1

SIMPLE REGRESSION MODEL I: Ordinary Least Squares Estimation

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics I

2

Simple Regression Model

Simple (Bivariate) Regression Model

$$y = \beta_0 + \beta_1 x + u$$

- y: dependent variable
- x: explanatory variable
- Also called "bivariate linear regression model", "two-variable linear regression model"
- Purpose: to explain the dependent variable y by the independent variable x

3

Simple Regression Model

Terminology

y	x
Dependent variable	**Independent variable**
Explained variable	**Explanatory variable**
Response variable	**Control variable**
Predicted variable	**Predictor variable**
Regressand	**Regressor**

4

Predictor variable: $y = \beta_0 + \beta_1 x + u$

u: Random Error term - Disturbance term

Represents factors other than x that affect y. These factors are treated as "unobserved" in the simple regression model.

Slope parameter β_1

- If the other factors in u are held fixed, i.e. $\Delta u = 0$, then the linear effect of x on y:

$$\Delta y = \beta_1 \Delta x$$

- β_1: slope term.

Intercept term (also called constant term): β_0

the value of y when $x = 0$.

Simple Regression Model: Examples

Agricultural production and fertilizer usage

$$yield = \beta_0 + \beta_1 fertilizer + u$$

yield: quantity of wheat production

Slope parameter β_1

$$\Delta yield = \beta_1 \Delta fertilizer$$

Ceteris Paribus, one unit change in fertilizer leads to β_1 unit change in wheat yield.

Random error term: u

Contains factors such as land quality, rainfall, etc, which are assumed to be unobserved.
Ceteris Paribus $\Leftrightarrow$ Holding all other factors fixed $\Leftrightarrow \Delta u = 0$

Simple Regression Model: Examples

A Simple Wage Equation

$$wage = \beta_0 + \beta_1 educ + u$$

wage: hourly wage (in dollars); educ: education level (in years)

Slope parameter β_1

$$\Delta wage = \beta_1 \Delta educ$$

β_1 measures the change in hourly wage given another year of education, holding all other factors fixed (ceteris paribus).

Random error term u

Other factors include labor force experience, innate ability, tenure with current employer, gender, quality of education, marital status, number of children, etc. Any factor that may potentially affect worker productivity.

Linearity

- The linearity of simple regression model means: a one-unit change in x has the same effect on y regardless of the initial value of x.
- This is unrealistic for many economic applications.
- For example, if there increasing or decreasing returns to scale then this model is inappropriate.
- In wage equation, the impact of the next year of education on wages has a larger effect than did the previous year.
- An extra year of experience may also have similar increasing returns.
- We will see how to allow for such possibilities in the following classes.

Assumptions for Ceteris Paribus conclusions

1. Expected value of the error term u is zero

- If the model includes a constant term (β_0) then we can assume:
$$\mathrm{E}(u) = 0$$
- This assumption is about the distribution of u (unobservables). Some u terms will be $+$ and some will be $-$ but on average u is zero.
- This assumption is always guaranteed to hold by redefining β_0.

Assumptions for Ceteris Paribus conclusions

2. Conditional mean of u is zero

- How can we be sure that the ceteris paribus notion is valid (which meas that $\Delta u = 0$?
- For this to hold, x and u must be uncorrelated. But since correlation coefficient measures only the linear association between two variables it is not enough just to have zero correlation.
- u must also be uncorrelated with the functions of x(e.g. x^2, $\sqrt{x}$ etc.)
- *Zero Conditional Mean* assumption ensures this:

$$\mathsf{E}(u|x) = \mathsf{E}(u) = 0$$

- This equation says that the average value of the unobservables is the same across all slices of the population determined by the value of x.

Zero Conditional Mean Assumption

Conditional mean of u given x is zero

$$\mathsf{E}(u|x) = \mathsf{E}(u) = 0$$

- Both u and x are random variables. Thus, we can define the conditional distribution of u given a value of x.
- A given value of x represents a slice in the population. The conditional mean of u for this specific slice of the population can be defined.
- This assumption means that the average value of u does not depend on x.
- For a given value of x unobservable factors have a zero mean. Also, unconditional mean of unobservables is zero.

Zero Conditional Mean Assumption: Example

Wage equation:

$$wage = \beta_0 + \beta_1 educ + u$$

- Suppose that u represents innate ability of employees, denoted abil.
- $\mathsf{E}(u|x)$ assumption implies that innate ability is the same across all levels of education in the population:

$$\mathsf{E}(abil|educ = 8) = \mathsf{E}(abil|educ = 12) = \ldots = 0$$

- If we believe that average ability increases with years of eduction this assumption will not hold.
- Since we cannot observe ability we cannot determine if average ability is the same for all education levels.

Zero Conditional Mean Assumption: Example

Agricultural production-fertilizer model:

- Recall the fertilizer experiment: the land is divided into equal plots and different amounts of fertilizer is applied to each plot.
- If the amount of fertilizer is assigned to land plots independent of the quality of land then the zero-conditional-mean assumption will hold.
- However, if larger amounts of fertilizer is assigned to land plots with higher quality then the expected value of the error term will increase with the amount of fertilizer.
- In this case zero conditional mean assumption is false.

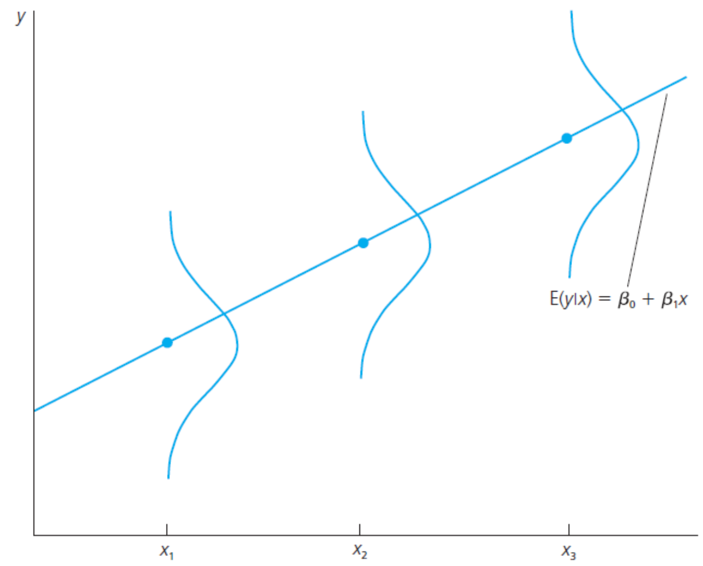
Population Regression Function (PRF)

- Expected value of y given x:

$$\begin{aligned} \mathsf{E}(y|x) &= \beta_0 + \beta_1 x + \underbrace{\mathsf{E}(u|x)}_{=0} \\ &= \beta_0 + \beta_1 x \end{aligned}$$

- This is called PRF. Obviously, conditional expectation of the dependent variable is a linear function of x.
- Linearity of PRF: for a one-unit change in x conditional expectation of y changes by β_1.
- The center of the conditional distribution of y for a given value of x is $\mathsf{E}(y|x)$.

Population Regression Function (PRF)



Systematic and Unsystematic Parts of Dependent Variable

- In the simple regression model

$$y = \beta_0 + \beta_1 x + u$$

under $\mathsf{E}(u|x) = 0$ the dependent variable y can be decomposed into two parts:
- Systematic part: $\beta_0 + \beta_1 x$. This is the part of y explained by x.
- Unsystematic part: u. This is the part of y that cannot be explained by x.

Estimation of Unknown Parameters

- How can we estimate the unknown population parameters (β_0, β_1) given a cross-sectional data set.?
- Suppose that we have a random sample of n observations:

$$\{y_i, x_i : i = 1, 2, \ldots, n\}$$

- Regression model can be written for each observation as follows:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n$$

- Now we have a system of n equations with two unknowns.

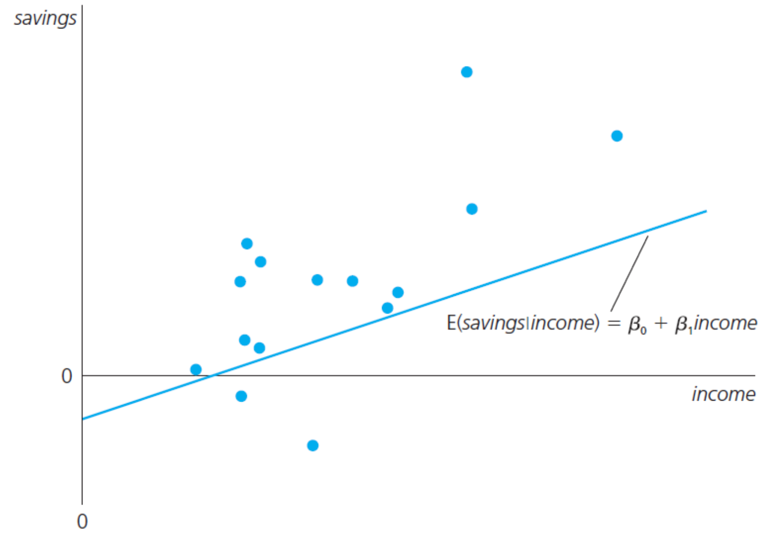
Estimation of unknown population parameters (β_0, β_1)

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \dots, n$$

n equations with 2 unknowns:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + u_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + u_2 \\ y_3 &= \beta_0 + \beta_1 x_3 + u_3 \\ \vdots &= \vdots \\ y_n &= \beta_0 + \beta_1 x_n + u_n \end{aligned}$$

Random Sample Example: Savings and income for 15 families



Estimation of unknown population parameters (β_0, β_1): Method of Moments

We just made two assumptions for ceteris paribus conclusions to be valid:

$$\begin{aligned} \mathsf{E}(u) &= 0 \\ \mathsf{Cov}(x, u) &= \mathsf{E}(xu) = 0 \end{aligned}$$

NOTE: If $\mathsf{E}(u|x) = 0$ then by definition $\mathsf{Cov}(x, u) = 0$ but the reverse may not be true. Since $\mathsf{E}(u) = 0$ then by definition $\mathsf{Cov}(x, u) = \mathsf{E}(xu)$. Using these assumptions and $u = y - \beta_0 - \beta_1 x$ **Population Moment Conditions** can be written as:

$$\begin{aligned} \mathsf{E}(y - \beta_0 - \beta_1 x) &= 0 \\ \mathsf{E}[x(y - \beta_0 - \beta_1 x)] &= 0 \end{aligned}$$

Now we have 2 equations with 2 unknowns.

Method of Moments: Sample Moment Conditions

Population moment conditions:

$$\begin{aligned} \mathsf{E}(y - \beta_0 - \beta_1 x) &= 0 \\ \mathsf{E}[x(y - \beta_0 - \beta_1 x)] &= 0 \end{aligned}$$

Replacing these with their sample analogs we obtain:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \frac{1}{n}\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \end{aligned}$$

This system can easily be solved for $\hat{\beta}_0$ and $\hat{\beta}_1$ using sample data . Note that $\hat{\beta}_0$ and $\hat{\beta}_1$ have hats on them, they are not fixed quantities. They change as the data change.

Method of Moments: Sample Moment Conditions

Using the properties of the summation operator, from the first sample moment condition:

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

where $\bar{y}$ and $\bar{x}$ sample means.
Using this we can write

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

Substituting this into the second sample moment condition we can solve for $\hat{\beta}_1$.

Method of Moments

Substituting $\hat{\beta}_0$ into second moment condition after multiplying it with $1/n$:

$$\sum_{i=1}^{n} x_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) = 0$$

This expression can be written as

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

Slope Estimator

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

The following properties have been used in deriving the expression above:

$$\sum_{i=1}^{n} x_i (x_i - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\sum_{i=1}^{n} x_i (y_i - \bar{y}) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Slope Estimator

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

1. Slope estimator is the ratio of the sample covariance between x and y to the sample variance of x.
2. The sign of $\hat{\beta}_1$ depends on the sign of sample covariance. If x and y are positively correlated in the sample, $\hat{\beta}_1$ is positive; if x and y are negatively correlated then $\hat{\beta}_1$ is negative.
3. To be able to calculate $\hat{\beta}_1$ x must have enough variability:

   $$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

   If all x values are the same then the sample variance will be 0. In this case, $\hat{\beta}_1$ will be undefined. For example, if all employees have the same level of education, say 12 years, then it is not possible to measure the impact of eduction on wages.

Ordinary Least Squares (OLS) Estimation

Fitted values of y can be calculated after $\hat{\beta}_0$ and $\hat{\beta}_1$ are found:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Residuals are the difference between observed and fitted values:

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \end{aligned}$$

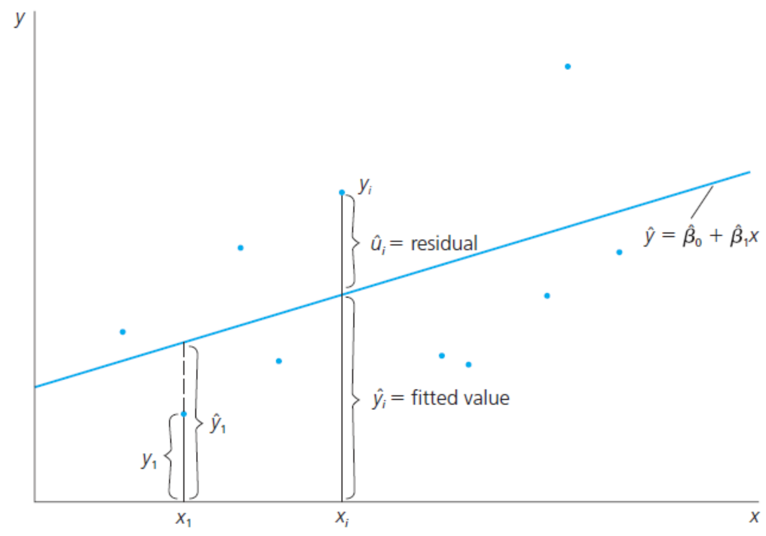
Residual is not the same as error term. The random error term u is unobserved whereas $\hat{u}$ is estimated given a sample of observations.

OLS Objective Function

OLS estimators are found by making the **sum of squared residuals** (SSR) as small as possible:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2$$

Residuals



Ordinary Least Squares (OLS) Estimators

OLS Problem

$$\min_{\hat{\beta}_0, \hat{\beta}_1} SSR = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

OLS First Order Conditions

$$\begin{aligned} \frac{\partial SSR}{\partial \hat{\beta}_0} &= -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \frac{\partial SSR}{\partial \hat{\beta}_1} &= -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{aligned}$$

The solution of this system is the same as the solution of the system obtained using the method of moments. Notice that if we multiply sample moment conditions by $-2n$ we obtain OLS first order conditions.

Population and Sample Regression Functions

Population Regression Function - PRF

$$\mathrm{E}(y|x) = \beta_0 + \beta_1 x$$

PRF is unique and unknown.

Sample Regression Function - SRF

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

SRF may be thought of as the estimated version of PRF. Interpretation of slope coefficient:

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x}$$

or

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x$$

Example: CEO Salary and Firm Performance

- We want to model the relationship between CEO salary and firm performance:

$$salary = \beta_0 + \beta_1 roe + u$$

- salary: annual CEO salary (1000 US$), roe: average return on equity for the last three years, %
- Using $n = 209$ firms in ceosal1.gdt data set in GRETL the SRF is estimated as follows:

$$\widehat{salary} = 963.191 + 18.501 roe$$

- $\hat{\beta}_1 = 18.501$. Interpretation: If the return of equity increases by one percentage point, i.e. $\Delta roe = 1$, then salary is predicted to increase by 18.501 or 18,501 US$ (ceteris paribus).

CEO Salary and Firm Performance: R application

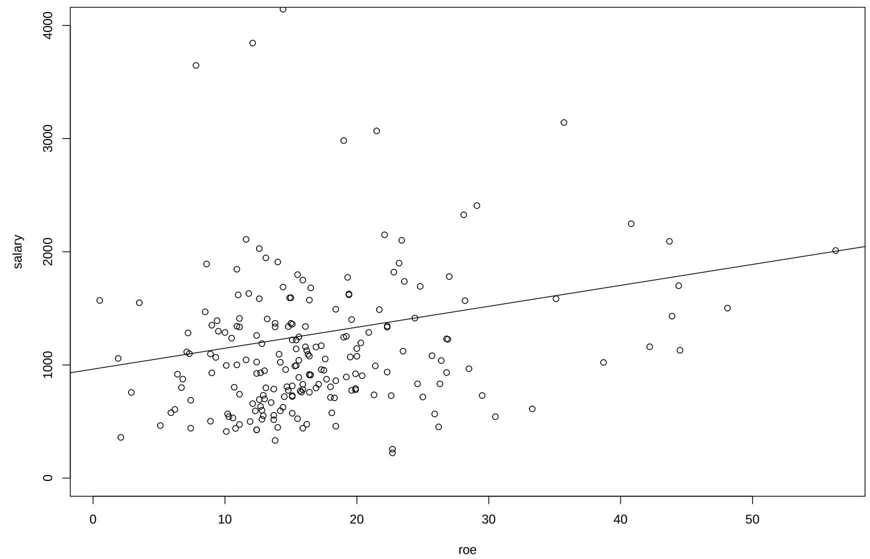
```
> library(wooldridge)
> data("ceosal1")
> View(ceosal1)
> lm(salary ~ roe, data = ceosal1)

Call:
lm(formula = salary ~ roe, data = ceosal1)

Coefficients:
(Intercept)          roe
      963.2         18.5
```

$$\widehat{salary} = 963.191 + 18.501 roe$$
$$salary = 963.191 + 18.501 roe + residual$$

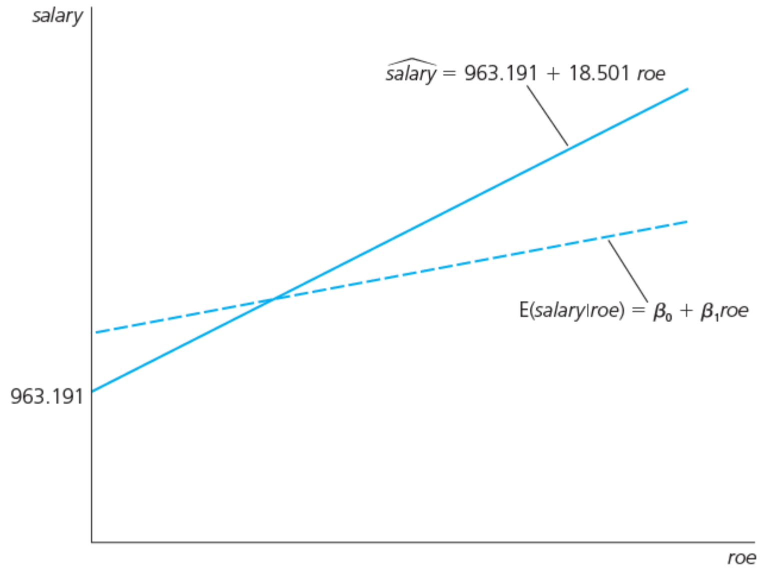
CEO Salary and Firm Performance: R application

```
> results1 <- lm(salary ~ roe, data = ceosal1)
> summary(results1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   963.19     213.24   4.517 1.05e-05 ***
roe            18.50      11.12   1.663   0.0978 .
---
Residual standard error: 1367 on 207 degrees of freedom
Multiple R-squared:  0.01319,Adjusted R-squared:  0.008421
F-statistic: 2.767 on 1 and 207 DF,  p-value: 0.09777

> attach(ceosal1)
> plot(roe, salary, ylim = c(0,4000))
> abline(results1)
```

CEO Salary Model - SRF



CEO Salary Model - SRF



Fitted Values and Residuals

```
> results1 <- lm(salary ~ roe, data = ceosal1)
> salaryhat <- fitted(results1)
> uhat <- resid(results1)
> table2.2 <- cbind(roe, salary, salaryhat, uhat)
> head(table2.2,n=15)

    roe salary salaryhat        uhat
1  14.1   1095  1224.058 -129.058071
2  10.9   1001  1164.854 -163.854261
3  23.5   1122  1397.969 -275.969216
4   5.9    578  1072.348 -494.348338
5  13.8   1368  1218.508  149.492288
6  20.0   1145  1333.215 -188.215063
7  16.4   1078  1266.611 -188.610785
8  16.3   1094  1264.761 -170.760660
9  10.5   1237  1157.454   79.546207
10 26.3    833  1449.773 -616.772523
11 25.9    567  1442.372 -875.372056
12 26.8    933  1459.023 -526.023116
13 14.8   1339  1237.009  101.991102
14 22.3    937  1375.768 -438.767778
15 56.3   2011  2004.808    6.191886
```

CEO Salary Model - Fitted values, Residuals (table 2.2 in the text)

TABLE 2.2 Fitted Values and Residuals for the First 15 CEOs

obsno	roe	salary	salaryhat	uhat
1	14.1	1095	1224.058	–129.0581
2	10.9	1001	1164.854	–163.8542
3	23.5	1122	1397.969	–275.9692
4	5.9	578	1072.348	–494.3484
5	13.8	1368	1218.508	149.4923
6	20.0	1145	1333.215	–188.2151
7	16.4	1078	1266.611	–188.6108
8	16.3	1094	1264.761	–170.7606
9	10.5	1237	1157.454	79.54626
10	26.3	833	1449.773	–616.7726
11	25.9	567	1442.372	–875.3721
12	26.8	933	1459.023	–526.0231
13	14.8	1339	1237.009	101.9911
14	22.3	937	1375.768	–438.7678
15	56.3	2011	2004.808	6.191895

Algebraic Properties of OLS Estimators

- Sum of OLS residuals, as well as their sample mean is zero:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \bar{\hat{u}} = 0$$

  This follows from the first sample moment condition.
- Sample covariance between x and residuals is zero:

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

  This follows from the second sample moment condition.
- The point $(\bar{x}, \bar{y})$ is always on the regression line.
- Sample average of the fitted values is equal to the sample average of observed y values: $\bar{y} = \bar{\hat{y}}$

Sum of Squares

- For each observation i we have
$$y_i = \hat{y}_i + \hat{u}_i$$
Summing both sides of this equation we obtain the following quantities:
- SST: Total Sum of Squares
$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- SSE: Explained Sum of Squares
$$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- SSR: Residual Sum of Squares
$$SSR = \sum_{i=1}^{n}\hat{u}_i^2$$

Sum of Squares

- SST gives the total variation in y:
$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
Note that Var$(y) = SST/(n-1)$.
- Similarly, SSE measures the variation in the fitted values.
$$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- SSR measures the sample variation in the residuals.
$$SSR = \sum_{i=1}^{n}\hat{u}_i^2$$
- Total sample variation in y can be written as
$$SST = SSE + SSR$$

Goodness-of-fit

- By definition total sample variation in y can be decomposed into two parts:
$$SST = SSE + SSR$$
- Dividing both sides by SST we obtain:
$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$
- The ratio of explained variation to the total variation is called the **coefficient of determination** and denoted by R^2:
$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = \frac{\text{Var}(\hat{y})}{\text{Var}(y)} = 1 - \frac{\text{Var}(\hat{u})}{\text{Var}(y)}$$
- Since SSE can never be larger than SST we have $0 \leq R^2 \leq 1$
- R^2 is interpreted as the fraction of the sample variation in y that is explained by x. After multiplying by 100 it can be interpreted as the percentage of the sample variation in y explained by x.
- R^2 can also be calculated as follows: $R^2 = Corr(y, \hat{y})^2$

R Example: College GPA and High School GPA

```
> gpareg <- lm(colGPA ~ hsGPA,data = gpa1)
> summary(gpareg)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.41543    0.30694   4.611 8.98e-06 ***
hsGPA        0.48243    0.08983   5.371 3.21e-07 ***
---
Residual standard error: 0.34 on 139 degrees of freedom
Multiple R-squared:  0.1719,Adjusted R-squared:  0.1659
F-statistic: 28.85 on 1 and 139 DF,  p-value: 3.211e-07
# obtaining R-squared manually:
> var(fitted(gpareg))/var(gpa1$colGPA)
[1] 0.1718563
# or
> cor(fitted(gpareg), gpa1$colGPA)^2
[1] 0.1718563
```

In equation form:

$$\widehat{colGPA} = 1.42 + 0.48\ hsGPA, \quad R^2 = 0.1719$$

Almost 17.19% of the variation in college GPA can be explained by the variations in high school GPA.