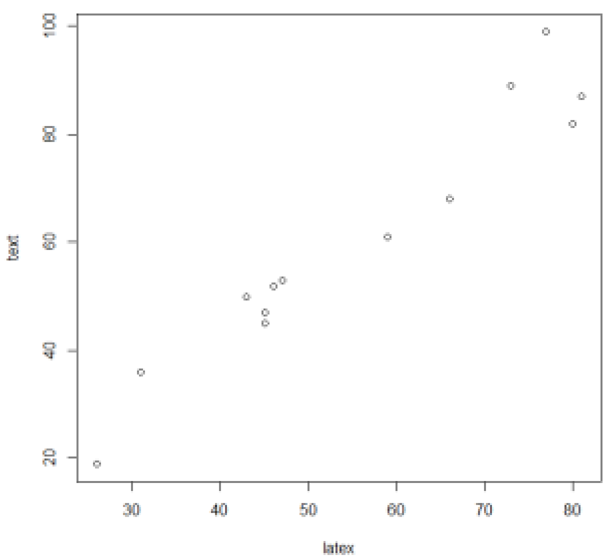
STAT 350 Review Set (#3)

Make sure to study Lab 8 software output and questions.

1. The editor of a statistics textbook would like to plan for the next edition. A key variable is the number of pages that will be in the final version. Text files are prepared by the authors using LaTeX, and separate files contain figures and tables. For the previous edition of the textbook, the number of pages in the LaTeX files can easily be determined, as well as the number of pages in the final version of the textbook. Here are the data:

	Chapter												
	1	2	3	4	5	6	7	8	9	10	11	12	13
LaTeX pages	77	73	59	80	45	66	81	45	47	43	31	46	26
Text pages	99	89	61	82	47	68	87	45	53	50	36	52	19

Scatter Plot



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-6.20176	5.71233	-1.086	0.301
latex	1.20810	0.09828	xxx	0

Multiple R-squared: 0.9321, Adjusted R-squared: 0.926

STAT 350 Review Set (#3)

a) What is the explanatory variable? Response variable?

LaTex Page

b) Describe the overall pattern of the scatter plot above.

Strong Linear Positive relationship between textbook page and LaTex Page

c) Find the equation of the least-squares regression line based on the software output.

TextPages = - 6.20 + 1.21 LaTeXPages

d) Interpret the slope of the regression line.

c) Find the predicted number of pages for the next edition if the number of LaTeX pages for a chapter is 62.

TextPages = - 6.20 + 1.21 *62 = 68.82 => 69

d) What proportion of the variation in Text pages is explained by LaTex pages?

Multiple R-squared: 0.9321, **2 = 93.2%**

The proportion of the variation in Text pages is explained by LaTex pages is 93.2%

e) What is the value of the correlation coefficient r?

0.97

f) Find a 90% confidence interval for the slope β_1

df = 13 – 2 = 11. T* = 1.796

1.2081 +/- 1.796*0.09828

g) Suppose there are 80 LaTex pages in a chapter for the new edition. Raj obtained a 95% confidence interval for the mean pages and a 95% prediction interval for the pages of the final version, but did not label them. Which of the following two is the prediction interval?

______Prediction_______ Interval (75, 106)

______Confidence_______ Interval (84, 97)

h) Which interval should the editor use to plan for the next edition of the statistics textbook?
Prediction Interval

i) One new chapter has 200 LaText pages. Discuss if it is appropriate use the regression line obtained in part (c) to predict the final textbook pages.
Not really. We need to be careful about extrapolation.

STAT 350 Review Set (#3)

2. A study shows that there is a positive correlation between the size of a hospital (measured by its number of beds x) and the median number of days y that patients remain in the hospital. Does this mean that you can shorten a hospital stay by choosing a small hospital? Why?

No. Correlation does not equal to Causation. Sicker patients tend to go to larger hospitals, and stay longer.

3. Data show that married, divorced, and widowed men earn quite a bit more than men the same age who have never been married. This does not mean that a man can raise his income by getting married, because men who have never been married are different from married men in many ways other than marital status. Suggest several lurking variables that might help explain the association between marital status and income.

Age.

4. (10.44) Can a pretest on mathematics skills predict success in a statistics course? The 62 students in an introductory statistics class took a pretest at the beginning of the semester. The least-squares regression line for predicting the score y on the final exam from the pretest score x was = 13.8 + 0.81x. The standard error of b1 was 0.43.

(a) Test the null hypothesis that there is no linear relationship between the pretest score and the score on the final exam against the two-sided alternative.

1) H0: slope = 0 vs H1: slope != 0

2) test statistic t = 0.81/0.43 = 1.88, with DF = 60

3) P-value in (0.025*2 , 0.05*2) = (0.05, 0.10)

4) At significance level 0.05, we fail to reject the null. Concluding that there is no linear relationship between the two scores.

(b) Would you reject this null hypothesis versus the one-sided alternative that the slope is positive? Explain your answer.

Yes, we would reject the null for one-sided alternative. Because the p-value is between 0.025 and 0.05.

STAT 350 Review Set (#3)

5. (10.39, 10.40, 10.41) The Leaning Tower of Pisa is an architectural wonder. Engineers concerned about the tower's stability have done extensive studies of its increasing tilt. Measurements of the lean of the tower over time provide much useful information. The following table gives measurements for the years 1975 to 1987. The variable "lean" represents the difference between where a point on the tower would be if the tower were straight and where it actually is. The data are coded as tenths of a millimeter in excess of 2.9 meters, so that the 1975 lean, which was 2.9642 meters, appears in the table as 642. Only the last two digits of the year were entered into the computer.[17]

Year	75	76	77	78	79	80	81	82	83	84	85	86	87
Lean	642	644	656	667	673	688	696	698	713	717	725	742	757

A. Does the trend in lean over time appear to be linear?

Yes

B. What is the equation of the least-squares line? What percent of the variation in lean is explained by this line?

Lean = -61.12 + 9.32 Year

C. Give a 99% confidence interval for the average rate of change (tenths of a millimeter per year) of the lean.

DF = 13 – 2 = 11 → t* = 3.1058

9.3187 +/- 3.1058 x 0.3099 = (8.36, 10.28)

D. In 1918 the lean was 2.9071 meters. (The coded value is 71.) Using the least-squares equation for the years 1975 to 1987, calculate a predicted value for the lean in 1918. (Note that you must use the coded value 18 for year.)

Lean(18) = 107

107 mm beyond 2.9 m → 2.9107 m

E. Although the least-squares line gives an excellent fit to the data for 1975 to 1987, this pattern did not extend back to 1918. Write a short statement explaining why this conclusion follows from the information available. Use numerical and graphical summaries to support your explanation.

This is an example of extrapolation.

F. How would you code the explanatory variable for the year 2013?

113 = 2013 - 1900

G. The engineers working on the Leaning Tower of Pisa were most interested in how much the tower would lean if no corrective action was taken. Use the least-squares equation to predict the tower's lean in the year 2013. (*Note:* The tower was renovated in 2001 to make sure it does not fall down.)

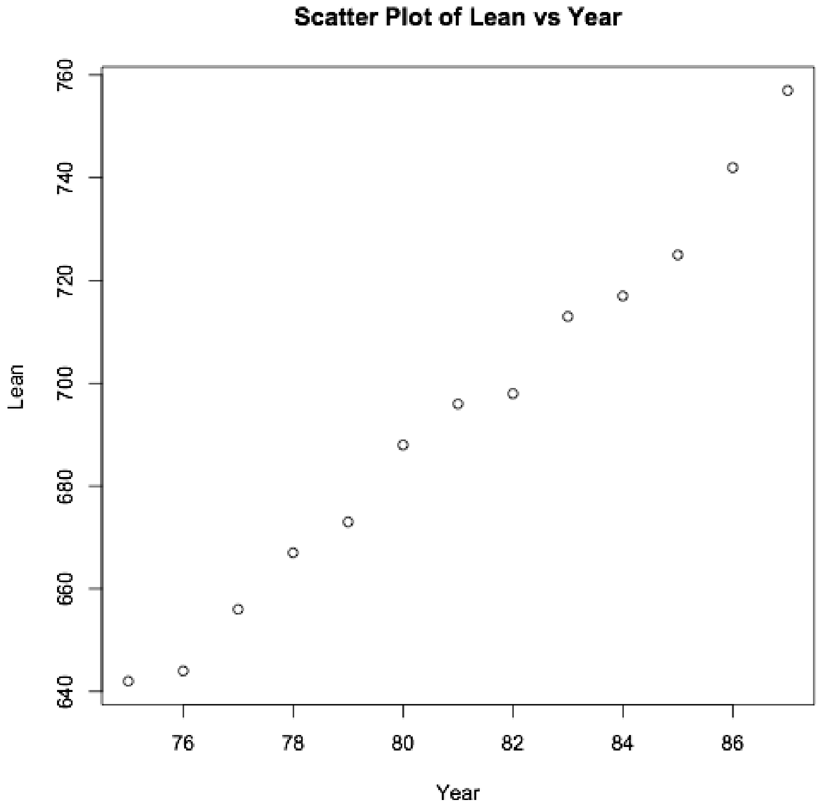
Lean(113) = 992

992 mm beyond 2.9 m → 3.892 m

H. To give a margin of error for the lean in 2013, would you use a confidence interval for a mean response or a prediction interval? Explain your choice.

Prediction Interval, since we are predicting for one specific year.

STAT 350 Review Set (#3)



Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-61.1209	25.1298	xxxx	0.0333 *
year	9.3187	0.3099	xxxx	6.5e-12 ***

s = 4.181, R-squared = 0.988