STA 108: Applied Statistical Methods: Regression Analysis

Course Material Summary

University of California at Davis

Last Edit Date: 12/07/2021

STA 108 Math-related Summary

	Simple Linear Regression	Matrix Approach of Linear Regression
Model	$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $i = 1, 2, 3, \dots, n$ where we have n pairs of observations $(X_1, Y_1) \dots (X_n, Y_n)$	$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1}$, $\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}_{n\times(1+p)}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(1+p)\times 1}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{p\times 1}$ 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) $E(\varepsilon_i) = 0,\ i = 1, 2, 3, \dots, n$ 3) $Var(\varepsilon_i) = \sigma^2,\ i = 1, 2, 3, \dots, n$	$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$ 1) $E(\boldsymbol{\varepsilon}) = 0$ 2) $Var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$ *3) $\sigma^2 \boldsymbol{I} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$ 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	$\varepsilon_i \sim N(0, \sigma^2), Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$	$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}), \boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$
Expectation and variance of Y	$E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i)$ $= E(\beta_0) + E(\beta_1 X_i) + E(\varepsilon_i)$ $= \beta_0 + \beta_1 X_i + 0$ $= \boxed{\beta_0 + \beta_1 X_i}$ $Var(Y_i) = Var(\beta_0 + \beta_1 X_i + \varepsilon_i)$ $= Var(\beta_0) + Var(\beta_1 X_i) + Var(\varepsilon_i)$ $= 0 + 0 + \sigma^2$ $= \boxed{\sigma^2}$	$E(\boldsymbol{Y}) = E(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})$ $= E(\boldsymbol{X\beta}) + E(\boldsymbol{\varepsilon})$ $= \boldsymbol{X\beta} + 0$ $= \boxed{\boldsymbol{X\beta}}$ 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 $\boldsymbol{Y}$ is: $E(\boldsymbol{Y}) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$ $Var(\boldsymbol{Y}) = Var(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})$ $= Var(\boldsymbol{X\beta}) + Var(\boldsymbol{\varepsilon})$ $= 0 + \sigma^2 \boldsymbol{I}$ $= \boxed{\sigma^2 \boldsymbol{I}}$ The variance-covariance matrix of the random vector $\boldsymbol{Y}$ is:

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

		$= \boxed{\begin{bmatrix} \frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix}_{2\times 2}}$ ***$\boldsymbol{X^T Y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1+Y_2+\cdots+Y_n \\ X_1Y_1+X_2Y_2+\cdots+X_nY_n \end{bmatrix} = \boxed{\begin{bmatrix} \sum Y_i \\ \sum X_iY_i \end{bmatrix}_{n\times 1}}$ ****$\boldsymbol{\widehat{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} = \begin{bmatrix} \frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_iY_i \end{bmatrix} = \begin{bmatrix} \sum Y_i\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right)+\sum X_iY_i\left(\frac{-\bar{X}}{\sum(X_i-\bar{X})^2}\right) \\ \sum Y_i\left(\frac{-\bar{X}}{\sum(X_i-\bar{X})^2}\right)+\sum X_iY_i\left(\frac{1}{\sum(X_i-\bar{X})^2}\right) \end{bmatrix}$ $\sum Y_i\left(\frac{-\bar{X}}{\sum(X_i-\bar{X})^2}\right)+\sum X_iY_i\left(\frac{1}{\sum(X_i-\bar{X})^2}\right) = \frac{-n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2}+\frac{\sum X_iY_i}{\sum(X_i-\bar{X})^2} = \frac{\sum X_iY_i-n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2} = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} = \frac{S_{XY}}{S_{XX}} = \hat{\beta}_1$ $\sum Y_i\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right)+\sum X_iY_i\left(\frac{-\bar{X}}{\sum(X_i-\bar{X})^2}\right) = \frac{n\bar{Y}}{n}+\bar{X}\frac{n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2}-\bar{X}\frac{\sum X_iY_i}{\sum(X_i-\bar{X})^2} = \bar{Y}+\bar{X}\left(\frac{n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2}-\frac{\sum X_iY_i}{\sum(X_i-\bar{X})^2}\right)$ $= \bar{Y}+\bar{X}\cdot(-1)\cdot\left(\frac{\sum X_iY_i}{\sum(X_i-\bar{X})^2}-\frac{n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2}\right) = \bar{Y}-\bar{X}\left(\frac{\sum X_iY_i-n\bar{X}\bar{Y}}{\sum(X_i-\bar{X})^2}\right) = \bar{Y}-\bar{X}\hat{\beta}_1 = \hat{\beta}_0$ $\boxed{\boldsymbol{\widehat{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T Y} = \begin{bmatrix} \bar{Y}-\bar{X}\hat{\beta}_1 \\ \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}_{2\times 1}}$ *****Variance-covariance of $\boldsymbol{\widehat{\beta}}$ We know that $Var(\boldsymbol{\widehat{\beta}}) = (\boldsymbol{X^T X})^{-1}\sigma^2$, and $(\boldsymbol{X^T X})^{-1} = \begin{bmatrix} \frac{1}{n}+\frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \frac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix}$. As a result, variance-covariance of $\boxed{\boldsymbol{\widehat{\beta}} = \begin{bmatrix} \frac{\sigma^2}{n}+\frac{\bar{X}^2\sigma^2}{\sum(X_i-\bar{X})^2} & \frac{-\bar{X}\sigma^2}{\sum(X_i-\bar{X})^2} \\ \frac{-\bar{X}\sigma^2}{\sum(X_i-\bar{X})^2} & \frac{\sigma^2}{\sum(X_i-\bar{X})^2} \end{bmatrix}}$
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 $\boldsymbol{Y}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ in matrix form: $f_{\boldsymbol{Y}}(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n(\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}}\exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})}{2}\right]$, where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$. $Y \in$

	mean for Y, σ^2 is the variance of Y. $-\infty < Y < \infty$. Likelihood function: $\boxed{L\left(\beta_0, \beta_1, \sigma^2 \middle\| (Y_i, X_i)\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{{\varepsilon_i}^2}{2\sigma^2}\right]}$, where $\varepsilon_i \sim N(0, \sigma^2)$ independent and identically distributed. $L\left(\beta_0, \beta_1, \sigma^2 \middle\| (Y_i, X_i)\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{{\varepsilon_i}^2}{2\sigma^2}\right]$ $= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right]$ We use the log-likelihood function, which is easier: $l\left(\beta_0, \beta_1, \sigma^2 \middle\| (Y_i, X_i)\right)$ $= \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right]\right)$ $= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2\right]\right)$ $= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n\right) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$ $\frac{\partial l}{\partial \beta_0} = 0 - 0 - \frac{1}{2\sigma^2} \sum -2(Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum (Y_i - \beta_0 - \beta_1 X_i) = 0$ $\sum Y_i - \sum \beta_0 - \beta_1 \sum X_i = 0$ $n\bar{Y} - n\beta_0 - n\beta_1 \bar{X} = 0$ $n\bar{Y} - n\beta_1 \bar{X} = n\beta_0$ $\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$	R^p We know that $\boldsymbol{\mu} = \boldsymbol{X\beta}, \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$: $f_{\boldsymbol{Y}}(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y}-\boldsymbol{\mu})}{2}\right]$ $= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 \boldsymbol{I}))^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T (\sigma^2 \boldsymbol{I})^{-1} (\boldsymbol{Y}-\boldsymbol{X\beta})}{2}\right]$ $= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 \boldsymbol{I}))^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T (\boldsymbol{Y}-\boldsymbol{X\beta})}{2\sigma^2}\right]$ We can see that $\det(\sigma^2 \boldsymbol{I}) = det \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = (\sigma^2)^n$, $(\sigma^2 \boldsymbol{I})^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{\sigma^2} \end{bmatrix} = \frac{1}{\sigma^2} \boldsymbol{I}$. Likelihood function: $\boxed{L(\boldsymbol{\beta}, \sigma^2 \| (\boldsymbol{Y}, \boldsymbol{X})) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T (\boldsymbol{Y}-\boldsymbol{X\beta})}{2\sigma^2}\right]}$, where the model error is $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. We use the log-likelihood function, which is easier: $l(\boldsymbol{\beta}, \sigma^2 \| (\boldsymbol{Y}, \boldsymbol{X})) = \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n (\det(\sigma^2 \boldsymbol{I}))^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T (\boldsymbol{Y}-\boldsymbol{X\beta})}{2\sigma^2}\right]\right)$ $= \log\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n\right) - \frac{n}{2}\log(\sigma^2) - \frac{(\boldsymbol{Y}-\boldsymbol{X\beta})^T (\boldsymbol{Y}-\boldsymbol{X\beta})}{2\sigma^2}$ $\frac{\partial l}{\partial \boldsymbol{\beta}} = 0 - 0 - \frac{1}{2\sigma^2} (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{Y} = 0$ $\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X\beta} = \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{Y}$ $(\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X\beta} = (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{Y}$ $\boxed{\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{Y}}$ $\frac{\partial l}{\partial \sigma^2} = 0 - \frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \frac{(-1)}{(\sigma^2)^2} \cdot (\boldsymbol{Y} - \boldsymbol{X\beta})^T (\boldsymbol{Y} - \boldsymbol{X\beta}) = 0$

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

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

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

	$= n\beta_0 + n\beta_1\bar{X} + 0 - n\hat{\beta}_0 + n\hat{\beta}_1\bar{X}$ $= n\beta_0 + n\beta_1\bar{X} - n\bar{Y} - n\hat{\beta}_1\bar{X} + n\hat{\beta}_1\bar{X}$ $= n\beta_0 + n\beta_1\bar{X} - n\bar{Y}$ $= n(\beta_0 + \beta_1\bar{X} - \bar{Y})$ ##$(\bar{X}, \bar{Y})$ is on the line $= n(\bar{Y} - \bar{Y})$ $= 0$ 4) $\sum \hat{\varepsilon}_i^2 \le \sum (Y_i - \mu_0 - \mu_1 X_i)^2$ 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: $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$, notice that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$, let's substitute $\hat{\beta}_0$, and we will get: $\bar{Y} = \bar{Y} - \hat{\beta}_1\bar{X} + \hat{\beta}_1\bar{X}$ $\bar{Y} = \bar{Y}$	Idempotent: $\boldsymbol{HH} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ $= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ $= \boldsymbol{H}$ **4) The matrix $(\boldsymbol{I} - \boldsymbol{H})$ is also symmetric and idempotent Symmetric: $\dim(\boldsymbol{I} - \boldsymbol{H}) = n \times n$, which is a square matrix, and: $(\boldsymbol{I} - \boldsymbol{H})^T = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T$ $= \boldsymbol{I}^T - (\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)^T$ $= \boldsymbol{I} - (\boldsymbol{X}^T)^T((\boldsymbol{X}^T\boldsymbol{X})^{-1})^T\boldsymbol{X}^T$ $= \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ $= \boldsymbol{I} - \boldsymbol{H}$ Idempotent: $(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{I}^2 - \boldsymbol{IH} - \boldsymbol{HI} + \boldsymbol{H}^2$ $= \boldsymbol{I} - \boldsymbol{H} - \boldsymbol{H} + \boldsymbol{H}$ $= \boldsymbol{I} - \boldsymbol{H}$ ***5) $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{H} = \boldsymbol{IH} - \boldsymbol{HH} = \boldsymbol{H} - \boldsymbol{H}^2 = \boldsymbol{H} - \boldsymbol{H} = 0$ $\boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{HI} - \boldsymbol{HH} = \boldsymbol{H} - \boldsymbol{H}^2 = \boldsymbol{H} - \boldsymbol{H} = 0$ ****6) $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ $= \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ $= (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ $\hat{\boldsymbol{\varepsilon}} \perp \hat{\boldsymbol{Y}} \Leftrightarrow \hat{\boldsymbol{\varepsilon}} \cdot \hat{\boldsymbol{Y}} = 0$ since $\hat{\boldsymbol{Y}}$ is the projection of $\boldsymbol{Y}$ then $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$, which should be perpendicular to $\hat{\boldsymbol{Y}}$. $\hat{\boldsymbol{\varepsilon}} \cdot \hat{\boldsymbol{Y}} = \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{Y}} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\hat{\boldsymbol{Y}} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{HY} = \boldsymbol{Y}^T\boldsymbol{IHY} - \boldsymbol{Y}^T\boldsymbol{HHY} = \boldsymbol{Y}^T\boldsymbol{HY} - \boldsymbol{Y}^T\boldsymbol{HY} = 0$ $\hat{\boldsymbol{\varepsilon}} \perp \boldsymbol{X} \Leftrightarrow \hat{\boldsymbol{\varepsilon}} \cdot \boldsymbol{X} = 0$ since $\hat{\boldsymbol{Y}}$ is the projection of $\boldsymbol{Y}$ onto the column space of $\boldsymbol{X}$ and $\hat{\boldsymbol{\varepsilon}} \perp \hat{\boldsymbol{Y}}$, so $\hat{\boldsymbol{\varepsilon}} \perp \boldsymbol{X}$. $\hat{\boldsymbol{\varepsilon}} \cdot \boldsymbol{X} = \hat{\boldsymbol{\varepsilon}}^T\boldsymbol{X} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{Y}^T\boldsymbol{IX} - \boldsymbol{Y}^T\boldsymbol{HX} = \boldsymbol{Y}^T\boldsymbol{X} - \boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{Y}^T\boldsymbol{X} - \boldsymbol{Y}^T\boldsymbol{X} = 0$
Estimators of Var and SD (Replace σ^2 with MSE)	$s^2(\hat{\beta}_1) = \frac{MSE}{S_{XX}}$ $s(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{XX}}}$ $s^2(\hat{\beta}_0) = MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)$ $s(\hat{\beta}_0) = \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right)}$	$s^2(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}MSE$ $s(\hat{\boldsymbol{\beta}}) = \sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}MSE}$ If $s^2(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}MSE = \begin{bmatrix} \frac{MSE}{n} + \frac{\bar{X}^2 MSE}{\sum(X_i - \bar{X})^2} & \frac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} \\ \frac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} & \frac{MSE}{\sum(X_i - \bar{X})^2} \end{bmatrix}$, 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	$SSTO = S_{YY} = \sum(Y_i - \bar{Y})^2$ $SSE = \sum \hat{\varepsilon}_i^2 = \sum\left(Y_i - \hat{Y}_i\right)^2$ $SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2$ (SSTO = SSE + SSR) $MSE = \frac{SSE}{df(SSE)} = \frac{SSE}{n-2}$ (unbiased estimator of σ^2) $MSR = \frac{SSR}{df(SSR)} = SSR$ $MSTO = \frac{SSTO}{df(SSTO)} = \frac{SSTO}{n-1}$ $E(SSE) = (n-2)\sigma^2$ $E(MSE) = \sigma^2$ $E(MSR) = \sigma^2 + {\beta_1}^2 \sum(X_i - \bar{X})^2$	$SSTO = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 + n\bar{Y}^2 = \boldsymbol{Y^T Y} - n\bar{Y}^2$ $SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 = \left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right)^T\left(\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right) = \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{Y^T Y} - \hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^T Y}$ $SSR = \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = SSTO - SSE = \hat{\boldsymbol{\beta}}^{\boldsymbol{T}}\boldsymbol{X^T Y} - n\bar{Y}^2$ (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 $\lvert t^* \rvert > t(1-\frac{\alpha}{2}; n-2)$. 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 < α, 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 c chi-squared distribution with df=v; Z and V are independent. Since $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, and it is a linear combination of $\boldsymbol{Y}$, so $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$. For a single $\hat{\boldsymbol{\beta}}_i, i = 1, \dots, p$, $\hat{\boldsymbol{\beta}}_i = N(\boldsymbol{\beta}_i, \sigma^2 {(\boldsymbol{X}^T\boldsymbol{X})_{ii}}^{-1})$ Then the Z-score (standardization) is $\boxed{\frac{\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i}{\sqrt{\sigma^2 {(\boldsymbol{X}^T\boldsymbol{X})_{ii}}^{-1}}} \sim N(0,1)}$, $i = 1, \dots, p$(or $p+1$ if you include the intercept) *2) Chi-square distribution $\boxed{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{\sigma^2} \sim X^2(df = n-p)}$, where $n-p = trace(\boldsymbol{I}-\boldsymbol{H})$. Let $V = \frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{\sigma^2}$. Now we can get $\boxed{\boldsymbol{T} = \frac{Z}{\sqrt{\frac{V}{v}}} = \frac{\frac{\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i}{\sqrt{\sigma^2 {(\boldsymbol{X}^T\boldsymbol{X})_{ii}}^{-1}}}}{\sqrt{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{\sigma^2(n-p)}}} = \frac{\frac{\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i}{\sqrt{{(\boldsymbol{X}^T\boldsymbol{X})_{ii}}^{-1}}}}{\sqrt{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y}}{(n-p)}}} = \frac{\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i}{\sqrt{MSE{(\boldsymbol{X}^T\boldsymbol{X})_{ii}}^{-1}}} \sim t(df = n-p)}$, $i = 1, \dots, p$. 3) F-distribution with df_1 and df_2 degrees of freedom is the distribution of $\boxed{F = \frac{\frac{S_1}{d_1}}{\frac{S_2}{d_2}} = \frac{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{Y}}{\sigma^2(n-m)}}{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H}_2)\boldsymbol{Y}}{\sigma^2(n-p)}} = \frac{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{Y}}{(n-m)}}{\frac{\boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{H}_2)\boldsymbol{Y}}{(n-p)}} = \frac{MSE_1}{MSE_2} \sim F(df_1 = n-m, df_2 = n-p)}$ (when two models are independent)
Confidence Interval (CI)	A $(1-\alpha)100\%$ confidence interval for β_1: $\boxed{\hat{\beta}_1 \pm t\left(1 - \frac{\alpha}{2}; n-2\right) s(\hat{\beta}_1)}$ A $(1-\alpha)100\%$ confidence interval for β_0: $\boxed{\hat{\beta}_0 \pm t\left(1 - \frac{\alpha}{2}; n-2\right) s(\hat{\beta}_0)}$ Interpretation: We are $(1-\alpha)100\%$ confident that the true value of $\hat{\beta}_0/\hat{\beta}_1$ is between xx and xx.

	## By default, $\alpha = 0.05$	
Estimation of Mean Response at $\boldsymbol{X} = \boldsymbol{X_h}$ (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 $s(\hat{Y}_h)$: $\boxed{s^2(\hat{Y}_h) = \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}}\right] MSE}$ => $s(\hat{Y}_h) = \sqrt{s^2(\hat{Y}_h)}$ 3) A $(1-\alpha)100\%$ confidence interval for the mean response at X_h is given by: $\boxed{\hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-2) s(\hat{Y}_h)}$ 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 $\boldsymbol{X} = \boldsymbol{X_h} = c$, we let $\boldsymbol{X_h} = \begin{bmatrix} 1 \\ c \end{bmatrix}$ and plug its transpose $\boldsymbol{X_h^T}$ into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \boldsymbol{X_h^T}\boldsymbol{\hat{\beta}}$ 2) Now we should get $s(\hat{Y}_h)$: $\boxed{s^2(\hat{Y}_h) = \boldsymbol{X_h^T} s^2(\boldsymbol{\hat{\beta}}) \boldsymbol{X_h} = \boldsymbol{X_h^T}(\boldsymbol{X^T X})^{-1} MSE \boldsymbol{X_h}}$ => $s(\hat{Y}_h) = \sqrt{s^2(\hat{Y}_h)}$ 3) A $(1-\alpha)100\%$ confidence interval for the mean response at X_h is given by: $\boxed{\hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-2) s(\hat{Y}_h)}$ 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 $s(\hat{Y}_h)$: $\boxed{s^2(pred) = \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{XX}}\right] MSE}$ $= \boxed{MSE + s^2(\hat{Y}_h)}$ $=> s^2(pred) = \sqrt{s^2(pred)}$ 3) A $(1-\alpha)100\%$ confidence interval for the mean response at X_h is given by: $\boxed{\hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-2) s(pred)}$ 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 $\boldsymbol{X} = \boldsymbol{X_h} = c$, we let $\boldsymbol{X_h} = \begin{bmatrix} 1 \\ c \end{bmatrix}$ and plug its transpose $\boldsymbol{X_h^T}$ into the fitted line: $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \boldsymbol{X_h^T}\boldsymbol{\hat{\beta}}$ 2) Now we should get $s(\hat{Y}_h)$: $\boxed{s^2(pred) = MSE + s^2(\hat{Y}_h) = MSE + \boldsymbol{X_h^T} s^2(\boldsymbol{\hat{\beta}}) \boldsymbol{X_h} = MSE + \boldsymbol{X_h^T}(\boldsymbol{X^T X})^{-1} MSE \boldsymbol{X_h}}$ => $s(pred) = \sqrt{s^2(pred)}$ 3) A $(1-\alpha)100\%$ confidence interval for the mean response at X_h is given by: $\boxed{\hat{Y}_h \pm t(1-\frac{\alpha}{2}; n-2) s(pred)}$ 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)	

<table>
<tr><td>Band</td><td></td></tr>
<tr><td>ANOVA</td><td>
For simple linear regression:

<table>
<tr><th></th><th>df</th><th>Sum of square</th><th>Mean square</th><th>F ratio</th><th>P-value</th></tr>
<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>
</table>
</td></tr>
<tr><td>Coefficient of Determination</td><td>

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) $0 \le R^2 \le 1$

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 ($R^2 = 1$⇔SSE=0⇔$\sum(Y_i - \hat{Y}_i)^2 = 0$⇔$Y_i = \hat{Y}_i$ for all i)

9) $R^2 = 0$ means no linear association between X and Y. However, a nonlinear association may exist. ($R^2 = 0$⇔SSR=0⇔$\sum(\hat{Y}_i - \bar{Y})^2 = 0$⇔$\hat{Y}_i = \bar{Y}$ for all i)
</td></tr>
<tr><td>MSE is an unbiased estimator for σ^2</td><td>

$SSE = \sum(Y_i - \hat{Y}_i)^2 = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}})$

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

Plug $\hat{\boldsymbol{Y}} = \boldsymbol{HY}$ into SSE:

$SSE = (\boldsymbol{Y} - \boldsymbol{HY})^T(\boldsymbol{Y} - \boldsymbol{HY})$
$\quad = [(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}]^T[(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}]$
$\quad = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ ## (I-H) is idempotent
$\quad = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$

Get the expectation of SSE:

$E(SSE) = E(\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y})$
$\quad = tr[(\boldsymbol{I} - \boldsymbol{H})\sigma^2\boldsymbol{I}] + (\boldsymbol{X\beta})^T(\boldsymbol{I} - \boldsymbol{H})(\boldsymbol{X\beta})$
$\quad = \sigma^2 tr(\boldsymbol{I} - \boldsymbol{H}) + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{IX\beta} - \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{HX\beta}$

Expected Value of Quadratic Form
Let $\boldsymbol{Y}$ be a random vector, such that $E(\boldsymbol{Y}) = \mu$, $Var(\boldsymbol{Y}) = \Sigma$, and $\boldsymbol{A}$ is a symmetric matrix, then:
$E(\boldsymbol{Y}^T\boldsymbol{AY}) = tr(\boldsymbol{A}\Sigma) + \mu^T\boldsymbol{A}\mu$
</td></tr>
</table>

	$= \sigma^2\big(tr(\boldsymbol{I}) - tr(\boldsymbol{H})\big) + \boldsymbol{\beta^T X^T X \beta} - \boldsymbol{\beta^T X^T X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X^T X \beta}$ $= \sigma^2(n-p) + \boldsymbol{\beta^T X^T X \beta} - \boldsymbol{\beta^T X^T X \beta}$ $= \sigma^2(n-p)$ 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.