Lecture 7

Hypothesis Testing

Last week we learned

Least squares estimation

Properties of the least squares estimator

Estimating variance of the random error term

Confidence intervals

Multiple Linear Regression

- The matrix form of the MLR model is

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

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The matrix X is called the design matrix.

- The least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = CX^T\boldsymbol{y} \sim N_{p+1}(\boldsymbol{\beta}, \sigma^2 C) \quad \text{where} \quad C = (X^TX)^{-1}$$

- The least squares estimate of σ^2 is

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-p-1} = \frac{\boldsymbol{e}^T\boldsymbol{e}}{n-p-1}$$

where $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ is the vector of residuals and $\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}}$ is the vector of fitted values.

Confidence and Prediction Intervals

- A $100(1-\alpha)\%$ confidence interval for the regression coefficient, β_j, is given by

$$\mathrm{CI}(\beta_j) = \left[\hat{\beta}_j - t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{\beta}_j),\ \hat{\beta}_j + t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{\beta}_j)\right]$$

where $\hat{\beta}_j = (CX^T\boldsymbol{y})_{jj}$ and $\mathrm{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 c_{jj}}$

- A $100(1-\alpha)\%$ CI on the mean response μ_0 at the level $\boldsymbol{x}_0$ is given by

$$\mathrm{CI}(\mu_0) = \left[\hat{\mu}_0 - t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{\mu}_0),\ \hat{\mu}_0 + t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{\mu}_0)\right]$$

where $\hat{\mu}_0 = \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}}$ and $\mathrm{se}(\hat{\mu}_0) = \sqrt{\hat{\sigma}^2\boldsymbol{x}_0^T C\boldsymbol{x}_0}$.

- A $100(1-\alpha)\%$ PI on a new observation y_0 at $\boldsymbol{x}_0$ is

$$\mathrm{PI}(y_0) = \left[\hat{y}_0 - t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{y}_0),\ \hat{y}_0 + t_{\alpha/2,n-p-1} \cdot \mathrm{se}(\hat{y}_0)\right]$$

where $\hat{y}_0 = \boldsymbol{x}_0^T\hat{\boldsymbol{\beta}}$ and $\mathrm{se}(\hat{y}_0) = \sqrt{\hat{\sigma}^2(1+\boldsymbol{x}_0^T C\boldsymbol{x}_0)}$.

Lecture 7

Hypothesis Testing

Aim: to understand hypothesis tests in MLR models

1. Decision making and the hat matrix
2. F-test for significance of regression
3. t-test for individual regression coefficients
4. F-test for a group of predictors
5. Coefficient of multiple determination, R^2

- After fitting a MLR model and computing the parameter estimates, $\hat{\boldsymbol{\beta}}$, we have to make some decisions about the model:

  1. *Is the model a good fit for the data?*
  2. *Do we really need all the predictors in the model?*

- Generally, a model with fewer predictors and about the same "predictive power" is better.

There are several hypothesis tests that we can utilise to answer these questions:

1. An F-test for significance of regression: checks the significance of the whole regression model.
2. A t-test for individual regression coefficients: checks the significance of individual regression coefficients.
3. An F-test for a group of regression coefficients: simultaneously checks the significance of a number of regression coefficients; it can also be used to test individual coefficients.
4. The general linear F-test: checks the significance of a hypothesis of a general linear type applied to all regression coefficients.

We need to introduce an object called the hat matrix before we can proceed with testing hypothesis.

The hat matrix

- The vector of fitted values $\hat{\boldsymbol{y}}$ can be written as

$$\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}} = X(X^TX)^{-1}X^T\boldsymbol{y} = H\boldsymbol{y}$$

  where H is called the hat matrix.

- It is a symmetric $n \times n$ matrix:

$$H^T = (X(X^TX)^{-1}X^T)^T = X(X^TX)^{-1}X^T = H$$

- It is an idempotent:

$$H^2 = (X(X^TX)^{-1}X^T)^2 = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = H$$

- It acts as an identity operator on X:

$$HX = X(X^TX)^{-1}X^TX = X$$

- The trace of H and its rank equal $p+1$:

$$\mathrm{tr}H = \mathrm{tr}(X(X^TX)^{-1}X^T) = \mathrm{tr}((X^TX)^{-1}X^TX) = \mathrm{tr}I_{p+1} = p+1 = \mathrm{rank}H$$

- $H\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^TH = \mathbf{1}^T$, where $\mathbf{1}$ is an n-dimensional vector of 1's

Lecture 7

Hypothesis Testing

1. Decision making and the hat matrix
2. F-test for significance of regression
3. t-test for individual regression coefficients
4. F-test for a group of predictors
5. Coefficient of multiple determination, R^2

F-test for significance of regression

- The test for significance of regression is a test to determine if there is a linear relationship between the response and any of the predictors:

$$Y \overset{?}{\sim} X_1, X_2, \ldots, X_p$$

- This procedure is often thought of as an overall or global test of model adequacy.

- The appropriate hypotheses are

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \beta_j \neq 0$$

- The test procedure is a generalization of the analysis of variance used in the simple linear regression model.

- The total sum of squares, SS_T, is partitioned into the sum of squares due to regression, SS_R, and the residual sum of squares, SS_E, where

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2, \qquad SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2, \qquad SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Claim. The analysis of variance identity holds true,

$$SS_T = SS_R + SS_E$$

Proof. We first rewrite SS_R as

$$\begin{aligned}
SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 &= (\hat{\boldsymbol{y}} - \mathbf{1}\bar{y})^T (\hat{\boldsymbol{y}} - \mathbf{1}\bar{y}) \\
&= \hat{\boldsymbol{y}}^T \hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}^T \mathbf{1}\bar{y} - \bar{y}\mathbf{1}^T \hat{\boldsymbol{y}} + \bar{y}^2 \mathbf{1}^T \mathbf{1} \\
&= (H\boldsymbol{y})^T (H\boldsymbol{y}) - (H\boldsymbol{y})^T \mathbf{1}\bar{y} - \bar{y}\mathbf{1}^T H\boldsymbol{y} + n\bar{y}^2 \\
&= \boldsymbol{y}^T H^T H\boldsymbol{y} - \boldsymbol{y}^T H\mathbf{1}\bar{y} - \bar{y}\mathbf{1}^T H\boldsymbol{y} + n\bar{y}^2 \\
&= \boldsymbol{y}^T H\boldsymbol{y} - \boldsymbol{y}^T \mathbf{1}\bar{y} - \bar{y}\mathbf{1}^T \boldsymbol{y} + n\bar{y}^2 \\
&= \boldsymbol{y}^T H\boldsymbol{y} - n\bar{y}^2
\end{aligned}$$

since $H\mathbf{1} = \mathbf{1}$, $\mathbf{1}^T H = \mathbf{1}^T$ and $\boldsymbol{y}^T \mathbf{1} = \mathbf{1}^T \boldsymbol{y} = \sum_{i=1}^{n} y_i = n\bar{y}$.

Claim. The analysis of variance identity holds true,

$$SS_T = SS_R + SS_E$$

Proof. We have shown that

$$SS_R = \mathbf{y}^T H \mathbf{y} - n\bar{y}^2$$

In a similar way, we rewrite SS_E as

$$SS_E = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - H\mathbf{y})^T(\mathbf{y} - H\mathbf{y}) = \mathbf{y}^T(I-H)^T(I-H)\mathbf{y} = \mathbf{y}^T(I-H)\mathbf{y}.$$

Therefore

$$SS_R + SS_E = \mathbf{y}^T H \mathbf{y} - n\bar{y}^2 + \mathbf{y}^T\mathbf{y} - \mathbf{y}^T H \mathbf{y} = \mathbf{y}^T\mathbf{y} - n\bar{y}^2$$

On the other hand,

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2 = \mathbf{y}^T\mathbf{y} - n\bar{y}^2$$

which is the wanted result.

F-test for significance of regression

- If the null hypothesis H_0 is true, then the sampling distributions of SS_R and SS_E are

$$\frac{SS_R}{\sigma^2} \sim \chi^2_p \qquad \frac{SS_E}{\sigma^2} \sim \chi^2_{n-p-1}$$

- Moreover, they are independent, giving

$$F = \frac{SS_R/p}{SS_E/(n-p-1)} = \frac{MS_R}{MS_E} \sim F_{p,n-p-1}$$

- Analysis of variance table

Source of variation	d.o.f.	SS	MS	F	P-value
Regression	p	SS_R	$MS_R = \frac{SS_R}{p}$	$F = \frac{MS_R}{MS_E}$	α_{cal}
Residual	$n-p-1$	SS_E	$MS_E = \frac{SS_E}{n-p-1}$		
Total	$n-1$	SS_T			

- We reject the null hypothesis H_0 at significance level α if

$$F_{cal} > F_{crit} = F_{\alpha,p,n-p-1} \iff \alpha_{cal} < \alpha_{crit}$$

Lecture 7

Hypothesis Testing

1. Decision making and the hat matrix
2. F-test for significance of regression
3. t-test for individual regression coefficients
4. F-test for a group of predictors
5. Coefficient of multiple determination, R^2

t-test for individual regression coefficients

- Once we have determined that at least one of the predictors is important, a logical question becomes which one(s).
- Adding a variable to a regression model always causes the regression sum of squares, SS_R, to increase, and the residual sum of squares, SS_E, to decrease.
- We must decide whether the increase in SS_R is sufficient to warrant using the additional predictor in the model.
- We must be careful to include only predictors that are of real value in explaining the response.

t-test for individual regression coefficients

- The hypotheses for significance of any individual regression coefficient, such as β_j, are

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0.$$

They test whether β_j is significantly different from zero.

- The test statistic for the null hypothesis is

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj}) \quad \Longrightarrow \quad T = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} = \frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)} \sim t_{n-p-1}$$

where c_{jj} is the jjth matrix element of $C = (X^T X)^{-1}$.

- The null hypothesis $H_0 : \beta_j = 0$ is rejected if the calculated value t_{cal} of T is

$$|t_{cal}| > t_{crit} = t_{\alpha/2, n-p-1} \quad \Longleftrightarrow \quad \alpha_{cal} < \alpha_{crit}$$

t-test for individual regression coefficients

- The t-test is really a partial or marginal test. This means that the test statistic depends not on the j-th predictor only, but also on all other predictors that are included in the model at the same time. This is because

$$\hat{\boldsymbol{\beta}} \sim N_{p+1}(\boldsymbol{\beta}, \sigma^2 C) \quad \Longrightarrow \quad \mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c_{ij}$$

- Thus, if any predictor is added or removed from a regression model, hypothesis tests for individual slopes need to be repeated.

- If the null hypothesis is rejected, we conclude that the j-th predictor has a significant influence on the response, given the other predictors in the model at the same time.

Lecture 7

Hypothesis Testing

1. Decision making and the hat matrix
2. F-test for significance of regression
3. t-test for individual regression coefficients
4. F-test for a group of predictors
5. Coefficient of multiple determination, R^2

F-test for a group of predictors

- Checking significance of regression variables one-by-one may sometimes be not the most effective approach in model building.
- We might want to simultaneously check the significance of a subset of regression coefficients instead.
- The extra sum of squares method allows us to directly determine the significance of a subset of regression coefficients.

F-test for a group of predictors

- Suppose our model has p predictors. We can then partition predictors into two groups,

$$(X_1, \ldots, X_{p-r}) \quad \text{and} \quad (X_{p-r+1}, \ldots, X_p).$$

We want to simultaneously test, whether the latter group of r predictors can be removed from the model.

- Suppose we partition the vector of regression coefficients accordingly into two parts

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

where

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-r} \end{pmatrix} \qquad \boldsymbol{\beta}_1 = \begin{pmatrix} \beta_{p-r+1} \\ \vdots \\ \beta_p \end{pmatrix}$$

- We want to test the hypothesis

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

- The full model may be written as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = X^{red}\boldsymbol{\beta}_1 + X^{extra}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

- The X^{red} is the $n \times (p-r+1)$ reduced design matrix consisting of the columns of X associated with $\boldsymbol{\beta}_1$

$$X^{red} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1\,p-r} \\ 1 & x_{21} & \dots & x_{2\,p-r} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{n\,p-r} \end{bmatrix} \qquad \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-r} \end{bmatrix}$$

- The X^{extra} is the $n \times r$ extra matrix consisting of the columns of X associated with $\boldsymbol{\beta}_2$

$$X^{extra} = \begin{bmatrix} x_{1\,p-r+1} & \dots & x_{1p} \\ x_{2\,p-r+1} & \dots & x_{2p} \\ \vdots & & \vdots \\ x_{n\,p-r+1} & \dots & x_{np} \end{bmatrix} \qquad \boldsymbol{\beta}_2 = \begin{bmatrix} \beta_{p-r+1} \\ \vdots \\ \beta_p \end{bmatrix}$$

F-test for a group of predictors

- The full model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is described by the "full" hat matrix

$$H^{full} = H = X(X^T X)^{-1} X^T$$

and the "full" sums of squares

$$SS_R^{full} = \boldsymbol{y}^T H \boldsymbol{y} - n\bar{y}^2 \qquad \nu_R^{full} = p$$
$$SS_E^{full} = \boldsymbol{y}(I - H)\boldsymbol{y} \qquad \nu_E^{full} = n - p - 1$$

- The reduced model $\boldsymbol{y} = X^{red}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$ is described by the "reduced" hat matrix

$$H^{red} = X^{red}(X^{red\,T} X^{red})^{-1} X^{red\,T}$$

and the "reduced" sums of squares

$$SS_R^{red} = \boldsymbol{y}^T H^{red} \boldsymbol{y} - n\bar{y}^2 \qquad \nu_R^{red} = p - r$$
$$SS_E^{red} = \boldsymbol{y}(I - H^{red})\boldsymbol{y} \qquad \nu_E^{red} = n - p + r - 1$$

- The total sum of squares is the same for both models

$$SS_T^{full} = SS_T^{red} = SS_T = \boldsymbol{y}^T \boldsymbol{y} - n\bar{y}.$$

F-test for a group of predictors

- The regression sum of squares due to $\boldsymbol{\beta}_2$ given that $\boldsymbol{\beta}_1$ is already in the model is

$$\begin{aligned} SS_R^{extra} &= SS_R^{full} - SS_R^{red} \\ &= (\mathbf{y}^T H \mathbf{y} - n\bar{y}^2) - (\mathbf{y}^T H^{red} \mathbf{y} - n\bar{y}^2) \\ &= \mathbf{y}^T (H - H^{red}) \mathbf{y} \end{aligned}$$

  is called the extra sum of squares ($\nu_R^{extra} = r$) due to $\boldsymbol{\beta}_2$.

- It measures the increase in the regression sum of squares that results from adding predictors $X_{p-r+1}, X_{p-r+2}, \dots, X_p$ to a model that already contains $X_1, X_2, \dots, X_{p-r}$.

- The extra sum of squares is sometimes also written as

$$SS_R(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - SS_R(\boldsymbol{\beta}_1)$$

- The Analysis of Variance identity implies that

$$SS_R^{extra} = SS_R^{full} - SS_R^{red} = SS_E^{red} - SS_E^{full}$$

F-test for a group of predictors

- Assuming the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is true, the sampling distributions of SS_R^{extra} and SS_E^{full} are

$$\frac{SS_R^{extra}}{\sigma^2} \sim \chi^2_r \qquad \frac{SS_E^{full}}{\sigma^2} \sim \chi^2_{n-p-1}$$

- Moreover, they are independent, giving

$$F = \frac{SS_R^{extra}/r}{SS_E^{full}/(n-p-1)} = \frac{MS_R^{extra}}{MS_E^{full}} \sim F_{r,n-p-1}$$

- We reject H_0 if $F_{cal} > F_{crit} = F_{\alpha,r,n-p-1}$ concluding that at least one of the parameters in $\boldsymbol{\beta}_2$ is non-zero, and consequently at least one of the predictors $X_{p-r+1},\ldots,X_{p-1},X_p$ contributes significantly to the regression model.

- This test is also called a partial F-test because it measures the contribution of the predictors in X^{extra} given that the other predictors in X^{red} are already in the model.

- ANOVA table for the extra sum of squares analysis:

Source of variation	d.o.f.	SS	MS	F
Residual Reduced	$\nu_E^{red} = n - p + r - 1$	SS_E^{red}		
Residual Full	$\nu_E^{full} = n - p - 1$	SS_E^{full}	$MS_E^{full} = \frac{SS_E^{full}}{\nu_E}$	
Extra	$\nu_R^{extra} = r$	SS_R^{extra}	$MS_R^{extra} = \frac{SS_R^{extra}}{\nu_R^{extra}}$	$F = \frac{MS_R^{extra}}{MS_E^{full}}$

Lecture 7

Hypothesis Testing

1. Decision making and the hat matrix
2. F-test for significance of regression
3. t-test for individual regression coefficients
4. F-test for a group of predictors
5. Coefficient of multiple determination, R^2

Coefficient of multiple determination, R^2

- The coefficient of multiple determination

$$R^2 = 1 - \frac{SS_E}{SS_T} \in [0, 1]$$

indicates the amount of total variability explained by the model.

- The positive square root of R^2 is called the multiple correlation coefficient and measures the linear association between Y and predictors, $X_1, X_2, \ldots, X_p$.

- The value of R^2 increases (or remains the same) as more predictors are added to the model, even if the new predictors do not contribute significantly to the model.

- An increase in the value of R^2 cannot be taken as a sign to conclude that the new model is superior to the older model.

Adjusted coefficient of multiple determination, R^2

- A better statistic to use is the adjusted R^2

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T} = 1 - \frac{SS_E/(n-p-1)}{SS_T/(n-1)}$$

- The adjusted R^2_{adj} only increases when significant terms are added to the model.
- Addition of unimportant predictors may lead to a decrease in the value of R^2_{adj}.
- Removal of unimportant predictors will generally lead to increase in the value of R^2_{adj}.

Summary

- The overall F-test tests if there is a linear relationship between the response Y and any of the predictors $X_1, X_2, \ldots, X_p$. The appropriate hypotheses are

$$H_0 : \text{all slopes } \beta_j = 0 \quad \text{vs.} \quad H_1 : \text{at least one slope } \beta_j \neq 0$$

- The test statistic is

$$F = \frac{SS_R/p}{SS_E/(n-p-1)} = \frac{MS_R}{MS_E} \sim F_{p,n-p-1}$$

where

$$SS_R = \boldsymbol{y}^T H \boldsymbol{y} - n\bar{y}^2 \qquad SS_E = \boldsymbol{y}^T (I-H)\boldsymbol{y} \qquad H = X(X^TX)^{-1}X^T$$

- An individual t-test tests the significance of any individual regression coefficient, including the intercept. The appropriate hypotheses are

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0$$

- The test statistic is

$$T_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \sim t_{n-p-1}$$

where

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-p-1} \qquad c_{jj} = (C)_{jj} = ((X^TX)^{-1})_{jj}$$

Summary

- A partial F-test determines, whether a group of r predictors can be removed from the model. The appropriate hypotheses are

$$H_0 : \beta_{p-r+1} = \ldots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \beta_j \neq 0$$

- The test statistic is

$$F = \frac{SS_R^{extra}/r}{SS_E^{full}/(n-p-1)} = \frac{MS_R^{extra}}{MS_E^{full}} \sim F_{r,n-p-1}$$

  where SS_R^{extra} is the extra sum of squares

$$SS_R^{extra} = SS_R^{full} - SS_R^{reduced} = SS_E^{reduced} - SS_E^{full}$$

- The coefficient of multiple determination

$$R^2 = 1 - \frac{SS_E}{SS_T} \in [0,1]$$

- The adjusted coefficient of multiple determination

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T} = 1 - \frac{SS_E/(n-p-1)}{SS_T/(n-1)}$$

Next week

Diagnostics and Model Building