Section 3. Simple Regression – OLS Estimators

1. **Last time: Lines, Population Linear Regression**
2. **Estimators of β_0, β_1** are b_0 and b_1
3. **Least Squares Estimators**
4. **OLS Assumptions**

PARAMETERS

RANDOM VARIABLES

1. Review: Lines, Clouds and Pop. Linear Regression

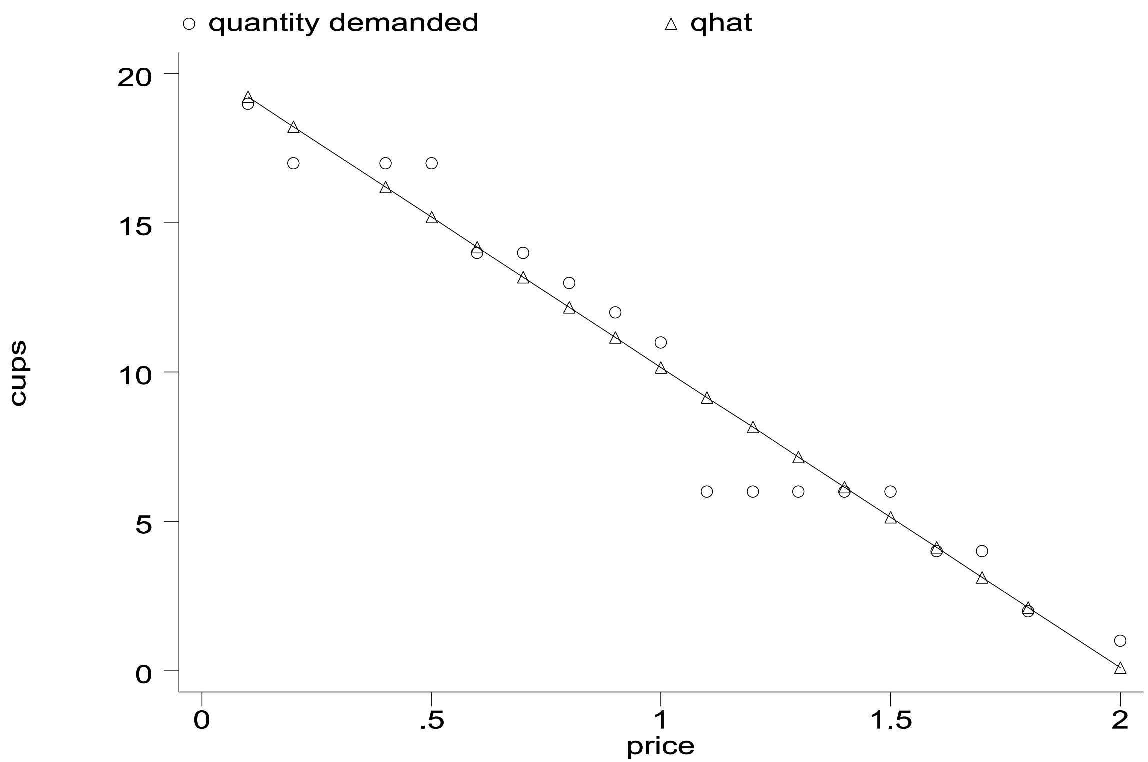
- e.g. Demand for Coffee, Global warming, CA test scores and student-teacher ratios
- Decide which parameters in population we care about (β_0,β_1)

  - just like we did with μ
- Draw a sample and estimate parameters

  - just like we did with μ
- Construct CI for parameters, test hypotheses, make predictions.

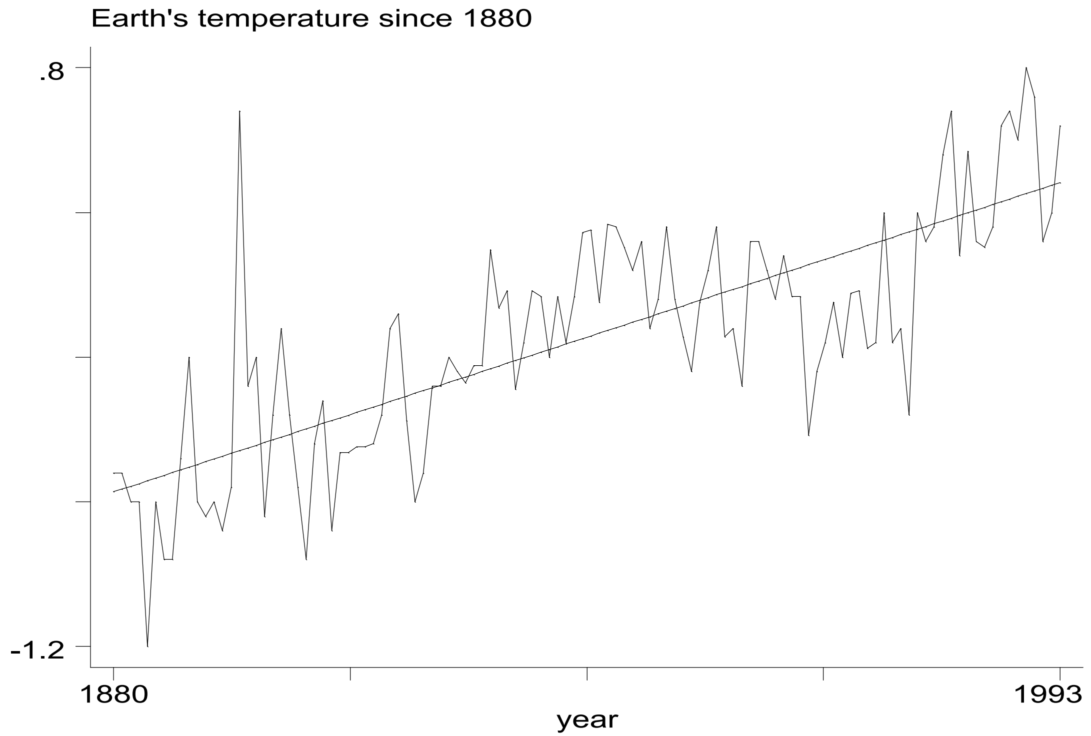
  - just like..

Coffee Example (with a line)



What's the slope of the line?
About how many cups would you sell at $1?

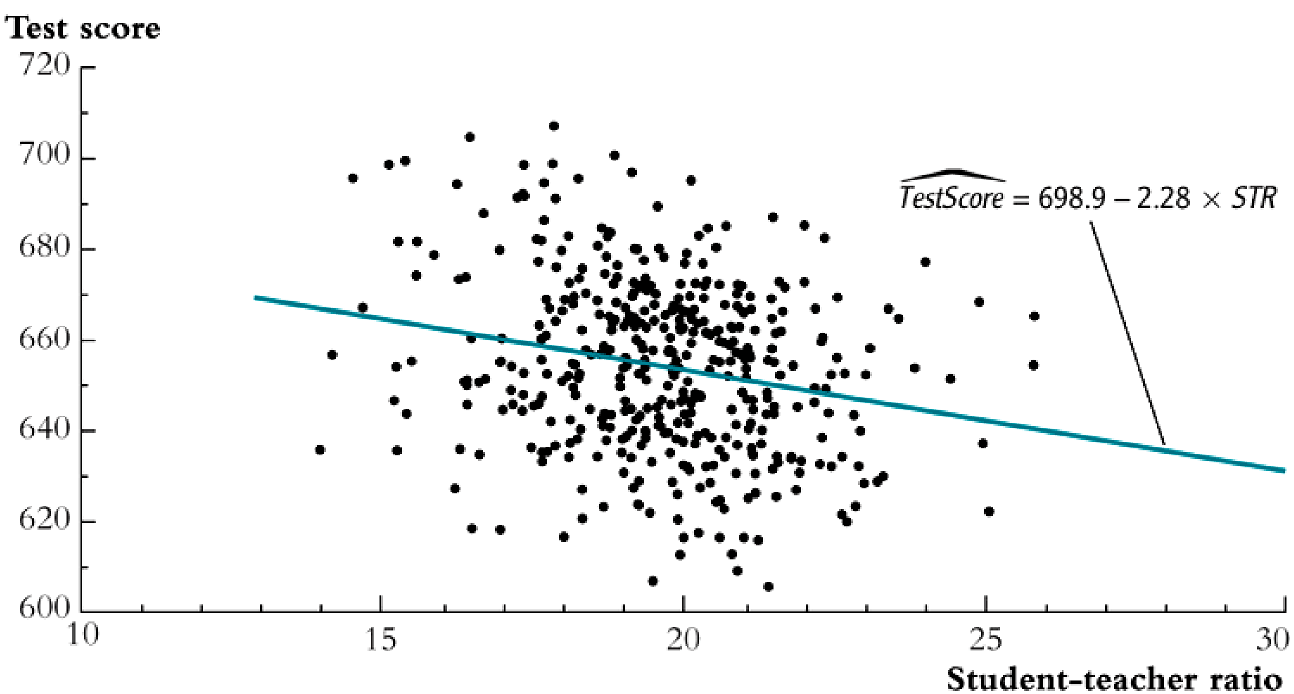
Global Warming Example



Is the slope statistically different from zero?

FIGURE 4.3 The Estimated Regression Line for the California Data

The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.



2. Estimators of β_0, β_1

- **Decide which parameters in population we care about (β_0, β_1)**
  - we chose Pop. Linear Regression
- **Draw a sample and estimate parameters**
- **Which estimates to use?**
- **We'd like something with nice properties:**
  - unbiased, $E(b_1) = \beta_1$
  - consistent, $b_1 \xrightarrow{p} \beta_1$
  - efficient, $V(b_1)$ is smallest
  - with an approximately normal distribution. $b_1 \overset{A}{\sim} N(\beta_1, V(b_1))$

$t = \frac{b_1 - \beta_1}{\sqrt{\hat{V}(b_1)}} \overset{A}{\sim} N(0,1)$ "ratio"

3. The Ordinary Least Squares (OLS) Estimators

- **Try b_0 and b_1 that minimize the sum of e_i^2 in**

$$\sum_{i=1}^{n} e_i^{\,2}$$

FIND b_0, b_1 that minimize Σe_i^2

where $e_i = Y_i - (b_0 + b_1 x)$

- **Why this one? Well, it's analogous to what we asked for in the population.**
- **And it seems like a nice property in the sample.**
- Deriving formulae for OLS estimators..

  (S&W p. 143)

Minimize $\sum e_i^2 = \sum (Y_i - (b_0 + b_1 X_i))^2$ by choice of b_0, b_1

First order Conditions $\frac{\partial \sum e_i^2}{\partial b_0} = 0$, $\frac{\partial \sum e_i^2}{\partial b_1} = 0$

$$0 = \frac{\partial \sum e_i^2}{\partial b_0} = \frac{\partial \sum (Y_i - (b_0 + b_1 X_i))^2}{\partial b_0} = \frac{\partial (Y_1 - (b_0 + b_1 X_1))^2 + (Y_2 - (b_0 + b_1 X_2))^2 + \ldots}{\partial b_0}$$

$$= 2e_1 \frac{\partial e_1}{\partial b_0} + \ldots$$

$$= 2e_1(-1) + 2e_2(-1) + \ldots 2e_N(-1) = 2(-1)\sum e_i = 0$$

$$\Leftrightarrow \boxed{\sum e_i = 0} \quad \#1$$

$$0 = \frac{\partial \sum e_i^2}{\partial b_1} = \frac{\partial \sum (Y_i - (b_0 + b_1 X_i))^2}{\partial b_1} = \frac{\partial (Y_1 - (b_0 + b_1 X_1))^2 + (Y_2 - (b_0 + b_1 X_2))^2 + \ldots}{\partial b_1}$$

$$= 2e_1 \frac{\partial e_1}{\partial b_1} + \ldots$$

$$= 2e_1(-X_1) + 2e_2(-X_2) + \ldots 2e_N(-X_N) = -2\sum e_i X_i$$

$$\Leftrightarrow \boxed{\sum e_i X_i = 0} \quad \#2$$

2 F.O.C. $\sum e_i = 0$, $\sum e_i X_i = 0$, now solve for b_0, b_1

$$\sum[Y_i - (b_0 + b_1 X_i)] = 0 \Rightarrow \frac{1}{N}\sum Y_i - \frac{1}{N}\sum(b_0 + b_1 X_i) = 0$$

$$\bar{Y} = \frac{1}{N} N b_0 + b_1 \frac{1}{N}\sum X_i$$

#1 $\bar{Y} = b_0 + b_1 \bar{X} \Rightarrow b_0 = \bar{Y} - b_1 \bar{X}$

$$0 = \sum(Y_i - (b_0 + b_1 X_i))X_i \Rightarrow \sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 = 0$$

LEMMA

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) \equiv S_{XY}$$

$$= \sum X_i Y_i - \sum \bar{X} Y_i - \sum X_i \bar{Y} + \sum \bar{X}\bar{Y}$$

$$= \sum X_i Y_i - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y}$$

$$= \sum X_i Y_i - n\bar{X}\bar{Y}$$

#2 $\frac{1}{N}\sum X_i Y_i - b_0 \bar{X} - b_1 \frac{\sum X_i^2}{N} = 0$

substitute in here

$$\frac{1}{N}\sum X_i Y_i - (\bar{Y} - b_1 \bar{X})\bar{X} - b_1 \frac{\sum X_i^2}{N} = 0$$

$$\frac{1}{N}\sum X_i Y_i - \bar{Y}\bar{X} = b_1 \frac{\sum X_i^2}{N} - b_1 \bar{X}^2$$

(use lemma again)

$$\frac{1}{N}\sum(X_i - \bar{X})(Y_i - \bar{Y}) = b_1\left(\frac{\sum X_i^2}{N} - \bar{X}^2\right) = b_1 \frac{\sum(X_i - \bar{X})^2}{N}$$

$$\Rightarrow b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})/N}{\sum(X_i - \bar{X})^2/N}$$

Formula for OLS estimators

The OLS Estimator, Predicted Values, and Residuals

The OLS estimators of the slope β_1 and the intercept β_0 are:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.8}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.9}$$

The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1, \ldots, n \tag{4.10}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{4.11}$$

The estimated intercept ($\hat{\beta}_0$), slope ($\hat{\beta}_1$), and residual ($\hat{u}_i$) are computed from a sample of n observations of X_i and Y_i, $i = 1, \ldots, n$. These are estimates of the unknown true population intercept (β_0), slope (β_1), and error term (u_i).

4. Assumptions

The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n, \text{ where:}$$

1. The error term u_i has conditional mean zero given X_i, that is, $E(u_i | X_i) = 0$;
2. (X_i, Y_i), $i = 1, \ldots, n$ are independent and identically distributed (i.i.d.) draws from their joint distribution; and
3. (X_i, u_i) have nonzero finite fourth moments.

Next time..

- Confidence intervals for β_0 and β_1
- Examples