Part I: Simple linear regression

L01 Simple linear regression model and least square method

1. Two simple linear regression models

   (1) Simple linear regression with intercept

   $$y = \beta_0 + \beta_1 x + \epsilon,\ \epsilon \sim N(0,\ \sigma^2)$$

   is a simple linear regression model where the mean of response (dependent variable) is a function of independent variable (predictor) called the regression function,

   $$E(y) = \beta_0 + \beta_1 x.$$

   It is simple since there is only one predictor, It is linear since $E(y) = (1,\ x)\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$ is a linear function of unknown parameter vector $\beta = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$. The regression function gives a line with intercept β_0 and slope β_1. Clearly

   $$y \sim N(\beta_0 + \beta_1 x,\ \sigma^2).$$

   (2) Simple linear regression without intercept

   $$y = \beta x + \epsilon,\ \epsilon \sim N(0,\ \sigma^2) \Longleftrightarrow y \sim N(\beta x,\ \sigma^2)$$

   is a simple linear regression model without intercept.

   (3) Sample and basic statistics

   A sample from a simple liner regression contains n pairs: $(x_i,\ y_i)$, $i = 1, ..., n$. For the model without intercept, all information are contained in six basic statistics

   $$n,\ \sum x,\ \sum y,\ \sum x^2,\ \sum y^2 \text{ and } \sum xy.$$

   Your calculator should allow you to enter the n pairs and to obtain the six statistics by key-pressing. Information in the model with intercept are contained in another set of six statistics

   $$\begin{aligned}&n,\ \overline{x} = \tfrac{\sum x}{n},\ \overline{y} = \tfrac{\sum y}{n},\ Sxx = \textstyle\sum(x-\overline{x})^2 = \sum x^2 - \frac{1}{n}(\sum x)^2\\ &Syy = \textstyle\sum(y-\overline{y})^2 = \sum y^2 - \frac{1}{n}(\sum y)^2\\ &\text{and } Sxy = \textstyle\sum(x-\overline{x})(y-\overline{y}) = \sum xy - \frac{1}{n}(\sum x)(\sum y).\end{aligned}$$

   Again, you should be able to get these statistics by few key-pressings.

   (4) Unified model expression on sample

   $$y = X\beta + \epsilon \text{ with } \epsilon \sim N(0,\ \sigma^2 I) \Longleftrightarrow y \sim N(X\beta,\ \sigma^2 I).$$

   When $y = \begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$, $X = \begin{pmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}$ and $\beta = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$, the above expression is for the sample from the model with intercept, $y_i \sim N(\beta_0 + \beta_1 x_i,\ \sigma^2)$, $i = 1, ..., n$. But for the sample from the model without intercept, $X = \begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$ and $\beta \in R$.

2. Point estimators for β and σ^2
in $y = X\beta + \epsilon \sim N(X\beta,\, \sigma^2 I)$.

   (1) Least square estimators for β.
   For $y = X\beta + \epsilon \sim N(X\beta,\, \sigma^2 I)$, if $Q(\beta) = \sum_i [y_i - E(y_i)]^2$ is minimized at $\widehat{\beta}$, then $\widehat{\beta}$ is a least square estimator (LSE) for β and $Q(\widehat{\beta})$ is the sum of squared errors (SSE).
   By linear algebra or by calculus methods one obtains

$$\widehat{\beta} = (X'X)^{-1}X'y \text{ and SSE} = y'[I_n - X(X'X)^{-1}X']y.$$

   (2) Maximum likelihood estimators of β and σ^2
   The probability density function (pdf) of $y \sim N(X\beta,\, \sigma^2)$ treated as a function of β and σ^2 is called the likelihood function

$$L(\beta,\, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right].$$

   If $L(\beta,\, \sigma^2) \le L(\widehat{\beta},\, \widehat{\sigma}^2)$ for all β and σ^2, then $\widehat{\beta}$ and $\widehat{\sigma}^2$ are called the maximum likelihood estimators (MLEs) for β and σ^2.
   By linear algebra and calculus methods one obtains

$$\widehat{\beta} = (X'X)^{-1}X'y,\, \widehat{\sigma}^2 = \frac{SSE}{n} \text{ and } AlsoL(\widehat{\beta},\, \widehat{\sigma}^2) = \left(\frac{n}{2\pi e}\right)^{n/2} (\text{SSE})^{-n/2}.$$

3. Formulas

   (1) For the model with intercept
   $\widehat{\beta}_1 = \frac{Sxy}{Sxx}$ and $\widehat{\beta}_0 = \overline{y} - \overline{x}\widehat{\beta}_1$ are LSEs and MLEs for β_1 and β_0.
   SSE $= Syy - \frac{(Sxy)^2}{Sxx}$ and $\frac{SSE}{n}$ is MLE for σ^2.
   So $E(y(x_0)) = \beta_0 + \beta_1 x_0$ is estimated by $\widehat{y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$

   (2) For the model without intercept
   $\widehat{\beta} = \frac{\sum xy}{\sum x^2}$ is LSE and MLE for β.
   $SSE = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2}$ and $\frac{SSE}{n}$ is MLE for σ^2.
   So $E(y(x_0)) = \beta x_0$ is estimated by $\widehat{y}(x_0) = \widehat{\beta} x_0$

Ex1: For simple linear regression with intercept and sample

x:	1	0	2	-1
y:	2	5	3	4

find $\widehat{\beta}_1$, $\widehat{\beta}_0$, SSE and $\widehat{y}(3)$.
Six statistics: $n = 4$, $\overline{x} = 0.5$, $\overline{y} = 3.5$, $Sxx = 5$, $Syy = 5$, $Sxy = -3$.
So $\widehat{\beta}_1 = \frac{Sxy}{Sxx} = \frac{-3}{5} = -0.6$, $\widehat{\beta}_0 = \overline{y} - \overline{x}\widehat{\beta}_1 = 3.5 + 0.5 \times 0.6 = 3.8$
$SSE = Syy - \frac{(Sxy)^2}{Sxx} = 5 - \frac{9}{5} = 3.2$
and $\widehat{y}(3) = \widehat{\beta}_0 + \widehat{\beta}_1(3) = 3.8 - 0.6 \times 3 = 2$.

Ex2: For simple linear regression without intercept and sample in Ex1,
find $\widehat{\beta}$, SSE and $\widehat{y}(3)$.
Six statistics: $n = 4$, $\sum x = 2$, $\sum y = 14$, $\sum x^2 = 6$, $\sum y^2 = 54$, $\sum xy = 4$.
So $\widehat{\beta} = \frac{\sum xy}{\sum x^2} = 0.6667$
$SSE = \sum y^2 - \frac{(\sum xy)^2}{\sum x^2} = 51.3333$
$\widehat{y}(3) = \widehat{\beta} \times 3 = 2$.

L02 Distributions of point estimators

1. Normal distribution of $\widehat{\beta}$

   (1) $\widehat{\beta} \sim N\left(\beta,\, \sigma^2(X'X)^{-1}\right)$
   By Theorem I: $x \sim N(\mu,\, \Sigma) \Longrightarrow Ax + b \sim N(A\mu + b,\, A\Sigma A')$,
   with $y \sim N\left(X\beta,\, \sigma^2 I\right)$ and $\widehat{\beta} = (X'X)^{-1}X'y$, $\widehat{\beta} \sim N\left(\beta,\, \sigma^2(X'X)^{-1}\right)$.
   Thus $\widehat{\beta}$ is an unbiased estimator (UE) for β since $E(\widehat{\beta}) = \beta$.

   (2) For simple linear regression with intercept
   $\widehat{\beta}_1 = (0,\, 1)\widehat{\beta} \sim N\left(\beta_1,\, \sigma^2_{\widehat{\beta}_1}\right)$ where $\sigma^2_{\widehat{\beta}_1} = \sigma^2 \frac{1}{Sxx}$.
   $\widehat{\beta}_0 = (1,\, 0)\widehat{\beta} \sim N\left(\beta_0,\, \sigma^2_{\widehat{\beta}_0}\right)$ where $\sigma^2_{\widehat{\beta}_0} = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{Sxx}\right)$.
   $\widehat{y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_0 = (1\, x_0)\widehat{\beta} \sim N\left(\beta_0 + \beta_1 x_0,\, \sigma^2_{\widehat{y}(x_0)}\right)$ where $\sigma^2_{\widehat{y}(x_0)} = \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{Sxx}\right)$.

   (3) For simple linear regression without intercept
   $\widehat{\beta} \sim N\left(\beta,\, \sigma^2_{\widehat{\beta}}\right)$ where $\sigma^2_{\widehat{\beta}} = \sigma^2 \frac{1}{\sum x^2}$.
   $\widehat{y}(x_0) = \widehat{\beta} x_0 \sim N\left(\beta x_0,\, \sigma^2_{\widehat{y}(x_0)}\right)$ where $\sigma^2_{\widehat{y}(x_0)} = \sigma^2 \frac{x_0^2}{\sum x^2}$.

2. χ^2-distribution associated with SSE

   (1) $\frac{SSE}{\sigma^2} \sim \chi^2(n - c)$ where c is the number of columns of X.
   Proof. By Theorem II:

   $$\text{If } x \sim N(\mu,\, \Sigma) \text{ and } A' = A = A\Sigma A, \text{ then } x'Ax \sim \chi^2(\mu' A\mu,\, r)$$
   $$\text{where } r = \text{rank}(A) = \text{rank}(\Sigma^{1/2}A\Sigma^{1/2}) = \text{tr}(\Sigma^{1/2}A\Sigma^{1/2}) = \text{tr}(A\Sigma),$$

   with $y \sim N(X\beta,\, \sigma^2 I)$, $\frac{SSE}{\sigma^2} = y'Ay$ where $A = \frac{I - X(X'X)^{-1}X'}{\sigma^2}$. One can check $A' = A = A\left(\sigma^2 I\right)A$, $(X\beta)'A(X\beta) = 0$ and $\text{tr}(A\sigma^2 I) = n - c$.
   So $\frac{SSE}{\sigma^2} \sim \chi^2(0,\, n - c) = \chi^2(n - c)$.

   Comment: Let $n - c = \text{rank}[I_n - X(X'X)^{-1}X']$ be DF of SSE, and MSE$= \frac{SSE}{n-c}$. Then MSE is an UE for σ^2 since $E\left(\frac{SSE}{\sigma^2}\right) = E(\chi^2(n - c)) = n0c \Longrightarrow E(MSE) = \sigma^2$.

   (2) For the model with intercept
   there is an SS table

Source	SS	DF	MS
Error	SSE	n-2	MSE

   $\sigma^2_{\widehat{\beta}_0} = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{Sxx}\right)$ has an UE $S^2_{\widehat{\beta}_0} = MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{Sxx}\right)$.
   $\sigma^2_{\widehat{\beta}_1} = \sigma^2 \frac{1}{Sxx}$ has an UE $S^2_{\widehat{\beta}_1} = MSE\, \frac{1}{Sxx}$.
   $\sigma^2_{\widehat{y}(x_0)} = \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{Sxx}\right)$ has an UE $S^2_{\widehat{y}(x_0)} = MSE\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{Sxx}\right)$.
   So there is a parameter table

Parameter	Estimator	Standard Error
β_0	$\widehat{\beta}_0$	$S_{\widehat{\beta}_0}$
β_1	$\widehat{\beta}_1$	$S_{\widehat{\beta}_1}$

(3) For the model without intercept
there is an SS table

Source	SS	DF	MS
Error	SSE	n-1	MSE

$\sigma^2_{\widehat{\beta}} = \sigma^2 \frac{1}{\sum x^2}$ has an UE $S^2_{\widehat{\beta}} = MSE\,\frac{1}{\sum x^2}$.
$\sigma^2_{\widehat{y}(x_0)} = \sigma^2 \frac{x_0^2}{\sum x^2}$ has an UE $S^2_{\widehat{y}(x_0)} = MSE\,\frac{x_0^2}{\sum x^2}$.
So there is a parameter table

Parameter	Estimator	Standard Error
β	$\widehat{\beta}$	$S_{\widehat{\beta}}$

3. Independence of $\widehat{\beta}$ and SSE

(1) $\widehat{\beta}$ and SSE are independent

Proof. By Theorem III: $x \sim N(\mu, \Sigma)$. then

$$\begin{array}{lcl} A\Sigma B' = 0 & \Longrightarrow & Ax \text{ and } Bx \text{ are independent} \\ B' = B \text{ and } A\Sigma B = 0 & \Longrightarrow & Ax \text{ and } x'Bx \text{ are independent} \\ A' = A,\ B' = B \text{ and } A\Sigma B = 0 & \Longrightarrow & x'Ax \text{ and } x'Bx \text{ are independent,} \end{array}$$

with $y \sim N(X\beta, \sigma^2 I)$, $\widehat{\beta} = Ay$ where $A = (X'X)^{-1}X'$ and $SSE = y'By$ where $B = I_n - X(X'X)^{-1}X'$. But $A\sigma^2 IB = 0$. So $\widehat{\beta}$ and SSE are independent.

(2) t-distributions in the models
For the model with intercept,
$\frac{\widehat{\beta}_0-\beta_0}{s_{\widehat{\beta}_0}} \sim t(n-2)$, $\quad \frac{\widehat{\beta}_1-\beta_1}{s_{\widehat{\beta}_1}} \sim t(n-2)$, $\quad \frac{\widehat{y}(x_0)-E(y(x_0))}{s_{\widehat{y}(x_0)}} \sim t(n-2)$.
For the model with out intercept,
$\frac{\widehat{\beta}-\beta}{s_{\widehat{\beta}}} \sim t(n-1)$ $\quad \frac{\widehat{y}(x_0)-E(y(x_0))}{S_{\widehat{y}(x_0)}} \sim t(n-1)$.

Proof. Show the first one. By the normal distribution of $\widehat{\beta}_0$, $\frac{\widehat{\beta}_0-\beta_0}{\sigma_{\widehat{\beta}_0}} \sim N(0, 1^2)$ which is independent to $\frac{SSE}{\sigma^2} \sim \chi 2(n-2)$.
By definition of t-distribution, $\frac{\widehat{\beta}_0-\beta_0}{\sigma_{\widehat{\beta}_0}} \div \sqrt{\frac{SSE}{\sigma^2(n-2)}} \sim t(n-2)$, i.e., $\frac{\widehat{\beta}_0-\beta_0}{S_{\widehat{\beta}_0}} \sim t(n-2)$.

Ex: A sample for a simple linear regression with intercept produced

$$n = 4,\ \overline{x} = 0.5,\ \overline{y} = 3.5,\ Sxx = 5,\ Syy = 5,\ Sxy = -3.$$

Fill out SSE table and Parameter table.

Source	SS	DF	MS
Error	$Syy - \frac{(Sxy)^2}{Sxx} = 3.2$	$n-2=2$	1.6

Parameter	Estimator	Standard Error
β_0	$\overline{y} - \overline{x}\widehat{\beta}_1 = 3.8$	$\sqrt{MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{Sxx}\right)} = 0.6928$
β_1	$\frac{Sxy}{Sxx} = -0.6$	$\sqrt{\frac{MSE}{Sxx}} = 0.5657$