STAT 3008 Applied Regression Analysis
Tutorial 5.

XU yongze

OCT 15&17

Contents

1 Introduction to Multiple Linear Regression Model

1.1 Brief introduction

♣ The Multiple Linear regression model is regarded as a generalized form of the simple linear regression model. Instead of using only one independent variable, we use several independent variables to predict the dependent variable.

1.2 Model building

♣ Model:
$y_i = \beta_0 = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$
$\mathrm{E}(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$
$\mathrm{Var}(Y|X) = \sigma^2$
Here $X_1, X_2, \cdots, X_p$ are columns vectors in X.

1.3 Matrix Notation of the Regression Model

♣ Matrix representation of Multiple regression is important.
♣ Model:
$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$
with:
$\mathrm{E}(\mathbf{e}) = \mathbf{0}$
$\mathrm{Var}(\mathbf{e}) = \sigma^2 \mathbf{I_n}$
$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I_n})$(Normality Assumption)

1.4 Estimation of the Parameter

This part is important in midterm test. You should know how to calculate the results.

♣ We estimate $\boldsymbol{\beta}$ by minimizing the sum of square of residuals.

$$RSS = \sum [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})]^2 \tag{1.1}$$

In matrix form:

$$RSS = \mathbf{e}^t \mathbf{e} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{t}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{1.2}$$

After the differentiation works(Refer to the lecture notes chapter 3), the estimation of $\boldsymbol{\beta}$is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^t X})^{-1}\mathbf{X^t Y}, \quad \mathbf{X^t X} \text{ is non-singular} \tag{1.3}$$

$$\hat{\sigma^2} = \frac{\widehat{RSS}}{n-(p+1)} \tag{1.4}$$

1.5 Properties of Estimates

♣ $\hat{\beta}$ is unbiased estimator of β.
$\mathrm{E}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X^t X})^{-1}\mathbf{X^t}\mathrm{E}(\mathbf{Y}) = (\mathbf{X^t X})^{-1}\mathbf{XX}\boldsymbol{\beta} = \boldsymbol{\beta}$

♣ $\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X^t X})^{-1}$
$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X^t X})^{-1}\mathbf{X^t}\sigma^2\mathbf{IX}(\mathbf{X^t X})^{-1} = \sigma^2(\mathbf{X^t X})^{-1}$
1>.The Var can be used to test components in $\boldsymbol{\beta}$.
2>.You may use this method to calculate Var for β_0 & β_1 in simple linear regression.

♣ Fitted Value
$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X^t X})^{-1}\mathbf{X^t Y} = \mathbf{HY}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X^t X})^{-1}\mathbf{X^t}$ is called the hat matrix.
$\mathbf{H}$ is symmetric.

♣ Residual
$\hat{\mathrm{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$

♣ Sum of Square of Residuals
$\widehat{RSS} = \hat{\mathrm{e}}^t\hat{\mathrm{e}} = \mathbf{Y^t}(\mathbf{I}-\mathbf{H})\mathbf{Y}$(As we have $\mathbf{HH} = \mathbf{H}$, $(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H}) = \mathbf{I}-\mathbf{H}$)

♣ $\mathrm{E}(\hat{\mathbf{Y}}) = \mathbf{X}\boldsymbol{\beta}$

♣ $\mathrm{Var}(\hat{\mathbf{Y}}) = \mathbf{H}\sigma^2\mathbf{IH}^t = \mathbf{H}\sigma^2\mathbf{IH} = \sigma^2\mathbf{H}$

♣ $\mathrm{E}(\hat{\mathrm{e}}) = (\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$

♣ $\mathrm{Var}(\hat{\mathrm{e}}) = (\mathbf{I}-\mathbf{H})\boldsymbol{\sigma^2}\mathbf{I}(\mathbf{I}-\mathbf{H})^{\mathbf{t}} = (\mathbf{I}-\mathbf{H})\boldsymbol{\sigma^2}\mathbf{I}(\mathbf{I}-\mathbf{H}) = \sigma^2(\mathbf{I}-\mathbf{H})$

♣ $\mathrm{E}(\hat{\sigma^2}) = \sigma^2$ (Refer to page 18 in lecture notes of chapter3)

2 Added-Variable plot

There are some equivalent expressions for an added-variable plot. Taking mean function $E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ as an example, added-variable plot for X_1, added-variable plot between Y and X_1 with the effect of X_2 removed, added-variable plot for X_1 after X_2, are all describing an added-variable plot whose vertical axis is the residual of model $y \sim X_2$ and horizontal axis is the residual of model $X_1 \sim X_2$.

2.1 Properties

- the estimated slope in the added-variable plot for $X_1 = \hat{\beta}_1 = \frac{\sum \hat{e}_{i_{\{Y|X_2\}}} \hat{e}_{i_{\{X_1|X_2\}}}}{\sum \hat{e}^2_{i_{\{X_1|X_2\}}}}$
- the residuals in the added-variable plot (for model $\hat{e}_{Y|X_2} \sim \hat{e}_{X_1|X_2}$) are identical to the residuals from the mean function with both predictors
- t-test for the coefficient is not quite the same from the added-variable plot and from the regression with both terms, because the degree of freedom in added-variable plot is more than that in regression with both terms by one
- if $X_1 = cX_2$, where c is a constant, the residuals for X_1 adjusted for X_2 will be zeros. The added variable plot will be a vertical line with x-intercept=0
- if $Y = cX_2$, which is an exact linear relationship without any random error, then the residuals for Y adjusted for X_2 will be zeros. The added variable plot will be a horizontal line with y-intercept=0
- If X_2 is exactly uncorrelated with both X_1 and Y, the added-variable plot for X_1 after X_2 has exactly the same shape as the scatter plot of Y versus X_1
- The vertical variation in an added-variable plot for X_1 after X_2 is always less than or equal to the vertical variation in a plot of Y versus X_1, since the vertical variable is the residuals from the regression of Y on X_2

2.2 An exercise to find the slope of added-variable plots (Example from 12-13 Midterm)

15. (10 marks). Given three variables $Y = (y_1, \ldots, y_n)$, $X1 = (x_1, \ldots, x_n)$, $X2 = (1, 1, 1, \ldots, 1)$.

  i) (5 marks) Find the slope of the added variable plot between Y and $X1$, with the effect of $X2$ removed.

ii) (5 marks) Find the slope of the added variable plot between Y and $X2$, with the effect of $X1$ removed.

(i)

$$\begin{aligned}
\hat{e}_{Y|X_2} &= Y - \bar{Y} && \text{(2 marks)} \\
\hat{e}_{X_1|X_2} &= X_1 - \bar{X}_1 && \text{(2 marks)} \\
slope &= \frac{\sum_{i=1}^{n}(y_i - \bar{Y})(x_i - \bar{X}_1)}{\sum_{i=1}^{n}(x_i - \bar{X}_1)^2} = \hat{\beta}_1 && \text{(1 mark)}
\end{aligned}$$

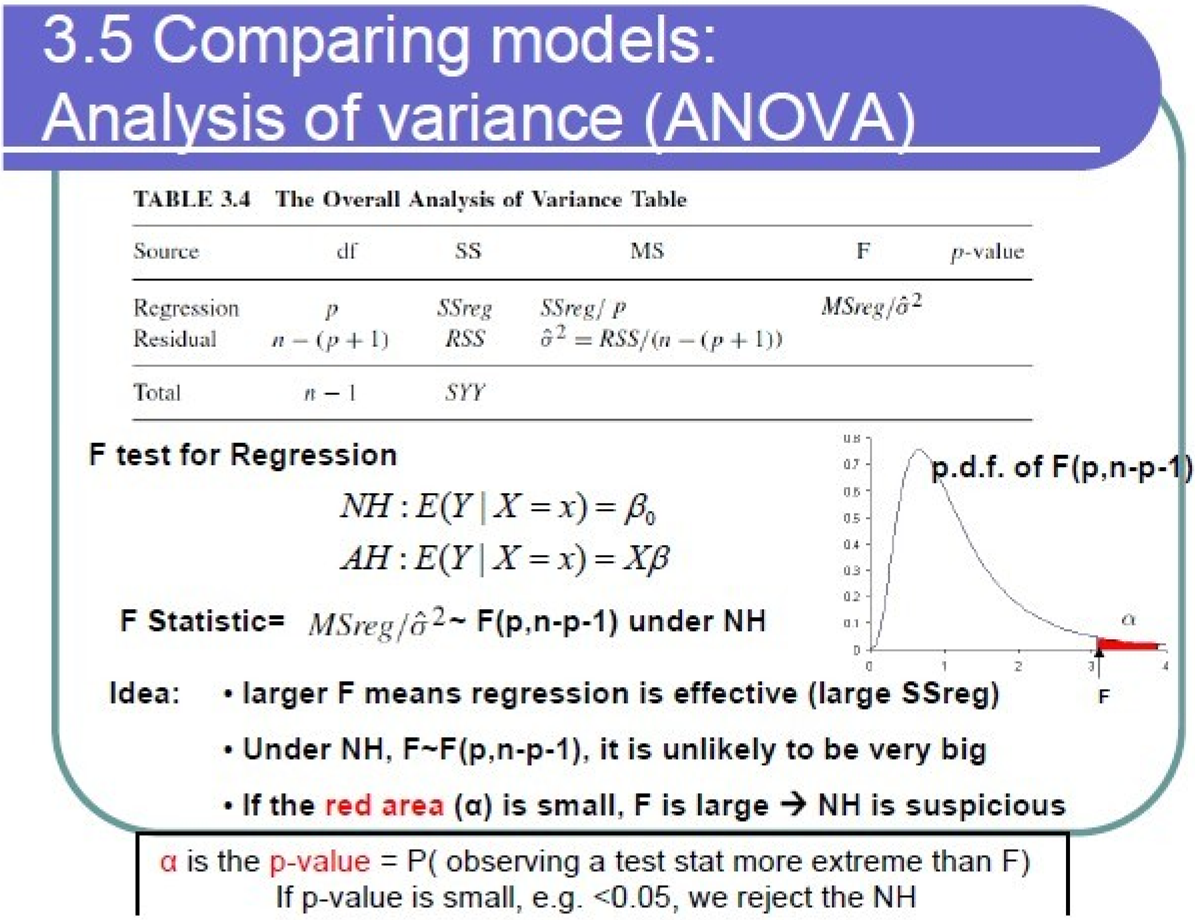
(ii)

$$\begin{aligned}
\hat{e}_{Y|X_1} &= Y - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} X_1 && \text{(2 marks)} \\
\hat{e}_{X_2|X_1} &= X_2 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} X_1 && \text{(2 marks)} \\
slope &= \frac{\sum_{i=1}^{n}(y_i - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x_i)(1 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} x_i)}{\sum_{i=1}^{n}(1 - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} x_i)^2} \\
&= \bar{Y} - \hat{\beta}_1 \bar{X} && \text{(1 mark)}
\end{aligned}$$

3 Parameter testing and Model Comparison

This section is also important in midterm test. Professor may let you conduct some tests or provide the confidence interval for some parameters.

3.1 ANOVA Table



This page is from our lecture notes, which is very important. It is used to test the joint significance of $\beta_1, \cdots, \beta_p$(You may also say the joint significance of $\mathbf{X}_1, \cdots, \mathbf{X}_p$).This is the only way we have learnt to do the joint significance testing.

RSS = $\mathbf{Y}^t(\mathbf{I} - \mathbf{H})\mathbf{Y}$

SSreg = SYY - RSS

♣ An example from 12-13 Midterm

Given $Y = (-2, -1, 6, 9)$, $X_1 = (1, 2, 3, 6)$ and $X_2 = (-2, 0, 0, 2)$.
Fill in the ANOVA Table: (6 marks)

ANOVA Table				
Source	Sum of Squares	d.f.	Mean Square	F-statistics
Regression	_____	_____	_____	_____
Residuals	12	_____	_____	
Total	_____			

$\bar{Y} = 12/4 = 3$, $SYY = 25 + 16 + 9 + 36 = 86$.

ANOVA Table				
Source	Sum of Squares	d.f.	Mean Square	F-statistics
Regression	74(half mark)	2(1 mark)	37(1 mark)	37/12(1 mark)
Residuals	12	1(1 mark)	12(1 mark)	
Total	86(half mark)			

3.2 Testing a Single Parameter

We have two ways to conduct such tests.
1>.T-test of coefficient.

- NH: $\beta_k = 0$
- AH: $\beta_k \neq 0$

T-stat:

$$T = \frac{\hat{\beta}_k}{sd(\hat{\beta}_k)} \sim t(n-p-1) \tag{3.1}$$

In almost all cases we see in this course, we have t(n-p-1) as our test distribution. This is important.

2>.F-test of coefficient:

- NH:$y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \underline{\quad} + \beta_{k+1}x_{k+1} + \cdots + e_i$
- AH:$y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k x_k + \beta_{k+1}x_{k+1} + \cdots + e_i$

F-stat:

$$F = \frac{(RSS_{NH} - RSS_{AH})/(df_{NH} - df_{AH})}{RSS_{AH}/df_{AH}} = \frac{SSreg/1}{\hat{\sigma^2}} \sim F(1, n-p-1) \tag{3.2}$$

The F-test is equivalent to the t-test. You master either one of these two methods will be OK.

P.S. You may also need to review the concept of the Confidence Interval for the prediction and fitted value. You may focus more on the results then on the derivation procedures, as the derivation procedures contain some concepts out of the range of this course.

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-y)^2 \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(y_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(y_i-\bar{y})^2+2\sum_{i=1}^{n}\hat{e}_i(\hat{y}_i-\bar{y}) \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(y_i-\bar{y})^2+2\sum_{i=1}^{n}\hat{e}_i\hat{y}_i-2\bar{y}\sum_{i=1}^{n}\hat{e}_i \\
&= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(y_i-\bar{y})^2 \\
&\quad (\because \sum_{i=1}^{n}\hat{e}_i\hat{y}_i=0 \;\&\; \sum_{i=1}^{n}\hat{e}_i=0)
\end{aligned}
$$