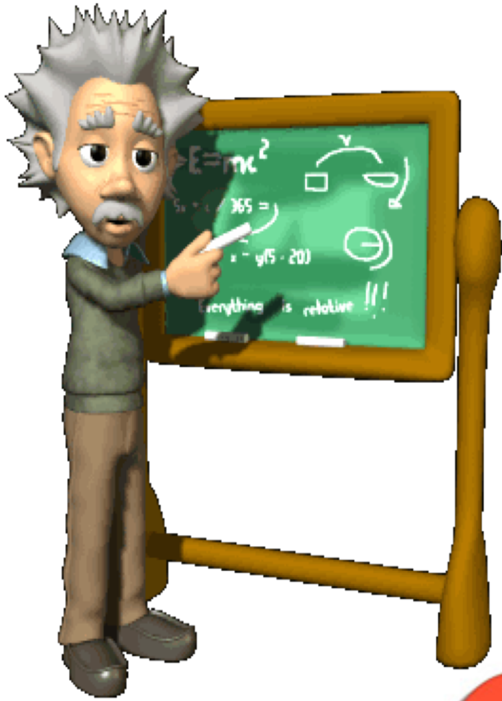
Regression Analysis
--Final Review



Chapter 2：The Simple Linear Regression Model

Basic model with one regressor:

SLR Model Form

$$y = \beta_0 + \beta_1 x + \epsilon$$

The least squares procedure and SLR have the following assumptions:

1. x_i's are nonrandom
2. $\epsilon_i \sim N(0, \sigma^2)$ and *independent*
3. ϵ_i are uncorrelated

Chapter 2：The Simple Linear Regression Model

- Derive least squares estimation of $\beta_0 \ and \ \beta_1$.
- Obtain the equation of the least squares regression line.
- Provide the interpretation of $\hat{\beta}_0$ and $\hat{\beta}_1$.
- Derive Maximum likelihood estimation of $\beta_0, \ \beta_1$ and σ^2.
- Obtain the unbiased estimation of σ^2.
- Perform test for the significance of the slope.
- Find and interpret the coefficient of determination R^2.
- Find confidence intervals for the mean value of response.
- Find prediction intervals for the mean value of response.
- Determine and interpret SSE, SST, SSR.
- Construct ANOVA table and determine the value of the F statistic.

Chapter 3
The Multiple Linear Regression Model

General Regression Model in terms of Matrix:

$$
y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.
$$

$$
\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}
$$

$$
E(\boldsymbol{\varepsilon}) = 0, \qquad Cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n
$$

where

$$
\boldsymbol{y} : n\times 1, \quad \boldsymbol{X} : n\times p, \quad \boldsymbol{\beta} : p\times 1,
$$

$\boldsymbol{\varepsilon} : n\times 1$ with iid components.

Chapter 3 The Multiple Linear Regression Model

- Derive least squares estimation of β.
- Provide the interpretation of $\hat{\beta}$.
- Derive Maximum likelihood estimation of β and σ^2.
- Obtain the unbiased estimation of σ^2.
- Determine and interpret the coefficient of determination R^2.
- Test whether the variable x_j gives a significant contribution to the model.
- Construct ANOVA table and determine the value of the F statistic.
- Conduct two statistical tests: the lack of fit test and regression model test.
- Explain what is multicollinearity and the influence of multicollinearity to the model. How to detect multicollinearity?

Chapter 4
Criteria For Choice of Best Model

- What is difference between R^2 and $R^2{}_{adj}$? explain how to use R^2 and $R^2{}_{adj}$ to compare models.
- Describe the procedure of Cross Validation.
- What is PRESS statistics? explain how to use PRESS statistics to compare models. Show PRESS$=\sum_{i=1}^{n}[\frac{e_i}{1-h_{ii}}]^2$
- What is Cp statistics? explain how to use Cp statistics to compare models.
- What is meaning of overfitting and underfitting?
- Describe the following variable selection procedures (Forward selection, Backward elimination, Stepwise regression, MAXR, Best subset)

Chapter 5 Statistical Diagnostics

- What is regression diagnostic？ Why we need to conduct model diagnostics? List statistical diagnostics in regression model.
- Define the studentized residuals
- When Non-linearity is detected，give some suggestions on remedies.
- When Non-constant residual variance is diagnosed, list two remedies you could come up with.
- Give comments on four kinds of diagnostics plots.
- Give the definition of DFFITS, DFBETAS and Cook distance. Explain how to use them to identify the influential cases. Based on their definition, derive the formula of DFFITS, DFBETAS and Cook distance.