

STA 108: Applied Statistical Methods: Regression Analysis

Course Material Summary

University of California at Davis

Last Edit Date: 02/03/2023

Disclaimer and Term of Use:

1. We do not guarantee the accuracy and completeness of the summary content. Some of the course material may not be included, and some of the content in the summary may not be correct. You should use this file properly and legally. We are not responsible for any results from using this file.
2. Although most of the content in this summary is originally written by the creator, there may be still some of the content that is adapted (derived) from the slides and codes from *Professor Xueheng Shi*. We use those as references and quotes in this file. Please [contact us](#) to delete this file if you think your rights have been violated.
3. This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

STA 108 Math-related Summary

	Simple Linear Regression	Matrix Approach of Linear Regression
Model	$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, 3, \dots, n$ where we have n pairs of observations $(X_1, Y_1) \dots (X_n, Y_n)$	$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}$, $\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}_{n \times (1+p)}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(1+p) \times 1}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$ when we have n observations and p inputs in each observation.
Prerequisite	$\varepsilon_i \sim N(0, \sigma^2)$ independent and identically distributed (iid) 1) $\varepsilon_1, \dots, \varepsilon_n$ are independent 2) $E(\varepsilon_i) = 0, \quad i = 1, 2, 3, \dots, n$ 3) $Var(\varepsilon_i) = \sigma^2, \quad i = 1, 2, 3, \dots, n$	$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$ 1) $E(\boldsymbol{\varepsilon}) = 0$ 2) $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ $*3) \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$ The variance-covariance matrix tells the ε_i are independent and have constant variance σ^2. All covariance terms are zero, only diagonal is σ^2, which means they are uncorrelated. As a result, under normality, this indicating independence.
Distribution	$\varepsilon_i \sim N(0, \sigma^2), Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$	$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}), \mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$
Expectation and variance of Y	$E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i)$ $= E(\beta_0) + E(\beta_1 X_i) + E(\varepsilon_i)$ $= \beta_0 + \beta_1 X_i + 0$ $= \boxed{\beta_0 + \beta_1 X_i}$ $Var(Y_i) = Var(\beta_0 + \beta_1 X_i + \varepsilon_i)$ $= Var(\beta_0) + Var(\beta_1 X_i) + Var(\varepsilon_i)$ $= 0 + 0 + \sigma^2$ $= \boxed{\sigma^2}$	$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$ $= E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\varepsilon})$ $= \mathbf{X}\boldsymbol{\beta} + 0$ $= \boxed{\mathbf{X}\boldsymbol{\beta}}$ If the means of $Y_1, Y_2, \dots, Y_n$ are $\mu_1, \mu_2, \dots, \mu_n$ then the mean of the random vector $\mathbf{Y}$ is: $E(\mathbf{Y}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$ $Var(\mathbf{Y}) = Var(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$ $= Var(\mathbf{X}\boldsymbol{\beta}) + Var(\boldsymbol{\varepsilon})$ $= 0 + \sigma^2 \mathbf{I}$ $= \boxed{\sigma^2 \mathbf{I}}$ The variance-covariance matrix of the random vector $\mathbf{Y}$ is:

		$Var(\mathbf{Y}) = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{bmatrix}$
S_{XX}, S_{YY}, S_{XY}	$S_{XX} = \sum (X_i - \bar{X})^2$ $S_{YY} = \sum (Y_i - \bar{Y})^2$ $S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$	$S_{XX} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2 = \mathbf{X}^T \mathbf{X} - n\bar{X}^2$ $S_{YY} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = \mathbf{Y}^T \mathbf{Y} - n\bar{Y}^2$ $*S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$
Ordinary Least Square (OLS)	$Q(\beta_0, \beta_1) = \sum_{1 \leq i \leq n} (Y_i - \beta_0 - \beta_1 X_i)^2$ $\min_{\beta_0, \beta_1} \sum_{1 \leq i \leq n} (Y_i - \beta_0 - \beta_1 X_i)^2$ $\frac{\partial Q}{\partial \beta_0} = \sum -2(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum (Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum Y_i - \sum \beta_0 - \beta_1 \sum X_i = 0$ $n\bar{Y} - n\beta_0 - n\beta_1 \bar{X} = 0$ $n\bar{Y} - n\beta_1 \bar{X} = n\beta_0$ $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ $\frac{\partial Q}{\partial \beta_1} = \sum -2X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum X_i^2 = 0$ $\sum X_i Y_i - \bar{Y} \sum X_i + \beta_1 \bar{X} \sum X_i - \beta_1 \sum X_i^2 = 0$ $\hat{\beta}_1 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$ $\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$	$Q(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ $= \min_{\boldsymbol{\beta}} (\mathbf{Y}^T \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta})$ $= \min_{\boldsymbol{\beta}} (\mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta})$ $\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = 0$ $\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y}$ $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ $*\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_i \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \end{bmatrix}$ $= \begin{bmatrix} 1+1+\cdots+1 & X_1+X_2+\cdots+X_i \\ X_1+X_2+\cdots+X_i & X_1^2+X_2^2+\cdots+X_i^2 \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}_{2 \times 2}$ $**(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{n \sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} = \frac{1}{n(\sum X_i^2 - \frac{n^2 \bar{X}^2}{n})} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix}$ $= \frac{1}{n(\sum X_i^2 - n\bar{X}^2)} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} = \frac{1}{n \sum (X_i - \bar{X})^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} & \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} \\ \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} & \frac{n}{n \sum (X_i - \bar{X})^2} \end{bmatrix}$

		$= \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix}_{2 \times 2}$ $*** \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1 + Y_2 + \cdots + Y_n \\ X_1 Y_1 + X_2 Y_2 + \cdots + X_n Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}_{n \times 1}$ $**** \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} \sum Y_i \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right) + \sum X_i Y_i \left(\frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \right) \\ \sum Y_i \left(\frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \right) + \sum X_i Y_i \left(\frac{1}{\sum (X_i - \bar{X})^2} \right) \end{bmatrix}$ $\sum Y_i \left(\frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \right) + \sum X_i Y_i \left(\frac{1}{\sum (X_i - \bar{X})^2} \right) = \frac{-n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} + \frac{\sum X_i Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{s_{XY}}{s_{XX}} = \hat{\beta}_1$ $\sum Y_i \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right) + \sum X_i Y_i \left(\frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \right) = \frac{n\bar{Y}}{n} + \bar{X} \frac{n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} - \bar{X} \frac{\sum X_i Y_i}{\sum (X_i - \bar{X})^2} = \bar{Y} + \bar{X} \left(\frac{n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} - \frac{\sum X_i Y_i}{\sum (X_i - \bar{X})^2} \right)$ $= \bar{Y} + \bar{X} \cdot (-1) \cdot \left(\frac{\sum X_i Y_i}{\sum (X_i - \bar{X})^2} - \frac{n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} \right) = \bar{Y} - \bar{X} \left(\frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum (X_i - \bar{X})^2} \right) = \bar{Y} - \bar{X} \hat{\beta}_1 = \hat{\beta}_0$ $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} \bar{Y} - \bar{X} \hat{\beta}_1 \\ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}_{2 \times 1}$ *****Variance-covariance of $\hat{\boldsymbol{\beta}}$ We know that $Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2$, and $(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix}$. As a result, variance-covariance of $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} & \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{bmatrix}$
Maximum Likelihood Estimation (MLE)	Probability density function (pdf) in normal distribution $N(0, \sigma^2)$: $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$, where μ is the	Probability density function (pdf) for $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in matrix form:

mean for Y , σ^2 is the variance of Y . $-\infty < Y < \infty$.

Likelihood function:

$$L(\beta_0, \beta_1, \sigma^2 | (Y_i, X_i)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right],$$

where $\varepsilon_i \sim N(0, \sigma^2)$ independent and identically distributed.

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2 | (Y_i, X_i)) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_i^2}{2\sigma^2}\right] \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right] \end{aligned}$$

We use the log-likelihood function, which is easier:

$$\begin{aligned} l(\beta_0, \beta_1, \sigma^2 | (Y_i, X_i)) &= \log\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right]\right) \\ &= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2\right]\right) \\ &= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n\right) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \end{aligned}$$

$$\frac{\partial l}{\partial \beta_0} = 0 - 0 - \frac{1}{2\sigma^2} \sum -2(Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\sum (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\sum Y_i - \sum \beta_0 - \beta_1 \sum X_i = 0$$

$$n\bar{Y} - n\beta_0 - n\beta_1 \bar{X} = 0$$

$$n\bar{Y} - n\beta_1 \bar{X} = n\beta_0$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$f_Y(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\Sigma))^{-\frac{1}{2}} \exp\left[-\frac{(Y-\mu)^T \Sigma^{-1} (Y-\mu)}{2}\right], \text{ where } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}. Y \in \mathbb{R}^n$$

$\mathbb{R}^n$

We know that $\mu = X\beta, \Sigma = \sigma^2 I$:

$$\begin{aligned} f_Y(y) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\Sigma))^{-\frac{1}{2}} \exp\left[-\frac{(Y-\mu)^T \Sigma^{-1} (Y-\mu)}{2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 I))^{-\frac{1}{2}} \exp\left[-\frac{(Y-X\beta)^T (\sigma^2 I)^{-1} (Y-X\beta)}{2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 I))^{-\frac{1}{2}} \exp\left[-\frac{(Y-X\beta)^T (Y-X\beta)}{2\sigma^2}\right] \end{aligned}$$

$$\text{We can see that } \det(\sigma^2 I) = \det \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = (\sigma^2)^n, (\sigma^2 I)^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{\sigma^2} \end{bmatrix} = \frac{1}{\sigma^2} I.$$

Likelihood function:

$$L(\beta, \sigma^2 | (Y, X)) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(Y-X\beta)^T (Y-X\beta)}{2\sigma^2}\right], \text{ where the model error is } \varepsilon \sim N(0, \sigma^2 I).$$

We use the log-likelihood function, which is easier:

$$\begin{aligned} l(\beta, \sigma^2 | (Y, X)) &= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 I))^{-\frac{1}{2}} \exp\left[-\frac{(Y-X\beta)^T (Y-X\beta)}{2\sigma^2}\right]\right) \\ &= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n\right) - \frac{n}{2} \log(\sigma^2) - \frac{(Y-X\beta)^T (Y-X\beta)}{2\sigma^2} \end{aligned}$$

$$\frac{\partial l}{\partial \beta} = 0 - 0 - \frac{1}{2\sigma^2} (X^T X)^{-1} X^T Y = 0$$

$$X^T X \beta = X^T Y$$

$$(X^T X)^{-1} X^T X \beta = (X^T X)^{-1} X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

	$\frac{\partial l}{\partial \beta_1} = 0 - 0 - \frac{1}{2\sigma^2} \sum -2X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum X_i^2 = 0$ $\sum X_i Y_i - (\bar{Y} - \beta_1 \bar{X}) \sum X_i - \beta_1 \sum X_i^2 = 0$ $\sum X_i Y_i - \bar{Y} \sum X_i + \beta_1 \bar{X} \sum X_i - \beta_1 \sum X_i^2 = 0$ $\hat{\beta}_1 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = \frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$ $\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$ $\frac{\partial l}{\partial \sigma^2} = 0 - \frac{\partial}{\partial \sigma^2} \left[\frac{n}{2} \log(\sigma^2) \right] - \frac{\partial}{\partial \sigma^2} \left[\frac{1}{2\sigma^2} \sum -2X_i(Y_i - \beta_0 - \beta_1 X_i)^2 \right]$ $= -\frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{(-1)}{2(\sigma^2)^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$ $= -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 = 0$ $\hat{\sigma}^2 = \frac{1}{n} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$	$\frac{\partial l}{\partial \sigma^2} = 0 - \frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \frac{(-1)}{(\sigma^2)^2} \cdot (Y - X\beta)^T (Y - X\beta) = 0$ $-\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot (Y - X\beta)^T (Y - X\beta) = 0$ $-\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \cdot (Y - X\beta)^T (Y - X\beta) = 0$ $\frac{n}{\sigma^2} = \frac{1}{(\sigma^2)^2} \cdot (Y - X\beta)^T (Y - X\beta)$ $\sigma^2 = \frac{1}{n} (Y - X\beta)^T (Y - X\beta)$ $\hat{\sigma}^2 = \frac{1}{n} (Y - X\hat{\beta})^T (Y - X\hat{\beta})$
Interpret $\hat{\beta}$	$\hat{\beta}_1$: For one-unit increase in X, the Y will increase/decrease by $ \hat{\beta}_1 $ on average . $\hat{\beta}_0$: When X equals to zero, Y equals to $\hat{\beta}_0$.	
Expectation of β (show unbiased)	$E(\hat{\beta}_1) = E\left(\frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2}\right)$ $= \frac{\sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2} E(Y_i)$ $= \frac{\sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2} (\beta_0 + \beta_1 X_i)$ $= \frac{\sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2} \beta_0 + \beta_1 \frac{\sum (X_i - \bar{X}) X_i}{\sum (X_i - \bar{X})^2}$ $= 0 + \beta_1$	$E(\hat{\beta}) = E((X^T X)^{-1} X^T Y)$ $= (X^T X)^{-1} X^T E(Y)$ $= (X^T X)^{-1} X^T X \beta$ $= I \beta$ $= \boxed{\beta}$ ## If $W = AY$, then $E(W) = AE(Y)$

	$= \boxed{\beta_1}$ $E(\hat{\beta}_0) = E(\bar{Y} - \hat{\beta}_1 \bar{X})$ $= E(\bar{Y}) - E(\hat{\beta}_1 \bar{X})$ $= E\left(\frac{1}{n} \sum Y_i\right) - \bar{X} E(\hat{\beta}_1)$ $= \frac{1}{n} \sum E(Y_i) - \beta_1 \bar{X}$ $= \frac{1}{n} \sum E(\beta_0 + \beta_1 X_i) - \beta_1 \bar{X}$ $= \frac{1}{n} (\sum \beta_0 + \beta_1 \sum X_i) - \beta_1 \bar{X}$ $= \frac{1}{n} (n\beta_0 + n\beta_1 \bar{X}) - \beta_1 \bar{X}$ $= \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X}$ $= \boxed{\beta_0}$	
Variance of β	Let $c = \sum (X_i - \bar{X})^2$, $c_i = X_i - \bar{X}$ $Var(\hat{\beta}_1) = Var\left(\frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2}\right)$ $= \frac{1}{c^2} Var(\sum c_i Y_i)$ $= \frac{1}{c^2} \sum c_i^2 Var(Y_i)$ $= \frac{\sum c_i^2}{c^2} Var(Y_i)$ $= \frac{\sum (X_i - \bar{X})^2}{[\sum (X_i - \bar{X})^2]^2} \sigma^2$ $= \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$ $= \boxed{\frac{\sigma^2}{S_{XX}}}$	$Var(\hat{\beta}) = Var((X^T X)^{-1} X^T Y)$ $= (X^T X)^{-1} X^T Var(Y) (X^T X)^{-1} X^T$ $= (X^T X)^{-1} X^T X (X^T X)^{-1} \sigma^2 I$ $= \boxed{(X^T X)^{-1} \sigma^2}$ ## If $W = AY$, then $Var(W) = AVar(Y)A^T$

	$ \begin{aligned} Var(\hat{\beta}_0) &= Var(\bar{Y} - \hat{\beta}_1 \bar{X}) \\ &= Var(\bar{Y}) + Var(\hat{\beta}_1 \bar{X}) \\ &= Var\left(\frac{1}{n} \sum Y_i\right) + \bar{X}^2 Var(\hat{\beta}_1) \\ &= \frac{1}{n^2} \sum Var(Y_i) + \bar{X}^2 \frac{\sigma^2}{s_{XX}} \\ &= \frac{1}{n^2} n \sigma^2 + \bar{X}^2 \frac{\sigma^2}{s_{XX}} \\ &= \frac{1}{n} \sigma^2 + \bar{X}^2 \frac{\sigma^2}{s_{XX}} \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{s_{XX}} \right) \end{aligned} $	
Fitted Model	$ \begin{aligned} \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, 2, 3, \dots, n \\ \hat{\varepsilon}_i &= Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i), \quad i = 1, 2, 3, \dots, n \end{aligned} $	$ \begin{aligned} \hat{Y} &= X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY \\ \hat{\varepsilon} &= Y - X\hat{\beta} = Y - X(X^T X)^{-1} X^T Y = (I - H)Y, \quad \hat{\beta} \sim N(\beta, (X^T X)^{-1} \sigma^2) \end{aligned} $
Properties of $\hat{\varepsilon}$	$ \begin{aligned} 1) \quad E(\hat{\varepsilon}_i) &= E(Y_i - \hat{Y}_i) \\ &= E(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \\ &= E(Y_i) - E(\hat{\beta}_0) - E(\hat{\beta}_1 X_i) \\ &= \beta_0 + \beta_1 X_i - \beta_0 - \beta_1 X_i \\ &= 0 \\ 2) \quad Var(\hat{\varepsilon}_i) &= Var(Y_i - \hat{Y}_i) \\ &= Var(Y_i + \hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= Var(Y_i) + Var(\hat{\beta}_0) + Var(\hat{\beta}_1 X_i) \\ &= \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{s_{XX}} \right) + \frac{\sigma^2}{s_{XX}} Var(X_i) \\ &= \sigma^2 + \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{s_{XX}} + 0 \\ &= \sigma^2 \left(1 + \frac{1}{n} + \frac{\bar{X}^2}{s_{XX}} \right) \\ 3) \quad \sum \hat{\varepsilon}_i &= \sum (Y_i - \hat{Y}_i) \\ &= \sum Y_i - \sum \hat{Y}_i \\ &= \sum (\beta_0 + \beta_1 X_i + \varepsilon_i) - \sum (\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= \sum \beta_0 + \beta_1 \sum X_i + \sum \varepsilon_i - \sum \hat{\beta}_0 - \hat{\beta}_1 \sum X_i \end{aligned} $	$ \begin{aligned} 1) \quad E(\hat{\varepsilon}) &= E(Y - X\hat{\beta}) \\ &= E(Y) - E(X\hat{\beta}) \\ &= E(Y) - X(X^T X)^{-1} X^T E(Y) \\ &= X\beta - X(X^T X)^{-1} X^T X\beta \\ &= X\beta - X\beta \\ &= 0 \\ 2) \quad Var(\hat{\varepsilon}) &= Var(Y - X\hat{\beta}) \\ &= Var(Y) - Var(X\hat{\beta}) \\ &= \sigma^2 I - X(X^T X)^{-1} X^T Var(Y) (X(X^T X)^{-1} X^T)^T \\ &= \sigma^2 I - X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T \sigma^2 I \\ &= \sigma^2 I - X(X^T X)^{-1} X^T \sigma^2 I \\ &= (I - H) \sigma^2 \\ *3) \quad &\text{The hat matrix } H = X(X^T X)^{-1} X^T \text{ is symmetric and idempotent.} \\ &\text{Symmetric: } \dim(H) = n \times n, \text{ which is a square matrix, and:} \\ &H^T = (X(X^T X)^{-1} X^T)^T \\ &= (X^T)^T ((X^T X)^{-1})^T X^T \\ &= X(X^T X)^{-1} X^T \\ &= H \end{aligned} $

	$= n\beta_0 + n\beta_1\bar{X} + 0 - n\hat{\beta}_0 + n\hat{\beta}_1\bar{X}$ $= n\beta_0 + n\beta_1\bar{X} - n\bar{Y} - n\hat{\beta}_1\bar{X} + n\hat{\beta}_1\bar{X}$ $= n\beta_0 + n\beta_1\bar{X} - n\bar{Y}$ $= n(\beta_0 + \beta_1\bar{X} - \bar{Y}) \quad \text{##}(\bar{X}, \bar{Y}) \text{ is on the line}$ $= n(\bar{Y} - \bar{Y})$ $= 0$ 4) $\sum \hat{\varepsilon}_i^2 \leq \sum (Y_i - \mu_0 - \mu_1 X_i)^2$ for all real numbers μ_0 and μ_1. Fitted line is based on OLS, thus always has the smallest error than other lines. 5) $\sum X_i \hat{\varepsilon}_i = 0$ 6) $\sum \hat{Y}_i \hat{\varepsilon}_i = 0$ 7) The fitted regression line passes through the point of averages $(\bar{X}, \bar{Y})$. The fitted regression line is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, we plug $(\bar{X}, \bar{Y})$ into the fitted line and get: $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}, \text{ notice that } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \text{ let's substitute } \hat{\beta}_0, \text{ and we will get:}$ $\bar{Y} = \bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 \bar{X}$ $\bar{Y} = \bar{Y}$	Idempotent: $HH = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T$ $= X(X^T X)^{-1} X^T$ $= H$ **4) The matrix $(I - H)$ is also symmetric and idempotent Symmetric: $\dim(I - H) = n \times n$, which is a square matrix, and: $(I - H)^T = (I - X(X^T X)^{-1} X^T)^T$ $= I^T - (X(X^T X)^{-1} X^T)^T$ $= I - (X^T)^T ((X^T X)^{-1})^T X^T$ $= I - X(X^T X)^{-1} X^T$ $= I - H$ Idempotent: $(I - H)(I - H) = I^2 - IH - HI + H^2$ $= I - H - H + H$ $= I - H$ ***5) $(I - H)H = IH - HH = H - H^2 = H - H = 0$ $H(I - H) = HI - HH = H - H^2 = H - H = 0$ ****6) $\hat{\varepsilon} = Y - X\hat{\beta}$ $= Y - X(X^T X)^{-1} X^T Y$ $= (I - H)Y$ $\hat{\varepsilon} \perp \hat{Y} \Leftrightarrow \hat{\varepsilon} \cdot \hat{Y} = 0$ since $\hat{Y}$ is the projection of Y then $\hat{\varepsilon} = Y - \hat{Y}$, which should be perpendicular to $\hat{Y}$. $\hat{\varepsilon} \cdot \hat{Y} = \hat{\varepsilon}^T \hat{Y} = Y^T (I - H) \hat{Y} = Y^T (I - H) H Y = Y^T I H Y - Y^T H H Y = Y^T H Y - Y^T H Y = 0$ $\hat{\varepsilon} \perp X \Leftrightarrow \hat{\varepsilon} \cdot X = 0$ since $\hat{Y}$ is the projection of Y onto the column space of X and $\hat{\varepsilon} \perp \hat{Y}$, so $\hat{\varepsilon} \perp X$. $\hat{\varepsilon} \cdot X = \hat{\varepsilon}^T X = Y^T (I - H) X = Y^T I X - Y^T H X = Y^T X - Y^T X (X^T X)^{-1} X^T X = Y^T X - Y^T X = 0$
Estimators of Var and SD (Replace σ^2 with MSE)	$s^2(\hat{\beta}_1) = \frac{MSE}{S_{XX}}$ $s(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{XX}}}$ $s^2(\hat{\beta}_0) = MSE \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)$ $s(\hat{\beta}_0) = \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)}$	$s^2(\hat{\beta}) = (X^T X)^{-1} MSE$ $s(\hat{\beta}) = \sqrt{(X^T X)^{-1} MSE}$ If $s^2(\hat{\beta}) = (X^T X)^{-1} MSE = \begin{bmatrix} \frac{MSE}{n} + \frac{\bar{X}^2 MSE}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X} MSE}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} MSE}{\sum (X_i - \bar{X})^2} & \frac{MSE}{\sum (X_i - \bar{X})^2} \end{bmatrix}$, then $s^2(\hat{\beta}_0) = \frac{MSE}{n} + \frac{\bar{X}^2 MSE}{\sum (X_i - \bar{X})^2}$, $s^2(\hat{\beta}_1) = \frac{MSE}{\sum (X_i - \bar{X})^2}$, $s(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X} MSE}{\sum (X_i - \bar{X})^2}$

Sum Square and Mean Square	$SSTO = S_{YY} = \sum(Y_i - \bar{Y})^2$ $SSE = \sum \hat{\epsilon}_i^2 = \sum(Y_i - \hat{Y}_i)^2$ $SSR = \sum(\hat{Y}_i - \bar{Y})^2$ (SSTO = SSE + SSR) $MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{n-2}$ (unbiased estimator of σ^2) $MSR = \frac{SSR}{df(SSR)} = SSR$ $MSTO = \frac{SSTO}{df(SSTO)} = \frac{SSTO}{n-1}$ $E(SSE) = (n-2)\sigma^2$ $E(MSE) = \sigma^2$ $E(MSR) = \sigma^2 + \beta_1^2 \sum(X_i - \bar{X})^2$	$SSTO = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 + n\bar{Y}^2 = \mathbf{Y}^T \mathbf{Y} - n\bar{Y}^2$ $SSE = \sum(Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^T (\mathbf{Y} - \hat{\mathbf{Y}}) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}$ $SSR = \sum(\hat{Y}_i - \bar{Y})^2 = SSTO - SSE = \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} - n\bar{Y}^2$ (SSTO = SSE + SSR) $MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{n-p}$ (unbiased estimator of σ^2) $MSR = \frac{SSR}{df(SSR)}$ $MSTO = \frac{SSTO}{df(SSTO)}$ $E(SSE) = (n-p)\sigma^2$ $E(MSE) = \frac{E(SSE)}{n-p} = \sigma^2$
Central Limit Theorem	If the data are non-normal, we need the sample size $n \geq 30$ to make inference. If the data are normal, no requirement on sample size.	
Hypothesis Testing	1) $H_0: \beta = c, H_1: \beta \neq c, \alpha = 0.05$ (two-sided) Step 1: Compute the t-statistic: $t^* = \frac{\hat{\beta} - c}{s(\hat{\beta})}$. Step 2: Rejection Region rule: Reject H_0 if $ t^* > t(1 - \frac{\alpha}{2}; n-2)$. Step 3: Make conclusion, whether reject H_0 . 2) $H_0: \beta = c, H_1: \beta > c, \alpha = 0.05$ (one-sided) Step 1: Compute the t-statistic: $t^* = \frac{\hat{\beta} - c}{s(\hat{\beta})}$. Step 2: Rejection Region rule: Reject H_0 if $t^* > t(1 - \alpha; n-2)$. Step 3: Make conclusion, whether reject H_0 . 3) $H_0: \beta = c, H_1: \beta < c, \alpha = 0.05$ (one-sided) Step 1: Compute the t-statistic: $t^* = \frac{\hat{\beta} - c}{s(\hat{\beta})}$. Step 2: Rejection Region rule:	

	Reject H_0 if $t^* < -t(1 - \alpha; n - 2)$. Step 3: Make conclusion, whether reject H_0. ## When p-value $< \alpha$, we can reject H_0. ## When calculating p-value for two-sided test, we should double the value we got from the table.
P-value	P-value is the probability you observe the test statistic and more extreme than test statistic under the null hypothesis H_0 .
F-test	How to test whether two models differ or not? Use F-test. Comparing the MSE from these two models by using F-test. 1) t-distribution t-distribution with $df=v$ defined as the distribution of random variable T: $T = \frac{Z}{\sqrt{\frac{V}{v}}}$, where Z is a standard normal with mean 0 and variance 1; V has a chi-squared distribution with $df=v$; Z and V are independent. Since $\hat{\beta} = (X^T X)^{-1} X^T Y$, and it is a linear combination of Y, so $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$. For a single $\hat{\beta}_i, i = 1, \dots, p$, $\hat{\beta}_i \sim N(\beta_i, \sigma^2 (X^T X)^{-1}_{ii})$ Then the Z-score (standardization) is $\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 (X^T X)^{-1}_{ii}}} \sim N(0, 1), i = 1, \dots, p$ (or $p + 1$ if you include the intercept) *2) Chi-square distribution $\frac{Y^T (I - H) Y}{\sigma^2} \sim \chi^2(df = n - p)$, where $n - p = \text{trace}(I - H)$. Let $V = \frac{Y^T (I - H) Y}{\sigma^2}$. Now we can get $T = \frac{Z}{\sqrt{\frac{V}{v}}} = \frac{\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 (X^T X)^{-1}_{ii}}}}{\sqrt{\frac{Y^T (I - H) Y}{\sigma^2 (n - p)}}} = \frac{\frac{\hat{\beta}_i - \beta_i}{\sqrt{(X^T X)^{-1}_{ii}}}}{\sqrt{\frac{Y^T (I - H) Y}{(n - p)}}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{MSE(X^T X)^{-1}_{ii}}} \sim t(df = n - p), i = 1, \dots, p$. 3) F-distribution with df_1 and df_2 degrees of freedom is the distribution of $F = \frac{\frac{s_1}{d_1}}{\frac{s_2}{d_2}} = \frac{\frac{Y^T (I - H_1) Y}{\sigma^2 (n - m)}}{\frac{Y^T (I - H_2) Y}{\sigma^2 (n - p)}} = \frac{\frac{(n - m)}{MSE_1}}{\frac{(n - p)}{MSE_2}} \sim F(df_1 = n - m, df_2 = n - p)$ (when two models are independent)
Confidence Interval (CI)	A $(1 - \alpha)100\%$ confidence interval for β_1: $\hat{\beta}_1 \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(\hat{\beta}_1)$ A $(1 - \alpha)100\%$ confidence interval for β_0: $\hat{\beta}_0 \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(\hat{\beta}_0)$ Interpretation: We are $(1 - \alpha)100\%$ confident that the true value of β_0/β_1 is between xx and xx.

	## By default, $\alpha = 0.05$	
Estimation of Mean Response at $X = X_h$ (CI)	1) When $X = X_h$, we plug it into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h$ 2) Now we should get $s(\hat{Y}_h)$: $s^2(\hat{Y}_h) = \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}} \right] MSE \Rightarrow s(\hat{Y}_h) = \sqrt{s^2(\hat{Y}_h)}$ 3) A $(1 - \alpha)100\%$ confidence interval for the mean response at X_h is given by: $\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(\hat{Y}_h)$ Interpretation: We are $(1 - \alpha)100\%$ confident that the mean value of Y when $X = X_h$ is between xx and xx. ## By default, $\alpha = 0.05$	1) When $X = X_h = c$, we let $X_h = \begin{bmatrix} 1 \\ c \end{bmatrix}$ and plug its transpose X_h^T into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = X_h^T \hat{\beta}$ 2) Now we should get $s(\hat{Y}_h)$: $s^2(\hat{Y}_h) = X_h^T s^2(\hat{\beta}) X_h = X_h^T (X^T X)^{-1} MSE X_h \Rightarrow s(\hat{Y}_h) = \sqrt{s^2(\hat{Y}_h)}$ 3) A $(1 - \alpha)100\%$ confidence interval for the mean response at X_h is given by: $\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(\hat{Y}_h)$ Interpretation: We are $(1 - \alpha)100\%$ confident that the mean value of Y when $X = X_h$ is between xx and xx. ## By default, $\alpha = 0.05$
Prediction Interval (PI)	1) When $X = X_h$, we plug it into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h$ 2) Now we should get $s(\hat{Y}_h)$: $s^2(pred) = \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}} \right] MSE$ $= MSE + s^2(\hat{Y}_h)$ $\Rightarrow s^2(pred) = \sqrt{s^2(pred)}$ 3) A $(1 - \alpha)100\%$ confidence interval for the mean response at X_h is given by: $\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(pred)$ Interpretation: We are $(1 - \alpha)100\%$ confident that the predicted value of Y when $X = X_h$ is between xx and xx. ## By default, $\alpha = 0.05$	1) When $X = X_h = c$, we let $X_h = \begin{bmatrix} 1 \\ c \end{bmatrix}$ and plug its transpose X_h^T into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = X_h^T \hat{\beta}$ 2) Now we should get $s(\hat{Y}_h)$: $s^2(pred) = MSE + s^2(\hat{Y}_h) = MSE + X_h^T s^2(\hat{\beta}) X_h = MSE + X_h^T (X^T X)^{-1} MSE X_h$ $\Rightarrow s(pred) = \sqrt{s^2(pred)}$ 3) A $(1 - \alpha)100\%$ confidence interval for the mean response at X_h is given by: $\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(pred)$ Interpretation: We are $(1 - \alpha)100\%$ confident that the predicted value of Y when $X = X_h$ is between xx and xx. ## By default, $\alpha = 0.05$
Confidence	A $(1 - \alpha)100\%$ confidence band for the regression line is given by $\hat{Y}_h \pm W s(\hat{Y}_h)$, where $W = \sqrt{2F(1 - \alpha; 2; n - 2)}$. (F-distribution)	

Band																									
ANOVA	For simple linear regression: <table><thead><tr><th></th><th>df</th><th>Sum of square</th><th>Mean square</th><th>F ratio</th><th>P-value</th></tr></thead><tbody><tr><td>Regression</td><td>1</td><td>SSR</td><td>MSR</td><td>$\frac{MSR}{MSE}$ or $(t^*)^2$</td><td>xxx</td></tr><tr><td>Error</td><td>n</td><td>SSE</td><td>MSE</td><td></td><td></td></tr><tr><td>Total</td><td>n+1</td><td>SSTO</td><td>MSTO</td><td></td><td></td></tr></tbody></table>		df	Sum of square	Mean square	F ratio	P-value	Regression	1	SSR	MSR	$\frac{MSR}{MSE}$ or $(t^*)^2$	xxx	Error	n	SSE	MSE			Total	n+1	SSTO	MSTO		
	df	Sum of square	Mean square	F ratio	P-value																				
Regression	1	SSR	MSR	$\frac{MSR}{MSE}$ or $(t^*)^2$	xxx																				
Error	n	SSE	MSE																						
Total	n+1	SSTO	MSTO																						
Coefficient of Determination	1) $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$ 2) Interpretation: We can say $R^2\%$ of the variability in Y can be explained by X. 3) $R_{adj}^2 = \frac{MSR}{MSTO} = 1 - \frac{MSE}{MSTO}$ 4) R^2 is at least as large as R_{adj}^2. 5) $0 \leq R^2 \leq 1$ 6) $r^2 = [corr(X, Y)]^2$ 7) R^2 is unit free (does not depend on X or Y) 8) $R^2 = 1$ means perfect linear association between X and Y ($R^2 = 1 \Leftrightarrow SSE=0 \Leftrightarrow \sum(Y_i - \hat{Y}_i)^2 = 0 \Leftrightarrow Y_i = \hat{Y}_i$ for all i) 9) $R^2 = 0$ means no linear association between X and Y. However, a nonlinear association may exist. ($R^2 = 0 \Leftrightarrow SSR=0 \Leftrightarrow \sum(\hat{Y}_i - \bar{Y})^2 = 0 \Leftrightarrow \hat{Y}_i = \bar{Y}$ for all i)																								
MSE is an unbiased estimator for σ^2	$SSE = \sum(Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^T (\mathbf{Y} - \hat{\mathbf{Y}})$ $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y}$ Plug $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ into SSE: $SSE = (\mathbf{Y} - \mathbf{H}\mathbf{Y})^T (\mathbf{Y} - \mathbf{H}\mathbf{Y})$ $= [(\mathbf{I} - \mathbf{H})\mathbf{Y}]^T [(\mathbf{I} - \mathbf{H})\mathbf{Y}]$ $= \mathbf{Y}^T (\mathbf{I} - \mathbf{H})^T (\mathbf{I} - \mathbf{H}) \mathbf{Y}$ ## (I-H) is idempotent $= \mathbf{Y}^T (\mathbf{I} - \mathbf{H}) \mathbf{Y}$ Get the expectation of SSE: $E(SSE) = E(\mathbf{Y}^T (\mathbf{I} - \mathbf{H}) \mathbf{Y})$ $= tr[(\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}] + (\mathbf{X}\boldsymbol{\beta})^T (\mathbf{I} - \mathbf{H}) (\mathbf{X}\boldsymbol{\beta})$ $= \sigma^2 tr(\mathbf{I} - \mathbf{H}) + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{I} \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{H} \mathbf{X} \boldsymbol{\beta}$ ## Expected Value of Quadratic Form Let $\mathbf{Y}$ be a random vector, such that $E(\mathbf{Y}) = \boldsymbol{\mu}$, $Var(\mathbf{Y}) = \boldsymbol{\Sigma}$, and $\mathbf{A}$ is a symmetric matrix, then: $E(\mathbf{Y}^T \mathbf{A} \mathbf{Y}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$																								

$$\begin{aligned}
&= \sigma^2(tr(\mathbf{I}) - tr(\mathbf{H})) + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\
&= \sigma^2(n - p) + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\
&= \sigma^2(n - p)
\end{aligned}$$

We know $MSE = \frac{SSE}{n-p}$, so $E(MSE) = \frac{E(SSE)}{n-p} = \frac{\sigma^2(n-p)}{n-p} = \sigma^2$. As a result, we know that MSE is an unbiased estimator of σ^2 .

NOTICE:

1. Do not guarantee 100% correctness of this document.
2. Texts in **red** mean the correctness of the content is uncertain.
3. *, **, ***, ... mean something additional.
4. Texts begin with ## and in **green** mean notation.