Stats 100C – Linear Models

University of California, Los Angeles

Duc Vu

Fall 2021

This is stats 100C – Linear Models taught by Professor Christou. There is not an official textbook used for the course. Instead, handouts and reference materials are distributed and can be accessed through the class website. You can find other math/stats lecture notes through my personal blog. Let me know through my email if you notice something mathematically wrong/concerning. Thank you!

Contents

List of Theorems

List of Definitions

§1 Lec 1: Sep 27, 2021

§1.1 Simple Linear Regression Models

Consider

$$Y_i = \mu + \varepsilon_i$$

with $\varepsilon_i \overset{\text{i.i.d}}{\sim} N(0,\sigma)$; specifically, $Y_1,\ldots,Y_n \overset{\text{i.i.d}}{\sim} N(\mu,\sigma)$. We want to estimate μ and σ^2 using least squares or method of maximum likelihood (MML).
Method of Least Squares (OLS – Ordinary Least Squares):

$$\begin{aligned}
\min Q &= \sum_{i=1}^{n}(Y_i-\mu)^2 \\
\frac{\partial Q}{\partial \mu} &= -2\sum (Y_i-\mu) = 0 \\
\sum Y_i - n\hat{\mu} &= 0 \\
\implies \hat{\mu} &= \overline{Y}
\end{aligned}$$

Method of Maximum Likelihood (MML):

$$\begin{aligned}
f(y_i) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
&= \left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \\
L = f(y_1)\ldots f(y_n) &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2} \\
\ln L &= -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum (y_i-\mu)^2 \\
\frac{\partial \ln L}{\partial \mu} &= 0, \qquad \frac{\partial \ln L}{\partial \sigma^2} = 0
\end{aligned}$$

Solve the above, we obtain the MLE of μ and σ^2

$$\hat{\mu} = \hat{y}, \qquad \hat{\sigma}^2 = \frac{\sum(y_i-\hat{\mu})^2}{n} = \frac{\sum(y_i-\overline{y})^2}{n}$$

Notice that $\hat{\sigma}^2$ is biased and we adjust it to be unbiased as follows

$$S^2 = \frac{\sum(y_i-\overline{y})^2}{n-1}$$

§1.2 Prediction Problem

Given $Y_1,\ldots,Y_n$, we want to predict a new Y, e.g., Y_0. An educated guess here is

$$\hat{Y}_0 = \overline{Y}$$

1. Predictor assumption: $\hat{Y}_0 = \sum_{i=1}^{n} a_i Y_i$
2. We want $\hat{Y}_0$ to be unbiased, i.e., $E\hat{Y}_0 = \mu$

$$\begin{aligned}
E\sum a_i Y_i &= \mu \\
\sum a_i E Y_i &= \mu \\
\implies \sum a_i &= 1
\end{aligned}$$

3. Minimize the mean square error of prediction, i.e.,

$$E\left(Y_0 - \hat{Y}_0\right)^2 \qquad \text{s.t.} \qquad \sum a_i = 1$$

Notice that this is a constraint optimization problem, we use the method of Lagrange multiplier to obtain

$$\min Q = E\left(Y_0 - \hat{Y}_0\right)^2 - 2\lambda\left(\sum a_i - 1\right)$$

<u>Note</u>: $EW^2 = \text{var}(W) + (EW)^2$

$$\begin{aligned}
\min Q &= \text{var}\left(Y_0 - \hat{Y}_0\right) - 2\lambda\left[\sum a_i - 1\right] \\
&= \text{var}(Y_0) + \text{var}(\hat{Y}_0) - 2\operatorname{cov}\left(Y_0, \hat{Y}_0\right) - 2\lambda\left[\sum a_i - 1\right] \\
&= \sigma^2 + \sigma^2\sum a_i^2 - 2\lambda\left[\sum a_i - 1\right] \\
\frac{\partial Q}{\partial a_i} &= 2\sigma^2 a_i - 2\lambda = 0 \\
a_i &= \frac{\lambda}{\sigma^2}
\end{aligned}$$

Notice that $a_1 = a_2 = \ldots = a_n = \frac{\lambda}{\sigma^2}$. So

$$\sum a_i = \frac{n\lambda}{\sigma^2} = 1 \implies \lambda = \frac{\sigma^2}{n}$$

Thus, we can see that

$$a_i = \frac{1}{n}$$

and therefore since $\hat{Y}_0 = \sum a_i Y_i$, it follows that $\hat{Y}_0 = \overline{Y}$.

<u>Prediction Interval</u>:

$$Y_0 - \hat{Y}_0 \sim N\left(0, \sigma\sqrt{1 + \frac{1}{n}}\right)$$

Recall from 100B

$$\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

So,

$$\frac{\frac{Y_0 - \hat{Y}_0 - 0}{\sigma\sqrt{1+\frac{1}{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{Y_0 - \hat{Y}_0}{S\sqrt{1 + \frac{1}{n}}} \sim t_{n-1}$$

We can now construct the prediction interval for Y_0 as follows

$$P\left(-t_{\frac{\alpha}{2};n-1} \le \frac{Y_0 - \hat{Y}_0}{S\sqrt{1+\frac{1}{n}}} \le t_{\frac{\alpha}{2};n-1}\right) = 1 - \alpha$$

Finally, $Y_0 \in \hat{Y}_0 \pm t_{\frac{\alpha}{2};n-1} S\sqrt{1 + \frac{1}{n}}$.

> **Remark 1.1.** Compare this to the confidence interval for μ : $\mu \in \overline{Y} \pm t_{\frac{\alpha}{2};n-1}\frac{S}{\sqrt{n}}$.

§2 Lec 2: Sep 29, 2021

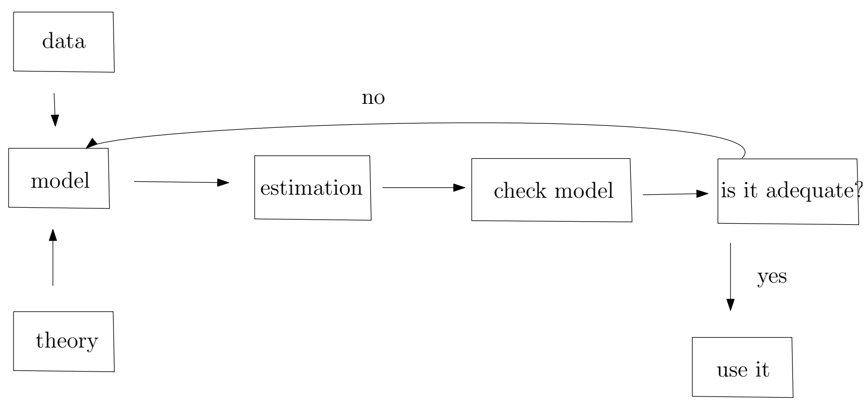
§2.1 Linear Regression

Consider a simple regression model

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\ \text{or } Y_i &= \beta_1 X_i + \varepsilon_i \end{aligned}$$

Data:

y	x
y_1	x_1
$\vdots$	$\vdots$
y_n	x_n



where the parameters are

$$\begin{cases} \beta_0: & \text{intercept} \\ \beta_1: & \text{slope} \end{cases}$$

and $X_1, \ldots, X_n$ are predictors that are not random; $\varepsilon_1, \ldots, \varepsilon_n$ are random error terms/disturbance/stochastic terms, and $Y_1, \ldots, Y_n$ are random response variable.
Assumption (Gauss-Markov Conditions):

$$E(\varepsilon_i) = 0, \quad \operatorname{var}(\varepsilon_i) = \sigma^2$$

$\varepsilon_1, \ldots, \varepsilon_n$ are independent. Using the Gauss-Markov conditions,

$$\begin{aligned} EY_i &= \beta_0 + \beta_1 X_i \\ \operatorname{var}(Y_i) &= \sigma^2 \\ \min Q &= \sum \varepsilon_i^2 \\ \min Q &= \sum \left(Y_i - \beta_0 - \beta_1 X_i\right)^2 \\ \frac{\partial Q}{\partial \beta_0} &= -2 \sum \left(Y_i - \beta_0 - \beta_1 X_i\right) = 0 \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum \left(Y_i - \beta_0 - \beta_1 X_i\right) X_i = 0 \end{aligned}$$

So,

$$\begin{cases} \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \\ \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0 \end{cases}$$

$$\implies \begin{cases} n\beta_0 + \beta_1 \sum x_i = \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 = \sum x_i y_i \end{cases} \quad \text{– normal equations}$$

We can solve the above to get $\hat{\beta}_0, \hat{\beta}_1$.

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Determinant of the matrix:

$$\begin{aligned} n \sum x_i^2 - \left(\sum x_i \right)^2 &= n \left[\sum x_i^2 - \frac{\left(\sum x_i \right)^2}{n} \right] \\ &= n \sum (x_i - \overline{x})^2 \geq 0 \end{aligned}$$

If $x_1 = x_2 = \ldots = x_n = \overline{x}$ then $\sum (x_i - \overline{x})^2 = 0$. From normal equations we get

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \quad \text{from (1)}$$

and plug (1) into (2) to obtain

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{1}{n} (\sum x_i)(\sum y_i)}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

or

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

or

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} \tag{*}$$

or

$$\hat{\beta}_1 = \frac{\sum (y_i - \overline{y}) x_i}{\sum (x_i - \overline{x})^2}$$

or

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

Note: From (*), we have

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} \\ &= \frac{(x_1 - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} + \ldots + \frac{(x_n - \overline{x}) y_n}{\sum (x_i - \overline{x})^2} \\ &= k_1 y_1 + \ldots + k_n y_n = \sum_{i=1}^n k_i y_i \end{aligned}$$

where $k_i = \frac{x_i - \overline{x}}{\sum(x_i - \overline{x})^2}$. Notice that

$$\begin{aligned}\sum k_i &= 0\\ \sum k_i^2 &= \frac{1}{\sum(x_i-\overline{x})^2}\\ \sum k_i x_i &= \frac{\sum(x_i-\overline{x})x_i}{\sum(x_i-\overline{x})^2} = 1\end{aligned}$$

<u>Properties of $\hat{\beta}_1$:</u>

$$\begin{aligned}E\hat{\beta}_1 &= E\sum k_i y_i = \sum k_i E y_i\\ &= \sum k_i\left(\beta_0 + \beta_1 x_i\right)\\ &= \beta_0 \sum k_i + \beta_1 \sum k_i x_i\\ &= \beta_1 \text{ – unbiased}\end{aligned}$$

For the variance,

$$\begin{aligned}\operatorname{var}(\hat{\beta}_1) &= \operatorname{var}\left(\sum k_i y_i\right)\\ &= \sum k_i^2 \operatorname{var}(Y_i)\\ &= \frac{\sigma^2}{\sum(x_i-\overline{x})^2}\end{aligned}$$

<u>Properties of $\hat{\beta}_0$:</u>

$$\begin{aligned}\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1\overline{x} = \frac{\sum y_i}{n}\\ &= \sum \frac{y_i}{n} - \overline{x}\sum k_i y_i\\ &= \sum\left(\frac{1}{n} - \overline{x}k_i\right) y_i\\ &= \sum_{i=1}^{n} l_i y_i\end{aligned}$$

where $l_i = \frac{1}{n} - \overline{x}k_i$ and the properties of l_i are

$$\begin{aligned}\sum l_i &= 1\\ \sum l_i^2 &= \sum\left(\frac{1}{n} - \overline{x}k_i\right)^2 = \sum\left(\frac{1}{n^2} + \overline{x}^2 k_i^2 - \frac{2}{n}\overline{x}k_i\right)\\ &= \frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i-\overline{x})^2}\\ \sum l_i x_i &= 0\end{aligned}$$

Now, we can easily show that $\hat{\beta}_0$ is unbiased

$$\begin{aligned}E\hat{\beta}_0 &= E\sum l_i y_i = \sum l_i E y_i\\ &= \sum l_i\left(\beta_0 + \beta_1 x_i\right) = \beta_0\sum l_i + \beta_1 \sum l_i x_i\\ &= \beta_0\end{aligned}$$

Thus,

$$\operatorname{var}\left(\hat{\beta}_0\right) = \operatorname{var}\left(\sum l_i y_i\right) = \sigma^2 \sum l_i^2 = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum (x_i - \overline{x})^2}\right)$$

The fitted value is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \overline{y} + \hat{\beta}_1 \left(x_i - \overline{x}\right)$$

and the residual is defined as

$$e_i = y_i - \hat{y}_i$$

with properties

$$\begin{aligned} \sum e_i &= 0 \\ \sum e_i x_i &= 0 \\ \sum e_i \hat{y}_i &= 0 \end{aligned}$$

<u>Estimation Using MML:</u>

Assume $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. Then $Y_i \sim N\left(\beta_0 + \beta_1 X_i, \sigma\right)$. The log-likelihood function is

$$\ln L = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

So, we need to solve

$$\frac{\partial \ln L}{\partial \beta_0} = 0, \quad \frac{\partial \ln L}{\partial \beta_1} = 0$$

to get $\hat{\beta}_0, \hat{\beta}_1$ which are the same as least squares method.

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - \beta_0 - \beta_1 x_i)^2 = 0 \\ \hat{\sigma}^2 &= \frac{\sum e_i^2}{n} \end{aligned}$$

Then,

$$\sum (y_i - \overline{y})^2 = \sum \left(\underbrace{y_i - \hat{y}_i}_{e_i} + \hat{y}_i + \overline{y}\right)^2$$

in which we expand to get

$$\underbrace{\sum (y_i - \overline{y})^2}_{\text{SST}} = \underbrace{\sum e_i^2}_{\text{SSE}} + \underbrace{\sum (\hat{y}_i - \overline{y})^2}_{\text{SSR}}$$

in which

$$\begin{cases} \text{SST: sum of squares total} \\ \text{SSE: sum of squares error} \\ \text{SSR: sum of squares regression} \end{cases}$$

§3 Lec 3: Oct 1, 2021

§3.1 Gauss-Markov Theorem

Recall

$$\hat{\beta}_1 = \sum k_i Y_i$$

where $k_i = \frac{x_i - \overline{x}}{\sum(x_i - \overline{x})^2}$. Consider now

$$b_1 = \sum a_i Y_i$$

which is another unbiased estimator of β_1. Then $Eb_1 = \beta_1$ or $E \sum a_i Y_i = \beta_1$. So

$$\begin{aligned} \beta_1 &= \sum a_i EY_i \\ &= \sum a_i \left(\beta_0 + \beta_1 X_i\right) \\ &= \beta_0 \sum a_i + \beta_1 \sum a_i X_i \end{aligned}$$

Thus,

$$\begin{cases} \sum a_i = 0 \\ \sum a_i x_i = 1 \end{cases}$$

and we know that

$$\operatorname{var}(b_1) = \operatorname{var}\left(\sum_{i=1}^n a_i Y_i\right) = \sigma^2 \sum a_i^2$$

and

$$\operatorname{var}(\hat{\beta}_1) = \sigma^2 \sum k_i^2 = \frac{\sigma^2}{\sum(x_i - \overline{x})^2}$$

Now let $a_i = k_i + d_i$. Then,

$$\begin{aligned} \operatorname{var}(b_1) &= \sigma^2 \sum (k_i + d_i)^2 \\ &= \sigma^2 \sum k_i^2 + \sigma^2 \sum d_i^2 + 2\sigma^2 \sum k_i d_i \end{aligned}$$

We need to show $\sum k_i d_i = 0$.

$$\begin{aligned} \sum k_i(a_i - k_i) &= \sum k_i a_i - \sum k_i^2 \\ &= \frac{\sum(x_i - \overline{x})a_i}{\sum(x_i - \overline{x})^2} - \frac{1}{\sum(x_i - \overline{x})^2} \\ &= \frac{\sum x_i a_i}{\sum(x_i - \overline{x})^2} - \frac{\overline{x}\sum a_i}{\sum(x_i - \overline{x})^2} - \frac{1}{\sum(x_i - \overline{x})^2} \\ &= 0 \end{aligned}$$

So $\operatorname{var}(b_1) \geq \operatorname{var}(\hat{\beta}_1)$ and therefore $\hat{\beta}_1$ is the best linear unbiased estimator (BLUE).

§3.2 Estimation of σ^2

Using MML

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n}$$

Is it unbiased?

$$E\hat{\sigma}^2 = \frac{\sum Ee_i^2}{n} = \frac{\sum \left[\operatorname{var}(e_i) + (Ee_i)^2\right]}{n}$$

<u>*Note*</u>: $e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$. So

$$Ee_i = E\left[Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right] = (\beta_0 + \beta_1 X_i) - (\beta_0 + \beta_1 X_i) = 0$$

Then,

$$E\hat{\sigma}^2 = \frac{\sum \text{var}(e_i)}{n}$$

Notice that

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

or

$$e_i = Y_i - \overline{Y} - \hat{\beta}_1 (X_i - \overline{X})$$

where $\hat{Y}_i = \overline{Y} + \hat{\beta}_1 (X_i - \overline{X})$. Substitute in and we get

$$\begin{aligned}
\text{var}(e_i) &= \text{var}\left[Y_i - \overline{Y} - \hat{\beta}_1 (X_i - \overline{X})\right] \\
&= \text{var}(Y_i) + \text{var}(\overline{Y}) + (X_i - \overline{X})^2 \text{var}(\hat{\beta}_0) - 2\,\text{cov}(Y_i, \overline{Y}) - 2(X_i - \overline{X})\,\text{cov}(Y_i, \hat{\beta}_1) \\
&\quad + 2(X_i - \overline{X})\,\text{cov}(\overline{Y}, \hat{\beta}_1)
\end{aligned}$$

Let's compute each term there.

$$\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\
\text{var}(Y_i) &= \sigma^2 \\
\overline{Y} &= \beta_0 + \beta_1 \overline{X} + \frac{\sum \varepsilon_i}{n} \\
\text{var}(\overline{Y}) &= \frac{\sigma^2}{n} \\
\text{cov}(Y_i, \overline{Y}) &= \text{cov}\left(Y_i, \frac{Y_1 + \ldots + Y_i + \ldots + Y_n}{n}\right) \\
&= \frac{1}{n}\text{cov}(Y_i, Y_1) + \ldots + \frac{1}{n}\text{cov}(Y_i, Y_i) + \ldots + \frac{1}{n}\text{cov}(Y_i, Y_n) \\
&= \frac{\sigma^2}{n} \\
\text{cov}(Y_i, \hat{\beta}_1) &= \text{cov}(Y_i, \sum k_i Y_i) \\
&= \text{cov}(Y_i, k_1 Y_1) + \ldots + \text{cov}(Y_i, k_i Y_i) + \ldots + \text{cov}(Y_i, k_n Y_n) \\
&= k_1\,\text{cov}(Y_i, Y_1) + \ldots + k_i\,\text{cov}(Y_i, Y_i) + \ldots + k_n\,\text{cov}(Y_1, Y_n) \\
&= \sigma^2 k_i = \sigma^2 \frac{x_i - \overline{x}}{\sum (x_i - \overline{x})^2}
\end{aligned}$$

<u>*Note*</u>: A property of covariance

$$\text{cov}(aY, bQ) = ab\,\text{cov}(Y, Q)$$

And for the last term,

$$\begin{aligned}
\text{cov}(\overline{Y}, \hat{\beta}_1) &= \text{cov}\left(\frac{Y_1 + \ldots + Y_n}{n}, k_1 Y_1 + \ldots + k_n Y_n\right) \\
&= \text{cov}(\frac{Y_1}{n}, k_1 Y_1 + \ldots + k_n Y_n) + \ldots + \text{cov}(\frac{Y_n}{n}, k_1 Y_1 + \ldots + k_n Y_n) \\
&= \frac{\sigma^2}{n} k_1 + \frac{\sigma^2}{n} k_2 + \ldots + \frac{\sigma^2}{n} k_n \\
&= \frac{\sigma^2}{n} \sum k_i = 0
\end{aligned}$$

Now, we're ready to compute the variance

$$\begin{aligned}\operatorname{var}(e_i) &= \sigma^2 + \frac{\sigma^2}{n} + \frac{\sigma^2(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2} - \frac{2\sigma^2}{n} - \frac{2\sigma^2(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\\ &= \sigma^2\left(1-\frac{1}{n}-\frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)\end{aligned}$$

Therefore,

$$\begin{aligned}E\hat{\sigma}^2 &= \frac{\sum \operatorname{var}(e_i)}{n} = \sigma^2\frac{\sum_{i=1}^n\left(1-\frac{1}{n}-\frac{(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)}{n}\\ &= \frac{(n-2)}{n}\sigma^2\end{aligned}$$

It follows that the unbiased estimator of σ^2 is

$$S_e^2 = \frac{n}{n-2}\sigma^2 = \frac{\sum e_i^2}{n-2}$$

§3.3 Distribution Theory

Let $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and we assume $\varepsilon_1,\dots,\varepsilon_n \overset{\text{i.i.d}}{\sim} N(0,\sigma)$

$$\begin{aligned}\hat{\beta}_1 &= \sum k_i Y_i \implies \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum(x_i-\overline{x})^2}}\right)\\ \hat{\beta}_0 &= \sum l_i Y_i \implies \hat{\beta}_0 \sim N\left(\beta_0, \sigma\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{\sum(x_i-\overline{x})^2}}\right)\end{aligned}$$

We will show $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$ in the next lecture.