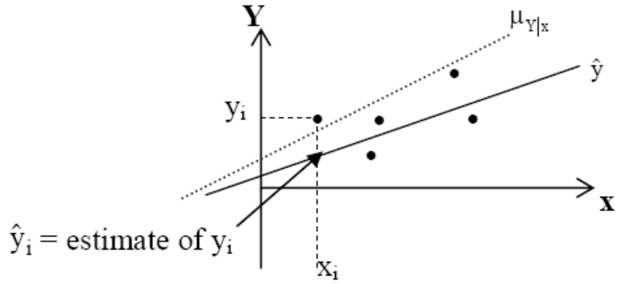
ENGG 319 Section #11 Summary
Simple Linear Regression and Correlation

Simple Linear Regression

- Regressor variable x is fixed
- Response variable Y is a random variable



Estimator: Fitted Regression Line: $\hat{y} = a + bx$

True Population Regression Line: $\mu_{Y|x} = \alpha + \beta x$

$y_i = \alpha + \beta x_i + \varepsilon_i$ and $y_i = a + bx_i + e_i$

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

Minimize $SSE = \sum_{i=1}^{n} e_i^2$ to get the *least squares estimates of* ***a*** *and slope* ***b*** *of the linear regression model*

$b = \dfrac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$	$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$
$a = \bar{y} - b\bar{x}$ $b = S_{xy}/S_{xx}$ $SSE = S_{yy} - bS_{xy}$	$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \dfrac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$

Point Estimates

$\hat{y}_o$ is a point estimate for $\mu_{Y\|x_o}$	$s^2 = \dfrac{SSE}{n-2}$ is a point estimate for $\sigma^2_{Y\|x_i} = \sigma^2$

Confidence and Prediction Intervals

A (1-α)100% confidence interval for the mean response $\mu_{Y\|x_0}$ is $\hat{y}_0 \pm t_{\alpha/2} s\sqrt{\dfrac{1}{n} + \dfrac{(x_0 - \bar{x})^2}{S_{xx}}}$ where $t_{\alpha/2}$ is a value of the t-distribution with n-2 degrees of freedom.	A (1-α)100% prediction interval for a single response y_0 is $\hat{y}_0 \pm t_{\alpha/2} s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x_0 - \bar{x})^2}{S_{xx}}}$ where $t_{\alpha/2}$ is a value of the t-distribution with n-2 degrees of freedom.

Correlation

Sample coefficient of determination $r^2 = \dfrac{SSR}{SST} = \dfrac{S_{xy}^2}{S_{xx}S_{yy}} \quad 0 \le r^2 \le 1$	Sample correlation coefficient $r = b\sqrt{\dfrac{S_{xx}}{S_{yy}}} = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ $-1 < r < 1$