Example 10.8: Metatarsus adductus (call it MA) is a turning in of the front part of the foot that is common in adolescents and usually corrects itself. Hallux abducto valgus (call it HAV) is a deformation of the big toe that is not common in youth and often requires surgery. Perhaps the severity of MA can help predict the severity of HAV. Table 2.3 (page 120) gives data on **38** consecutive patients who came to a medical center for HAV surgery. Using X-rays, doctors measured the angle of deformity for both MA and HAV. They speculated that there is a positive association – more serious MA is associated with more serious HAV.

Give a 95% confidence interval for the slope. Explain how this interval can tell you what to conclude from a significance test for this parameter.

Solution:

The least squares estimate for β_1 is $b_1 = 0.338835$.

The value of $SE(b_1)$ is
$s/\sqrt{(Sxx)} = 7.2237 / \sqrt{(1643.711)} = .17818$.

The t-value we need is $t_{.025,36} = 2.031$.

So a 95% CI for β_1 is
$.338835 \pm (.17818)(2.031) = (-.023, .700)$.

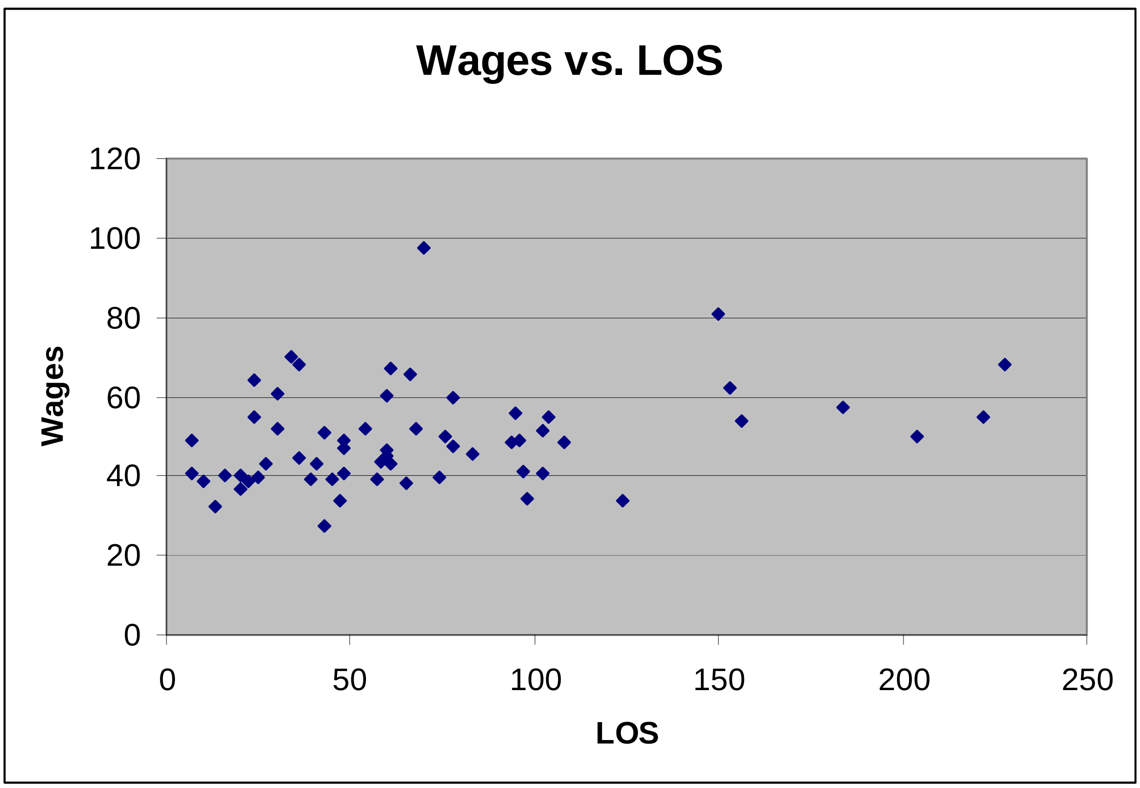
Since 0 is in this interval, we can conclude that the p-value is greater than .05.

Example 10.9: We assume that our wages will increase as we gain experience and become more valuable to our employers. Wages also increase because of inflation. By examining a sample of employees at a given point in time, we can look at part of the picture. How does length of service (LOS) relate to wages? Table 10.1 gives data on the LOS on the LOS in months and wages for 60 women who work in Indiana banks. Wages are yearly total income divided by the number of weeks worked. We have multiplied wages by a constant for reasons of confidentiality.

(a) Plot wages versus LOS. Describe the relationship. There is one woman with relatively high wages for her length of service. Circle this point and do not use it in the rest of this exercise.
(b) Find the least-squares line. Summarize the significance test for the slope. What do you conclude?
(c) State carefully what the slope tells you about the relationship between wages and length of service.
(d) Give a 99% confidence interval for the slope.

Solution:

(a) The relationship seems to be weak and positive from looking at the plot (graph on next page)



(b) Sxx = 157856.7, Sxy = 11564.37,
so b_1 = Sxy/Sxx = 11564.37/157856.7 = 0.0733.
b_0 = Ybar - b_1Xbar = 48.54636 - 0.073*70.49153 = 43.4.

So Wages = 43.4 + 0.0733 LOS.

The test statistic is $t = b_1/SE(b_1) = 2.85$ with df = 57.

This yields a p-value of 0.006, which tells us that we have reason to believe that the slope parameter is significantly different from 0.

(c) This slope tells us that wages increase by 0.0733 for every one unit increase in LOS.

(d) $SE(b_1) = .0257$. $t_{.005,57} \approx 2.66$.
So a 99% CI for the slope parameter is
0.073 +/- (.0257)(2.66) = (.005, .14).

Example 10.10: Refer to the previous exercise. Analyze the data with the outlier included.

(a) How does this change the estimates of the parameters β_0, β_1, and σ?
(b) What effect does the outlier have on the results of the significance test for the slope?
(c) How has the width of the 99% confidence interval changed?

Solution:

(a) The new estimated equation is Wages = 44.2 + .0731 LOS. The value of s changed from 10.21 to 11.98 with the outlier included.

(b) The new test statistic is $t = b_1/(s/\sqrt{(Sxx)}) = 2.42$ with df = 58, which yields a p-value of 0.018.

(c) The width of the confidence interval will change by the same factor as the change in s.

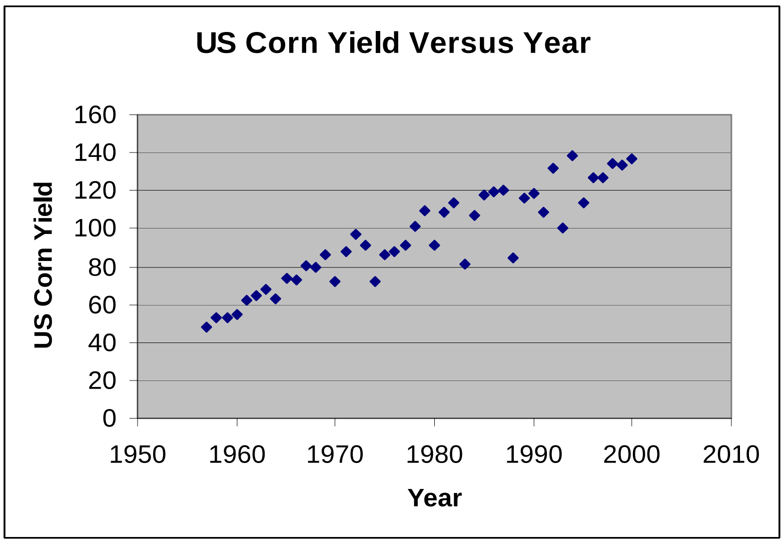
Example 10.11: In Example 10.8 we examined the yield in bushels per acre of corn for the years 1966, 1976, 1986, and 1996. Data for all years between 1957 and 2000 appear in Table 10.2.

(a) Plot yield versus year. Describe the relationship. Are there any outliers or unusual years.

(b) Perform the regression analysis and summarize the results. How rapidly has yield increased over time?

Solution:

(a)



The relationship seems to be positive and strong. There don't seem to be any outliers. There were some low yields around the mid-80's though.

(b) The equation we obtain is Yield = -3571.67 + 1.8533 year. We obtain a t-statistic of t = 15.92 with df = 42, which yields a p-value that is essentially 0. The value of s is 9.808. The yield has grown an estimated 1.8533 per year.

Example 10.12: Utility companies need to estimate the amount of energy that will be used by their customers. The consumption of natural gas required for heating homes depends on the outdoor temperature. When the weather is cold, more gas will be consumed. A study of one home recorded the average daily gas consumption y (in hundreds of cubic feet) for each month during one heating season. The explanatory variable x is the average number of heating degree-days per day during the month. One heating degree-day is accumulated for each degree a day's average temperature falls below 65 degrees Fahrenheit. An average temperature of 50 degrees, for example, corresponds to 15 degree-days. The data for October through June are given in the table on page 695.

(a) Find the equation of the least-squares line.
(b) Test the null hypothesis that the slope is zero and describe your conclusion.
(c) Give a 90% confidence interval for the slope.
(d) The parameter β_0 corresponds to natural gas consumption for cooking, hot water and other uses where there is no demand for heating. Give a 90% confidence interval for this parameter.

Solution:

(a) Gas Consumption = 1.232354 + .202212 Degree Days.

(b) We obtain a t statistic of t = .202212/.011448 = 17.7 with df = 7. The p-value is essentially 0, which gives of strong evidence of an association.

(c) A 90% CI for the slope is .202212 +/- (.011448)$t_{0.05,7}$ = .202212 +/- (.011448)(1.895) = (0.1805, 0.2239).

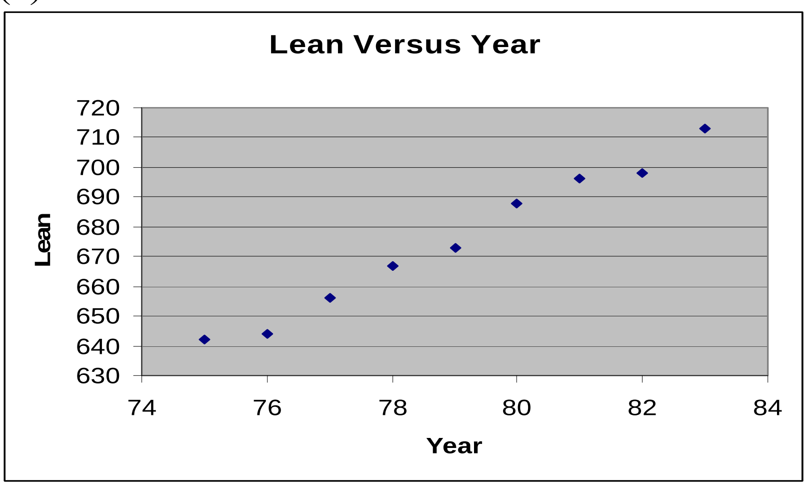
(d) A 90% CI for the intercept is 1.232354 +/- (1.895)(.286037) = (0.690, 1.774).

Example 10.13: 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 in the computer.

(a) Plot the data. Does the trend in lean over time appear to be linear?
(b) What is the equation of the least-squares line? What percent of the cariation in lean is explained by this line?
(c) Give a 99% confidence interval for the average rate of change (tenths of a millimeter per year) of the lean.

Solution:

(a)



This trend seems to be linear over time.

(b) Lean = -61.12 + 9.32 Year. The value of R-square is .988.

(c) A 99% confidence interval for the slope parameter is 9.318681 +/- (.309914)(3.106) = (8.356, 10.281).

Example 10.14: Refer to the previous exercise.

(a) 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).

(b) 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.

Solution:

(a) The predicted value is -62.12 + 9.32(18) = 106.64.

(b) This is called extrapolation – using the data to predict well outside the range of the x values.

A 95% prediction interval when X = 18 is 62.6 to 150.7.

The width tells how unreliable the prediction is.