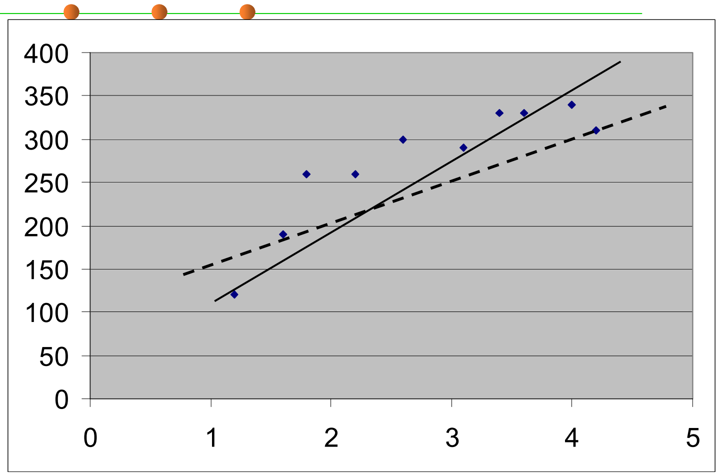
Introduction

- What does it mean when there is a strong positive correlation between x and y ?
- Regression analysis aims to find a precise formula to relate the movements of y to those of x
- The use of regression requires a good deal of thought and a good dose of skepticism

Example: Sales vs. Advertising

Advertising (mil.$)	Sales (mil.$)
1.2	120
1.6	190
1.8	260
2.2	260
2.6	300
3.1	290
3.4	330
3.6	330
4.0	340
4.2	310

Example: Sales vs. Advertising



Example: Sales vs. Advertising

- It is believed that sales, S, are tied to advertising, A, by a simple linear equation:

$$S = \beta_0 + \beta_1 A$$

- What do β_0 and β_1 represent?
- How can we find β_0 and β_1 ?

Sales vs. Advertising

- The linear relation is usually not exact.
- A more realistic model:

$$S_i = \beta_0 + \beta_1 A_i + \varepsilon_i$$

Where β_0 and β_1 are *regression coefficients.*

Measurement Error ε_i

- Properties of distribution
  - A mean of zero
  - Symmetry around zero
  - An assignment of greater probability to small errors than to larger ones
- Errors are assumed to be:
  - Independent
  - Have same variance (*homoscedasticity*)

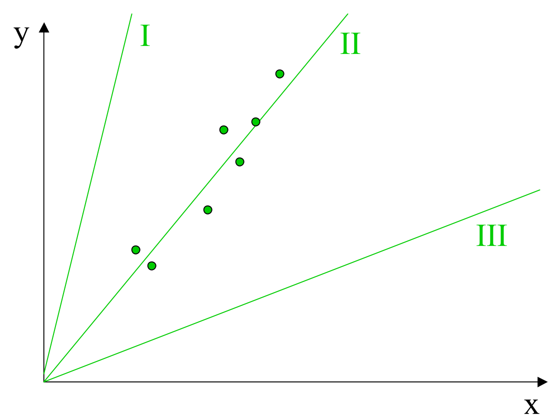
Method of Least Squares

- Consider the simple formula:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

- Where the measurement errors are independent samples from $N(0, \sigma_\varepsilon)$
- How to find the estimators of β_0 and β_1?

Choosing the best line



Least Squares Estimates

- Suggest an index to measure discrepancy between points and line
- Focus on vertical disparities between points and line
- Sum of the square of the deviations:

$$L(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Least Squares Estimates

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

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

$$\text{where } \bar{y} = (1/n)\sum_{i=1}^{n} y_i \text{ and } \bar{x} = (1/n)\sum_{i=1}^{n} x_i$$