

STAT 2: Modeling with Regression and ANOVA, Second Edition

Errata in the book

Page 15, intro to exercises 17-22: Line 5, “gender (male or female)” should read “gender (male=0, female=1).”

Page 18, exercise 0.32 (b): This should say that the Cubs lost 58 games, not 59.

Page 32: The y-scale for Figure 1.7 (b) should go from -2 to 2 rather than 0 to 4.

Page 38, example 1.8: The second line says there are 13 islands. This should read “14 islands”.

Page 53, exercise 1.22 (b): The sales price should be listed as \$94,000, not \$95,000.

Page 53, exercise 1.26: The values for SquareFeet are in 1000s of feet. Add “, in 1000s” after “of the home” at the beginning of the third line.

Pages 53-54, exercise 1.28: All references to the variable *BaseYear* should be changed to *t* (what actually appears in the data set).

Page 56, exercise 1.42: Last line should say “... use the second case for part (d)” rather than for “part (e)”.

Page 58, exercise 1.47: The word at the end of the second line should be “invest” rather than “investment.”

Page 69, example 2.5: *BatterNumber* is the number of batters already faced by this pitcher. Instead it is the batter number in the game, *including both teams*. So in the very first game in the data file, Kershaw faced batters 1-4, and then 10-12 (batters 5-9 were from his team). So the *BatterNumber* doesn't give “the number of batters he had already faced in the game” but does still give a sense of how far into the game he faced that batter.

Page 75, second paragraph before the Chapter Summary: This paragraph should refer to the prediction interval of (18.07, 19.19). The observations are individual years, so we are predicting for one observation. This also changes the last sentence to replace 18.4 with 18.1 and replacing 18.8 with 19.2.

Page 78-79, exercise 2.19: The distance is measured in miles, not feet. On the last line of part (a) the word “foot” should be “mile”. Also variable names are incorrect throughout the exercise. “*adj2007*” should be “*Adj2007*” (two times), and “*distance*” should be “*Distance*” (two times).

Page 79, exercise 2.20 (a): The variable names are incorrect. “*adj2007*” should be “*Adj2007*”, and “*squarefeet*” should be “*SquareFeet*”.

Page 79, exercise 2.26 (b): Replace the word “weight” with “width.”

Page 82, exercise 2.41: In both parts (a) and (b) “*LeafWidth*” should be “*Width*”.

Pages 82-83, exercise 2.46: The variable names are incorrect. “*adj2007*” should be “*Adj2007*” (three times), and “*squarefeet*” should be “*SquareFeet*” (three times).

Page 83, exercise 2.47: The variable names are incorrect throughout the exercise. “*adj2007*” should be “*Adj2007*” (two times), and “*distance*” should be “*Distance*” (three times).

Page 83, exercise 2.51: Line 5, “gender (male or female)” should read “gender (male=0, female=1).”

Page 85, exercise 2.58 (e): The prediction should be for a year that is not in the data set. The prediction should be for 2020, not 2010.

Page 144, exercise 3.18: “*adj2007*” should be “*Adj2007*” (3 times), “*distance*” should be “*Distance*” (3 times), “*squarefeet*” should be “*SquareFeet*” (2 times), and on the 7th line “(size of home)” should be “(size of home, in 1000s)”.

Page 146, exercise 3.28: The third line should say “28 specimens” not “27 specimens”.

Page 147, exercise 3.30 (a): On the second line, “*Treatment*” should be “*Noise*”.

Page 151, exercise 3.48: The variable names are incorrect throughout the exercise. “*adj2007*” should be “*Adj2007*” (two times), “*distance*” should be “*Distance*” (four times), and “*garagegroup*” should be “*GarageGroup*” (five times).

Page 151, exercise 3.49: On the fourth line “in this exercise” should read “in this dataset”. In second paragraph, first sentence “with logged variables” should be replaced with “by using the natural log of the variables”. Also all variable names are incorrect. “*adj2007*” should be “*Adj2007*”, “*distance*” should be “*Distance*”, “*garagegroup*” should be “*GarageGroup*”, “*logadj2007*” should be “*logAdj2007*”, “*logdistance*” should be “*logDistance*”, “*logsquarefeet*” should be “*logSquareFeet*”, and “*no_ful_baths*” should be “*NumFullBaths*.”

Page 172, last line: The studres_{13} value for Dade county is -3.28, not -3.38.

Page 190, exercise 4.8 (a): On the second line, $T\text{sqrt}MDs$ should be replaced with \sqrt{MDs} .

Page 217, ANOVA test box: In the first line, “*l*” should be italicized.

Page 245, exercise 5.7 (b): This should read “... and dotplot of residuals.” (The word “of” is missing.)

Page 249-251, exercises 5.38, 5.40, 5.46, and 5.48: In the data, method 1 = vertical suture, method 2 = meniscus arrow, and method 3 – FasT-Fix.

Page 250, exercise 5.42: The logs that are in the table are \log_{10} rather than natural logs.

Page 252, exercise 5.54 (d): This should read “Compare your answers to parts (b) and (c)” rather than “Compare your answers to parts (a) and (b).”

Page 270: The 9th line of text should start with “Rural = 20%” rather than Urban = 20%. The next two sentences should read “The fitted value is about 9 points higher, at 28.8%. Compare that point with another at Rural = 60%, with fitted value Urban = 58.4%, very similar to the value for Rural.” (Both fitted values are modified slightly.)

Page 296, exercise 6.30 (a): The model listed of $FE = River + Site + Error$ should be $FE = Grand\ Average + River + Site + Error$.

Page 296, exercise 6.31 (a): The model listed of $HeartRate = Shrew + Phase + Error$ should be $HeartRate = Grand\ Average + Shrew + Phase + Error$.

Page 298, exercise 6.38 (a): The wrong symbols are used. α_1 should be τ_1 and α_2 should be τ_2 .

Page 319: The equation for the Interaction sum of squares should be: $\sum_{i,j,k} (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2$

Page 323, first blue box: The last horizontal line should be between the lines for Error and Total rather than after the line for Total.

Page 334, exercises 7.3 and 7.4: These are multiple choice questions with one correct answer.

Page 335, exercise 7.13: Each part of this question is a multiple choice question.

Page 336, exercise 7.21: Delete the sentence “The following dotplot shows mean ATP...” Replace it with “The following dotplot shows mean iron retention for each of the four combinations of amount of zinc and amount of iron in fruit drinks.”

Page 390, ASSESS section: The first sentence should read: “Figure 8.20(a) shows the normal probability plot of the residuals; Figure 8.20(b) shows the residual versus fits plot for this model.”

Page 409, exercise 8.60: The dataset referenced should be FFlies2 nor FFlies. (Note the Chapter 8 folder of datasets has the wrong dataset in it named ex08-60FFlies – again, this should be based on the FFlies2 data, not the FFlies data.)

Page 415, table 9.2: The label for the last row should be “Proportion getting 7 hours or more” rather than “Proportion getting less than 7 hours”.

Page 443: The second paragraph after the number 2, should say “If the predictor is binary, e raised to the fitted slope equals the odds ratio.” Also, left hand side of the formula in the next line should be e^{β_1} rather than β_1 . The next paragraph should begin “If the predictor is quantitative, e raised to the fitted slope gives the odds ratio for two values of X spaces one unit apart. The left hand side of the following equation should be e^{β_1} rather than β_1 .”

Page 486, exercise 10.4: There should be a blue triangle at $x=2, y=0.5$

Page 491, exercise 10.29 (b): The first sentence should read “... 87-year-old woman” rather than “... 87-year-old man”.

Page 491, exercise 10.30 (c): The first sentence should read “... 87-year-old woman” rather than “... 87-year-old man”.

Page 494, exercise 10.40 (b): The word “nest” should be “nested.”

Page 589, exercise 0.21 answers: Currently they are labeled as parts (a), (b), (b), and (c). Should be labeled as parts (a), (b), (c), and (d).