Workshop Statistics: Discovery with Data, Second Edition

Topic 10: Least Squares Regression I

Activity 10-5: Cars' Fuel Efficiency (*cont.*)

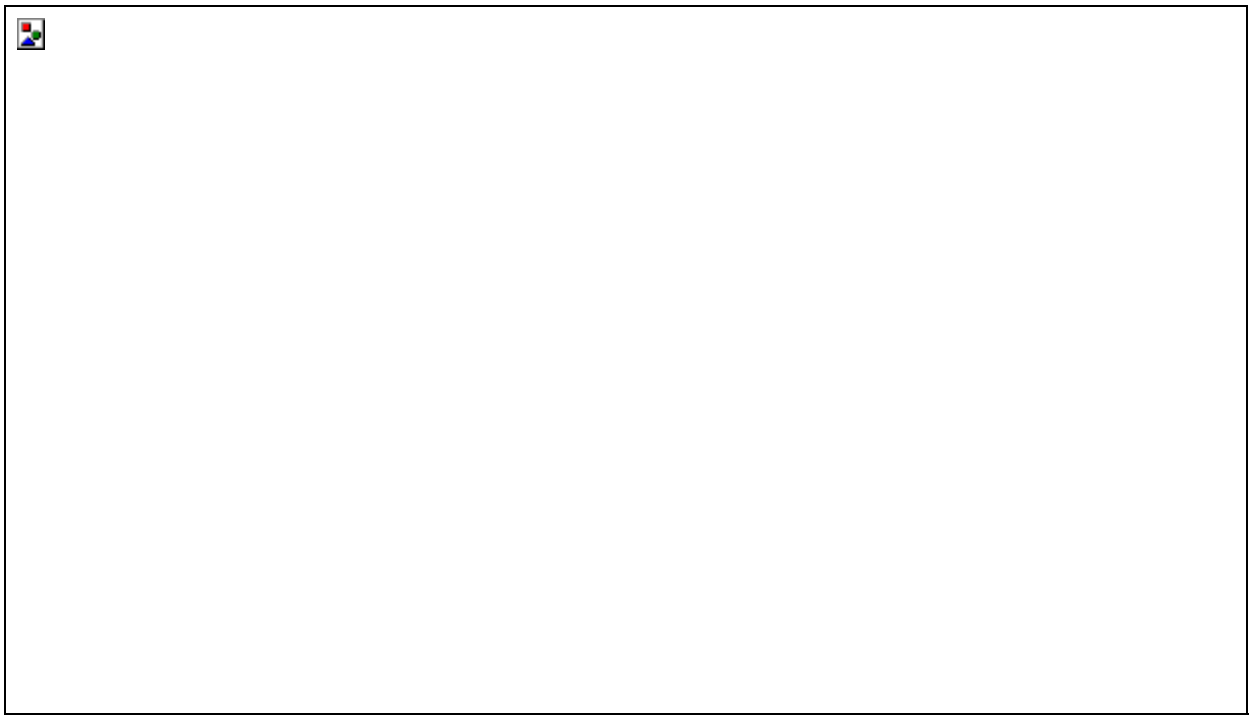
(a) mpg = 41.684 - .007 * weight
(b) 23.099 MPG
(c) .7 MPG
(d) .666

Activity 10-6: Governors' Salaries (*cont.*)

(a) governor's salary = $60,775 + $1.46 * per capita income
(b) 8.4%
(c) California has the largest positive residual, which signifies that their governor has a salary that is much higher than least squares regression would predict.
(d) Connecticut has the largest (in absolute value) negative residual, which signifies that their governor has a salary that is much less than least squares regression would predict.
(e) Connecticut has the largest fitted value because it has the largest per capita income.

Activity 10-7: College Tuitions (*cont.*)

(a) tuition = $113,732 - $54.40 * founded
(b) The slope of the line is negative. For every year one adds to the founding date, least squares regression predicts a decrease in tuition by $54.40.
(c) $9,012
(d)



The line seems to do a poor job of summarizing the relationship between tuition and founding year for public schools. All public school points appear to be well below the line, where as private school points are both above and below the line, albeit more private school points are above the line. We should probably analyze the public and private schools separately.

Activity 10-8: College Tuitions (*cont.*)

(a)

- private 2-year: tuition = -$92,398 + $53.20 * founded
- private 4-year: tuition = $84,719 - $37.10 * founded
- public 2-year: tuition = $154,315 - $77.10 * founded
- public 4-year: tuition = -$13,138 + $9.59 * founded

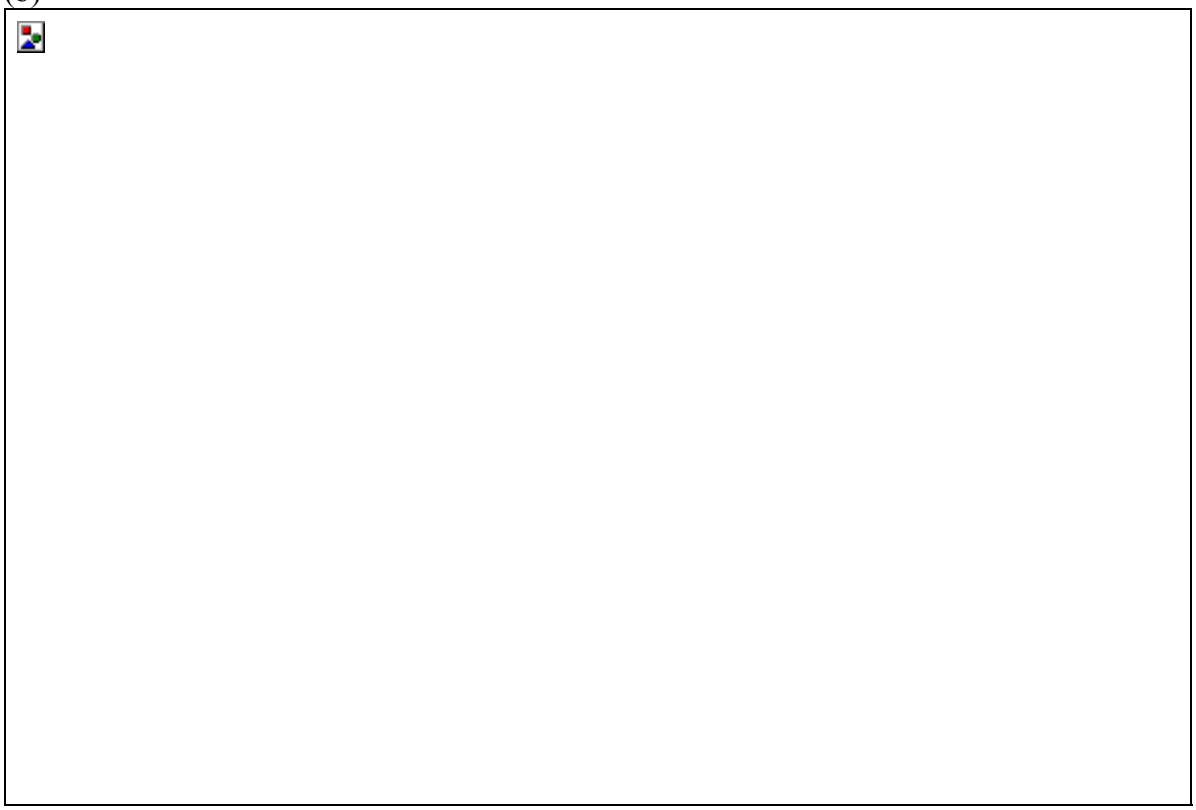
(b) The private 2-year and public 4-year lines have positive slopes, while the other two lines have negative slopes. The y-intercepts vary greatly.
(c)

- private 2-year: $10,012
- private 4-year: $13,301.50
- public 2-year: $5,897.50
- public 4-year: $5,322.75

Note: The answers given for (a) and (b) above are the answer for (a) in the Calculator version, and the answer for (c) is the answer to (b).

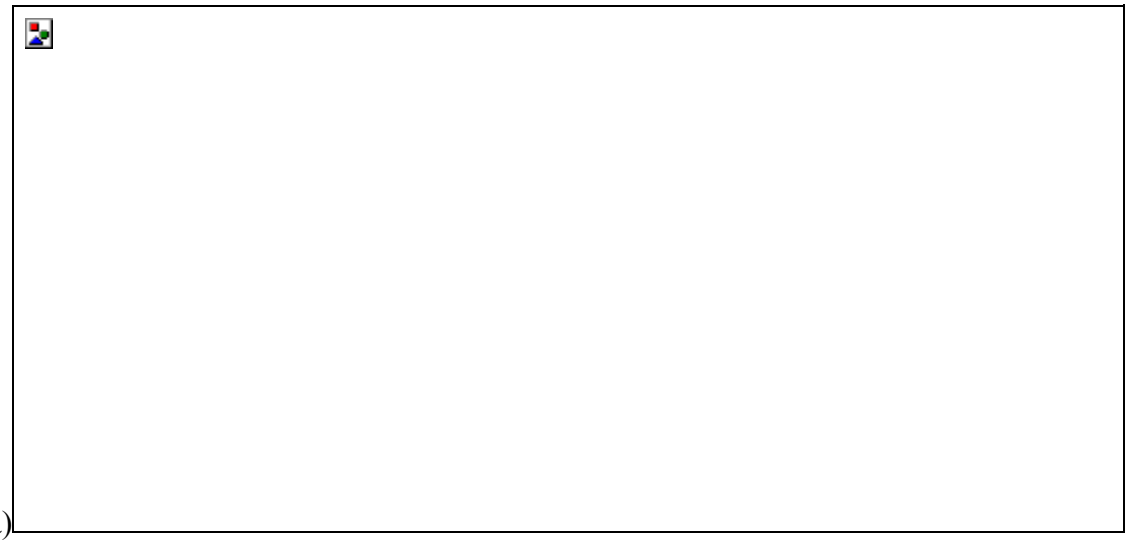
Activity 10-9: Fast Food Sandwiches (*cont.*)

(a) calories = -14.2 + 65.7 * serv ozs
(b)

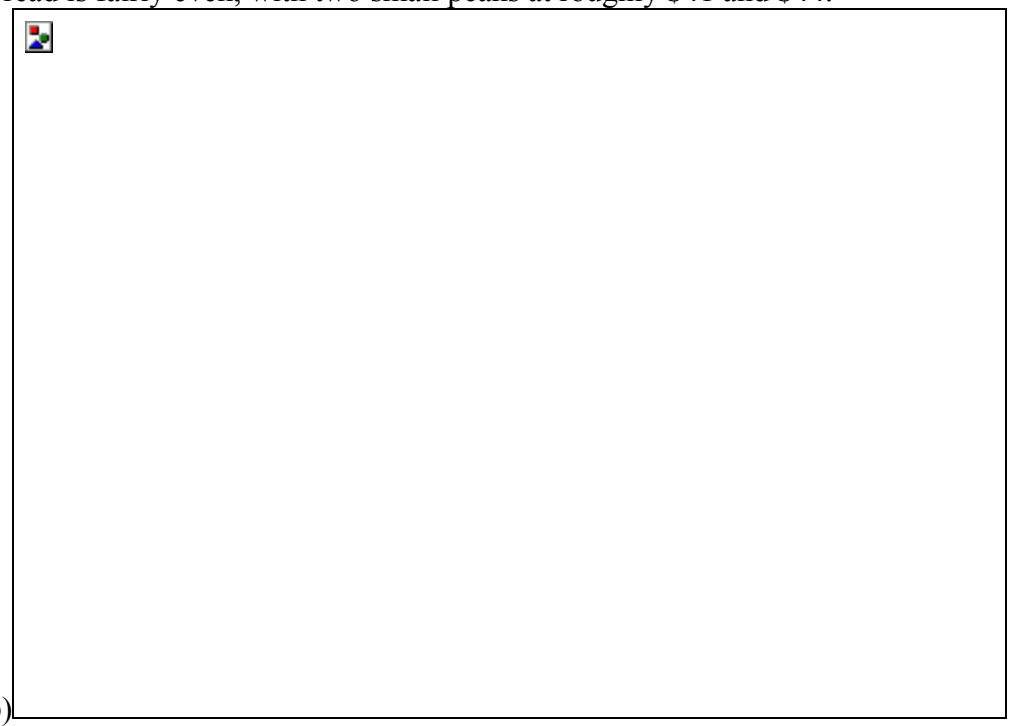


Roast beef has the most concentrated distribution of residuals. Chicken has the largest spread of residuals.
(c) positive: 10; negative: 2; The actual amount of calories in roast beef sandwiches is higher than the fitted value for 10 out of 12 roast beef sandwiches.

Activity 10-10: Electricity Bills

(a) 

The spread is from about $36 to about $51, with one outlier at about $55. The spread is fairly even, with two small peaks at roughly $41 and $44.

(b) 

There appears to be a moderately strong negative association.

(c) bill = 55.13 - .2138 * temp

(d) The bill will drop about $.21 for each degree rise in temperature.

(e) fitted value: $63.91; residual: -$19.48

(f) The month with the largest fitted value will be the month with the smallest average temperature, namely March of 1993 (30 degrees).

(g) .483

Activity 10-11: Signature Measurements (*cont.*)

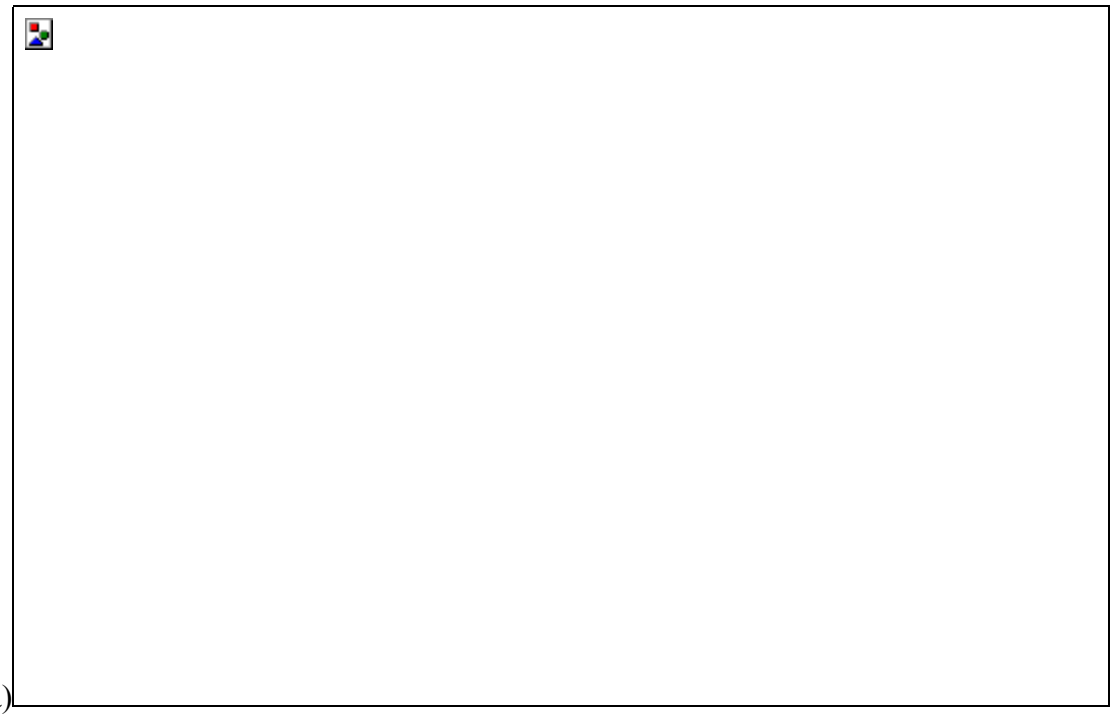
Answers will vary from class to class.

Activity 10-12: Turnpike Tolls

The parts of this activity should be labeled as follows, not (j), (k), and (l), as they appear in the text.

(a) .998
(b) $5.91
(c) approximately $.04 (exactly $.0402)
(d) approximately 25 miles (24.8756218905...)

Activity 10-13: Broadway Shows (*cont.*)

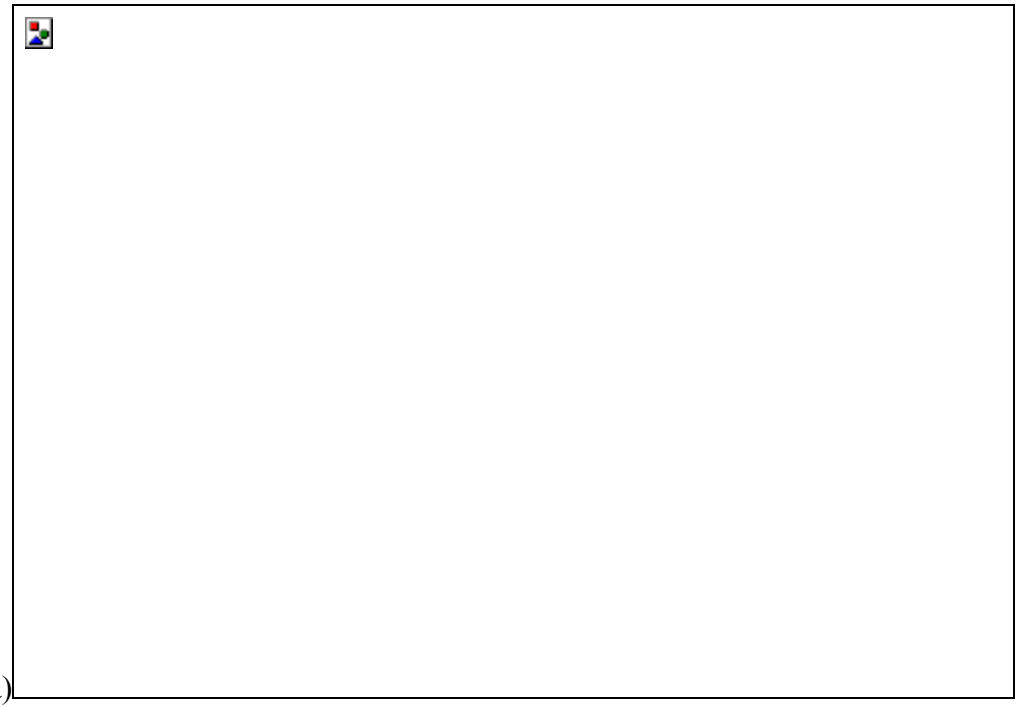
(a) 

gross = -$84,835 + $65.40 * attendance
(b) $405,665
(c) fitted value: $342,292.40; residual: $8,789.60
(d) The Lion King

Activity 10-14: Climatic Conditions (*cont.*)

Answers will vary from student to student.

Activity 10-15: Birth and Death Rates (*cont.*)

(a) 

(b) birth rate = 19.8 - .665 * death rate
(c) Utah, Arizona, Maine, Indiana
(d) The following states have negative residuals: Maine, Indiana Vermont, New Hampshire, Montana, Colorodo, Maryland, Virginia, Wyoming, Alaska, Wisconsin, Hawaii, Pennsylvania, West Virginia, Iowa, Connecticut, Minnesota, Rhode Island, Massachusetts, North Dakota, Oregon, Washington, Michigan, South Carolina, Delaware. These states had death rates which were lower than what the least squares regression line predicted.

Activity 10-16: College Football Players (*cont.*)

(a) weight = 181 + .85 * number
(b) 223.5 lbs.
(c) for each addition of 1 to a jersey number, the player weighs .85 lb. more on average than the player with the previous number
(d) We would expect him to neither gain nor lose weight because weight does not depend on jersey number.
(e) There is a moderate positive association between a player's weight and his jersey number, but one is certainly not dependent on the other. Therefore, knowing a player's jersey number isn't particularly useful in predicting his weight.

Activity 10-17: Incorrect Conclusions

(a) The mean of the residuals must equal zero because one must sum up the residuals, then divide by the number of residuals in order to calculate the mean. Since the sum of the residuals will always equal zero, the mean must always be zero.
(b) The median of the residuals need not be zero, though it is possible. Finding the median does not involve summing up the residuals, so the fact that this sum would always be zero has no affect on the median.
(c) No, if the slope is negative, then the observation with the largest value of the predictor variable would have the smallest fitted value.
(d) Yes. An observation can occur at but have an extreme y value. As long as there as many other points near this observation may not have much affect on the location of the regression line.

Activity 10-18: Airfares (*cont.*)

Note: original regression equation: fare = $83.30 + $.117 * distance
(a) fare = $583.30 + $.117 * distance; The y-intercept increased by $500.
(b) fare = $167 + $.235 * distance; Both the y-intercept and the slope doubled.
(c) fare = $83.30 + $.235 * distance; The slope doubled.
(d) fare = -$34.10 + $.117 * distance; The y-intercept has decreased by 1,000 times the slope.