Advanced Quantitative Methods: Ordinary Least Squares

Jos Elkink

University College Dublin

February 2, 2011

Outline

Terminology

- **y** is the **dependent** variable
  - referred to also (by Greene) as a **regressand**

Terminology

- **y** is the **dependent** variable
  - referred to also (by Greene) as a **regressand**
- **X** are the **independent** variables
  - also known as **explanatory** variables
  - also known as **regressors** or **predictors** (or **factors**, **carriers**)
  - **X** is sometimes called the **design matrix** (or **factor space**)

Terminology

- $\mathbf{y}$ is the **dependent** variable
  - referred to also (by Greene) as a **regressand**
- $\mathbf{X}$ are the **independent** variables
  - also known as **explanatory** variables
  - also known as **regressors** or **predictors** (or **factors**, **carriers**)
  - $\mathbf{X}$ is sometimes called the **design matrix** (or **factor space**)
- $\mathbf{y}$ is **regressed on** $\mathbf{X}$

Terminology

- $\mathbf{y}$ is the **dependent** variable
  - referred to also (by Greene) as a **regressand**
- $\mathbf{X}$ are the **independent** variables
  - also known as **explanatory** variables
  - also known as **regressors** or **predictors** (or **factors**, **carriers**)
  - $\mathbf{X}$ is sometimes called the **design matrix** (or **factor space**)
- $\mathbf{y}$ is **regressed on** $\mathbf{X}$
- The **error** term ε is sometimes called a **disturbance**: $\varepsilon = \mathbf{y} - \mathbf{X}\beta$.
- The difference between the observed and predicted dependent variable is called the **residual**: $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$.

Linear model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon_i$$

Linear model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon_i$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Linear model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon_i$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$$

Linear model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon_i$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$$

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$$

Components

Two components of the model:

$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2)$	Stochastic
$\boldsymbol{\mu} = \mathbf{X}\beta$	Systematic

Components

Two components of the model:

$$\begin{array}{l|l} \mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2) & \text{Stochastic} \\ \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} & \text{Systematic} \end{array}$$

Generalised version (not necessarily linear):

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

(King 1998, 8)

Components

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \beta) & \text{Systematic} \end{array}$$

Stochastic component: varies over repeated (hypothetical) observations on the same unit.

Systematic component: varies across units, but constant given **X**.

(King 1998, 8)

Uncertainty

$$\begin{array}{c|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about α and β; can be reduced by increasing n.

Uncertainty

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; can be reduced by increasing n.
Fundamental uncertainty: represented by stochastic component and exists independent of researcher.

Ordinary Least Squares (OLS)

For the **linear** model, the most popular method of estimation is **ordinary least squares** (OLS).

$\hat{\beta}^{OLS}$ are those estimates of β that minimize the sum of squared residuals: $\mathbf{e}'\mathbf{e}$.

Ordinary Least Squares (OLS)

For the **linear** model, the most popular method of estimation is **ordinary least squares** (OLS).

$\hat{\boldsymbol{\beta}}^{OLS}$ are those estimates of $\boldsymbol{\beta}$ that minimize the sum of squared residuals: $\mathbf{e}'\mathbf{e}$.

OLS is the **best linear unbiased estimator** (BLUE).

Outline

Assumptions: specification

- Linear in parameters (i.e. $f(\mathbf{X}\beta) = \mathbf{X}\beta$ and $E(\mathbf{y}) = \mathbf{X}\beta$)

Note that this does not imply that you cannot include non-linearly transformed variables, e.g. $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$ can be estimated with OLS.

Assumptions: specification

- Linear in parameters (i.e. $f(\mathbf{X}\beta) = \mathbf{X}\beta$ and $E(\mathbf{y}) = \mathbf{X}\beta$)
- No extraneous variables in $\mathbf{X}$
- No omitted independent variables

Assumptions: specification

- Linear in parameters (i.e. $f(\mathbf{X}\beta) = \mathbf{X}\beta$ and $E(\mathbf{y}) = \mathbf{X}\beta$)
- No extraneous variables in $\mathbf{X}$
- No omitted independent variables
- Parameters to be estimated are constant

Assumptions: specification

- Linear in parameters (i.e. $f(\mathbf{X}\beta) = \mathbf{X}\beta$ and $E(\mathbf{y}) = \mathbf{X}\beta$)
- No extraneous variables in $\mathbf{X}$
- No omitted independent variables
- Parameters to be estimated are constant
- Number of parameters is less than the number of cases, $k < n$

Assumptions: errors

- Errors have an expected value of zero, $E(\varepsilon|\mathbf{X}) = 0$

Assumptions: errors

- Errors have an expected value of zero, $E(\varepsilon|\mathbf{X}) = 0$
- Errors are normally distributed, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$

Assumptions: errors

- Errors have an expected value of zero, $E(\varepsilon|\mathbf{X}) = 0$
- Errors are normally distributed, $\varepsilon \sim N(0, \sigma^2)$
- Errors have a constant variance, $var(\varepsilon|\mathbf{X}) = \sigma^2 < \infty$

Assumptions: errors

- Errors have an expected value of zero, $E(\boldsymbol{\varepsilon}|\mathbf{X}) = 0$
- Errors are normally distributed, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$
- Errors have a constant variance, $var(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2 < \infty$
- Errors are not autocorrelated, $cov(\varepsilon_i, \varepsilon_j|\mathbf{X}) = 0 \quad \forall \quad i \neq j$

Assumptions: errors

- Errors have an expected value of zero, $E(\varepsilon|\mathbf{X}) = 0$
- Errors are normally distributed, $\varepsilon \sim N(0, \sigma^2)$
- Errors have a constant variance, $var(\varepsilon|\mathbf{X}) = \sigma^2 < \infty$
- Errors are not autocorrelated, $cov(\varepsilon_i, \varepsilon_j|\mathbf{X}) = 0 \quad \forall \quad i \neq j$
- Errors and **X** are uncorrelated, $cov(\mathbf{X}, \varepsilon) = 0$

Assumptions: regressors

- **X** varies

Assumptions: regressors

- **X** varies
- **X** is of full column rank (note: requires $k < n$)

Assumptions: regressors

- **X** varies
- **X** is of full column rank (note: requires $k < n$)
- No measurement error in **X**

Assumptions: regressors

- **X** varies
- **X** is of full column rank (note: requires $k < n$)
- No measurement error in **X**
- No endogenous variables in **X**

Outline

1. Linear model
2. OLS assumptions
3. OLS algebra
4. OLS properties
5. R^2

$\mathbf{y} = \mathbf{X}\beta + \varepsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{k \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

$\mathbf{y} = \mathbf{X}\beta + \varepsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n\times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{k\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} \\ \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} \\ \vdots \\ \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} \end{bmatrix}_{n\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

Deriving $\hat{\beta}^{OLS}$

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\hat{\beta} + \mathbf{e}$$

Deriving $\hat{\boldsymbol{\beta}}^{OLS}$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{OLS} &= \underset{\hat{\boldsymbol{\beta}}}{\arg\min}\ \mathbf{e}'\mathbf{e} \\
&= \underset{\hat{\boldsymbol{\beta}}}{\arg\min}\,(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})
\end{aligned}$$

Deriving $\hat{\beta}^{OLS}$

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) \\
&= (\mathbf{y}' - (\mathbf{X}\hat{\beta})')(\mathbf{y} - \mathbf{X}\hat{\beta}) \\
&= \mathbf{y}'\mathbf{y} - (\mathbf{X}\hat{\beta})'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + (\mathbf{X}\hat{\beta})'\mathbf{X}\hat{\beta} \\
&= \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\
&= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}
\end{aligned}$$

Deriving $\hat{\boldsymbol\beta}^{OLS}$

$$\hat{\boldsymbol\beta}^{OLS} = \arg\min_{\hat{\boldsymbol\beta}} \mathbf{e}'\mathbf{e} \Longrightarrow$$

$$\frac{\partial(\mathbf{e}'\mathbf{e})}{\partial\hat{\boldsymbol\beta}^{OLS}} = 0$$

$$\frac{\partial(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\boldsymbol\beta} + \hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol\beta})}{\partial\hat{\boldsymbol\beta}^{OLS}} = 0$$

Deriving $\hat{\beta}^{OLS}$

$$\frac{\partial(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})}{\partial\hat{\beta}^{OLS}} = 0$$

$$2\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} - 2\mathbf{X}'\mathbf{y} = 0$$

$$2\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = 2\mathbf{X}'\mathbf{y}$$

$$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

$$\begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}'_{n \times k} \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k \times 1}$$

$$= \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}'_{n \times k} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

($\hat{\beta}$ here refers to OLS estimates.)

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ x_{13} & x_{23} & \cdots & x_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}_{k \times n} \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}_{n \times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k \times 1}$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ x_{13} & x_{23} & \cdots & x_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}_{k \times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

$$\begin{bmatrix} n & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i2} & \sum (x_{i2})^2 & \cdots & \sum x_{i2}x_{ik} \\ \sum x_{i3} & \sum x_{i3}x_{i2} & \cdots & \sum x_{i3}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i2} & \cdots & \sum (x_{ik})^2 \end{bmatrix}_{k\times k} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{k\times 1} = \begin{bmatrix} \sum y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}_{k\times 1}$$

($\sum$ refers to $\sum_i^n$.)

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

So this can be seen as a set of linear equations to solve:

$$\begin{aligned}
\hat{\beta}_1 n + \hat{\beta}_2 \sum x_{i2} \quad & + \cdots + \hat{\beta}_k \sum x_{ik} & = \sum y_i \\
\hat{\beta}_1 \sum x_{i2} + \hat{\beta}_2 \sum (x_{i2})^2 \quad & + \cdots + \hat{\beta}_k \sum x_{i2} x_{ik} & = \sum x_{i2} y_i \\
\hat{\beta}_1 \sum x_{i3} + \hat{\beta}_2 \sum x_{i3} x_{i2} \quad & + \cdots + \hat{\beta}_k \sum x_{i3} x_{ik} & = \sum x_{i3} y_i \\
& \vdots & \\
\hat{\beta}_1 \sum x_{ik} + \hat{\beta}_2 \sum x_{ik} x_{i2} \quad & + \cdots + \hat{\beta}_k \sum (x_{ik})^2 & = \sum x_{ik} y_i
\end{aligned}$$

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

When there is only one independent variable, this reduces to:

$$\hat{\beta}_1 n + \hat{\beta}_2 \sum x_i = \sum y_i$$
$$\hat{\beta}_1 \sum x_i + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$\mathbf{X}'\mathbf{X}\hat{\beta}^{OLS} = \mathbf{X}'\mathbf{y}$

When there is only one independent variable, this reduces to:

$$\hat{\beta}_1 n + \hat{\beta}_2 \sum x_i = \sum y_i$$

$$\hat{\beta}_1 \sum x_i + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_1 = \frac{\sum y_i - \hat{\beta}_2 \sum x_i}{n} = \frac{n\bar{y} - \hat{\beta}_2 n\bar{x}}{n} = \bar{y} - \hat{\beta}_2\bar{x}$$

$$(\bar{y} - \hat{\beta}_2\bar{x}) \sum x_i + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$n\bar{y}\bar{x} - \hat{\beta}_2 n\bar{x}^2 + \hat{\beta}_2 \sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i - n\bar{y}\bar{x}}{\sum x_i^2 - n\bar{x}^2}$$

Simple regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Simple regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

If $\beta_1 = 0$ and the only regressor is the intercept, then $\hat{\beta}_0 = \bar{y}$.

Simple regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

If $\beta_1 = 0$ and the only regressor is the intercept, then $\hat{\beta}_0 = \bar{y}$.

If $\beta_0 = 0$, so that there is no intercept and one explanatory variable $\mathbf{x}$, then $\hat{\beta}_1 = \frac{\sum_i^n x_i y_i}{\sum_i^n x_i^2}$.

Simple regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

If $\beta_1 = 0$ and the only regressor is the intercept, then $\hat{\beta}_0 = \bar{y}$.

If $\beta_0 = 0$, so that there is no intercept and one explanatory variable $\mathbf{x}$, then $\hat{\beta}_1 = \frac{\sum_i^n x_i y_i}{\sum_i^n x_i^2}$.

If there is an intercept and one explanatory variable, then

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

Demeaning

If the observations are expressed as deviations from their means, i.e.:

$$y_i* = y_i - \bar{y}$$
$$x_i^* = x_i - \bar{x},$$

then

$$\hat{\beta}_1 = \frac{\sum_i^n x_i^* y_i^*}{\sum_i^n x_i^{*2}}$$

Demeaning

If the observations are expressed as deviations from their means, i.e.:

$$y_i* = y_i - \bar{y}$$
$$x_i^* = x_i - \bar{x},$$

then

$$\hat{\beta}_1 = \frac{\sum_i^n x_i^* y_i^*}{\sum_i^n x_i^{*2}}$$

The intercept can be estimated as $\bar{y} - \hat{\beta}_1 \bar{x}$.

Standardized variables

The coefficient on a variable is interpreted as the effect on **y** given a one unit increase in **x** and thus the interpretation is dependent on the scale of measurement.

Standardized variables

The coefficient on a variable is interpreted as the effect on **y** given a one unit increase in **x** and thus the interpretation is dependent on the scale of measurement.

An option can be to standardize the variables:

$$\mathbf{y}^* = \frac{\mathbf{y} - \bar{y}}{\sqrt{\sigma_y^2}}$$

$$\mathbf{x}^* = \frac{\mathbf{x} - \bar{x}}{\sqrt{\sigma_x^2}}$$

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta}^* + \varepsilon^*$$

and interpret the coefficients as the effect on **y** expressed in standard deviations as **x** increase by one standard deviation.

Standardized variables

```
library(arm)
summary(standardize(lm(y ~ x)),
          binary.inputs="leave.alone")
```

Standardized variables

```
library(arm)
summary(standardize(lm(y ~ x)),
          binary.inputs="leave.alone")
```

See also Andrew Gelman (2007), "Scaling regression inputs by dividing by two standard deviations".

Exercise: teenage gambling

```
library(faraway)
data(teengamb)
summary(teengamb)
```

Using matrix formulas,

1. regress *gamble* on a constant
2. regress *gamble* on *sex*
3. regress *gamble* on *sex*, *status*, *income*, *verbal*

Exercise: teenage gambling

1. Regress *gamble* on *sex*, *status*, *income*, *verbal*
2. Which observation has the largest residual?
3. Compute mean and median of residuals
4. Compute correlation between residuals and fitted values
5. Compute correlation between residuals and *income*
6. All other predictors held constant, what would be the difference in predicted expenditure on gambling between males and females?

Deriving $var(\hat{\beta}^{OLS})$

$$\begin{aligned} var(\hat{\beta}^{OLS}) &= E[(\hat{\beta}^{OLS} - E(\hat{\beta}^{OLS}))(\hat{\beta}^{OLS} - E(\hat{\beta}^{OLS}))'] \\ &= E[(\hat{\beta}^{OLS} - \beta)(\hat{\beta}^{OLS} - \beta)'] \end{aligned}$$

Deriving $var(\hat{\beta}^{OLS})$

$$\begin{aligned}\hat{\beta}^{OLS} - \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon - \beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\end{aligned}$$

Deriving $var(\hat{\beta}^{OLS})$

$$
\begin{aligned}
var(\hat{\beta}^{OLS}) &= E[(\hat{\beta}^{OLS} - \beta)(\hat{\beta}^{OLS} - \beta)'] \\
&= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon)((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon)'] \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

Standard errors

$$var(\hat{\beta}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

The standard errors of $\hat{\beta}^{OLS}$ are then the square root of the diagonal of this matrix.

Standard errors

$$var(\hat{\beta}^{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

The standard errors of $\hat{\beta}^{OLS}$ are then the square root of the diagonal of this matrix.

In the simple case where $\mathbf{y} = \beta_0 + \beta_1\mathbf{x}$, this gives

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i^n (x_i - \bar{x})^2}$$

Note how an increase in variance in $\mathbf{x}$ leads a decrease in the standard error of $\hat{\beta}_1$.

Estimating σ^2

$$
\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\beta}^{OLS} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\
&= \mathbf{M}\mathbf{y}
\end{aligned}
$$

Estimating σ^2

$$
\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\beta}^{OLS} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\
&= \mathbf{M}\mathbf{y} \\
&= \mathbf{M}(\mathbf{X}\beta + \varepsilon) \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\beta + \mathbf{M}\varepsilon \\
&= \mathbf{X}\beta - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + \mathbf{M}\varepsilon \\
&= \mathbf{X}\beta - \mathbf{X}\beta + \mathbf{M}\varepsilon \\
&= \mathbf{M}\varepsilon
\end{aligned}
$$

Estimating σ^2

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= (\mathbf{M}\varepsilon)'(\mathbf{M}\varepsilon) \\
&= \varepsilon'\mathbf{M}'\mathbf{M}\varepsilon \\
&= \varepsilon'\mathbf{M}\varepsilon \\
E[\mathbf{e}'\mathbf{e}|\mathbf{X}] &= E[\varepsilon'\mathbf{M}\varepsilon|\mathbf{X}]
\end{aligned}$$

Estimating σ^2

$$\begin{aligned}
\mathbf{e}'\mathbf{e} &= (\mathbf{M}\varepsilon)'(\mathbf{M}\varepsilon) \\
&= \varepsilon'\mathbf{M}'\mathbf{M}\varepsilon \\
&= \varepsilon'\mathbf{M}\varepsilon \\
E[\mathbf{e}'\mathbf{e}|\mathbf{X}] &= E[\varepsilon'\mathbf{M}\varepsilon|\mathbf{X}]
\end{aligned}$$

$\varepsilon'\mathbf{M}\varepsilon$ is a scalar, so if you see it as a 1×1 matrix, it is equal to its trace, therefore

$$\begin{aligned}
E[\mathbf{e}'\mathbf{e}|\mathbf{X}] &= E[tr(\varepsilon'\mathbf{M}\varepsilon|\mathbf{X}] \\
&= E[tr(\mathbf{M}\varepsilon\varepsilon')|\mathbf{X}] \\
&= tr(\mathbf{M}E[\varepsilon\varepsilon'|\mathbf{X}]) \\
&= tr(\mathbf{M}\sigma^2\mathbf{I}) = \sigma^2 tr(\mathbf{M})
\end{aligned}$$

Estimating σ^2

$$\begin{aligned}
tr(\mathbf{M}) &= tr(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
&= tr(\mathbf{I}) - tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\
&= n - tr(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\
&= n - k \\
E(\mathbf{e}'\mathbf{e}|\mathbf{X}) &= (n-k)\sigma^2 \\
\hat{\sigma}^2 &= \frac{\mathbf{e}'\mathbf{e}}{n-k}
\end{aligned}$$

OLS in R

```
n <- dim(X)[1]
k <- dim(X)[2]
b.hat <- solve(t(X) %*% X) %*% t(X) %*% y
e <- y - X %*% b.hat
s2.hat <- 1/(n-k) * t(e) %*% e
v.hat <- s2.hat * solve(t(X) %*% X)
```

OLS in R

```
n <- dim(X)[1]
k <- dim(X)[2]
b.hat <- solve(t(X) %*% X) %*% t(X) %*% y
e <- y - X %*% b.hat
s2.hat <- 1/(n-k) * t(e) %*% e
v.hat <- s2.hat * solve(t(X) %*% X)
```

Or:

```
summary(lm(y ~ X))
```

Exercise: US wages

```
library(faraway)
data(uswages)
summary(uswages)
```

1. Using matrix formulas, regress *wage* on *educ*, *exper* and *race*.
2. Interpret the results
3. Plot residuals against fitted values and against *educ*
4. Repeat with $log(wage)$ as dependent variable

Outline

Unbiasedness of $\hat{\beta}^{OLS}$

From deriving $var(\hat{\beta}^{OLS})$ we have:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}
\end{aligned}$$

Unbiasedness of $\hat{\beta}^{OLS}$

From deriving $var(\hat{\beta}^{OLS})$ we have:

$$\begin{aligned}\hat{\beta}^{OLS} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

Since we assume $E(\boldsymbol{\varepsilon}) = 0$:

$$E(\hat{\beta}^{OLS} - \boldsymbol{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) = 0$$

Unbiasedness of $\hat{\beta}^{OLS}$

From deriving $var(\hat{\beta}^{OLS})$ we have:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

Since we assume $E(\boldsymbol{\varepsilon}) = 0$:

$$E(\hat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) = 0$$

Therefore, $\hat{\boldsymbol{\beta}}^{OLS}$ is an **unbiased** estimator of $\boldsymbol{\beta}$.

Efficiency of $\hat{\beta}^{OLS}$

The **Gauss-Markov theorem** states that there is no **linear unbiased estimator** of β that has a smaller sampling variance than $\hat{\beta}^{OLS}$, i.e. $\hat{\beta}^{OLS}$ is BLUE.

Efficiency of $\hat{\beta}^{OLS}$

The **Gauss-Markov theorem** states that there is no **linear unbiased estimator** of β that has a smaller sampling variance than $\hat{\beta}^{OLS}$, i.e. $\hat{\beta}^{OLS}$ is BLUE.

An estimator is **linear** iff it can be expressed as a linear function of the data on the dependent variable: $\hat{\beta}_j^{linear} = \sum_i^n f(x_{ij})y_i$, which is the case for OLS:

$$\begin{aligned}\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{Cy}\end{aligned}$$

(Wooldridge, pp. 101-102, 111-112)

Gauss-Markov Theorem

$$\begin{aligned}\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{C}\mathbf{y}\end{aligned}$$

Imagine we have another linear estimator: $\hat{\beta}^{*} = \mathbf{W}\mathbf{y}$, where $\mathbf{W} = \mathbf{C} + \mathbf{D}$.

Gauss-Markov Theorem

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{C}\mathbf{y}\end{aligned}$$

Imagine we have another linear estimator: $\hat{\boldsymbol{\beta}}^{*} = \mathbf{W}\mathbf{y}$, where $\mathbf{W} = \mathbf{C} + \mathbf{D}$.

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{*} &= (\mathbf{C} + \mathbf{D})\mathbf{y} \\ &= \mathbf{C}\mathbf{y} + \mathbf{D}\mathbf{y} \\ &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}\mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\varepsilon}\end{aligned}$$

Gauss-Markov Theorem

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}^*) &= E(\hat{\boldsymbol{\beta}}^{OLS} + \mathbf{DX}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\varepsilon}) \\
&= E(\hat{\boldsymbol{\beta}}^{OLS}) + \mathbf{DX}\boldsymbol{\beta} + E(\mathbf{D}\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + \mathbf{DX}\boldsymbol{\beta} + 0 \\
\Rightarrow \mathbf{DX}\boldsymbol{\beta} &= 0 \\
\Rightarrow \mathbf{DX} &= 0,
\end{aligned}
$$

because both $\hat{\boldsymbol{\beta}}^{OLS}$ and $\hat{\boldsymbol{\beta}}^*$ are unbiased, so $E(\hat{\boldsymbol{\beta}}^{OLS}) = E(\hat{\boldsymbol{\beta}}^*) = \boldsymbol{\beta}$.

(Hayashi, pp. 29-30)

Gauss-Markov Theorem

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}\boldsymbol{\varepsilon} \\ \hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= (\mathbf{C} + \mathbf{D})\boldsymbol{\varepsilon} \end{aligned}$$

Gauss-Markov Theorem

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^* &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}\varepsilon \\
\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} &= \hat{\boldsymbol{\beta}}^{OLS} + \mathbf{D}\varepsilon - \boldsymbol{\beta} \\
&= (\mathbf{C} + \mathbf{D})\varepsilon \\
var(\hat{\boldsymbol{\beta}}^*) &= E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})' \\
&= E((\mathbf{C} + \mathbf{D})\varepsilon)E((\mathbf{C} + \mathbf{D})\varepsilon)' \\
&= (\mathbf{C} + \mathbf{D})E(\varepsilon\varepsilon')(\mathbf{C} + \mathbf{D})' \\
&= \sigma^2(\mathbf{C} + \mathbf{D})(\mathbf{C} + \mathbf{D})' \\
&= \sigma^2(\mathbf{CC}' + \mathbf{DC}' + \mathbf{CD}' + \mathbf{DD}') \\
&= \sigma^2((\mathbf{X}'\mathbf{X})^{-1} + \mathbf{DD}') \\
&\geq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

Consistency of $\hat{\beta}^{OLS}$

$$
\begin{aligned}
\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\
&= \beta + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\varepsilon)
\end{aligned}
$$

Consistency of $\hat{\beta}^{OLS}$

$$\begin{aligned}
\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\
&= \beta + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\varepsilon)
\end{aligned}$$

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q} \quad \Longrightarrow \quad \lim_{n\to\infty} (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1} = \mathbf{Q}^{-1}$$

Consistency of $\hat{\beta}^{OLS}$

$$\begin{aligned}
\hat{\beta}^{OLS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\
&= \beta + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\varepsilon)
\end{aligned}$$

$$\lim_{n\to\infty} \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q} \quad \Longrightarrow \quad \lim_{n\to\infty} (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1} = \mathbf{Q}^{-1}$$

$$E(\frac{1}{n}\mathbf{X}'\varepsilon) = E(\frac{1}{n}\sum_{i}^{n} x_i\varepsilon_i) = 0$$

Consistency of $\hat{\beta}^{OLS}$

$$\begin{aligned} var(\frac{1}{n}\mathbf{X}'\varepsilon) &= E(\frac{1}{n}\mathbf{X}'\varepsilon(\frac{1}{n}\mathbf{X}'\varepsilon)') \\ &= \frac{1}{n}\mathbf{X}'E(\varepsilon\varepsilon')\mathbf{X}\frac{1}{n} \\ &= (\frac{\sigma^2}{n})(\frac{\mathbf{X}'\mathbf{X}}{n}) \\ \lim_{n\to\infty} var(\frac{1}{n}\mathbf{X}'\varepsilon) &= 0\mathbf{Q} = 0 \end{aligned}$$

Consistency of $\hat{\beta}^{OLS}$

$$\hat{\beta}^{OLS} = \beta + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\varepsilon)$$

$$\lim_{n\to\infty}(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1} = \mathbf{Q}^{-1}$$

$$E(\frac{1}{n}\mathbf{X}'\varepsilon) = 0$$

$$\lim_{n\to\infty} var(\frac{1}{n}\mathbf{X}'\varepsilon) = 0$$

imply that the sampling distribution of $\hat{\beta}^{OLS}$ "collapses" to β as n becomes very large, i.e.:

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^{OLS} = \beta$$

Outline

Sums of squares

SST Total sum of squares $\sum(y_i - \bar{y})^2$

SSE Explained sum of squares $\sum(\hat{y}_i - \bar{y})^2$

SSR Residual sum of squares $\sum e_i^2 = \sum(\hat{y}_i - y_i)^2 = \mathbf{e}'\mathbf{e}$

The key to remember is that **SST = SSE + SSR**

Sums of squares

SST Total sum of squares $\sum(y_i - \bar{y})^2$

SSE Explained sum of squares $\sum(\hat{y}_i - \bar{y})^2$

SSR Residual sum of squares $\sum e_i^2 = \sum(\hat{y}_i - y_i)^2 = \mathbf{e}'\mathbf{e}$

The key to remember is that $\mathbf{SST = SSE + SSR}$

Sometimes instead of "explained" and "residual", "regression" and "error" are used, respectively, so that the abbreviations are swapped (!).

R^2

Defined in terms of sums of squares:

$$\begin{aligned} R^2 &= \frac{SSE}{SST} \\ &= 1 - \frac{SSR}{SST} \\ &= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \end{aligned}$$

Interpretation: the proportion of the variation in **y** that is explained linearly by the independent variables

R^2

Defined in terms of sums of squares:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Interpretation: the proportion of the variation in **y** that is explained linearly by the independent variables

A much over-used statistic: it may not be what we are interested in at all

R^2 in matrix algebra

$$
\begin{aligned}
\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{y}}'\mathbf{e} + \mathbf{e}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\hat{\beta}'\mathbf{X}'\mathbf{e} + \mathbf{e}'\mathbf{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}
\end{aligned}
$$

R^2 in matrix algebra

$$\begin{aligned}
\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{y}}'\mathbf{e} + \mathbf{e}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + 2\hat{\beta}'\mathbf{X}'\mathbf{e} + \mathbf{e}'\mathbf{e} \\
&= \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\
R^2 &= 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{y}} = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}}
\end{aligned}$$

(Hayashi, p. 20)

R^2

When a model has no intercept, it is possible for R^2 to lie outside the interval $(0, 1)$

R^2

When a model has no intercept, it is possible for R^2 to lie outside the interval $(0, 1)$

R^2 rises with the addition of more explanatory variables. For this reason we often report the **adjusted** R^2:

$$1 - (1 - R^2)\frac{n-1}{n-k}$$

R^2

When a model has no intercept, it is possible for R^2 to lie outside the interval $(0, 1)$

R^2 rises with the addition of more explanatory variables. For this reason we often report the **adjusted** R^2:

$$1 - (1 - R^2)\frac{n-1}{n-k}$$

The R^2 values from different **y** samples *cannot be compared*

Exercise: US wages

```
library(faraway)
data(uswages)
summary(uswages)
```

1. Using matrix formulas, regress *wage* on *educ*, *exper* and *race*.
2. What proportion of the variance in *wage* is explained by these three variables?