Simple Linear Regression: Model Inference

Engineering Statistics II
Section 12.3

Josh Engwer

TTU

2023

PART I:

Simple Linear Regression

Estimation of β_0 & β_1

Model Assumptions

Linear Model

Sums

Centered Sums

OLS Point Estimators of β_0 & β_1

Sums of Squares

Partitioning Variation

Simple Linear Regression (Model Assumptions)

Proposition

(Simple Linear Regression Model Assumptions)

- *(**1 Numerical Response**) Response is not categorical.*
- *(**1 Numerical Regressor**) Regressor is not categorical.*

- *(**Regressors are Perfect**) No errors in regressor measurements.*
- *(**Balance around Fit Line**) Nearly equal scatter above & below fit line.*

- *(**Independence**) All measurements are independent.*
- *(**Normality**) All measurements are approximately normally distributed.*
- *(**Equal Variances**) All measurements have approx. same variance.*

Mnemonic: ***1NR(1NR)*** | ***RaP BaFL*** | ***I.N.EV***

Simple Linear Regression Models (Sums)

Several particular sums show up in Simple Linear Regression:

Definition

(Sums)

Let vectors $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$. Then:

Regressor Sum	$S_x := \sum_i x_i$
Response Sum	$S_y := \sum_i y_i$
Squared Regressor Sum	$S_{xx} := \sum_i x_i x_i$
Squared Response Sum	$S_{yy} := \sum_i y_i y_i$
Cross-termed Sum	$S_{xy} := \sum_i x_i y_i$

Several particular centered sums show up in Simple Linear Regression:

Definition

(Centered Sums)

Let vectors $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$. Then:

Squared Regressor Centered Sum	$SC_{xx} := \sum_i (x_i - \bar{x})^2$
Squared Response Centered Sum	$SC_{yy} := \sum_i (y_i - \bar{y})^2$
Cross-termed Centered Sum	$SC_{xy} := \sum_i (x_i - \bar{x})(y_i - \bar{y})$

Simple Linear Regression Models (Centered Sums)

It is useful to express the centered sums in terms of sums:

Lemma

(Centered Sum Lemma – CSL)

$$(a)\ SC_{xx} = S_{xx} - \frac{1}{n} S_x S_x \qquad (b)\ SC_{yy} = S_{yy} - \frac{1}{n} S_y S_y \qquad (c)\ SC_{xy} = S_{xy} - \frac{1}{n} S_x S_y$$

Simple Linear Regression (OLS Estimators)

Simple Regression Linear Model

Y_i	$\equiv$	rv for i^{th} response measurement
x_i	$\equiv$	Actual value of i^{th} regressor measurement
β_0	$\equiv$	Expected value of response when regressor is zero
β_1	$\equiv$	Expected change in response per one-unit increase in regressor
E_i	$\equiv$	Effect of random error on i^{th} response

ASSUMPTIONS: $E_1, E_2, \cdots, E_n \overset{iid}{\sim}$ Normal $\left(0, \sigma^2\right)$

LINEAR MODEL: $Y_i = \beta_0 + \beta_1 x_i + E_i \qquad \forall\, i = 1, \ldots, n$

REALIZED MODEL: $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$

$$\hat{\beta}_1 = \frac{SC_{xy}}{SC_{xx}} = \frac{S_{xy} - \frac{1}{n} S_x S_y}{S_{xx} - \frac{1}{n} S_x S_x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{S_y - \hat{\beta}_1 S_x}{n}$$

Simple Linear Regression (Sums of Squares)

Definition

(Sums of Squares for Simple Linear Regression)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\ i = 1, \ldots, n$$

Then, there are three key sums of squares:

(Total Variation) $\quad SS_{total} := \sum_i (y_i - \bar{y})^2$

(Unexplained Variation) $\quad SS_{res} := \sum_i (y_i - \hat{y}_i)^2$

(Explained Variation) $\quad SS_{reg} := \sum_i (\hat{y}_i - \bar{y})^2$

Simple Linear Regression (Partitioning Variation)

Theorem

(Sums of Squares Partitioning Variation Theorem – SSPVT)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

Then, the three key sums of squares are partitioned as follows:

$$\underbrace{\textstyle\sum_i (y_i - \overline{y})^2}_{SS_{total}} = \underbrace{\textstyle\sum_i (y_i - \hat{y}_i)^2}_{SS_{res}} + \underbrace{\textstyle\sum_i (\hat{y}_i - \overline{y})^2}_{SS_{reg}}$$

Lemma

(Sums of Squares Lemma – SSL)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \ \text{ where } \ E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

Then:

$$(a)\ SS_{total} = SC_{yy} \qquad (b)\ SS_{res} = SC_{yy} - \hat{\beta}_1^2 \cdot SC_{xx} \qquad (c)\ SS_{reg} = \hat{\beta}_1^2 \cdot SC_{xx}$$

PART II:

Simple Linear Regression

Estimation of σ^2

Expectation of $\hat{\beta}_1$

Variance of $\hat{\beta}_1$

Expectation of SS_{res}

Mean Squared Residual, MS_{res}

Point Estimator of σ^2

Estimated Variance of $\hat{\beta}_1$

Simple Linear Regression $(\mathbb{E}[\hat{\beta}_1]$ & $\mathbb{V}[\hat{\beta}_1])$

The expectation & variance of $\hat{\beta}_1$ are needed in the estimation of σ^2.
The expectation & variance of $\hat{\beta}_1$ are also used in inference later.

Theorem

(Expectation & Variance of $\hat{\beta}_1$ Theorem – EVB1HT)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS point estimators $\hat{\beta}_0, \hat{\beta}_1$.

$$\text{Then:} \qquad (a)\ \mathbb{E}[\hat{\beta}_1] = \beta_1 \qquad (b)\ \mathbb{V}[\hat{\beta}_1] = \frac{\sigma^2}{SC_{xx}}$$

Simple Linear Regression (Expectation of SS_{res})

Lemma

(Estimation of SS_{res} Lemma – ESSRESL)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS estimators $\hat{\beta}_0, \hat{\beta}_1$.

Then:

$$\mathbb{E}\left[SS_{res}\right] = (n-2)\sigma^2$$

Definition

(Mean Squared Residual)

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\ i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS estimators $\hat{\beta}_0, \hat{\beta}_1$.

Then:

$$\text{MS}_{res} := \frac{\text{SS}_{res}}{n-2}$$

Simple Linear Regression (Point Estimator of σ^2)

Proposition

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \;\; \text{where} \;\; E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS point estimators $\hat{\beta}_0, \hat{\beta}_1$.

Then:

$$\mathbb{E}\left[MS_{res}\right] = \sigma^2 \implies \hat{\sigma}^2 = MS_{res}$$

i.e. MS_{res} is always an unbiased estimator of σ^2.

Simple Linear Regression (Estimated Variance of $\hat{\beta}_1$)

Proposition

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \textit{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS point estimators $\hat{\beta}_0, \hat{\beta}_1$.

Then:

$$\hat{\mathbb{V}}[\hat{\beta}_1] = \frac{\hat{\sigma}^2}{SC_{xx}} = \frac{MS_{res}}{SC_{xx}}$$

PART III:

Simple Linear Regression

Model Inference

Estimated Std Dev of $\hat{\beta}_1$

t-CI for β_1

Model Utility t-Test

Model Utility F-Test

Simple Linear Regression (Estimated Std. Dev. of $\hat{\beta}_1$)

Proposition

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \quad \textit{where} \quad E_1, \cdots, E_n \overset{IID}{\sim} \textit{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS point estimators $\hat{\beta}_0, \hat{\beta}_1$.

Then:

$$\hat{\mathbb{D}}[\hat{\beta}_1] = \sqrt{\hat{\mathbb{V}}[\hat{\beta}_1]} = \sqrt{\frac{\hat{\sigma}^2}{SC_{xx}}} = \sqrt{\frac{MS_{res}}{SC_{xx}}}$$

Simple Linear Regression (t-CI for β_1)

Proposition

Given a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + E_i \text{ where } E_1, \cdots, E_n \overset{IID}{\sim} \text{Normal}(0, \sigma^2)$$

and the corresponding realized model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad \forall\, i = 1, \ldots, n$$

...where parameters β_0, β_1 are estimated by OLS point estimators $\hat{\beta}_0, \hat{\beta}_1$.

*Then the $100(1-\alpha)\%$ **independent t-CI for β_1** is*

$$\hat{\beta}_1 \pm t^*_{\nu_{res};\alpha/2} \cdot \hat{\mathbb{D}}[\hat{\beta}_1]$$

$$\text{where} \quad \hat{\mathbb{D}}[\hat{\beta}_1] := \sqrt{\hat{\mathbb{V}}[\hat{\beta}_1]} = \sqrt{\frac{\hat{\sigma}^2}{SC_{xx}}} = \sqrt{\frac{MS_{res}}{SC_{xx}}}$$

Simple Linear Regression (Model Utility t-Test)

Proposition

	1NR(1NR) | ***RaP BaFL*** | ***I.N.EV***
Linear Model:	$Y_i = \beta_0 + \beta_1 x_i + E_i$ $E_1, \cdots, E_n \overset{IID}{\sim}$ *Normal*$(0, \sigma^2)$
Realized Linear Model:	$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, \cdots, n$
Model Parameter Estimates:	$\hat{\beta}_1 = SC_{xy}/SC_{xx}, \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$
Predicted Values & Residuals:	$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad y_i^{res} := y_i - \hat{y}_i$
Test Statistic Value:	$t = \dfrac{\hat{\beta}_1}{\hat{\mathbb{D}}[\hat{\beta}_1]}$ *s.t.* $\hat{\mathbb{D}}[\hat{\beta}_1] := \sqrt{\dfrac{\hat{\sigma}^2}{SC_{xx}}} = \sqrt{\dfrac{MS_{res}}{SC_{xx}}}$

HYPOTHESIS TEST:	***REJECTION REGION AT LVL*** α :
$H_0 : \beta_1 = 0$ *vs.* $H_A : \beta_1 \neq 0$	$t \leq -t^*_{\nu_{res};\alpha/2}$ *or* $t \geq t^*_{\nu_{res};\alpha/2}$

Simple Linear Regression (Model Utility t-Test)

Proposition

Linear Model:	***1NR(1NR)*** | ***RaP BaFL*** | ***I.N.EV*** $Y_i = \beta_0 + \beta_1 x_i + E_i$ $E_1, \cdots, E_n \overset{IID}{\sim}$ Normal$(0, \sigma^2)$
Realized Linear Model:	$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, \cdots, n$
Model Parameter Estimates:	$\hat{\beta}_1 = SC_{xy}/SC_{xx}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
Predicted Values & Residuals:	$\hat{y}_i := \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad y_i^{res} := y_i - \hat{y}_i$
Test Statistic Value:	$t = \frac{\hat{\beta}_1}{\hat{\mathbb{D}}[\hat{\beta}_1]}$ *s.t.* $\hat{\mathbb{D}}[\hat{\beta}_1] := \sqrt{\frac{\hat{\sigma}^2}{SC_{xx}}} = \sqrt{\frac{MS_{res}}{SC_{xx}}}$

HYPOTHESIS TEST:	***P-VALUE DETERMINATION:***
$H_0 : \beta_1 = 0$ *vs.* $H_A : \beta_1 \neq 0$	*P-value* $= 2 \cdot [1 - \Phi_t(\lvert t \rvert; \nu_{res})]$

DECISION RULE:
If P-value $\leq \alpha$ *then reject* H_0 *in favor of* H_A
If P-value $> \alpha$ *then accept* H_0 (*i.e. fail to reject* H_0)

Simple Linear Regression (Model Utility F-Test)

(Given Linear Model: $Y_i = \beta_0 + \beta_1 x_i + E_i$)

$$H_0: \beta_1 = 0$$
$$H_A: \beta_1 \neq 0$$

---

1. Determine degrees of freedom: $\nu_{total} = n - 1, \quad \nu_{res} = n - 2, \quad \nu_{reg} = 1$
2. Compute Sums: $S_x := \sum_i x_i, \quad S_y = \sum_i y_i, \quad S_{xx} = \sum_i x_i x_i, \quad S_{yy} = \sum_i y_i y_i, \quad S_{xy} = \sum_i x_i y_i$
3. Compute Centered Sums: $SC_{xx} = S_{xx} - \frac{1}{n} S_x S_x, \quad SC_{yy} = S_{yy} - \frac{1}{n} S_y S_y, \quad SC_{xy} = S_{xy} - \frac{1}{n} S_x S_y$
4. Compute Parameter Estimates: $\hat{\beta}_1 = \dfrac{SC_{xy}}{SC_{xx}}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
5. Compute Sums of Squares: $\text{SS}_{total} = SC_{yy}, \quad \text{SS}_{res} = SC_{yy} - \hat{\beta}_1^2 \cdot SC_{xx}, \quad \text{SS}_{reg} = \hat{\beta}_1^2 \cdot SC_{xx}$
6. Compute Mean Squares: $\text{MS}_{res} := \text{SS}_{res} / \nu_{res}, \quad \text{MS}_{reg} := \text{SS}_{reg} / \nu_{reg}$
7. Compute Test Statistic Value: $f_{reg} = \text{MS}_{reg} / \text{MS}_{res}$
8. Compute F-cutoff/P-value:
   - By hand, lookup $f^*_{\nu_{reg}, \nu_{res}; \alpha}$
   - By SW, compute $p_{reg} = 1 - \Phi_F(f_{reg}; \nu_{reg}, \nu_{res})$
9. Render Decision:
   - If $f_{reg} \geq f^*_{\nu_{reg}, \nu_{res}; \alpha}$, then reject H_0; else accept H_0.
   - If $p_{reg} \leq \alpha$, then reject H_0; else accept H_0.

$Y_i = \beta_0 + \beta_1 x_i + E_i$ where $E_i \overset{IID}{\sim}$ Normal$(0, \sigma^2)$ **Model Utility F-Test (Significance Level α)**						
Variation Source	**df**	**Sum of Squares**	**Mean Square**	**F Stat Value**	**P-value**	**Decision**
Regression	ν_{reg}	SS$_{reg}$	MS$_{reg}$	f_{reg}	p_{reg}	Acc/Rej H_0
Unknown	ν_{res}	SS$_{res}$	MS$_{res}$			
Total	ν_{total}	SS$_{total}$				

Fin.