MATH 424: Assignment # 2 Chapter 3

Due on Monday, October 2, 2017

Kafai 11:10am

Jonathan Dombrowski

Contents

Question 8 **3**
(a) . . . 3
(b) . . . 3
(c) . . . 3
(d) . . . 3
(e) . . . 3
(f) . . . 3

Question 14 **4**
(a) . . . 4
(b) . . . 4
(c) . . . 4
(d) . . . 4

Question 20 **4**
(a) . . . 5
(b) . . . 5

Question 22 **5**
(a) . . . 5
(b) . . . 6
(c) . . . 6
(d) . . . 6

Question 28 **6**
(a) . . . 7
(b) . . . 8

Question 32 **9**
(a) . . . 10
(b) . . . 11

Question 40 **11**
(a) . . . 11
(b) . . . 11

Question 54 **11**
(setup) . . . 12
(a) . . . 13
(b) . . . 13
(c) . . . 14

Question 60 **14**
(a) . . . 15
b . . . 17

Question 8

Predicting home sales price. Real estate investors, homebuyers, and homeowners often use the appraised (or market) value of a property as a basis for predicting sale price. Data on sale prices and total appraised values of 76 residential prop- erties sold in 2008 in an upscale Tampa, Florida, neighborhood named Tampa Palms are saved in the TAMPALMS file. The first five and last five observations of the data set are listed in the accom- panying table.

(a) Propose a straight-line model to relate the appraised property value x to the sale price y for residential properties in this neighborhood.
(b) A MINITAB scatterplot of the data is shown on the previous page. [Note: Both sale price and total market value are shown in thousands of dollars.] Does it appear that a straight-line model will be an appropriate fit to the data?
(c) A MINITAB simple linear regression print- out is also shown (p. 100). Find the equation of the best-fitting line through the data on the printout.
(d) Interpret the y-intercept of the least squares line. Does it have a practical meaning for this application? Explain.
(e) Interpret the slope of the least squares line. Over what range of x is the interpretation meaningful?
(f) Use the least squares model to estimate the mean sale price of a property appraised at $300,000.

(a)

Proposed linear fit $= \hat{y} = 1.408x_1 + 1.359$

(b)

The linear model proposed does appear to fit the graph

(c)

Proposed linear fit $= \hat{y} = 1.36x_1 + 1.40827$

(d)

The y-intercept in this case represents the transaction price for a sale with the market value being zero. ie: for any transaction, an expense of $1408.27 will be added.

(e)

The positive slope of 1.36 is interpreted as the preice relationship between the market price and the sales price of a house in the area given the range of the slopes

(f)

y=1.36(300) + 1.40827

Question 14

Extending the life of an aluminum smelter pot. An investigation of the properties of bricks used to line aluminum smelter pots was published in the Amer- ican Ceramic Society Bulletin (February 2005). Six different commercial bricks were evaluated. The life length of a smelter pot depends on the porosity of the brick lining (the less porosity, the longer the life); consequently, the researchers measured the apparent porosity of each brick specimen, as well as the mean pore diameter of each brick. The data are given in the accompanying table
(a) Find the least squares line relating porosity(y) to mean pore diameter (x).
(b) Interpret the y-intercept of the line.
(c) Interpret the slope of the line.
(d) Predict the apparent porosity percentage for a brick with a mean pore diameter of 10 micrometers.

(a)

Proposed least squares fit : $\hat{y} = 0.950x_1 + 6.518$

(b)

The intercept appears to represent the Porosity at which the lifespan reaches zero, or the porostiy threshold at which pots are no longer used.

(c)

The slope represents the ratio between the porosity of the brick meausured in small local points vs. the porosity of the entire brick.

(d)

$\hat{y} = 6.3518 + 0.9498 \times 10 = 15.8498$
A brick with the mean pore diameter of 10 micrometers is predicted to have a porosity of 15.8498%.

Question 20

Extending the life of an aluminum smelter pot. Refer to the American Ceramic Society Bulletin (February 2005) study of bricks that line aluminum smelter pots, Exercise 3.14 (p. 103). You fit the sim- ple linear regression model relating brick porosity (y) to mean pore diameter (x) to the data in the SMELTPOT file.

(a) Find an estimate of the model standard deviation, σ.
(b) In Exercise 3.14d, you predicted brick poros- ity percentage when x = 10 micrometers. Use the result, part a, to estimate the error of prediction.

(a)

$$SSE = \sigma^2 = \frac{\sum_1^i (y_i - \hat{y}_i)^2}{n-2} = \sigma^2 \rightarrow \sigma = \sqrt{SSE} \tag{1}$$

$$s^2 = \frac{SSE}{d.f.} = \frac{40.551}{4} = 10.13; s = \sqrt{(10.13)} = 3.184 \tag{2}$$

(b)

The estimate for a bricks porosity when the observed porosity is $10\mu m$

$$f(10\mu m) = (15.850 \pm 3.184)\% \tag{3}$$

Question 22

Thermal characteristics of fin-tubes. A study was conducted to model the thermal performance of integral-fin tubes used in the refrigeration and process industries (Journal of Heat Transfer, August 1990). Twenty-four specially manufac- tured integral-fin tubes with rectangular fins made of copper were used in the experiment. Vapor was released downward into each tube and the vapor-side heat transfer coefficient (based on the outside surface area of the tube) was measured. The dependent variable for the study is the heat transfer enhancement ratio, y, defined as the ratio of the vapor-side coefficient of the fin tube to the vapor-side coefficient of a smooth tube evalu- ated at the same temperature. Theoretically, heat transfer will be related to the area at the top of the tube that is unflooded by condensation of the vapor. The data in the table are the unflooded area ratio (x) and heat transfer enhancement (y) values recorded for the 24 integral-fin tubes.

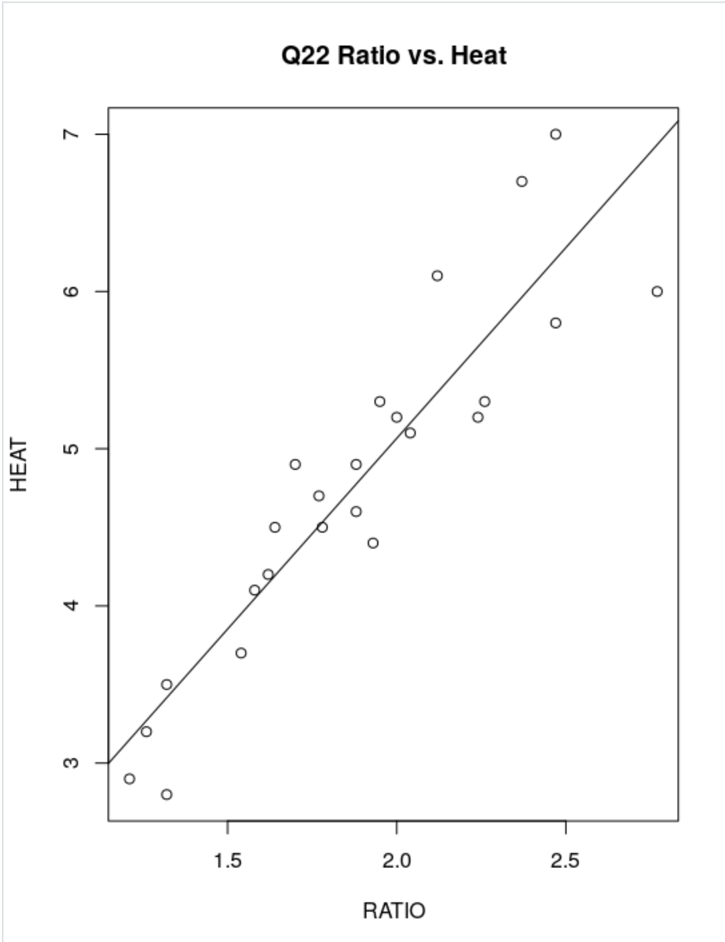
(a) Fit a least squares line to the data.
(b) Plot the data and graph the least squares line as a check on your calculations.
(c) Calculate SSE and s^2.
(d) Calculate s and interpret its value.

(a)

Coefficients:
(Intercept) ratio
0.2134 2.4264

$\hat{y} = 2.4264x_1 + 0.2134$

(b)



(c)

SSE = 4.531
$s2 = 0.206$

(d)

$s = \sqrt{s^2} = 0.454$ is the standard deviation of the linear model.
68% of values will fall within one σ of the mean, and 95% will fall within 2

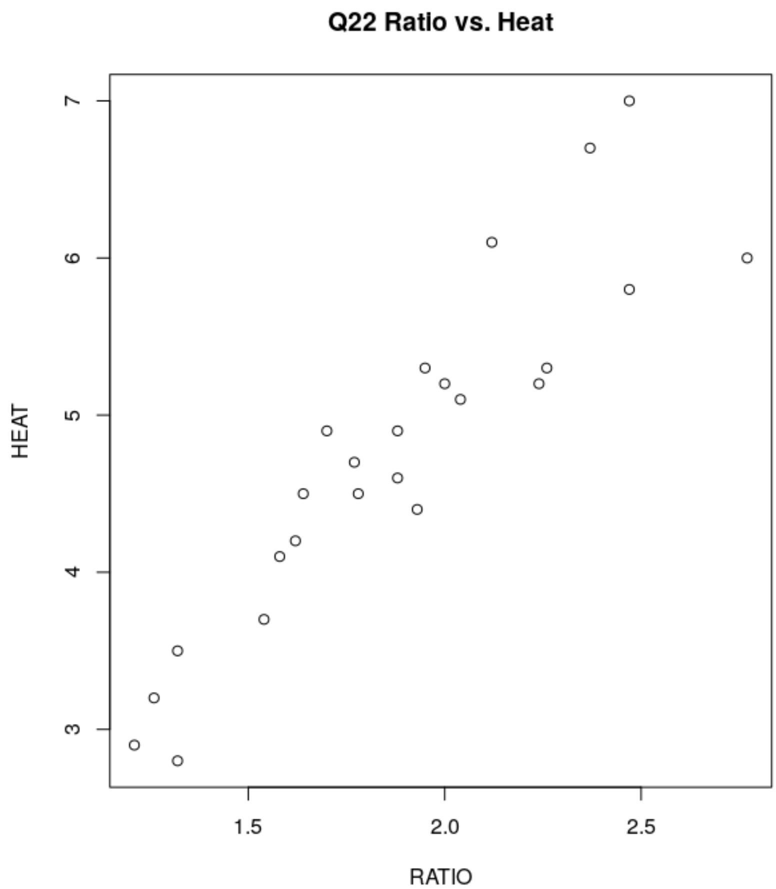
Question 28

Massage therapy for boxers. The British Jour- nal of Sports Medicine (April 2000) published a study of the effect of massage on boxing perfor- mance. Two variables measured on the boxers were blood lactate concentration (mM) and the boxers perceived recovery (28-point scale). Based on information provided in the article, the data in the table were obtained for 16 five-round boxing performances, where a massage was given to the boxer between rounds. Conduct a test to deter- mine whether blood lactate level (y) is linearly related to perceived recovery (x). Use $\alpha = .10$.

Question 28 continued on next page. . .

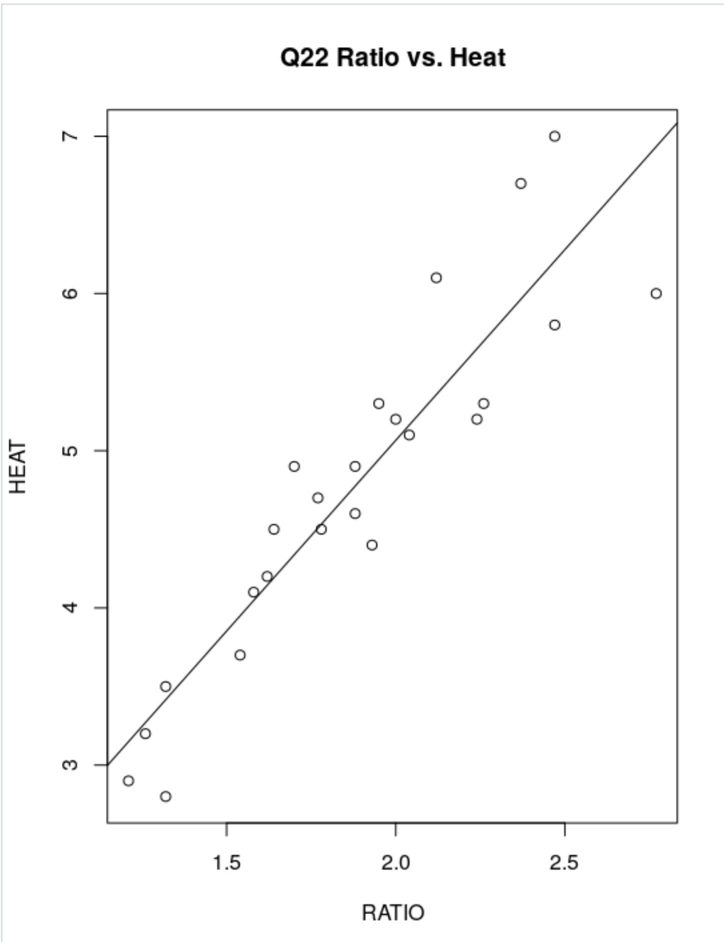
(a)

Based on the scatterplot, the data seems to have a weak positive trend between lactation and recovery rate.



(b)

$$H_o: \; \hat{R} = 0$$
$$H_a: \; \hat{R} \neq 0$$



Tests indicated that

$$R = 0.3251$$
$$R^2 = 0.2769$$
$$p - value : 0.0211 \implies$$

The p-value is small ($p - value \leq (\alpha = 0.1)$) which means that we must reject H_0.

Conclusion:
The correlation between the two variables is non-zero and estimated to be 90% certain that the level of correlation is between 0.189±

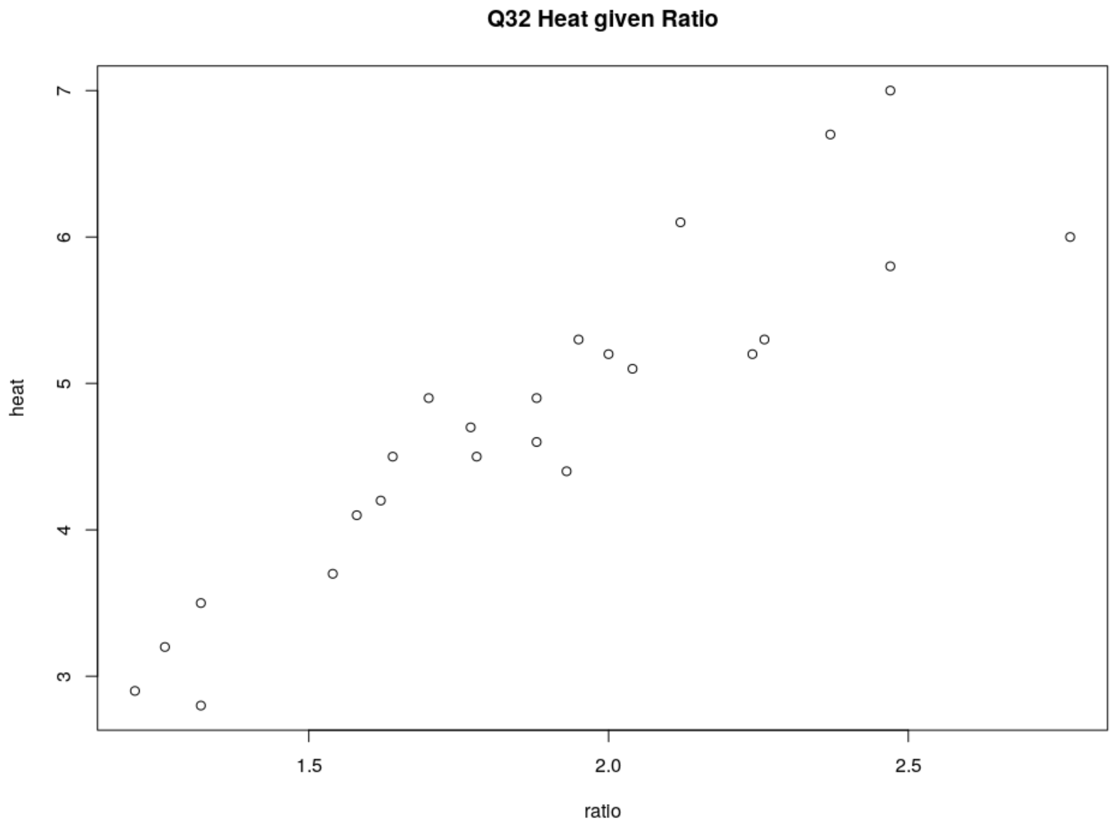
Question 32

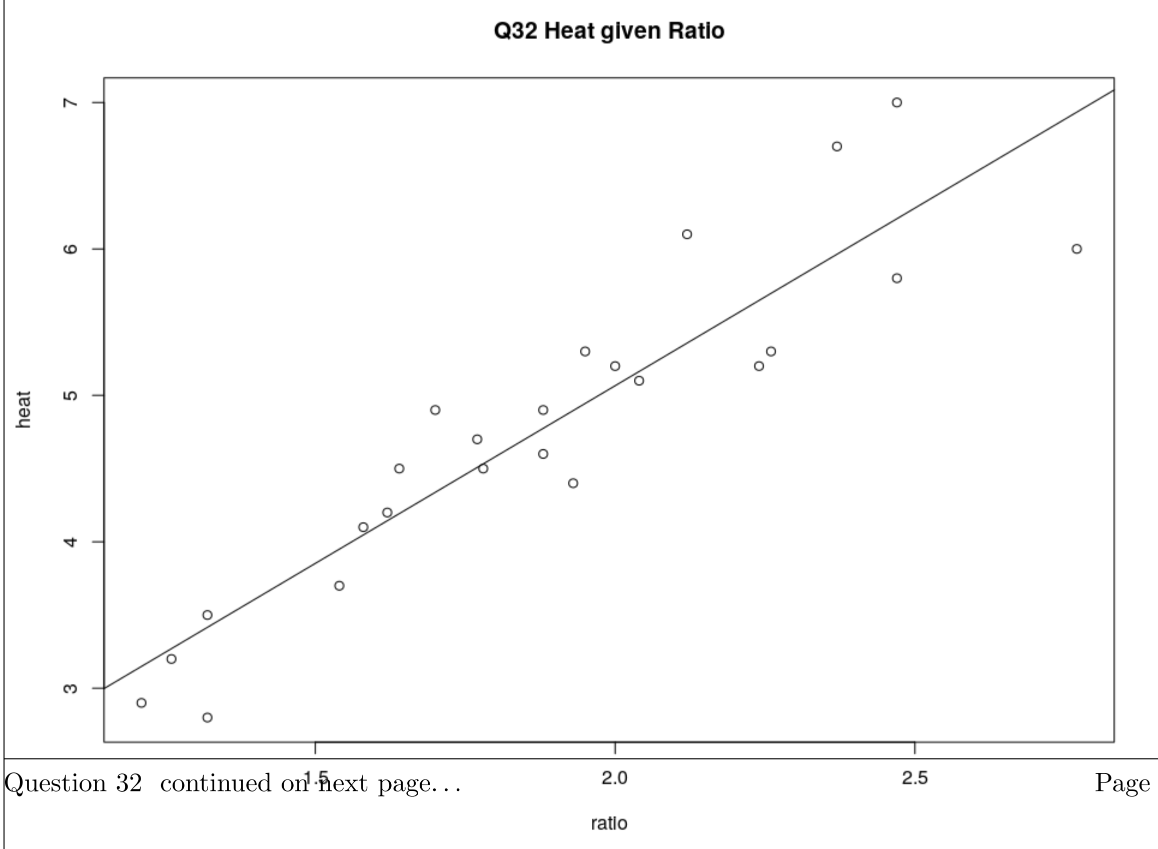
Thermal characteristics of fin-tubes. Refer to the Journal of Heat Transfer study of the straight-line relationship between heat transfer enhancement (y) and unflooded area ratio (x), Exercise 3.22 (p. 109). Construct a 95% confidence interval
Data set : HEAT

Question 32 continued on next page. . .

(a)

Plotting heat on the y-axis and ratio on the x-axis, a clear, positive correlation can be inferred. Testing for $\hat{R} = (0.81, 0.96)$, which is reasonable enough to continure.





(b)

The linear model approximation yields
$\beta 0 = 0.233 \pm 0.159$
$\beta 1 = 0.345 \pm 0.0325$
95% Confidence interval for $\beta_1 : (0.313, 0.378)$

Question 40

Predicting home sales price. Refer to the data on sale prices and total appraised values of 76 residential properties recently sold in an upscale Tampa, Florida, neighborhood, Exercise 3.8 (p. 100). The MINITAB simple linear regression printout relat- ing sale price (y) to appraised property (market) value (x) is reproduced on the next page, followed by a MINITAB correlation printout.
(a) Find the coefficient of correlation between appraised property value and sale price on the printout. Interpret this value.
(b) Find the coefficient of determination between appraised property value and sale price on the printout. Interpret this value.

(a)

$H_0 : \hat{\rho} = 0$
$H_a : \hat{\rho} \neq 0$
p-value from test: 2.2e-16.

P is incredibly small therefore, we must reject H_0
Proposed Correlation Coefficient: 0.9755
Proposed 95% confidence interval for $\hat{R}$

(b)

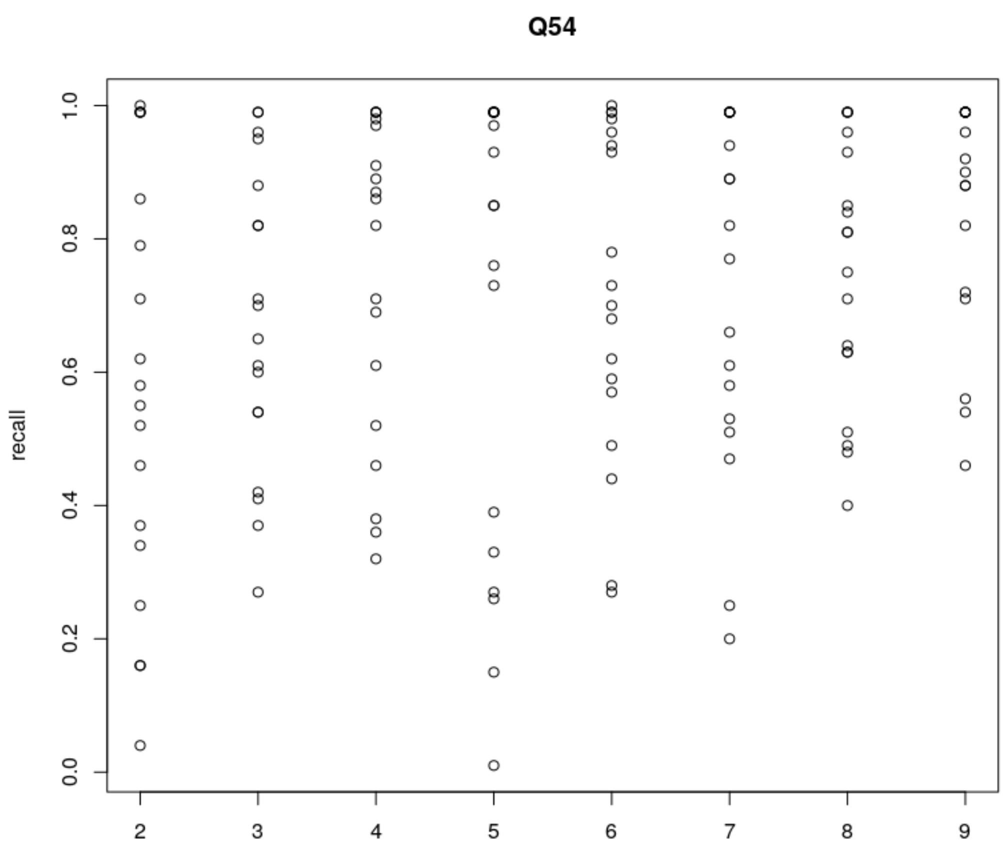
Coefficient of Determination $= R^2 = 0.9755^2 = 0.9516$

Question 54

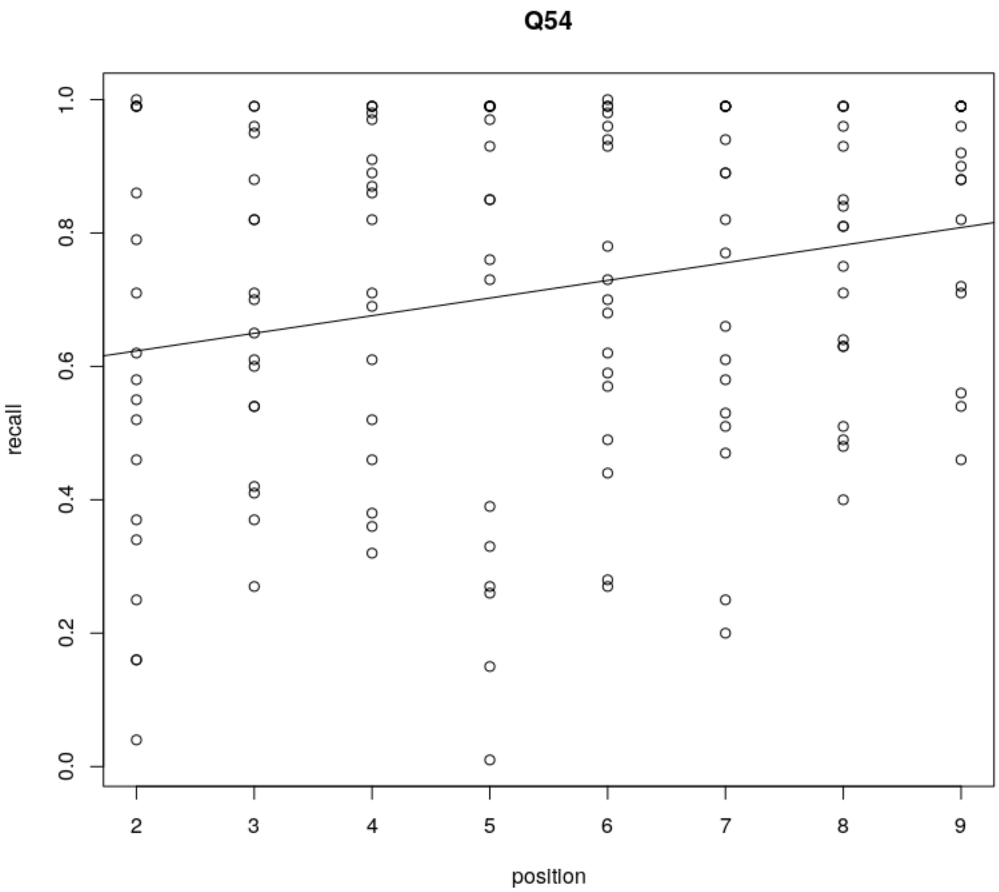
Recalling student names. Refer to the Journal of Experimental PsychologyApplied (June 2000) name retrieval study, Exercise 3.15 (p. 103).
(a) Find a 99% confidence interval for the mean recall proportion for students in the fifth position during the name game. Interpret the result.
(b) Find a 99% prediction interval for the recall proportion of a particular student in the fifth position during the name game. Interpret the result.
(c) Compare the two intervals, parts a and b. Which interval is wider? Will this always be the case? Explain.

Question 54 continued on next page. . .

(setup)



a- The initial scatter plot does not seem to have any kind of meaningful correlation.



b- After constructing a linear regression, the line still does not seem to correlate.
Testing for correlation ($\hat{R}$): $H_0 : \rho = 0$
$H_a : \rho \neq 0$
p-value = 0.0049, therefore we must reject H_0
Conclusion from correlation test is that ρ is not equal to 0, but after constructing a 99% confidence interval for the value: (0.0207, 0.4256), the correlation is weak at best

(a)

$\boxed{\text{Confidence interval for students mean recall proportion position 5 is : } (0.453, 0.929);\ 0.691 \pm 0.238}$

(b)

$\boxed{\text{Prediction interval for students mean recall proportion in position 5 is : } (0.198, 1.207);\ 0.703 \pm 0.501}$

(c)

The prediction interval is wider by a $2(E_2 - E_1) = 2(0.263) = 0.526$, which comes from the 1 that appears in the prediction interval and not the confidence interval

$$Confidence\ Interval \; = \; \hat{y} \pm (t_{\alpha/2})s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \tag{4}$$

and

$$Prediction\ Interval \; = \; \hat{y} \pm (t_{\alpha/2})s\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \tag{5}$$

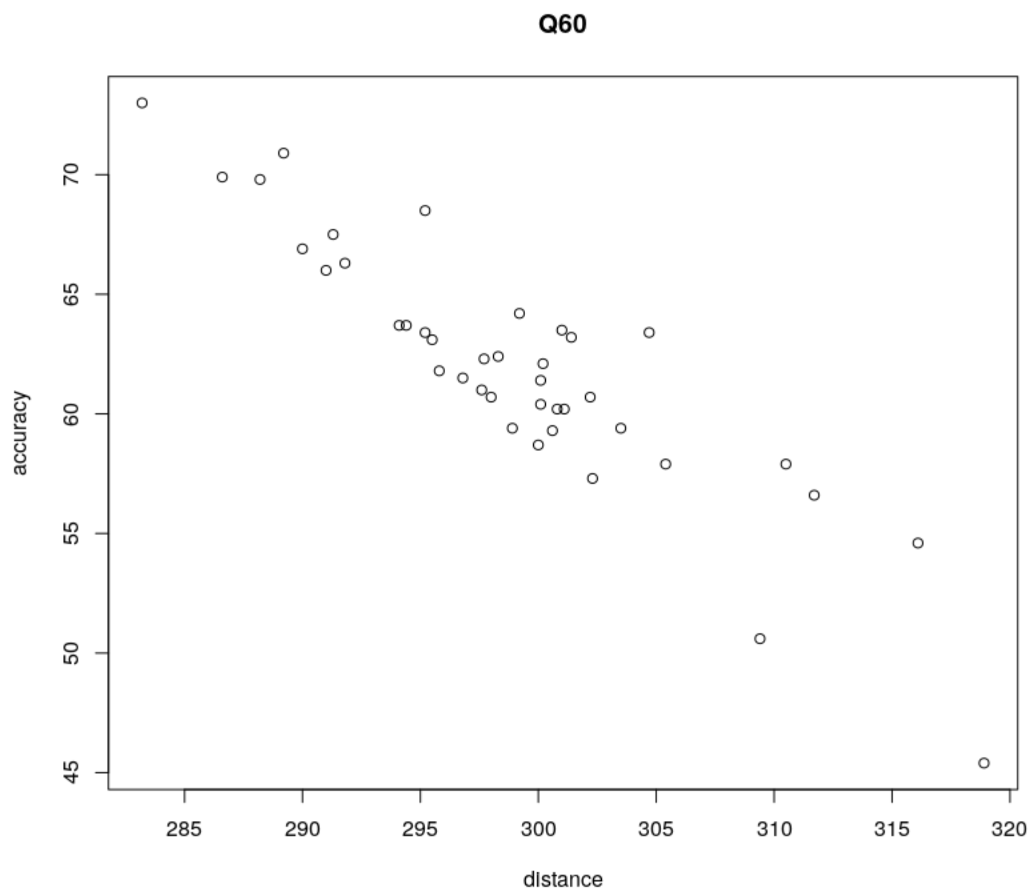
Therefore, the prediction interval will always be larger than the confidence interval.

Question 60

Ranking driving performance of professional golfers. A group of Northeastern University researchers developed a new method for rank- ing the total driving performance of golfers on the Professional Golf Association (PGA) tour (Sport Journal, Winter 2007). The method requires know- ing a golfers average driving distance (yards) and driving accuracy (percent of drives that land in the fairway). The values of these two variables are used to compute a driving performance index. Data for the top 40 PGA golfers (as ranked by the new method) are saved in the PGADRIVER file. (The first five and last five observations are listed in the table below.) A professional golfer is practicing a new swing to increase his average driving distance. However, he is concerned that his driving accuracy will be lower. Is his concern a valid one? Use simple linear regression, where y = driving accuracy and x = driving distance, to answer the question.

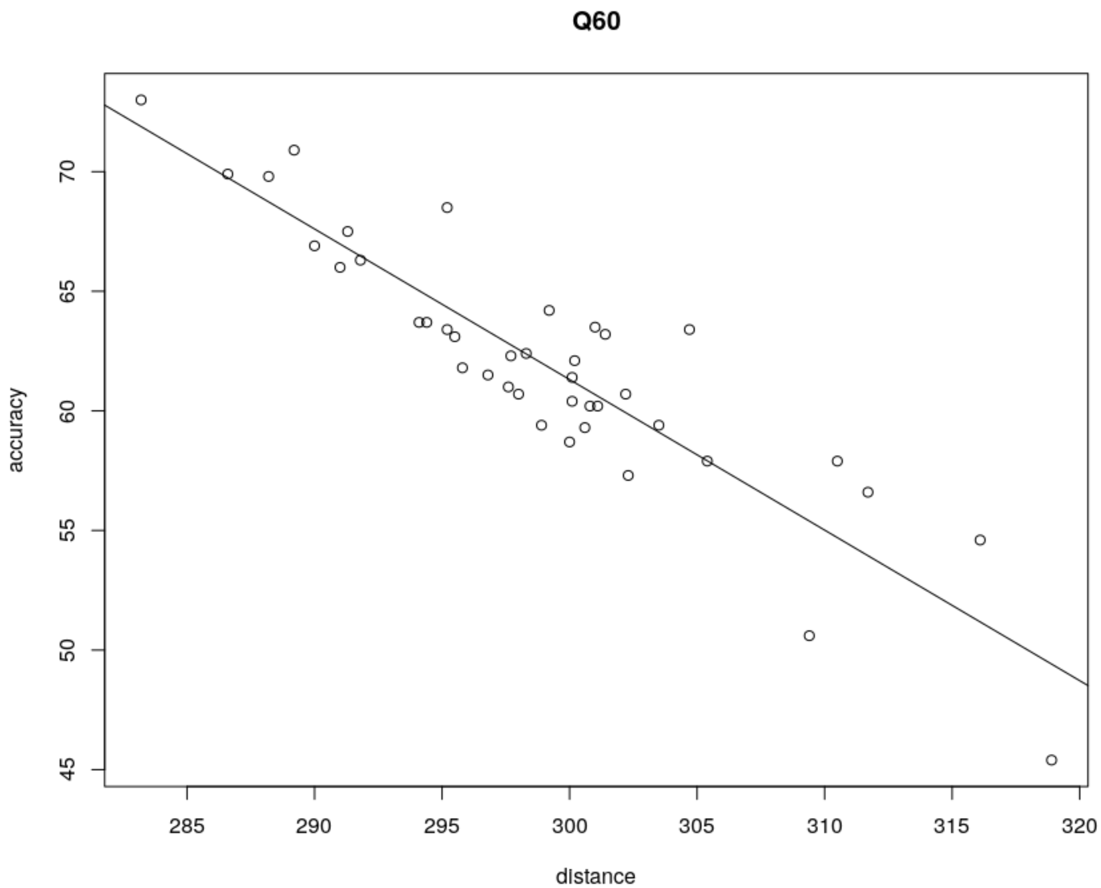
Question 60 continued on next page. . .

(a)



The Scatter plot appears to have a clear negative correlation
$H_0\ R = 0$
$H_a\ R \neq 0$
Test results show a p-value = 8.478e-16, therefore very small, which means it is necessary to reject H_0. The 95% confidence interval for this correlation is (-0.950, -0.829). With this degree of certainty in correlation, We can proceed to anaylsis.

Linear Regression:



Estimates:
$\hat{\beta}_0 = 250.142 \pm 14.231$
$\hat{\beta}_1 = -0.629 \pm 0.048$
$\hat{R}^2 = 0.8216$
$\hat{R} = (-0.950, -0.829)$

Assuming for Error:
(1) $\text{E}(\epsilon) = 0$
(2) $\text{Var}(\epsilon) =$ 2 is constant for all x-values
(3) ϵ has a normal distribution
(4) ϵ s are independent

$$s^2 = \frac{SSE}{d.f.} = \frac{874.99}{38} = 23.03 \tag{6}$$

s = 4.799

b

We are 95% confident that the following interval includes the mean decrease of accuracy per unit distance for golfing accuracies.
$\hat{\beta}_1 = -0.629 \pm 0.048$

In conclusion:
The golfers' concern of increasing his average driving distance, only to have his accuracy fall is a valid one. The highly correlated linear model that explains 95% of the values depicts a model in which:
$\hat{y} = (-0.629x_1 + 250.142)\%$, which corresponds to driving accuracy for a swing; accurate between 283.2 yards and 318.2 yards. For every yard that the ball travels, the golfer loses 0.629% accuracy. The concern is valid.