STATS 2107
Statistical Modelling and Inference II
Lecture notes
Chapter 3: Linear models part 1

Jono Tuke

School of Mathematical Sciences, University of Adelaide

Semester 2 2017

Recap of Simple Linear Regression (SLR)

Setup

Consider data of form

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

The **linear regression** model is

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

with

$$\varepsilon_i \sim N(0, \sigma^2)$$

independently for $i = 1, 2, \ldots, n$.

Least squares estimation

The least squares estimates of β_0 and β_1 are the values that jointly minimise

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

The least squares estimates of β_0 and β_1 are denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively.

Theorem

The least squares estimates for β_0 and β_1 are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \text{ and } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}},$$

where

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \text{ and } S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Estimation of σ^2

To estimate σ^2, we use the residual variance

$$s_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

Theorem

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with

$$E[Y_i] = \beta_0 + \beta_1 x_i \text{ and } var(Y_i) = \sigma^2,$$

then

$$E[\hat{\beta}_0] = \beta_0 \text{ and } E[\hat{\beta}_1] = \beta_1,$$

$$var[\hat{\beta}_0] = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \text{ and } var[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}},$$

$$cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}\sigma^2}{S_{xx}}, \text{ and}$$

$$E[S_e^2] = \sigma^2.$$

Proof (outline)

Theorem (cont.)

If $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ independently for $i = 1, 2, \ldots, n$, then

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)\right),$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right), \text{ and}$$

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2_{n-2}.$$

Prediction

Consider the regression model:

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

independently for $i = 1, 2, \ldots, n$.

How do we predict for an additional independent random variable:

$$Y_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$$

Theorem

Suppose $Y_1, Y_2, \ldots, Y_N$ are independent with

$$E[Y_i] = \beta_0 + \beta_1 x_i \text{ and } var(Y_i) = \sigma^2,$$

then

$$E[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \beta_0 + \beta_1 x_0,$$

$$var[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right).$$

Theorem

If

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2),$$

then

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \sim N\left(\beta_0 + \beta_1 x_0, \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\right)$$

$100(1-\alpha)\%$ confidence interval for $E[\beta_0 + \beta_1 x_0]$

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2} S_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}.$$

$100(1-\alpha)\%$ prediction interval for Y_0

$$\hat{\beta}_0 + \hat{\beta}_1 x_i \pm t_{\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}.$$

Residuals

The **residuals** are defined as

$$\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \;\; i = 1, 2, \ldots, n.$$

Properties of residuals

$$\sum_{i=1}^{n} \hat{e}_i = 0,$$

$$\sum_{i=1}^{n} \hat{e}_i x_i = 0,$$

$$E[\hat{E}_i] = 0, \text{ and}$$

$$var(\hat{E}_i) = \sigma^2 \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}} \right).$$

Standardized residuals

$$\tilde{e}_i = \frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{\sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}}$$

Studentized residuals

$$e_i^* = \frac{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{s_e \sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}}$$

Multiple linear regression

Setup

Consider data

$$
\begin{gathered}
(y_1, x_{11}, x_{12}, \ldots, x_{1r}) \\
(y_2, x_{21}, x_{22}, \ldots, x_{2r}) \\
\vdots \\
(y_n, x_{n1}, x_{n2}, \ldots, x_{nr})
\end{gathered}
$$

So we have n subjects with r predictors.

MLR model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_r x_{ir} + \varepsilon_i,$$

where

$$\varepsilon_i \sim i.i.d. N(0, \sigma^2),$$

for $i = 1, 2, \ldots, n$.

Matrix formulation

Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},\ X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1r} \\ 1 & x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nr} \end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix} \text{ and } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

The multiple regression model can then be formulated as

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Definition

A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$ is said to be **linearly independent** if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \ldots + \alpha_p \boldsymbol{v}_p = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0.$$

Otherwise it is said to be **linearly dependent**.

Linear independence and X

The columns of X in

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

must be linearly independent.

Why?

Least squares estimation of β

Lemma

If $X_{n \times p}$ is a matrix with linearly independent columns then the symmetric, $p \times p$ matrix $X^T X$ is invertible.

Proof

Theorem

If the columns of X are linearly independent then the least squares estimates of β are given uniquely by

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

Proof

Estimation of σ^2

The residual variance is

$$S_e^2 = \frac{1}{n-p} \| \boldsymbol{Y} - X\hat{\boldsymbol{\beta}} \|^2,$$

where $p = r + 1$, *i.e.* the number of β's.

Least Square Estimates (LSE)

Lemma

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} \text{ and } \boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

and let $V = \boldsymbol{a}^T \boldsymbol{Y}$. Then

$$E(V) = \boldsymbol{a}^T \boldsymbol{\eta},$$

$$var(V) = \sigma^2 \boldsymbol{a}^T \boldsymbol{a}$$

If, furthermore, $Y_i \sim N(\eta_i, \sigma^2)$ independently, then

$$V \sim N(\boldsymbol{a}^T \boldsymbol{\eta}, \sigma^2 \boldsymbol{a}^T \boldsymbol{a})$$

Theorem

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$, where

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} = X\boldsymbol{\beta}$$

where X is an $n \times p$ matrix with linearly independent columns and let $\boldsymbol{\lambda}$ be a constant vector, then,

$$E(\boldsymbol{\lambda}^T\hat{\beta}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$$

$$var(\boldsymbol{\lambda}^T\hat{\beta}) = \sigma^2\boldsymbol{\lambda}^T(X^TX)^{-1}\boldsymbol{\lambda}$$

$$E(S_e^2) = \sigma^2$$

If, furthermore, $Y_i \sim N(\eta_i, \sigma^2)$, then

$$\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\lambda}^T \boldsymbol{\beta}, \sigma^2 \boldsymbol{\lambda}^T (X^T X)^{-1} \boldsymbol{\lambda}) \quad \text{and} \quad \frac{(n-p)S_e^2}{\sigma^2} \sim \chi^2_{n-p}$$

independently.

Proof

Inference

It follows that

$$\frac{\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}^T\boldsymbol{\beta}}{S_e\sqrt{\boldsymbol{\lambda}^T(X^TX)^{-1}\boldsymbol{\lambda}}} \sim t_{n-p}$$

Confidence interval

A $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is given by

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} \pm t_{n-p}(\alpha/2)s_e\sqrt{\boldsymbol{\lambda}^T(X^TX)^{-1}\boldsymbol{\lambda}}.$$

Hypothesis test

To test $H_0: \boldsymbol{\lambda}^T \boldsymbol{\beta} = \delta_0$ at the α level of significance, calculate

$$t = \frac{\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - \delta_0}{s_e \sqrt{\boldsymbol{\lambda}^T (X^T X)^{-1} \boldsymbol{\lambda}}}$$

and reject H_0 if

$$|t| \geq t_{n-p}(\alpha/2)$$

P-value

The P-value is given by

$$\text{P-value} = P(|T| \geq |t|)$$

where t is the observed value of the test statistic and $T \sim t_{n-p}$.

BLUE for Multiple linear regression

Best linear unbiased estimation (Gauss-Markov theorem)

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent observations with $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}$$

and suppose $\boldsymbol{\eta} = X\beta$, where X is an $n \times p$ matrix whose columns are linearly independent.

If $\boldsymbol{a}^T \boldsymbol{Y}$ is an unbiased linear estimator for $\boldsymbol{\lambda}^T \beta$ then

$$var(\boldsymbol{a}^T \boldsymbol{Y}) \geq var(\boldsymbol{\lambda}^T \hat{\beta})$$

with equality if and only if

$$\boldsymbol{a} = X(X^T X)^{-1} \boldsymbol{\lambda}.$$

Hypothesis testing for several parameters

Setup

Suppose now that we wish to test a hypothesis of the form

$$H_0 : \beta_p = \beta_{p-1} = \ldots = \beta_{p-k+1} = 0.$$

That is, the last k components of the parameter vector $\boldsymbol{\beta}$ are all zero.

Let X_0 be the matrix containing the first $p - k$ columns of X and let

$$\boldsymbol{\beta}_0 = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-k} \end{pmatrix}.$$

Observe that H_0 can be expressed equivalently as

$$H_0: \ \boldsymbol{\eta} = X_0 \boldsymbol{\beta}_0.$$

Now let

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (X^T X)^{-1} X^T \boldsymbol{y} \\
\hat{\boldsymbol{\eta}} &= X \hat{\boldsymbol{\beta}} \\
\hat{\boldsymbol{\beta}}_0 &= (X_0^T X_0)^{-1} X_0^T \boldsymbol{y} \\
\hat{\boldsymbol{\eta}}_0 &= X_0 \hat{\boldsymbol{\beta}}_0
\end{aligned}$$

Lemma

$$\sum_{i=1}^{n}(y_i - \hat{\eta}_{0i})^2 = \sum_{i=1}^{n}(y_i - \hat{\eta}_i)^2 + \sum_{i=1}^{n}(\hat{\eta}_i - \hat{\eta}_{0i})^2.$$

That is,

$$\|\boldsymbol{y} - X_0\hat{\boldsymbol{\beta}}_0\|^2 = \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2 + \|X\hat{\boldsymbol{\beta}} - X_0\hat{\boldsymbol{\beta}}_0\|^2.$$

Proof

Expected values

If H_0 is true, then

$$E\left(\frac{1}{n-p_0}\|\boldsymbol{y}-X_0\hat{\boldsymbol{\beta}}_0\|^2\right)=\sigma^2,$$

where $p_0 = p - k$.

If H_0 is true, then so is the full regression model $\boldsymbol{\eta} = X\boldsymbol{\beta}$, and so

$$E\left(\frac{1}{n-p}\|\boldsymbol{y}-X\hat{\boldsymbol{\beta}}\|^2\right)=\sigma^2,$$

Hence what is

$$E\left(\frac{1}{p-p_0}\|X\hat{\boldsymbol{\beta}}-X_0\hat{\boldsymbol{\beta}}_0\|^2\right)?$$

Null not correct

If the full model is correct, but the null is not, then it can be shown that

$$E\left(\frac{1}{p-p_0}\|X\hat{\beta}-X_0\hat{\beta}_0\|^2\right) > \sigma^2$$

Test statistic

Hence we can test H_0 by calculating

$$F = \frac{\|X\hat{\beta} - X_0\hat{\beta}_0\|^2/(p-p_0)}{\|\boldsymbol{y} - X\hat{\beta}\|^2/(n-p)}$$

and rejecting if F is 'large'.

Definition

Suppose $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ independently and let

$$W = \frac{X_1/k_1}{X_2/k_2}.$$

Then W is said to follow the F-distribution with k_1, k_2 degrees of freedom and we write $W \sim F_{k_1,k_2}$.

Theorem

Suppose $\boldsymbol{Y} = X\beta + \boldsymbol{\epsilon}$ with $\epsilon_i \sim N(0, \sigma^2)$ independently for $i = 1, 2, \ldots, n$. If $H_0: \ \boldsymbol{\eta} = X_0\boldsymbol{\beta}_0$ is true, then

$$F = \frac{\|X\hat{\beta} - X_0\hat{\boldsymbol{\beta}}_0\|^2/(p - p_0)}{\|\boldsymbol{y} - X\hat{\beta}\|^2/(n - p)} \sim F_{p-p_0, n-p}.$$

Anova table

Source	SS	df	MS	F
H_0 vs M	$Q_0 - Q$	$p - p_0$	$(Q_0 - Q)/(p - p_0)(*)$	$F = \frac{*}{\dagger}$
Error	Q	$n - p$	$Q/(n-p)(\dagger)$	
Total	Q_0	$n - p_0$		

where

$$Q = \|\boldsymbol{y} - X\hat{\beta}\|^2 \text{ and } Q_0 = \|\boldsymbol{y} - X_0\hat{\beta}_0\|^2.$$

and

$$H_0 : \boldsymbol{\eta} = X_0\beta_0$$
$$M : \boldsymbol{\eta} = X\beta$$