 7 (pp. 164 – 165) for a possible math supplement exercise regarding logs being considered 0 on the ladder of powers.
  - The table contains a lot of useful information compactly.
  - The ActivStats activity Slide Data Through a Range of Powers might be useful for showing how the ladder of powers is a continuous range of transformations. There may be a better demo on the web, but I couldn’t find one in a few minutes’ searching. If you’re interested, I can show you how to get transformation slidebars using the software Arc (on the math department Windows server), to show transformations of either histograms or scatter plots (OK for demos, but if you want to use your own data, it could be a pain.)
- p. 247
  - Take a minute to explain why the model in the third line of the table is called “power”:

If  $\log y \approx b + m \log x$ , then (assuming logs are base 10)

$$10^{\log y} \approx 10^{b + m \log x}$$

$$y \approx 10^b x^m = c x^m$$

- The comment “Keep in mind: We seek a *useful* model, not perfection (or even “the best”).” is worth stressing – but perhaps with the caveat, “for understanding the data,” not “for arguing for what we want to argue for.”
- pp. 248 – 249 The “What Can Go Wrong” section provides good warnings about mistakes beginning (and sometimes advanced) students of statistics often make.
  - However, I have one caveat on the comment (second bullet) “It’s better to stick to powers between -2 and 2;” namely, that if one variable measures a linear quantity and another measures a volume, using the third power of the linear quantity (or the 1/3 power of the volume quantity) may be most appropriate.
- Exercises (pp. 252 – 258): I might assign 1, 3, 7, 11, 13 not to hand in (self-check), and perhaps 18 and 24 to hand in.

Review Exercises (pp. 258 – 266): Lots of good choices. Ones that stand out to me: 2(e), 3, 4, 6, 8, 11, 15, 17, 18, 21-23, 26, 28-32, 34, 39-41.

*Last revised Oct. 14, 2012*