Formula Sheet L0101/L0201 Notation

Simple Linear Regression	Multiple Linear Regression
$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$	$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	$\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{-1}\mathbf{X}^T\mathbf{Y}$
$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^n (x_i-\bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$	
$s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{n-2} = \frac{\sum_{i=1}^n (y_i-\hat{y}_i)^2}{n-2}$	$s^2 = \hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{n-p-1}$
$Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)$	$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$
$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}$	
$Var(\hat{y}_0 \mid x_0, X) = \sigma^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)$	$Var(\hat{y}_0 \mid \mathbf{X}, \mathbf{x}_0) = \sigma^2\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0$
$\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-2}\sqrt{\widehat{Var}(\hat{\beta}_j)}$	$\hat{\boldsymbol{\beta}}_j \pm t_{\frac{\alpha}{2},n-p-1} s\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{(j+1,j+1)}}$
$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\frac{\alpha}{2},n-2}\sqrt{\widehat{Var}(\hat{y}_0)}$	$\mathbf{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1} s\sqrt{\mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$
$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\frac{\alpha}{2},n-2} s\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}$	$\mathbf{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1} s\sqrt{1 + \mathbf{x}_0^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0}$
$T = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\widehat{Var}(\hat{\beta}_j)}} \sim T_{n-2}$	$T = \frac{\hat{\boldsymbol{\beta}}_j - \beta_j^0}{s\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{(j+1,j+1)}}} \sim T_{n-p-1}$

$$\mathbf{X^T X} = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & \ldots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \ldots & \sum x_{i1}x_{ip} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 & \ldots & \sum x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \ldots & \sum x_{ip}^2 \end{pmatrix}, \quad VIF_j = \frac{1}{1-R_j^2}$$

$$RSS = \sum_{i=1}^n (y_i - \hat{E}(Y_i \mid \mathbf{X}))^2 \qquad R^2 = \frac{SS_{reg}}{SST} = 1 - \frac{RSS}{SST} \qquad R^2_{adj} = 1 - \frac{RSS/n-p-1}{SST/n-1}$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad df: n-1 = n-p-1+p$$

$$F = \frac{SS_{reg}/p}{RSS/n-p-1} \sim F(p, n-p-1) \qquad F = \frac{(RSS_{reduced} - RSS_{full})/k}{RSS_{full}/n-p-1} \sim F(k, n-p-1)$$

$$h_{ii} = \frac{1}{n} + \frac{(x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2} > 2\frac{p+1}{n}, \quad r_i = \frac{\hat{e}_i}{s\sqrt{1-h_{ii}}} \notin [-2, 2]$$

$$DFBETA_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{s_{(i)}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{j+1,j+1}}}, \quad \mid DFBETA_{j(i)} \mid > \frac{2}{\sqrt{n}}$$

$$AIC \propto n\ln\left(\frac{RSS}{n}\right) + 2p, \quad BIC \propto n\ln\left(\frac{RSS}{n}\right) + (p+2)\ln(n)$$

Formula Sheet LEC5101 Notation

Simple Linear Regression	Multiple Linear Regression
$Y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \dots, n$	$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}$
$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$	$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$
$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$	
$s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}$	$s^2 = \hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{n-p-1}$
$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$	$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$
$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$	
$Var(\hat{Y}^* \mid x^*, X) = \sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$	$Var(\hat{Y}^* \mid \mathbf{X}, \mathbf{x}^*) = \sigma^2 \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*$
$\hat{\beta}_j \pm t_{1-\frac{\alpha}{2}, n-2} \hat{se}(\hat{\beta}_j)$	$\hat{\beta}_j \pm t_{1-\frac{\alpha}{2}, n-p-1} \hat{\sigma} \sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{(j+1, j+1)}}$
$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{1-\frac{\alpha}{2}, n-2} \hat{se}(\hat{Y}^*)$	$\mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \pm t_{1-\frac{\alpha}{2}, n-p-1} \hat{\sigma} \sqrt{\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}$
$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{1-\frac{\alpha}{2}, n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$	$\mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \pm t_{1-\frac{\alpha}{2}, n-p-1} \hat{\sigma} \sqrt{1 + \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}$
$\frac{\hat{\beta}_j - \beta_j^0}{\hat{se}(\hat{\beta}_j)} \sim t_{n-2}$	$\frac{\hat{\beta}_j - \beta_j^0}{\hat{\sigma}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{(j+1, j+1)}}} \sim t_{n-p-1}$

$$\mathbf{X}^{\mathbf{T}}\mathbf{X} = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \dots & \sum x_{i1}x_{ip} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 & \dots & \sum x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix}, \quad VIF_j = \frac{1}{1 - R_j^2}$$

$$RSS = \sum_{i=1}^n (Y_i - \hat{E}(Y_i \mid \mathbf{X}))^2 \qquad R^2 = \frac{SS_{reg}}{SST} = 1 - \frac{RSS}{SST} \qquad R^2_{adj} = 1 - \frac{RSS/n-p-1}{SST/n-1}$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2, \quad df: n-1 = n-p-1+p$$

$$F = \frac{SS_{reg}/p}{RSS/n-p-1} \sim F(p, n-p-1) \qquad F = \frac{(RSS_{reduced} - RSS_{full})/k}{RSS_{full}/n-p-1} \sim F(k, n-p-1)$$

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} > 2\frac{p+1}{n}, \quad r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} \notin [-2, 2]$$

$$DFBETA_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\hat{\sigma}_{(i)}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{j+1, j+1}}}, \quad \mid DFBETA_{j(i)} \mid > \frac{2}{\sqrt{n}}$$

$$AIC = n\log\left(\frac{RSS}{n}\right) + 2(p+1), \quad BIC = n\log\left(\frac{RSS}{n}\right) + \log(n)(p+1)$$