

Chapter 11 Regression Analysis

Manatees are large, gentle sea creatures that live along the Florida coast. Many manatees are killed or injured by powerboats. Here are data on powerboat registrations (in thousands) and the number of manatees killed by boats in Florida in the years 1977 to 1990:

<u>year</u>	<u>boats</u>	<u>deaths</u>
1977	447	13
1978	460	21
1979	481	24
1980	498	16
1981	513	24
1982	512	20
1983	526	15
1984	559	34
1985	585	33
1986	614	33
1987	645	39
1988	675	43
1989	711	50
1990	719	47

1. Use your calculator to draw a scatterplot of the number of manatees killed by boats against the number of powerboat registrations, in thousands. Is the pattern roughly linear? Are there any clear outliers?
2. Use your calculator to find the correlation.
3. Use your calculator to find the regression line where the explanatory variable is the number of powerboats registered and the response variable is the number of manatees killed.
4. Use your model from question 3 to find the number of manatees killed by boats a year in which there are 716,000 powerboats registered.

5. Is there are positive relationship between manatees killed and number of powerboats registered? Run a test of significance at the 5% significance level to answer this question.

- ▶ Enter the data in L1 and L2.
- ▶ Go to the linear regression test on the stat test menu (letter E).
- ▶ Setup the window as shown on the right (top).
- ▶ Read the results as shown on the right (bottom).

```
LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
 $\beta$  &  $\rho$ : $\neq 0 < 0$  
RegEQ:
Calculate
```

```
LinRegTTest
y=a+bx
 $\beta > 0$  and  $\rho > 0$ 
t=9.675470539
P=2.5545306E-7
df=12
↓a=-41.43043895
```

Here is the result of asking Minitab to do prediction for $x^* = 716$:

Fit	SE Fit	95% C.I.	95% P.I.
47.97	2.23	(43.11, 52.83)	(37.46, 58.48)

6. Give a 95% interval for the number of manatees that would be killed in a future year if boat registrations remained at 716,000.
7. Give a 95% interval for the mean number of manatees killed in all years where boat registrations were 716,000.
8. Give a 90% interval for the mean number of manatees killed in all years where boat registrations were 716,000.

Homework Problems

1. Is there a positive relationship between the length of time of an eruption (or duration) of Old Faithful and the time between eruptions? To answer this question, the following data was obtained. (All the numbers represent minutes.)

Duration	4.4	3.9	4.0	4.0	3.5	4.1	2.3	4.7	1.7	4.9	1.7	4.6	1.7
Time Between	78	74	68	76	80	84	50	93	55	76	58	74	56

- a) Find a regression equation where duration is the explanatory variable and the time between eruptions is the response variable.
- b) Do a test of significance to see if there is a positive relationship between the length of time of an eruption (or duration) and the time between eruptions. Test this at the $\alpha = 0.05$ level by writing out the hypotheses, finding a test statistic, finding a P-value, and writing a conclusion.

We wish to predict the time between eruptions for all durations that lasted 4 minutes. We did this using Minitab with the following output.

Duration	Fit	SE Fit	95.0% CI	95.0% PI
4.0	75.39	2.12	(70.73, 80.06)	(59.24, 91.55)

- c) Suppose we wish to estimate the mean amount of time between eruptions for all durations that lasted 4 minutes. Give a 95% interval for this estimate.
- d) Suppose we wish to estimate the amount of time between eruptions for all durations that lasted 4 minutes. Give a 95% interval for this estimate.

2. Does the time it takes students to take a test help predict the score on the test? Here are data on 13 test takers in a general education mathematics class. Time is given in minutes and the score is their number of points out of 100. Assume this is a simple random sample of all students taking a general education mathematics course.

Time	56	48	43	58	54	41	56	58	61	41	62	78	79
Score	100	99	90	89	84	84	65	83	92	94	74	68	72

- a) Find a regression equation for predicting the score on the test given the time spent taking it.
- b) Is there a negative relationship between the time spent taking a test and the score on the test for students in general education mathematics courses? Test this at the 5% level and give your hypotheses, test statistics, P -value, and conclusion.

We wish to predict the score for students that take 75 minutes to take the test. We did this using Minitab with the following results.

Fit	SE Fit	95% CI	95% PI
73.81	4.99	(62.83, 84.79)	(50.00, 97.62)

- c) Suppose we wish to estimate the score for a student that takes 75 minutes to take the test. Give a 95% interval for this.
- d) Suppose we wish to estimate the mean score for students that take 75 minutes to take the test. Give a 95% interval for this.