Statistics 333: Applied Regression Analysis

January 25, 2007

Announcements:

- Some course materials will be uploaded in the course website soon. http://www.stat.wisc.edu/Department/Courses/2006.2.html
- The issue with the course enrollment will be handled soon.
- Any other issues? Please let Prof. Johnson know your issues by email rich@stat.wisc.edu.

Announcements:

- Some course materials will be uploaded in the course website soon. http://www.stat.wisc.edu/Department/Courses/2006.2.html
- The issue with the course enrollment will be handled soon.
- Any other issues? Please let Prof. Johnson know your issues by email rich@stat.wisc.edu.

Announcements:

- Some course materials will be uploaded in the course website soon. http://www.stat.wisc.edu/Department/Courses/2006.2.html
- The issue with the course enrollment will be handled soon.
- Any other issues? Please let Prof. Johnson know your issues by email rich@stat.wisc.edu.

Announcements:

- Some course materials will be uploaded in the course website soon. http://www.stat.wisc.edu/Department/Courses/2006.2.html
- The issue with the course enrollment will be handled soon.
- Any other issues? Please let Prof. Johnson know your issues by email rich@stat.wisc.edu.

Agenda for Today's Lecture:

Plan to cover:

1. Overview for Simple Regression
2. Introduction to R

Agenda for Today's Lecture:

Plan to cover:

1. Overview for Simple Regression
2. Introduction to R

Agenda for Today's Lecture:

Plan to cover:

1. Overview for Simple Regression
2. Introduction to R

Agenda for Today's Lecture:

Plan to cover:

1. Overview for Simple Regression
2. Introduction to R

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable

Regression with a Single Predictor:

Notation:

- x: independent variable, predictor variable or input variable
- y: dependent variable or response variable

Regression with a Single Predictor:

Example:

Dosage x (in Milligrams) and the Number of Days of Relief y from Allergy for Ten Patients.

Dosage x	Duration of Relief y
3	9
3	5
4	12
5	9
6	14
6	16
7	22
8	18
8	24
9	22

Regression with a Single Predictor:

Example:

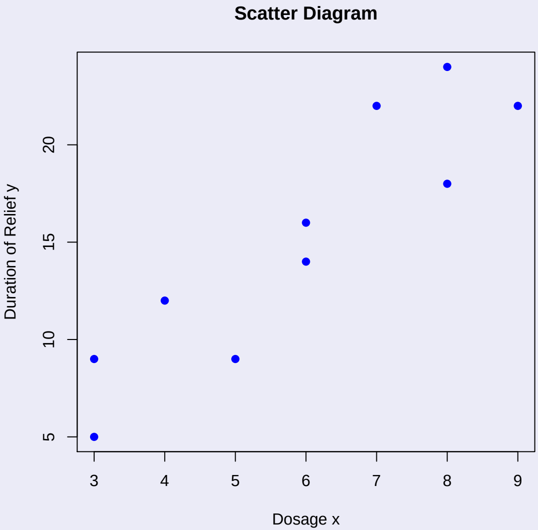
Dosage x (in Milligrams) and the Number of Days of Relief y from Allergy for Ten Patients.

Dosage x	Duration of Relief y
3	9
3	5
4	12
5	9
6	14
6	16
7	22
8	18
8	24
9	22

What would you do as a first step to analyze the relationship?

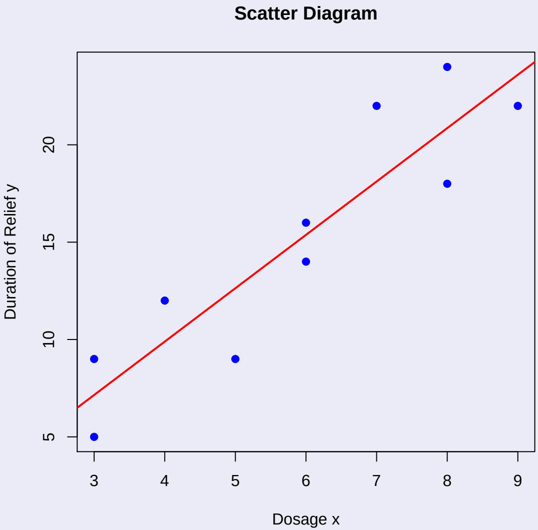
Regression with a Single Predictor:

Scatter Diagram:



Regression with a Single Predictor:

Scatter Diagram:



Straight Line Regression Model:

Statistical Model for a Straight Line Regression:

We assume that the response Y is a random variable that is related to the input variable x by

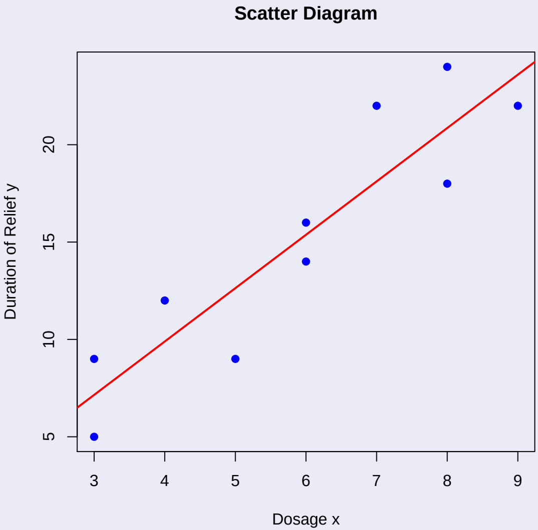
$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \ldots, n$$

where

1. Y_i denotes the response corresponding to the ith experimental run in which the input variable x is set at the value x_i.
2. $e_1, \ldots, e_n$ are the unobservable random errors which we assume are independently and normally distributed with mean zero and an unknown standard deviation σ.
3. The parameters β_0 and β_1, which together locate the straight line, are unknown.

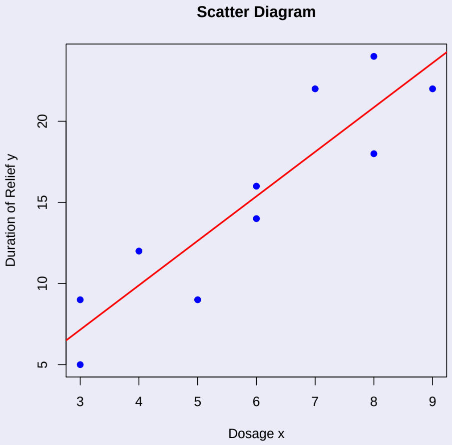
Straight Line Regression Model:

The statistical model for a straight line regression:



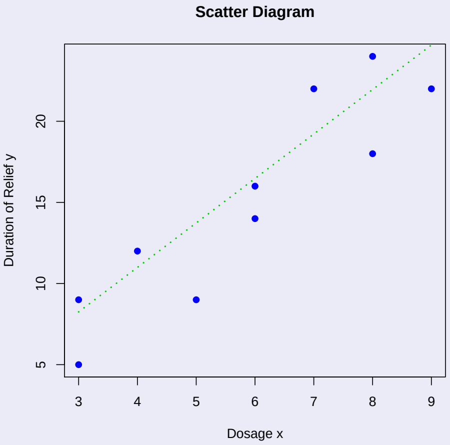
Straight Line Regression Model:

Assume the previous statistical model is correct. Then how to fit the best straight line of the y to x relationship in the scatter diagram?



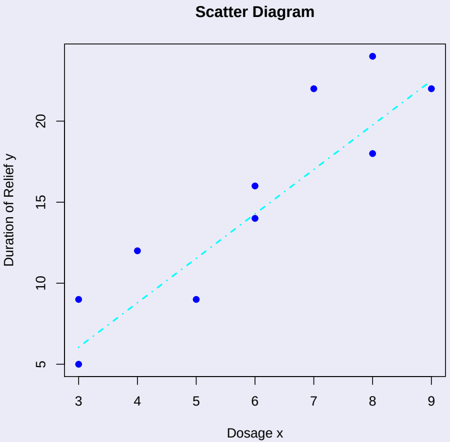
Straight Line Regression Model:

Assume the previous statistical model is correct. Then how to fit the best straight line of the *y* to *x* relationship in the scatter diagram?



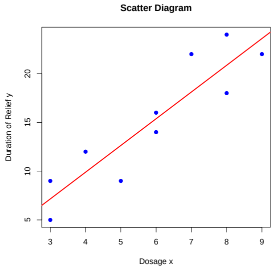
Straight Line Regression Model:

Assume the previous statistical model is correct. Then how to fit the best straight line of the *y* to *x* relationship in the scatter diagram?



Method of Least Squares:

Suppose that an arbitrary line $y = b_0 + b_1 x$ is drawn on the scatter diagram.



Principle of Least Squares:

Determine the values for the parameters so that the overall discrepancy

$$D = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

is minimized. The parameter values thus determined are called the least squares estimates.

Method of Least Squares:

- Least Squares Estimator of β_0:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

- Least Squares Estimator of β_1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

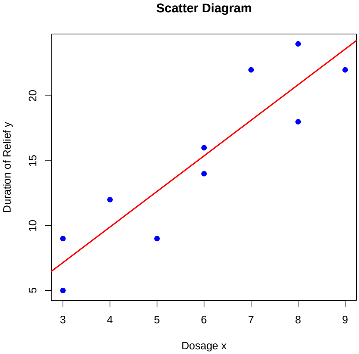
- Fitted (or estimated) regression line:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Residuals:

$$\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, \ldots, n$$

Method of Least Squares:



$$\hat{y} = -1.071 + 2.741x$$

By using the fitted regression line, we can predict a response y for a specified x value.

Inferences for β_1 and β_0:

```
> Summary(lmfit)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6333 -2.0128 -0.3741  2.0428  3.8851 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.0709     2.7509  -0.389 0.707219    
x             2.7408     0.4411   6.214 0.000255 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 2.821 on 8 degrees of freedom
Multiple R-Squared: 0.8284,	Adjusted R-squared: 0.8069 
F-statistic: 38.62 on 1 and 8 DF,  p-value: 0.0002555
```

Regression Analysis:

Regression analysis concerns the study of relationships between variables with the objectives of

1. identifying the relationship
2. estimating the parameters
3. validating the estimates
4. predicting the response variable