Stat 312: Lecture 17
Simple Linear Model

Moo K. Chung
mchung@stat.wisc.edu

March 27, 2003

Concepts

1. Deterministic model:
$$y = \beta_0 + \beta_1 x.$$
For example, x is the speed of a car and y is the distance the car traveled in 1 hour. This is an unrealistic model because the car can not maintain the absolute constant speed. So we introduce a noise term ϵ in the above equation.

2. Stochastic model: let ϵ be a random noise with $\mathbb{E}\epsilon = 0$ and $\mathbf{Var}\epsilon = \sigma^2$. Instead of the deterministic model we will have $y = \beta_0 + \beta_1 x + \epsilon$. Since ϵ is a random variable, we use Y instead of y:
$$Y = \beta_0 + \beta_1 x + \epsilon.$$
Note that $\mathbb{E}Y = \beta_0 + \beta_1 x$ and $\mathbf{Var}Y = \sigma^2$.

3. Given paired data $(x_1, y_1), \cdots, (x_n, y_n)$, we have a linear stochastic relationship
$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$
where y_j is the observed value of a random variable Y_j and $\epsilon_j \sim \epsilon$. Note that $\mathbb{E}Y_j = \beta_0 + \beta_1 x_j$. Let $\hat\beta_0, \hat\beta_1$ be estimators of β_0, β_1. Then the *predicted values* or fitted values $\hat y_j = \hat\beta_0 + \hat\beta_1 x_j$ are estimators of $\mathbb{E}Y_j = \beta_0 + \beta_1 x_j$. The differences between the observations y_j and the predicted values $\hat y_j$ are called the *residuals* (errors), i.e.
$$r_j = y_j - \hat y_j = y_j - \hat\beta_0 - \hat\beta_1 x_j.$$

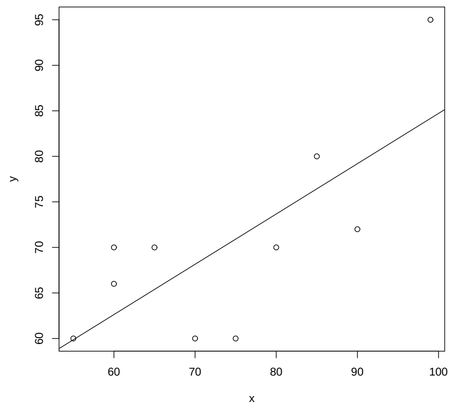
4. *Least squares estimation*. We estimate β_0 and β_1 by minimizing the *sum of the squared errors* (SSE):
$$SSE = \sum_{j=1}^n r_j^2 = \sum_{j=1}^n (y_j - \hat y_j)^2 = \sum_{j=1}^n (y_j - \beta_0 - \beta_1 x_j)^2.$$
Then the *regression line* is $y = \beta_0 + \beta_1 x$.

In-class problems

Example 1. 10 students took two midterm exams.

Student	01	02	03	04	05	06	07	08	09	10
Midterm 1	80	75	60	90	99	60	55	85	65	70
Midterm 2	70	60	70	72	95	66	60	80	70	60

Let's find the regression line.




```
> x<-c(80, 75, 60, 90, 99, 60, 55,
85, 65, 70)
> y<-c(70, 60, 70, 72, 95, 66, 60,
80, 70, 60)
> plot(x,y)
> rarara <-lm(y~x)
> rarara
Call: lm(formula = y ~ x)
Coefficients: (Intercept)      x
                29.4827      0.5523
> abline(rarara)

> library(Devore5)
> data(ex12.15)
> attach(ex12.15)
> ex12.15
   MoE  Strength
1  29.8      5.9
2  33.2      7.2
    ...      ...
> lm(Strength~MoE)
 Call: lm(formula = y ~ x)
 Coefficients:
 (Intercept)    x
 3.2925   0.1075
```

Self-study problems

Example 12.3., 12.4., 12.6., 12.7., 12.8.