MAT 1372 Statistics with Probability classwk 15 Fall 2012

# 12.2 SIMPLE LINEAR REGRESSION MODEL

Consider a pair of variables, one of which is called the *input variable* and the other the *response variable*. For a specified value *x* of the input variable, we express the value of the response variable

*Y* = A + B*x* + *e*.

The quantities A and B are parameters or constants that we would like to determine. The e is the *error* which varies for each data point and is the vertical signed distance between the data point and the line. Regression is the process of finding the A and B which minimize the errors in some way.

**2.** An area manager in a department store wants to study the relationship

between the number of workers on duty and the value of merchandise

lost to shoplifters. To do so, she assigned a different number of clerks

for each of 10 weeks. The results were as follows:

**Week #workers Loss**

1 9 420

2 11 350

3 12 360

4 13 300

5 15 225

6 18 200

7 16 230

8 14 280

9 12 315

10 10 410

**3.** The following data relate the traffic density, described in the number

of automobiles per mile, to the average speed of traffic on a moderately

large city thoroughfare. The data were collected at the same location at

10 different times within a span of 3 months.

**Density Speed**

69 25.4

56 32.5

62 28.6

119 11.3

84 21.3

74 22.1

73 22.3

90 18.5

38 37.2

22 44.6

**(a)** Which variable is the input and which is the response?

**(b)** Draw a scatter diagram.

**(c)** Does a simple linear regression model appear to be reasonable?

**4.** Repeat Prob. 3, but now let the square root of the speed, rather than the

speed itself, be the response variable.

**12.3 ESTIMATING THE REGRESSION PARAMETERS**

The most common way of collecting the errors is to sum their squares, called least-square estimation.



For given data pairs , we want to find values of A and B which minimize



It can be shown (via some multivariable calculus) that the slope



does the trick.

To find the y intercept A, it turns out that we want the line to pass through the point whose coordinates consist of the respective averages , i.e.,



Solving for A, we get

