Proof (we will check (i) & (ii))

1. We have $\underline{H}\underline{v} = \underline{X}\{(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{v}\} \in V$ and (i) follows.

2. Let $\underline{u} \in V$.

3. Because $\underline{X}$ is a basis matrix for V, we have $\underline{u} = \underline{X}\,\underline{c}$ ~~for~~, where $\underline{c} \in \mathbb{R}^{p+1}$.

4. Then

$$(\underline{v} - \underline{H}\underline{v})'\underline{u} = \underline{v}'(\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{X}\,\underline{c} = \underline{v}'(\underline{X} - \underline{X})\underline{c} = 0$$

5. (ii) is satisfied.

~~Conclusion~~ Corollary 1

The projection matrix $\underline{H}$ is idempotent (i.e. $\underline{H}^2 = \underline{H}$) and symmetric. All eigenvalues of $\underline{H}$ are either 0 or 1 and the rank of $\underline{H}$ is equal to its trace. The matrix $\underline{I} - \underline{H}$ is the projection matrix onto $V^{\perp}$.

Theorem 3 Consider the model $\underline{Y} = \eta + \varepsilon$, for $\eta \in V$. Let $\underline{H}$ be the projection matrix onto V. Let $\hat{\eta} = \underline{H}\underline{Y} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$.
Then $\hat{\eta}$ is the LS solution.

Proof

1. Let $\eta \in V$.

2. Then $\underline{H}\underline{Y} - \eta \in V$.

3. But $(\underline{I} - \underline{H})\underline{Y} \in V^{\perp}$.

4. Therefore,

$$\|\underline{Y} - \eta\|^2 = \|\underline{Y} - \underline{H}\underline{Y} + \underline{H}\underline{Y} - \eta\|^2 = \|(\underline{I} - \underline{H})\underline{Y} + (\underline{H}\underline{Y} - \eta)\|^2 = \underbrace{\|(\underline{I}-\underline{H})\underline{Y}\|^2}_{\text{does not depend on } \eta} + \|\underline{H}\underline{Y} - \eta\|^2$$

5. We minimize the left-side by taking $\eta = \underline{H}\underline{Y}$.

6. Hence, the LS solution is the projection $\underline{H}\underline{Y}$.

7. Uniqueness - exercise. [~~ex 16~~]

As a result, the LS estimate $\hat{\underline{b}}$ of $\underline{b}$ must satisfy

$$X\hat{b} = \cancel{X}\underline{H}\underline{Y} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \qquad /\cdot X'$$

$$X'X\hat{b} = X'Y \qquad \text{- estimating equation (normal equation) for the multiple regression model.}$$

$$(X'X)^{-1}/ \qquad \hat{b} = (X'X)^{-1}X'Y.$$

The estimate $\hat{Y} = X\hat{b}$ is called the fitted or predicted value of $\underline{Y}$. 55

6.05.14

The residual or estimate of the error vector is given by $\hat{\varepsilon} = \underline{Y} - \hat{\underline{Y}}$.

Since $\hat{Y} \in V$ and $\hat{\varepsilon} \in V^{\perp}$, we have $\hat{\underline{Y}} \perp \hat{\varepsilon}$.

Theorem 4

Consider the model $\underline{Y} = \underline{X}\,\underline{b} + \varepsilon$ and assume that $\varepsilon_1, .., \varepsilon_n$ are iid. and that $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2 < \infty$. Then

a) $E(\hat{b}) = \underline{b}$ and $Cov(\hat{b}) = \sigma^2 (X'X)^{-1}$.

b) $E(\hat{Y}) = \underline{X}\,\underline{b}$ and $Cov(\hat{Y}) = \sigma^2 H$

c) ~~$E(\hat{\varepsilon}) = \underline{0}$ and $Cov(\hat{\varepsilon}) = \sigma^2(I-H)$~~ $\hat{\varepsilon}$ ok

d) $E(\hat{\sigma}^2) = \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$

Proof

1. We have

$$\hat{b} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(Xb + \varepsilon) = b + (X'X)^{-1}X'\varepsilon$$

$$\hat{Y} = X\hat{b} = Xb + H\varepsilon$$

$$\hat{\varepsilon} = Y - \hat{Y} = Xb + \varepsilon - Xb - H\varepsilon = (I-H)\varepsilon.$$

2. Since $E(\varepsilon) = 0$ and $Cov(\varepsilon) = \sigma^2 I$, Then

$$E\hat{b} = b, \quad Cov(\hat{b}) = (X'X)^{-1}X'\sigma^2 I \cdot X \cdot (X'X)^{-1} = \sigma^2 (X'X)^{-1},$$

$$E\hat{Y} = Xb, \quad Cov(\hat{Y}) = H\,\sigma^2 I\,H' = \sigma^2 H \qquad (H' = H \text{ symetry})$$

correct ? but obvious ?

~~$E\hat{\eta} = EHY = EH(\hat{\varepsilon} + \hat{Y}) = H(I-H)E\varepsilon + E\hat{Y}$ $HE(Xb+\varepsilon) = HXb = X(X'X)^{-1}X'Xb = Xb$~~

~~$Cov(\hat{\eta}) = Cov(HY) = H\,Cov Y\,H' = H(Cov(Y - \hat{Y} + \hat{Y}))H' = H(Cov\,\hat{\varepsilon} + Cov\,\hat{Y})H' =$~~

~~$H(I-H)\sigma^2 I (I-H)'H' + H\sigma^2 HH' =$~~

~~$Cov(\hat{\eta}) = Cov(HY) = H\,Cov(Y)\,H' = H\sigma^2 I H = H\sigma^2$~~ $H\sigma^2$

$E\hat{\varepsilon} = 0, \; Cov(\hat{\varepsilon}) = \sigma^2(I-H)$

3. We have

$$(n-p-1)\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{\varepsilon}^T\hat{\varepsilon} = \varepsilon^T(I-H)(I-H)\varepsilon = \varepsilon^T(I-H)\varepsilon$$

4. Hence

$$E[(n-p-1)\hat{\sigma}^2] = E[\varepsilon^T(I-H)\varepsilon] = \cancel{E[tr(\varepsilon^T(I-H)\varepsilon)] = E[tr((I-H)\varepsilon\varepsilon^T)]} = \text{ok}$$

$$tr[(I-H)E[\varepsilon\varepsilon^T]] = tr[(I-H)\sigma^2 I] = tr[(I-H)\sigma^2] = (n-p-1)\sigma^2.$$

$X \sim \mu, \Sigma$

$E[X'AX] = tr(A\Sigma) + \mu'A\mu = tr((I-H)\sigma^2) = rank(\;)\sigma^2$ ($A = I-H$)

4.2 Basics of LS Inference under Normal Errors

6.05.14

Theorem 1

Consider the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and assume that ε has a $N_n(0, \sigma^2 I)$ distribution. Then the LS estimators satisfy the following:

(a) $\hat{b}$ has a $N(\underline{b}, \sigma^2 (X'X)^{-1})$ distribution.

(b) $\hat{Y}$ has a $N(\underline{X}b, \sigma^2 H)$ distribution.

(c) $\hat{\varepsilon}$ has a $N(0, \sigma^2 (I - H))$ — " — .

(d) $(n-p-1)\hat{\sigma}^2/\sigma^2$ has a $\chi^2(n-p-1)$ distribution.

(e) $\hat{Y}$ and $\hat{\varepsilon}$ are independent.

(f) $\hat{b}$ and $\hat{\sigma}^2$ are independent.

Proof

(a, b, c) obvious.

d) We have $\dfrac{(n-p-1)\hat{\sigma}^2}{\sigma^2} = \sigma^{-2} \varepsilon'(I - H)\varepsilon$ and $I - H$ is idempotent of rank $n-p-1$.

e) Since

$$\begin{bmatrix} \hat{Y} \\ \hat{\varepsilon} \end{bmatrix} = \begin{bmatrix} H \\ I - H \end{bmatrix} \varepsilon + \begin{bmatrix} X b \\ 0 \end{bmatrix}$$

Then $\hat{Y}$ and $\hat{\varepsilon}$ have a jointly normala distribution, while their covariance matrix has a form

$$\begin{bmatrix} H \\ 1 - H \end{bmatrix} \sigma^2 I \begin{bmatrix} H \\ 1 - H \end{bmatrix}' = \sigma^2 \begin{bmatrix} H & 0 \\ 0 & I - H \end{bmatrix}.$$

f) Since $\hat{Y} = X\hat{b}$ and $\hat{\sigma}^2 = \frac{1}{(n-p-1)} \hat{\varepsilon}^T \hat{\varepsilon}$, e) entails f).

652

Corollary 1

Consider the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and assume that ε has a $N_n(0, \sigma^2 I)$ distribution. Then the random variables

$$T_j = \frac{\hat{\beta}_j - \beta}{\hat{\sigma}\sqrt{(X_c' X_c)^{-1}_{jj}}}, \qquad j = 1, \dots, p,$$

where $(X_c' X_c)^{-1}_{jj}$ is the jth diagonal entry of $(X_c' X_c)^{-1}$ and X_c is the centered design matrix, have t-distributions with $n-p-1$ degrees of freedom.

Therefore, a level α test for the hypotheses

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_1 : \beta_j \neq 0 \qquad j = 1, \dots, p$$

is given by

reject H_0 if $|T_j| = \dfrac{|\hat{\beta}_j|}{\hat{\sigma}\sqrt{(X_c' X_c)^{-1}_{jj}}} > t_{1-\alpha/2, n-p-1}$, where $t_{1-\alpha/2, n-p-1}$ is the $(1-\alpha/2)$-quantile of the t-distribution

6. Tests of General Linear Hypotheses

Consider the model

$$\underline{Y} = \underline{X}\,\underline{b} + \underline{\varepsilon},$$

where $\underline{X}$ is an $n\times(p+1)$ design matrix, $\underline{b} = (\alpha, \underline{\beta}')'$. The above model we will call the full model.

We test a general linear hypothesis

$$H_0: \underline{A}\,\underline{b} = \underline{0},$$

against the alternative

$$H_1: \underline{A}\,\underline{b} \neq 0,$$

where $\underline{A}$ is a $q\times(p+1)$ specified matrix of full row rank $q < p+1$.

So, the rows of $\underline{A}$ provide the linear constraints.

Example 1

1) Suppose we are predicting Y based on a second degree polynomial model in x_1 and x_2 i.e,

$$E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2. \quad (*)$$

Suppose our null hypothesis is that the first-order terms suffice to predict Y.

The corresponding matrix $\underline{A}$ is

$$\underline{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

because, under H_0, $E(Y) = \alpha + \beta_1 x_1 + \beta_2 x_2$

2) Suppose for the model $(*)$, we think the slope parameters of x_1 and x_2 are the same. Then the null hypothesis can be expressed with the matrix

$$\underline{A} = [\,0 \;\; 1 \;\; -1 \;\; 0 \;\; 0 \;\; 0\,].$$

Let V_F, (where F stands for the full model), denote the column space of $\underline{X}$.

For the hypothesis H_0, the reduced model is the full model subject to H_0, i.e, the subspace given by

$$V_R = \{\underline{v} \in V_F : \underline{v} = \underline{X}\,\underline{b} \text{ and } \underline{A}\underline{b} = 0\}.$$

We will show (later) that $\dim(V_r) = (p+1) - q$.

Suppose we have a norm $\|\cdot\|$ for fitting models.

Let

$$\hat{\eta}_F = \underset{\underline{\eta} \in V_F}{\arg\min} \|\underline{Y} - \underline{\eta}\|.$$

Then, the distance between $\underline{Y}$ and the subspace V_F is

$$d_F = d(\underline{Y}, V_F) = \| Y - \hat{\eta}_F \| .$$

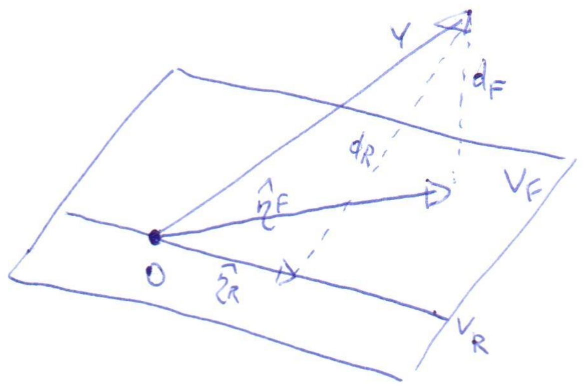
Let

$$\hat{\eta}_R = \arg\min_{\eta \in V_R} \| Y - \eta \|$$

and let

$d_R = d(\underline{Y}, V_R)$ denote the distance between $\underline{Y}$ and the subspace V_R.

We have $d(\underline{Y}, V_R) \geq d(\underline{Y}, V_F)$ (minimum over a larger set can be smaller)



An intuitive test statistic has a form

$$RD_{\|\cdot\|} = d(\underline{Y}, V_R) - d(\underline{Y}, V_F).$$

Small values of $RD_{\|\cdot\|}$ indicate ~~that~~ H_0 while large values indicate that H_1 is true.
Therefore, we will reject H_0 in favour of H_1 if $RD_{\|\cdot\|} \geq c$.
We will find c.

Assume that $\|\cdot\|$ is the Euclidean norm. Let $\underline{H}_F$ and $\underline{H}_R$ denote the projection matrices onto the subspaces V_F and V_R, respectively. Then

$$d^2_{LS}(\underline{Y}, V_F) = \| \underline{Y} - \underline{H}_F \underline{Y} \|^2_{LS} = \underline{Y}'(\underline{I} - \underline{H}_F)\underline{Y}$$

$$d^2_{LS}(\underline{Y}, V_R) = \| \underline{Y} - \underline{H}_R \underline{Y} \|^2_{LS} = \underline{Y}'(\underline{I} - \underline{H}_R)\underline{Y}.$$

Therefore

$$RD_{LS} = d^2_{LS}(\underline{Y}, V_R) - d^2_{LS}(\underline{Y}, V_F) = \underline{Y}'(\underline{H}_F - \underline{H}_R)\underline{Y}.$$

6.1 Distribution Theory for the LS Test for Normal Errors

Assume that $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$.

Definition 1
Let V_1 and V_2 be two subspaces of $\mathbb{R}^n$ and assume that $V_1 \subset V_2$.
Then the space V_2 mod V_1 ~~is def~~ has a form

$$V_2|V_1 = \{\underline{v} \in V_2 : \underline{v} \perp \underline{w}, \text{ for all } \underline{w} \in V_1\}.$$

Lemma 1
The matrix $\underline{H}_F - \underline{H}_R$ is the projection matrix onto the space $V_F|V_R$.

Proof
1. Let $\underline{U}_R$ be an o.n. basis matrix for V_R ~~and let~~
2. Let $[\underline{U}_R : \underline{U}_2]$ be an extention of $\underline{U}_R$ to an o.n. basis matrix of V_F.
3. Then $\underline{U}_2$ is a basis matrix for $V_F|V_R$ and $\underline{U}_2\underline{U}_2'$ is the projection matrix onto $V_F|V_R$.
   $\underline{U}_2(\underbrace{\underline{U}_2'\underline{U}}_{I_{p+1-q}})^{-1}\underline{U}_2'$
4. Also, $\underline{H}_R = \underline{U}_R\underline{U}_R'$ and

$$\underline{H}_F = [\underline{U}_R : \underline{U}_2][\underline{U}_R : \underline{U}_2]' = \underline{U}_R\underline{U}_R' + \underline{U}_2\underline{U}_2' = \underline{H}_R + \underline{U}_2\underline{U}_2'. \quad \square$$

Lemma 2
Let $\underset{n\times q}{\underline{C}} = \underline{X}(\underline{X}'\underline{X})^{-1}\underline{A}'$. Then $\underline{C}$ is a basis matrix for $V_F|V_R$. Further, the dimension of $V_F|V_R$ is q and V_R is $p+1-q$.

Proof (Ex)

$$C'C = A(X'X)^{-1}\cancel{X'X(X'X)^{-1}}A'$$

As a result $\underline{H}_F - \underline{H}_R = C(C'C)^{-1}C'$. Then

upisać

$$RD_{LS} = \underline{Y}'(\underline{H}_F - \underline{H}_R)\underline{Y} = Y'X(X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A\underbrace{(X'X)^{-1}X'Y}_{\hat{b}_{LS}}$$

$$= (A\hat{b}_{LS})'[A(X'X)^{-1}A']^{-1}A\hat{b}_{LS}.$$

So, the standardised thest statistic has a form

$$F_{LS} = \frac{\frac{1}{q}(A\hat{b}_{LS})'[A(X'X)^{-1}A']^{-1}A\hat{b}_{LS}}{\hat{\sigma}^2}, \qquad LRT$$

where $\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}\hat{e}_i^2$.

Theorem 1 Under the model $\underline{Y} = \underline{X}\underline{b} + \underline{\varepsilon}$ and the assumption that $\underline{\varepsilon}$ ~~has~~ has a $N_n(\underline{0}, \sigma^2 I)$ distribution, the statistic F_{LS} has an F-distribution with q and $n-p-1$ degrees of freedom and noncentrality parameter $\theta = (Ab)'[A(X'X)^{-1}A']^{-1}Ab/\sigma^2$.

Proof

1. ~~We have~~ $(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1)$.
2. $\hat{\sigma}^2$ is independent of $\hat{\underline{b}}_{LS}$
3. Hence numerator and denominator of F_{LS} are independent.
4. $\hat{\underline{b}}_{LS} \sim N_{\cancel{q}\,p+1}(\cancel{A}\underline{b}, \sigma^2(\underline{X}'X)^{-1})$
5. Hence $\underline{A}\hat{\underline{b}}_{LS} \sim N_q(\underline{A}\underline{b}, \sigma^2\underline{A}(\underline{X}'\underline{X})^{-1}A')$

and ~~numer~~ numerator $\times \frac{q}{\sigma^2} \sim \chi^2(q, \theta)$ $\square$.

Corollary 1

We reject H_0 in favour of H_1 if $F_{LS} \geq F_{1-\alpha, q, n-p-1}$

" now you can " find c ~~as~~ α