NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (14)

Multiple linear regression, properties of estimators [3.4-3.6]

Mette Langaas

To be lectured: February 18, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

Multiple linear regression model

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \text{ for } i = 1, ..., n$$

Matrix formulation:

$$\underset{(n\times 1)}{\boldsymbol{Y}} = \underset{(n\times p)}{\boldsymbol{X}}\underset{(p\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

$$E(\boldsymbol{\varepsilon}) = \underset{(n\times 1)}{\mathbf{0}} \quad \text{and} \quad Cov(\boldsymbol{\varepsilon}) = \underset{(n\times 1)}{\sigma^2 \boldsymbol{I}}$$

where

— $\boldsymbol{\beta}$ and σ^2 are unknown parameters and

— the design matrix $\boldsymbol{X}$ has ith row $[x_{i1}, x_{i2}, ..., x_{ip}]$.

Properties of LS-estimator

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

$$E(\hat{\beta}) = \beta \text{ and } \mathrm{Cov}(\hat{\beta}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

And, the LS estimator is BLUE.

ML estimation of σ

The MLE for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \frac{1}{n}(\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) = \frac{1}{n}\text{SSE}$$

where

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

$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$$

$$\text{SSE} = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$$

Plan MLR

1. E of SSE and the trace formula.
2. Normality. $\hat{\beta}$ and SSE properties, *t*-test.
3. Confidence interval for the regression plane and prediction interval for new response.
4. SST=SSR+SSE, notation and distribution
5. F and partial F tests.
6. R^2 and R^2_{adj}.

TMA4267 : lecture 14 (out of planned 26 with "new" material)

"HALF WAY" TODAY!

Aim now: E(SSE)

The trace formula [p78]

Let $\underset{n\times 1}{Y}$ be a random vector with $E(Y)=\mu$ and $Cov(Y)=\Sigma$,

and $\underset{n\times n}{A}$ a constant matrix. Then

$$E(Y^T A Y) = tr(A\Sigma) + \mu^T A \mu$$

(look kind of like a generalization of $E(X^2) = Var(X) + \mu^2$)

Proof:

$$E(Y^T A Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} E(Y_i \cdot Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \left(Cov(Y_i, Y_j) + E(Y_i)\cdot E(Y_j) \right)$$

$$Cov(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)\cdot E(Y_j)$$

$$E(Y^T A Y) = \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\Sigma_{ij}} + \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\mu_i\mu_j}_{\mu^T A \mu} = tr(A\Sigma) + \mu^T A \mu$$

$$tr(A\Sigma) = \sum_{i=1}^{n} \underbrace{(A\Sigma)_{ii}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\Sigma_{ij}$$

$$(A\Sigma)_{ii} = \sum_{j=1}^{n} a_{ij}(\Sigma_{ji})$$

$\Sigma_{ji} = \Sigma_{ij}$ symmetric Σ

A closer look at SSE

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (Y - \hat{Y})^T (Y - \hat{Y})$$

$$Y - \hat{Y} = Y - HY = (I - H)Y$$

$$SSE = Y^T \underbrace{(I-H)^T (I-H)}_{(I-H)} Y = Y^T (I-H) Y$$

$$E(SSE) = E\left(Y^T (I-H) Y \right) = \underline{tr\left((I-H)\sigma^2 I \right) + \beta^T X^T (I-H) X\beta}$$

Need $E(Y) = X\beta \leftarrow \mu$

$Cov(Y) = Cov(\varepsilon) = \sigma^2 I \leftarrow \Sigma$

$A = (I - H)$

$$\begin{cases} E(\varepsilon) = 0, \; Cov(\varepsilon) = \sigma^2 I \\ Y = X\beta + \varepsilon \end{cases}$$

$HX = X$

$$E(SSE) = \sigma^2 \underbrace{tr(I-H)} + \underbrace{\beta^T X^T X\beta - \beta^T X^T \overbrace{H X}\beta}_{0}$$

(equal)

$$tr(I - H) = tr(\underset{n\times n}{I}) - tr(\underset{n\times n}{H})$$

$$= n - tr\left(X (X^T X)^{-1} X^T \right) \quad \leftarrow tr(BC) = tr(CB)$$

$$= n - tr\left(\underbrace{X^T X \, (X^T X)^{-1}}_{\underset{p\times p}{I}} \right)$$

$$= n - p$$

$$\underline{\underline{E(SSE) = (n-p) \cdot \sigma^2}}$$

New unbiased estimator for σ^2

$$S^2 = \frac{1}{n-p} SSE$$

MLR - not assuming normality of errors

Assumptions:

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

$$\mathrm{E}(\varepsilon) = \mathbf{0} \text{ and } \mathrm{Cov}(\varepsilon) = \sigma^2 \boldsymbol{I} \text{ and } \varepsilon_i \text{ are independent } i = 1, \ldots, n$$

Unbiased estimators:

$$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \text{ and } S^2 = \frac{1}{n-p}\text{SSE}$$

And, we found that $\mathrm{Cov}(\hat{\beta}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.
For inference we need to assume a distribution for the errors, to then give us distributions for the estimators (and other statistics). Then we may construct hypothesis tests, confidence intervals and prediction intervals.

Assume normality

$$Y = X\beta + \varepsilon \quad \text{where } \varepsilon \sim N_n(0, \sigma^2 I)$$

$$\Rightarrow Y \sim N_n(X\beta, \underbrace{\sigma^2 I}_{\Sigma})$$

1) Homework: $\hat{Y} = HY \sim N_n(X\beta, H\sigma^2)$

$e = (I-H)Y \sim N_n(0, (I-H)\sigma^2)$

and $\hat{Y}$ and e are independent.

2) $\hat{\beta}$, SSE

a) $\hat{\beta} = (X^T X)^{-1} X^T Y = GY$ is a linear comb. of mvN Y's, so

$\hat{\beta}$ is mvN. $\hat{\beta} \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$

b) SSE

First theorem 3.26

If $\underset{n\times n}{A}$ is a symmetric end idempotent matrix of rank r

and $\underset{n\times 1}{Z} \sim N_n(0, \sigma^2 I)$. Then

$$Z^T A Z \sim \sigma^2 \cdot \chi^2_r$$

Because:

$$Z^T A Z = \underbrace{Z^T P}_{Z^{*T}} \Lambda P^T Z = Z^{*T} \Lambda Z^*$$

$A = P \Lambda P^T$ (P: eigen vector, Λ: eigenvalue)

$$\Lambda = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \quad \begin{matrix} r \\ n-r \end{matrix}$$

$$= Z_1^{*2} + Z_2^{*2} + \dots + Z_r^{*2} + 0 \cdot Z_{r+1}^{2} + \dots + 0 \cdot Z_n^{*2}$$

$$Z^* = P^T Z \sim N_n(0, P^T \sigma^2 I P) = N_n(0, \sigma^2 I)$$

$Z \sim N_n(0, \sigma^2 I)$

$P = [e_1 \; e_2 \; \dots \; e_n]$

$e_1^T e_2 = 0$

$e_1^T e_1 = 1$

thus $Z_i^{*2} \sim \chi^2_1$ and independent

How can this be used when $SSE = Y^T (I - H) Y$?

$SSE = Y^T (I-H) Y \quad (= Y^T (I-H)^T (I-H) Y)$

To use Th. 3.26 we need $Z \sim N_n(0, \sigma^2 I)$ but we have

$Y \sim N_n(X\beta, \sigma^2 I)$. What if we use $Y^* = (Y - X\beta) \sim N_n(0, \sigma^2 I)$

$$(I-H)Y^* = (I-H)(Y-X\beta) = (I-H)Y - (I-H)X\beta$$

$$= (I-H)Y - \underbrace{[X\beta - \underbrace{HX}_{X}\beta]}_{0} = (I-H)Y$$

$SSE = Y^{*T} (I-H) Y^*$ where $Y^* \sim N_n(0, \sigma^2 I)$

Using T3.26 we only need to find rank$(I-H)$.

E4P4 show that rank$(A) = tr(A)$ for A symmetric and idempotent.

$\Rightarrow$ rank$(I-H) = tr(I-H) = n-p$

Finally: $\dfrac{SSE}{\sigma^2} \sim \chi^2_{n-p}$

Properties of symmetric projection matrices

A projection $\boldsymbol{A}$ matrice is idempotent, $\boldsymbol{A}^2 = \boldsymbol{A}$. A symmetric projection matrix is orthogonal.

1. The eigenvalues of a projection matrix are 0 and 1.
2. The rank of a symmetric matrix (actually: a diagonalizable quadratic matrix) equals the number of nonero eigenvaluse of the matrix. Should be known from previous courses.
3. (Combining 1+2). If a $(n \times n)$ symmetric projection matrix $\boldsymbol{A}$ has rank r then r eigenvalues are 1 and $n - r$ are 0.
4. The trace and rank of a symmetric projection matrix are equal: $tr(\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$.

c) SSE and $\hat{\beta}$:

$Cov(Y) = \Sigma$

Let $\underline{A}Y = \underbrace{(X^TX)^{-1}X^T}Y = \hat{\beta}$

$\underline{B}Y = \underbrace{(I-H)}Y$ or $(I-H)(Y-X\beta)$

We know from Ch4 that if $A\Sigma B = 0$ then AY and BY are independent, and with $\Sigma = \sigma^2 I$ then $AB = 0$ implies independence.

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

$$= (X^TX)^{-1}X^T - \underbrace{(X^TX)^{-1}X^TX}_{I}(X^TX)^{-1}X^T = 0$$

Thus: $\underbrace{AY}_{\hat{\beta}}$ and $\underbrace{BY}_{(I-H)Y}$ are independent, and $\underbrace{AY}_{\hat{\beta}}$ and $\underbrace{Y^TB^TBY}_{SSE}$ must also be independent $\Rightarrow$ $\hat{\beta}$ and SSE are independent.

d) $\hat{\beta} \sim N_p(\beta, \underline{\sigma^2 (X^T X)^{-1}})$

and

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-p}$$

$S^2 = \frac{1}{n-p} SSE$

$$\frac{(n-p) S^2}{\sigma^2} \sim \chi^2_{n-p}$$

and independent.

$\hat{\beta}_i$ is ith element of $\hat{\beta}$

C_{ii}^{-1} is element (i,i) of $(X^T X)^{-1}$

Then (TMA4245)

$$\frac{\hat{\beta}_i - \beta}{\sqrt{C_{ii}^{-1}} \cdot S} \sim t_{n-p}$$

can be used to perform inference.

MLR - assuming normality of errors

Assumptions:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$
$$\varepsilon = N_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

Findings

1. $\hat{\boldsymbol{Y}}$ and $\boldsymbol{e}$. Both normally distributed, and independent of eachother.
2. $\hat{\beta}$ and SSE. $\hat{\beta}$ normally distributed, SSE/σ^2 chi-squared distributed, and independent of eachother.