Mathematical statistics

November 29^{th}, 2018

Lecture 23: Linear regression

Att.

- Final exam:

  Friday, 12/14/2018, 10:30am –12:30pm

  Gore Hall 115

- Course evaluation

Overview

Week 1	Probability reviews
Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	Chapter 7: Point Estimation
Week 7	Chapter 8: Confidence Intervals
Week 10	Chapters 9–10: Tests of Hypothesis
Week 14	Chapter 12: Linear regression

Key steps in statistical inference

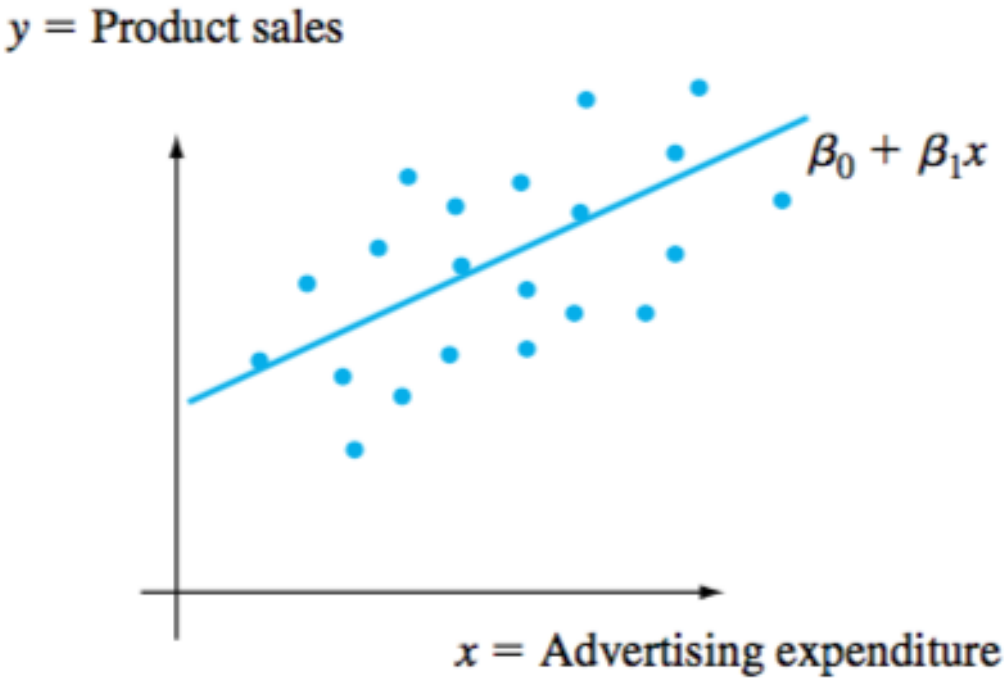
- Understand the statistical model [Chapter 6]
- Come up with reasonable estimates of the parameters of interest [Chapter 7]
- Quantify the confidence with the estimates [Chapter 8]
- Testing with the parameters of interest [Chapter 9]

Contexts

- The central mega-example: population mean μ
- Difference between two population means
- Linear regression

Linear regression

Linear regression



Mathematical model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Linear regression

Assumptions:

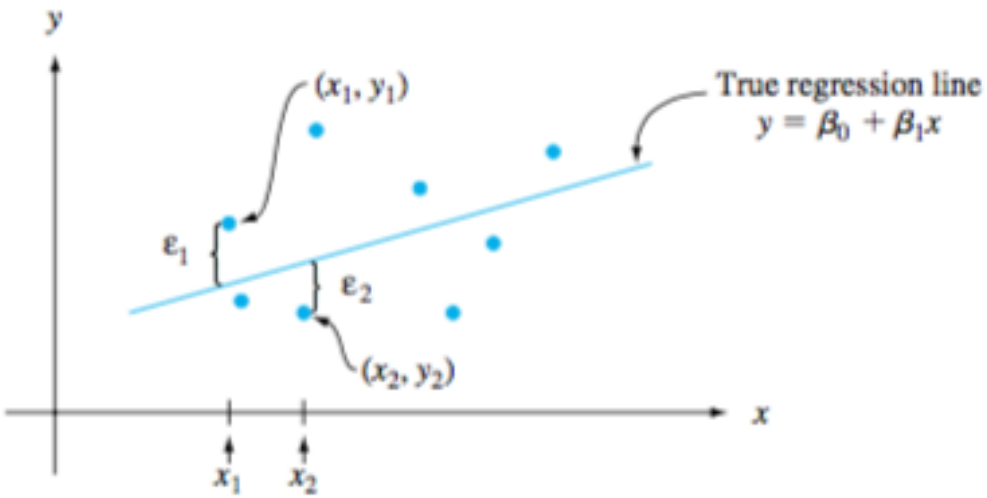
- There are parameters β_0, β_1 and σ such that for any fixed value of the independent variable x, the dependent variable Y is related to x through the model equation

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

  The random deviation (random variable) ϵ is assumed to be normally distributed with mean value 0 and variance σ^2

- The observed pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are regarded as having been generated independently of one another from the model equation

Linear regression



Mathematical model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

The principle of least squares

PRINCIPLE OF LEAST SQUARES

The vertical deviation of the point (x_i, y_i) from the line $y = b_0 + b_1 x$ is

$$\text{height of point} - \text{height of line} = y_i - (b_0 + b_1 x_i)$$

The sum of squared vertical deviations from the points $(x_1, y_1), \ldots, (x_n, y_n)$ to the line is then

$$f(b_0, b_1) = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2$$

The point estimates of β_0 and β_1, denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$ and called the **least squares estimates**, are those values that minimize $f(b_0, b_1)$. That is, $\hat{\beta}_0$ and $\hat{\beta}_1$ are such that $f(\hat{\beta}_0, \hat{\beta}_1) \leq f(b_0, b_1)$ for any b_0 and b_1. The **estimated regression line** or **least squares line** is then the line whose equation is $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

Least squares estimates

- Estimates

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Computing formulas

$$S_{xy} = \left(\sum x_i y_i\right) - \frac{(\sum x_i)(\sum y_i)}{n}$$

and

$$S_{xx} = \left(\sum x_i^2\right) - \frac{(\sum x_i)^2}{n}$$

Problem 2: Estimating σ by maximum likelihood

Problem

The joint density function of $(Y_1, Y_2, \ldots, Y_n)$ *is*

$$f_{joint}(y_1, y_2, \ldots, y_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_i [y_i - (\beta_0 + \beta_1 x_i)]^2}$$

What is the maximum likelihood estimator of σ*?*

Confidence intervals for β_1

Principles for deriving CIs

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution $f(x, \theta)$, then

- Find a random variable $Y = h(X_1, X_2, \ldots, X_n; \theta)$ such that the probability distribution of Y does not depend on θ or on any other unknown parameters.
- Find constants a, b such that

$$P\left[a < h(X_1, X_2, \ldots, X_n; \theta) < b\right] = 0.95$$

- Manipulate these inequalities to isolate θ

$$P\left[\ell(X_1, X_2, \ldots, X_n) < \theta < u(X_1, X_2, \ldots, X_n)\right] = 0.95$$

Normality of $\hat{\beta}_1$

- First, recall that

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

- On the other hand,

$$\sum (x_i - \bar{x})(\bar{Y}) = \bar{Y} \sum (x_i - \bar{x}) = \bar{Y} \cdot 0 = 0$$

- Thus

$$\hat{\beta}_1 = \sum c_i Y_i \quad \text{where} \quad c_i = \frac{(x_i - \bar{x})}{S_{xx}}$$

$\rightarrow$

is a linear combination of the independent r.v.'s $Y_1, Y_2, \ldots, Y_n$, each of which is normally distributed

Properties of $\hat{\beta}_1$

- To construct confidence intervals for β_1, we need to compute the expected value and the variance of $\hat{\beta}_1$ in terms of $(x_1, y_1), \ldots (x_n, y_n)$ and σ where

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

and

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) Y_i}{\sum (x_i - \bar{x})^2}$$

- Task: Compute $E[\hat{\beta}_1]$ and $Var[\hat{\beta}_1]$

Setting 1: If σ is known

Problem

Recall that

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{S_{xx}}}$$

follows the standard normal distribution.
Assuming that σ is known, construct the $100(1-\alpha)\%$ confidence interval for β_1.

Setting 2: σ is unknown

Theorem

If we define

$$S^2 = \frac{\sum [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2}{n-2}$$

then the random variable

$$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}}$$

follows the $t-$distribution with degrees of freedom $(n-2)$.

t distributions

Definition

Let Z be a standard normal rv and let W be a χ^2_ν rv independent of Z. Then the t distribution with degrees of freedom ν is defined to be the distribution of the ratio

$$T = \frac{Z}{\sqrt{W/\nu}}$$

Setting 2: σ is unknown

Our statistic:

$$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} = \frac{\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{S_{xx}}}}{\sqrt{\frac{S^2}{\sigma^2}}}$$

The theorem is a consequence of the following facts

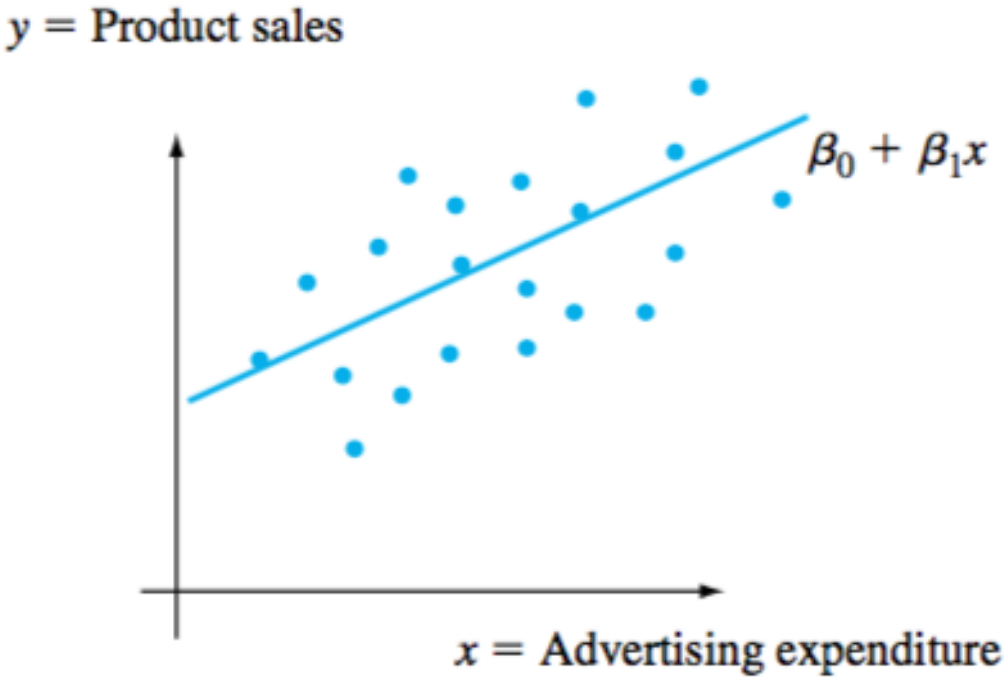
- $\hat{\beta}_1$ and S are independent
- The statistic

$$\frac{1}{\sigma^2} \sum [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

follows the χ^2-distribution with $(n-2)$ degrees of freedom.

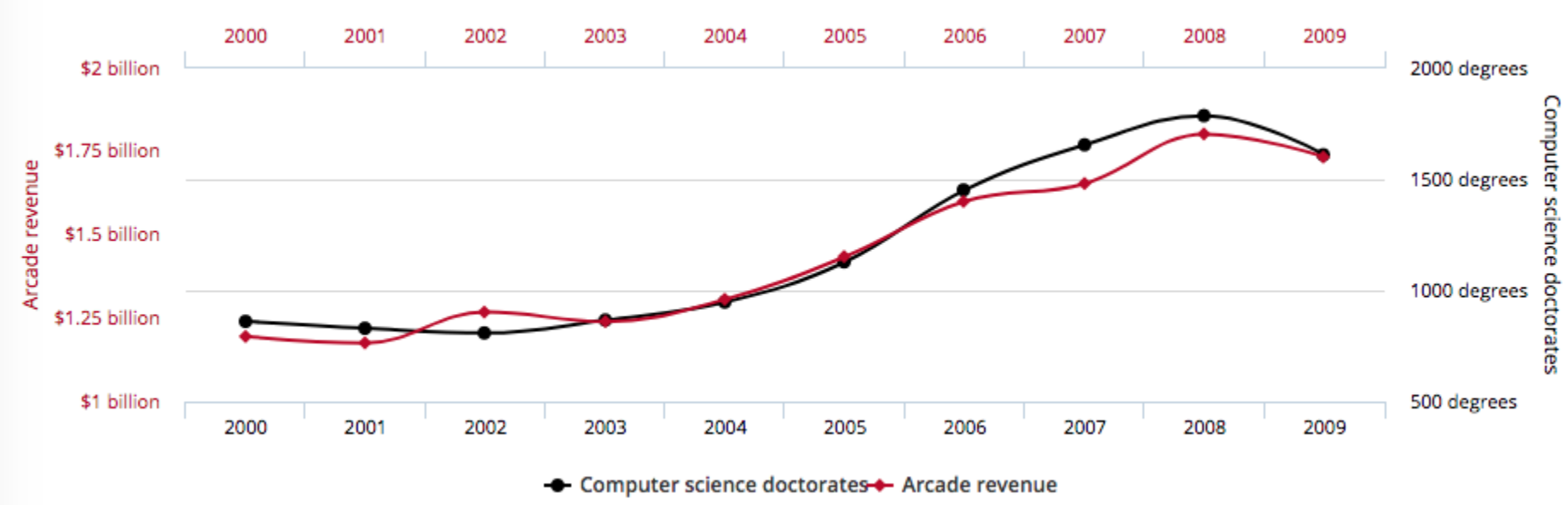
Testing with β_1

β_1 characterizes relation between x and Y



Question: Does increase advertising expense help increase sales?
$\rightarrow$ Testing $H_0 : \beta_1 = 0$ against $H_a : \beta_1 > 0$

β_1 characterizes relation between x and Y



Question: Do computer scientists spend too much time at arcades?

Hypothesis testing

Null hypothesis: H_0: $\beta_1 = \beta_{10}$

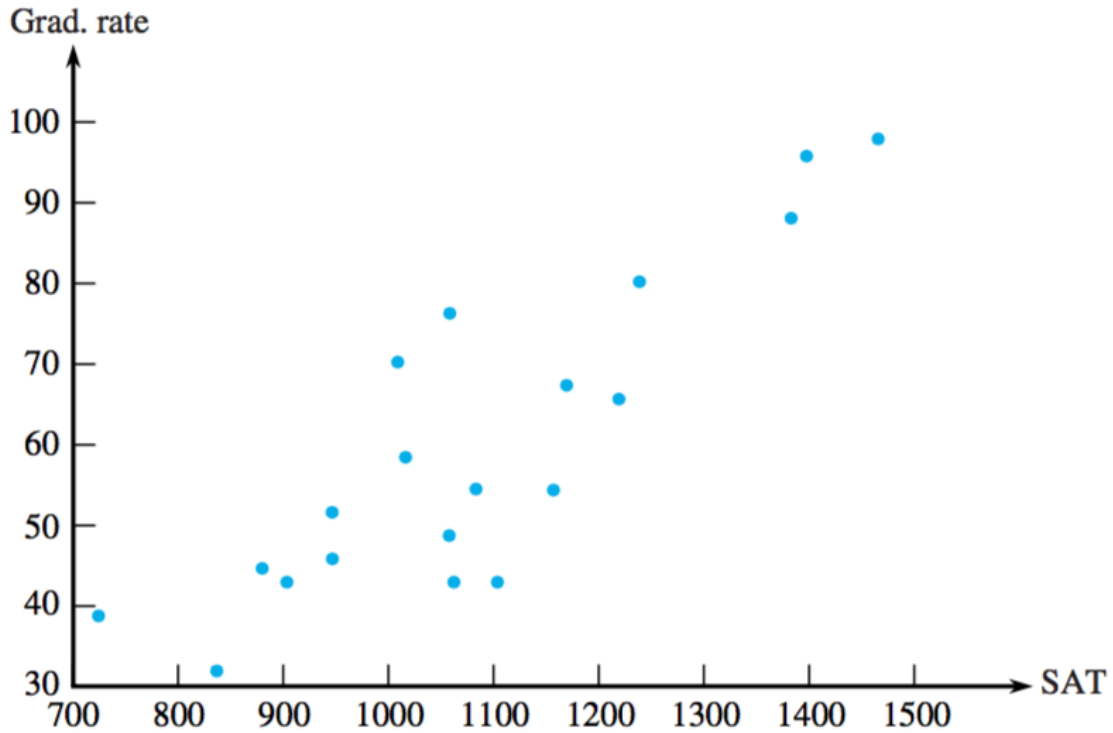
Test statistic value: $t = \dfrac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$

Alternative Hypothesis	Rejection Region for Level α Test
H_a: $\beta_1 > \beta_{10}$	$t \geq t_{\alpha,n-2}$
H_a: $\beta_1 < \beta_{10}$	$t \leq -t_{\alpha,n-2}$
H_a: $\beta_1 \neq \beta_{10}$	either $t \geq t_{\alpha/2,n-2}$ or $t \leq -t_{\alpha/2,n-2}$

A P-value based on $n - 2$ df can be calculated just as was done previously for t tests in Chapters 9 and 10.

Example 12.12

Is it possible to predict graduation rates from SAT scores?



Assume that

$$\hat{\beta}_1 = .08855; s = 10.29; S_{xx} = 704125; n = 20.$$