8 Hypothesis testing and Confidence Intervals

8.1 Hypothesis testing: General Hypothesis

Let $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\sigma}^2\mathbf{I})$, Then

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \qquad \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\sigma}^2)$$

$\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\beta_1} \\ \underset{\sim}{\beta_2} \end{pmatrix}_{(k+1)*1}$ $\underset{\sim}{\beta_1}$ — $(p+1)*1$

$\Rightarrow \dim(\underset{\sim}{\beta_2}) = k-p$

$\underset{\sim}{K}'$ — $q*(k+1)$

$\underset{\sim}{\beta}$ — $(k+1)*1$

Goal We are now interested in testing the following hypothesis:

$$H_0 : \mathbf{K}'\boldsymbol{\beta} = \boldsymbol{\mu} \tag{8.1}$$

where $\mathbf{K}'$ is $q \times (k+1)$, and $\mathbf{K}'$ is assumed to be full row rank.

$\operatorname{rank}(\underset{\sim}{K}) = q$

$H_0: \underset{\sim}{\beta_2} = 0 \Leftrightarrow \underset{\sim}{K}\underset{\sim}{\beta} = 0,\ \underset{\sim}{\mu} = 0$

$\underset{\sim}{K}' = (0 \ \ I_{k-p})$ $(k-p)*(k+1)$

$(k-p)*(k+1)$

$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} \sim N(\underset{\sim}{\beta}, \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1})$

Note that

1. We wish $\underset{\sim}{K}'\hat{\underset{\sim}{\beta}} = \underset{\sim}{\mu}$

$$\Rightarrow \underbrace{\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}}_{\text{越小越好}} \sim N[\mathbf{K}'\boldsymbol{\beta} - \boldsymbol{\mu}, \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}\boldsymbol{\sigma}^2]$$

σ^2未知 $\Rightarrow \hat{\sigma}^2$代替

2. $(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}$ is symmetric.

3. Let

$$Q = (\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})$$

(Q is sometimes denoted by SSH) then

$$\frac{Q}{\sigma^2} \sim \chi^2_{\{q, \frac{1}{2\sigma^2}(\mathbf{K}'\boldsymbol{\beta}-\boldsymbol{\mu})'[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}(\mathbf{K}'\boldsymbol{\beta}-\boldsymbol{\mu})\}}$$

$\hat{\sigma}^2 = \frac{SSE}{n-k-1}$

4. Q and SSE are independent

$SSE = (\underset{\sim}{y} - \hat{\underset{\sim}{y}})'(\underset{\sim}{y} - \hat{\underset{\sim}{y}}) = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{y}$, where $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$

Note that SSE and $\hat{\underset{\sim}{\beta}}$ are indep, $\Rightarrow$ SSE and Q are indep.
(Q is function of $\hat{\underset{\sim}{\beta}}$)

The test statistics

σ² 本身是未知的

$$
\begin{aligned}
F(H) &= \frac{Q/q}{SSE/[N-r(\mathbf{X})]} \\
&= \frac{Q}{q\hat{\sigma}^2} \\
&\sim F_{[q,N-r(\mathbf{X}),\underbrace{\frac{1}{2\sigma^2}(\mathbf{K}'\boldsymbol{\beta}-\boldsymbol{\mu})'[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}(\mathbf{K}'\boldsymbol{\beta}-\boldsymbol{\mu})}_{=\lambda}]}
\end{aligned}
$$

Here $\hat{\sigma}^2 = SSE/(N - r(\mathbf{X}))$ which is the unbiased estimator of σ^2.

Under $H_0 : \mathbf{K}'\mathbf{a} = \boldsymbol{\mu}$, we have

$$F(\mathbf{H}) \sim F_{\{q,\underbrace{N-r(\mathbf{X})}_{N-k-1}\}}.$$

Under H_0, $\lambda=0$

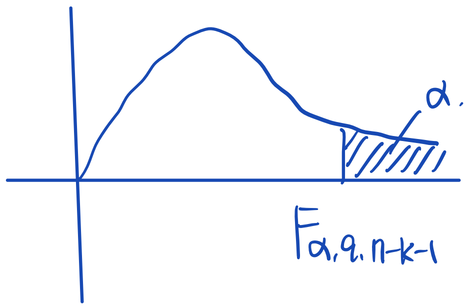
$$\boxed{F(H) = \frac{Q/q}{SSE/n-k-1} \overset{H_0}{\sim} F_{q,n-k-1}}$$

希望

Reject H_0 if $F(H) \geq F_{\alpha,q,n-k-1}$ (one-sided)

p-value $= P(F_{q,n-k-1} \geq F_H)$

rejection region:



Estimation under the constraint.

model: $\begin{cases} \underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{a} + \underset{\sim}{\varepsilon}, \ \underset{\sim}{\varepsilon} \sim N(0, \sigma^2 \underset{\sim}{I}) \\ \underset{\sim}{k}'\underset{\sim}{a} = \mu \end{cases}$

Note Estimation of $\mathbf{a}$ under the null hypothesis H_0 : $\boxed{\mathbf{K}'\mathbf{a} = \boldsymbol{\mu}}$ or under the constraint. Denote the LS estimator of $\mathbf{a}$ by $\tilde{\mathbf{a}}$. To obtain the least squares estimator of $\mathbf{a}$, need to minimize

$$\min_{\underset{\sim}{a}} (\mathbf{y} - \mathbf{X}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{a}) + 2\theta'(\mathbf{K}'\mathbf{a} - \boldsymbol{\mu})$$

with respect to $\mathbf{a}$ and θ. Note that 2θ is a vector of Lagrange multipliers. After the minimization,

$$\begin{aligned} \tilde{\mathbf{a}} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{K}(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \boldsymbol{\mu})) \\ &= \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}) \end{aligned} \tag{8.2}$$

- $\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})$; $\hat{\beta} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$ 无限制的情况
- $\tilde{\mathbf{a}}$ is the BLUE. (指在所有 under constrains 的 unbiased estimation 中最优)

Proof: Let $\begin{cases} \text{满足 } \underset{\sim}{k}'\underset{\sim}{a} = \mu \text{ (same with Gauss-Markov)} \\ \text{方差最小} \end{cases}$

$$\mathbf{L} = (\mathbf{y} - \mathbf{X}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{a}).$$

$\forall \mathbf{a}_0$ with constraint $\mathbf{K}'\mathbf{a}_0 = \boldsymbol{\mu}$

$$\begin{aligned} &\mathbf{L}_0 \\ =\ & (\mathbf{y} - \mathbf{X}\mathbf{a}_0)'(\mathbf{y} - \mathbf{X}\mathbf{a}_0) \\ =\ & (\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}} + \mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0)'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}} + \mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0) \\ =\ & (\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}}) + (\mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0)'(\mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0) + 2\underbrace{(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}})'(\mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0)}_{\shortparallel\ 0} \end{aligned}$$

But

$$\begin{aligned} &(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}})'(\mathbf{X}\tilde{\mathbf{a}} - \mathbf{X}\mathbf{a}_0) \\ =\ & (\mathbf{y}'\mathbf{X} - \tilde{\mathbf{a}}'\mathbf{X}'\mathbf{X})(\tilde{\mathbf{a}} - \mathbf{a}_0) \\ =\ & [\mathbf{y}'\mathbf{X} - \hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X}) + (\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})](\tilde{\mathbf{a}} - \mathbf{a}_0) \\ =\ & (\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'(\tilde{\mathbf{a}} - \mathbf{a}_0) \\ =\ & 0 \end{aligned}$$

Because $(\mathbf{K}'\tilde{\mathbf{a}} = \mathbf{K}'\mathbf{a}_0 = \boldsymbol{\mu})$, therefore,

$$\mathbf{L}_0 = (\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}}) + (\tilde{\mathbf{a}} - \mathbf{a}_0)'\mathbf{X}'\mathbf{X}(\tilde{\mathbf{a}} - \mathbf{a}_0)$$

Hence,

$$\mathbf{a}_0 = \tilde{\mathbf{a}} \quad \text{minimize} \quad \mathbf{L}_0$$

大空间中最优解好于子空间

Without the null hypothesis,

$$SSE = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Under the null hypothesis, the sum of squares of residual is (Reduced Model)

$$\begin{aligned}
SSE_{H_0} &= (\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}})'(\mathbf{y} - \mathbf{X}\tilde{\mathbf{a}}) \\
&= [\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\tilde{\mathbf{a}}]'[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\tilde{\mathbf{a}}] \\
&= [\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}})]'[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}})] \\
&= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}})
\end{aligned}$$

Since $((\hat{\boldsymbol{\beta}} - \tilde{\mathbf{a}})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0)$

From (8.2),

$$\begin{aligned}
SSE_{H_0} &= SSE + (\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad \mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}]^{-1}(\mathbf{K}\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}) \\
&= SSE + (\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu})'(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}) \\
&= SSE + Q \\
&\geq SSE
\end{aligned}$$

$H_0: \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{\mu}$

Special cases

1. $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0} \quad \Rightarrow \quad \mathbf{K}' = \mathbf{I}, q = k+1, \boldsymbol{\mu} = \boldsymbol{\beta_0}$

$K'\beta = \mu$

(a)

$$Q = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

(b) $F(H) = \frac{Q}{(k+1)\hat{\sigma}^2}$

(c) Under the null hypothesis,

$$F(H) \sim F_{\{k+1,n-(k+1)\}}$$

(d) $\tilde{\mathbf{a}} = \hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \underbrace{\cancel{\mathbf{a}_0}}_{\beta_0}) = \boldsymbol{\beta_0}$ under constraint: $\tilde{a} = \underset{\sim}{\beta_0}$

2. $H_0 : \boldsymbol{\lambda}'\boldsymbol{\beta} = m \quad \Rightarrow \quad \mathbf{K}' = \boldsymbol{\lambda}', q = 1, \boldsymbol{\mu} = m$

(a)

$$\begin{aligned} Q &= (\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - m)'[\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}]^{-1}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - m) \\ &= (\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - m)^2/\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda} \end{aligned}$$

(b) $F(H) = \frac{Q}{\hat{\sigma}^2}$

(c) Under the null hypothesis,

$$F(H) \sim F_{(1, n-r(\mathbf{X}))}$$

Note: $\sqrt{F(H)} = \frac{\sqrt{Q}}{\hat{\sigma}} \sim t_{n-r(\mathbf{X})}$

(d)

$$\begin{aligned} \tilde{\mathbf{a}} &= \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}[\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}]^{-1}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu}) \\ &= \hat{\boldsymbol{\beta}} - \frac{(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \overset{m}{\cancel{\boldsymbol{\mu}}})}{\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda} \end{aligned}$$

Note: $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\mu} \sim N(\boldsymbol{\lambda}'\boldsymbol{\beta} - \boldsymbol{\mu}, \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}\sigma^2)$

3. $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ where $\underset{k+1}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$ $\begin{matrix} k+1-h \\ h\text{-dimensional} \end{matrix}$

Remark 8.1

$\underset{h*(k+1)}{\underset{\sim}{K}'} = (\underset{(k+1-h)}{\underset{\sim}{0}} \,\vdots\, I_h)$

$H_0: \underset{\sim}{K}'\underset{\sim}{\beta} = 0.$

Remark 8.1: $\underset{\sim}{\beta} = \begin{pmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{pmatrix}$, $\underset{\sim}{X} = (\overset{\beta_1}{\underset{\sim}{X}_1} \;\; \overset{\beta_2}{\underset{\sim}{X}_2})$

$$\underset{\sim}{Q} = (\underset{\sim}{K}'\hat{\underset{\sim}{\beta}})' [\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{K}]^{-1}(\underset{\sim}{K}'\hat{\underset{\sim}{\beta}})$$

$$(\underset{\sim}{X}'\underset{\sim}{X}) = \begin{pmatrix} \underset{\sim}{X}_1'\underset{\sim}{X}_1 & \underset{\sim}{X}_1'\underset{\sim}{X}_2 \\ \underset{\sim}{X}_2'\underset{\sim}{X}_1 & \underset{\sim}{X}_2'\underset{\sim}{X}_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}_{21} & \underset{\sim}{A}_{22} \end{pmatrix}$$

$$(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \begin{pmatrix} A_{11}^{-1} + \underset{\sim}{A}_{11}^{-1}\underset{\sim}{A}_{12}\underset{\sim}{B}^{-1}A_{21}\underset{\sim}{A}_{11}^{-1} & -\underset{\sim}{A}_{11}^{-1}\underset{\sim}{A}_{12}\underset{\sim}{B}^{-1} \\ -\underset{\sim}{B}^{-1}A & \underset{\sim}{B}^{-1} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \underset{\sim}{B}_{11} & \underset{\sim}{B}_{12} \\ \underset{\sim}{B}_{21} & \underset{\sim}{B}_{22} \end{pmatrix}$$

$$\underset{\sim}{B} = \underset{\sim}{X}_2'\underset{\sim}{X}_2 - \underset{\sim}{X}_2'\underset{\sim}{X}_1(\qquad = \underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2$$

$$[\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{K}]^{-1} = \underset{\sim}{B} = \underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2$$

$$\begin{pmatrix} \hat{\underset{\sim}{\beta}}_1 \\ \hat{\underset{\sim}{\beta}}_2 \end{pmatrix} = \hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} = \begin{pmatrix} \underset{\sim}{B}_{11} & \underset{\sim}{B}_{12} \\ \underset{\sim}{B}_{21} & \underset{\sim}{B}_{22} \end{pmatrix} \begin{pmatrix} \underset{\sim}{X}_1'\underset{\sim}{y} \\ \underset{\sim}{X}_2'\underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{B}_{11}\underset{\sim}{X}_1'\underset{\sim}{y} + \underset{\sim}{B}_{12}\underset{\sim}{X}_2'\underset{\sim}{y} \\ \underset{\sim}{B}_{21}\underset{\sim}{X}_1'\underset{\sim}{y} + \underset{\sim}{B}_{22}\underset{\sim}{X}_2'\underset{\sim}{y} \end{pmatrix}$$

$$\hat{\underset{\sim}{\beta}}_2 = \underset{\sim}{B}_{21}\underset{\sim}{X}_1'\underset{\sim}{y} + \underset{\sim}{B}_{22}\underset{\sim}{X}_2'\underset{\sim}{y} = \underset{\sim}{B}^{-1}\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{y}$$

$$Q = \hat{\underset{\sim}{\beta}}_2'\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2\hat{\underset{\sim}{\beta}}_2 = \cdots = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2\underset{\sim}{B}^{-1}\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{y}$$

$$\frac{\underset{\sim}{Q}}{\sigma^2} \sim \chi^2\left(h, \tfrac{1}{2\sigma^2}(\underset{\sim}{\beta}_2\underset{\sim}{X}_2'(\underset{\sim}{I} - \underset{\sim}{H}_1)\underset{\sim}{X}_2\underset{\sim}{\beta}_2)\right), \quad F = \frac{Q}{h\hat{\sigma}^2} \overset{H_0: \beta_2 = 0}{\sim} F(h, n-k-1)$$

Likelihood Ratio Test

Theorem: If $\mathbf{y}$ is $N_n(\mathbf{xa}, \sigma^2\mathbf{I})$, where (rank of $\mathbf{x}$ is $k+1$), the likelihood ratio for H_0: $\mathbf{a} = \mathbf{0}$ can be based on

$$F = \frac{\hat{\mathbf{a}}'\mathbf{x}'\mathbf{y}/(k+1)}{(\mathbf{y}'\mathbf{y} - \hat{\mathbf{a}}'\mathbf{x}'\mathbf{y})/(n-k-1)}.$$

H_0 is rejected if $F > F_{\alpha,k+1,n-k-1}$.

8.2 Confidence intervals and prediction intervals

OUTLINE

1. Confidence region for $\boldsymbol{\beta}$
2. Confidence interval for β_j
3. Confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$
4. Confidence interval for $E(y^*)$ given $\boldsymbol{x} = \boldsymbol{x}^*$
5. Prediction interval for a future observation
6. Confidence interval for σ^2
7. Simultaneous intervals
   (a) Familywise confidence level
   (b) Bonferroni procedure
   (c) Scheffé procedure

Model: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

$\operatorname{rank}(X) = k+1$

Confidence region for $\boldsymbol{\beta}$

Since $P[\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{(k+1)\hat{\sigma}^2} \leq F_{\alpha,k+1,n-k-1}] = 1-\alpha$

a $100(1-\alpha)\%$ joint confidence region for $\beta_0, \beta_1, ..., \beta_k$ is defined to consist of all vectors $\boldsymbol{\beta}$ that satisfy

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \leq (k+1)\hat{\sigma}^2 F_{\alpha,k+1,n-k-1}$$

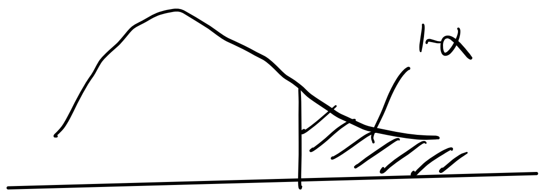
$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(X'X)^{-1})$

$\Rightarrow (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' \frac{(X'X)}{\sigma^2} (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \chi^2_{k+1}$

$\Rightarrow \frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(X'X)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{(k+1)\hat{\sigma}^2} \sim F(k+1, n-k-1)$

where $\hat{\sigma}^2 = \frac{SSE}{n-k-1}$, $SSE \perp\!\!\!\perp \hat{\boldsymbol{\beta}}$

$\Rightarrow P\left(\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{(k+1)\hat{\sigma}^2} \leq F_{\alpha,k+1,n-k-1}\right)$



Confidence interval for β_j

Since $P[-t_{\alpha/2,n-k-1} \le \frac{\hat{\beta}_j - \beta_j}{S\{\hat{\beta}_j\}} \le t_{\alpha/2,n-k-1}] = 1 - \alpha$

hence, a $100(1-\alpha)\%$ confidence interval for β_j is

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} S\{\hat{\beta}_j\}$$

$\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma^2)$

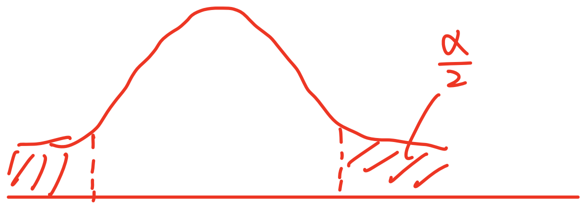
$\hat{\beta}_j \sim N(\beta_j, w_{jj}\sigma^2)$. w_{jj} – jth diagonal element of $(X'X)^{-1}$

$\Rightarrow \hat{\beta}_j - \beta_j / (w_{jj}\sigma^2)^{\frac{1}{2}} \sim N(0,1)$

$\Rightarrow \frac{\hat{\beta}_j - \beta_j}{g_{jj}^{\frac{1}{2}}\hat{\sigma}} \sim t_{n-k-1}$, $\hat{\sigma}^2 = \frac{SSE}{n-k-1}$

$\Rightarrow$ C.I.

$P\left(\left| \frac{\hat{\beta}_j - \beta_j}{w_{jj}^{\frac{1}{2}}\hat{\sigma}} \right| \le t_{\frac{\alpha}{2},n-k-1} \right) = 1-\alpha$



$-t_{\frac{\alpha}{2},n-k-1}$ $t_{\frac{\alpha}{2},n-k-1}$

Confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$

Note that

$$t = \frac{\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\hat{\sigma}\sqrt{\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\lambda}}} \sim t_{n-k-1},$$

hence, a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{\boldsymbol{\lambda}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\lambda}}$$

$\hat{\underset{\sim}{\beta}} \sim N(\underset{\sim}{\beta}, (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sigma^2)$

$\underset{\sim}{\lambda}'\hat{\underset{\sim}{\beta}} \sim N(\underset{\sim}{\lambda}'\underset{\sim}{\beta}, \underset{\sim}{\lambda}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{\lambda}\sigma^2)$

Confidence interval for $E(y^*)$ given $\boldsymbol{x} = \boldsymbol{x}^*$

$y^* = \underset{\sim}{x}^{*'}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}^*$

Given that $\boldsymbol{x} = \boldsymbol{x}^*$

$E(y^*) = \underset{\sim}{x}^{*}\underset{\sim}{\beta}$

$$E(y^*) = \boldsymbol{x}^{*'}\boldsymbol{\beta}$$

$$\widehat{E(y^*)} = \boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}}$$

$$Var(E(y^*) - \widehat{E(y^*)}) = [\boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*]\sigma^2$$

hence, a $100(1-\alpha)\%$ confidence interval for $E(y^*)$ is

$$\boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{\boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*}$$

Prediction interval for a future observation

Given that $\boldsymbol{x} = \boldsymbol{x}^*$. Let $y^* = \boldsymbol{x}^{'}\boldsymbol{\beta} + \epsilon_0$ be a future value of y when $\boldsymbol{x} = \boldsymbol{x}^*$ that needs to be predicted.

$$\hat{y}^* = \boldsymbol{x}^{*'}\hat{\boldsymbol{\beta}}$$

$$Var(y^* - \hat{y}^*) = [1 + \boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*]\sigma^2$$

hence, a $100(1-\alpha)\%$ prediction interval for y^* is

$$\hat{y}^* \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{1 + \boldsymbol{x}^{*'}(\boldsymbol{X'X})^{-1}\boldsymbol{x}^*}$$

$y^* - \hat{y}^* \sim N(0, \sigma^2 + \underset{\sim}{x}^{*'}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{x}^{*}\sigma^2)$

Confidence interval for σ^2

Note that $(n-k-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(n-k-1)}$. Therefore,

$$P[\chi^2_{1-\alpha/2,n-k-1} \leq \frac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-k-1}] = 1-\alpha$$

hence, a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{(n-k-1)\hat{\sigma}^2}{\chi^2_{\alpha/2,n-k-1}}, \frac{(n-k-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n-k-1}}\right)$$

Simultaneous intervals

1. Familywise confidence level: $1-\alpha_f$ implies that we are $100(1-\alpha_f)\%$ confident that every interval contains its respective parameter.

2. Bonferroni confidence intervals

   - Individual confidence level $1-\alpha_c$
   - m intervals
   - If we choose $\alpha_c = \alpha_f/m$, familywise confidence level $\geq 1-\alpha_f$
   - For m linear functions $\boldsymbol{\lambda_1'\beta}, \boldsymbol{\lambda_2'\beta}, ..., \boldsymbol{\lambda_m'\beta}$, the $100(1-\alpha)\%$ Bonferroni confidence intervals are

   $$\boldsymbol{\lambda_i'}\hat{\boldsymbol{\beta}} \pm t_{\alpha/2m,n-k-1}\hat{\sigma}\sqrt{\boldsymbol{\lambda_i'(X'X)^{-1}\lambda_i}}$$

   for $i = 1, ..., m$.

3. Scheffé confidence intervals for all possible linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$:

   The $100(1-\alpha)\%$ conservative confidence interval for any and all $\boldsymbol{a}'\boldsymbol{\beta}$ is

   $$\boldsymbol{\lambda'}\hat{\boldsymbol{\beta}} \pm \hat{\sigma}\sqrt{(k+1)F_{\alpha,k+1,n-k-1}\boldsymbol{\lambda'(X'X)^{-1}\lambda}}$$

Example Ch8: Multiple regression in R (Testing) 肺活量

R Package: GLMsData   Dataset: lungcap

A study of 654 youths in East Boston investigate the relationships between lung capacity (measured by forced expiratory volume in litres (FEV)) and several factors.

$n = 654$

Dependent variable: FEV — y

Independent variables:

- (Smoke status: 1 = smokers; 0 = non-smokers) factor
- age
- height
- gender — M, F

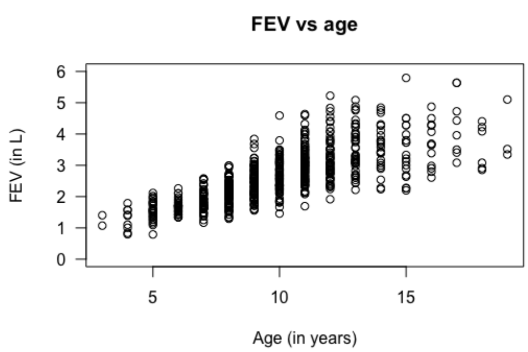
```
> library(GLMsData)
> data(lungcap)
> head(lungcap)
```

```
> head(lungcap)    #Show the first few lines of data
  Age   FEV Ht Gender Smoke
1   3 1.072 46      F     0
2   4 0.839 48      F     0
3   4 1.102 48      F     0
4   4 1.389 48      F     0
5   4 1.577 49      F     0
6   4 1.418 49      F     0
```

```
> summary(lungcap)
      Age              FEV               Ht          Gender        Smoke
 Min.   : 3.000   Min.   :0.791   Min.   :46.00   F:318    Min.   :0.00000
 1st Qu.: 8.000   1st Qu.:1.981   1st Qu.:57.00   M:336    1st Qu.:0.00000
 Median :10.000   Median :2.547   Median :61.50            Median :0.00000
 Mean   : 9.931   Mean   :2.637   Mean   :61.14            Mean   :0.09939
 3rd Qu.:12.000   3rd Qu.:3.119   3rd Qu.:65.50            3rd Qu.:0.00000
 Max.   :19.000   Max.   :5.793   Max.   :74.00            Max.   :1.00000
```

<u>Plotting the data</u>

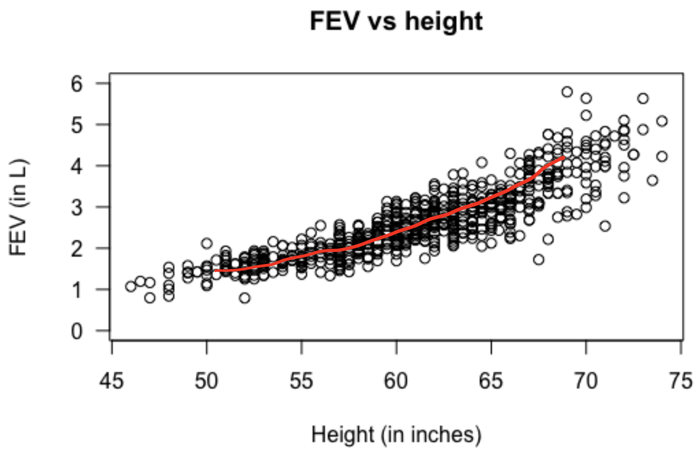
```
> plot(FEV ~ Age, data=lungcap, xlab="Age (in years)", ylab="FEV (in L)", main
= "FEV vs age", xlim=c(0,20), ylim=c(0,6), las=1)
```



基本上是线性关系.

和年龄关系不那么强烈

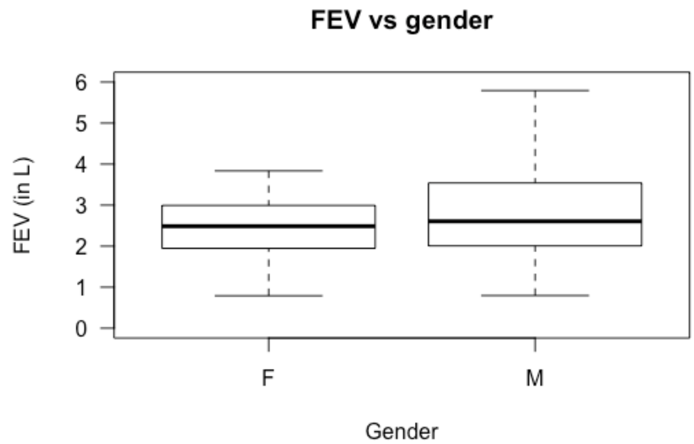
```
> plot(FEV ~ Ht, data=lungcap, main="FEV vs height", xlab="Height (in
inches)", ylab="FEV (in L)", las=1, ylim=c(0,6))
```



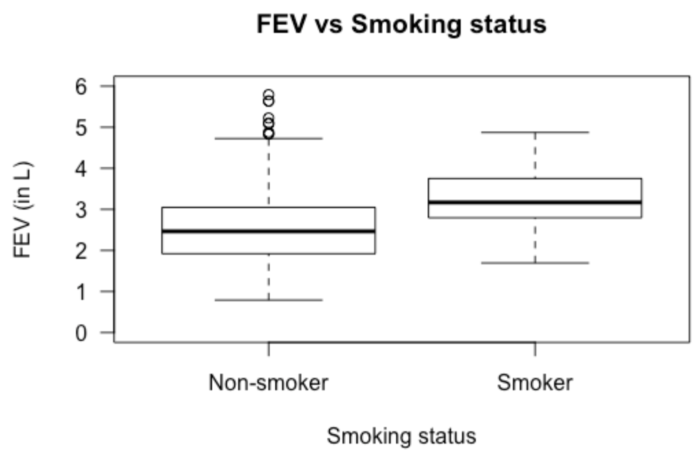
并不完全是线性关系

boxplot

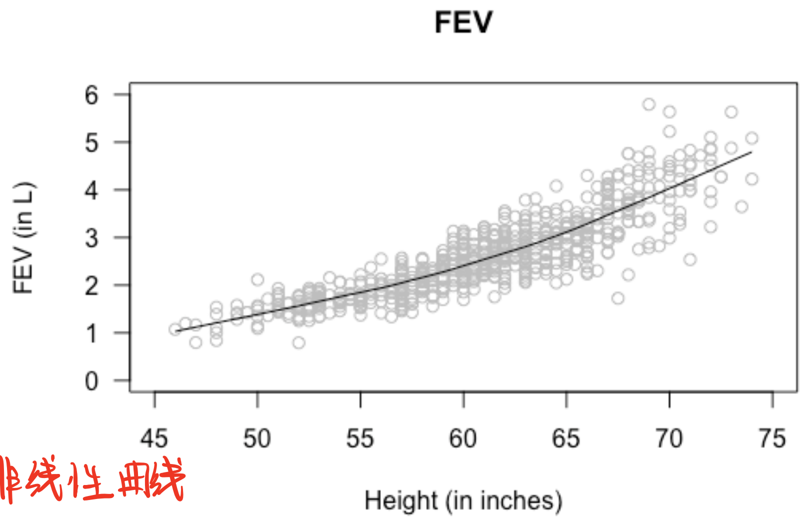
```
> plot(FEV ~ Gender, data=lungcap, main="FEV vs height", ylab="FEV (in L)",
las=1, ylim=c(0,6))
```




```
> lungcap$Smoke <- factor(lungcap$Smoke, levels=c(0,1), labels=c("Non-
smoker","Smoker"))  #change smoke from quantitative to factor
> plot(FEV~Smoke, data=lungcap, main ="FEV vs Smoking status", ylab = "FEV (in
L)", xlab="Smoking status", las = 1, ylim=c(0,6))
```



```
> scatter.smooth(lungcap$Ht, lungcap$FEV, las=1, col="grey",ylim=c(0,6),
xlim=c(45, 75), main="FEV", xlab="Height (in inches)", ylab="FEV)")
```



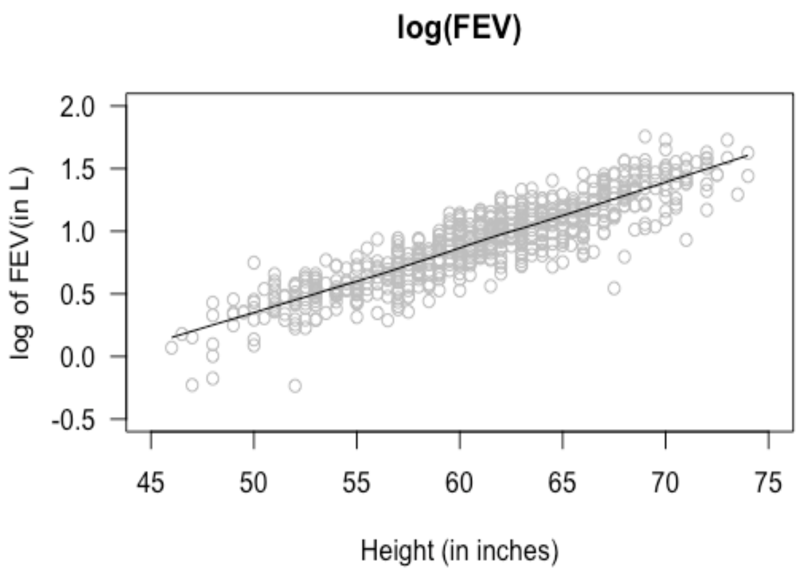
$y_i = f(x_i) + \varepsilon_i$

$lm: f(x_i) = \beta_0 + \beta_1 x_i$

$f(x)$

是x的非线性曲线

```
> scatter.smooth(lungcap$Ht, log(lungcap$FEV), las=1, col="grey",ylim=c(-
0.5,2), xlim=c(45, 75), main="log(FEV)", xlab="Height (in inches)",
ylab="FEV)")
```



非常像对数模型

model 1: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

model 2: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$ 推广

Multiple regression

Model A: with independent variables Age, Ht, Gender, and Smoke (full model)

```
> reg1 <- lm(log(FEV)~Age+Ht+Gender+Smoke,data=lungcap)
> summary(reg1)

Call:
lm(formula = log(FEV) ~ Age + Ht + Gender + Smoke, data = lungcap)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.63278 -0.08657  0.01146  0.09540  0.40701 
```

P-value for $H_0: \beta_j = 0$

$\hat{\beta}_j$ $\quad S(\hat{\beta}_j) = g_{jj}^{\frac{1}{2}}\hat{\sigma}$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.943998   0.078639 -24.721  < 2e-16 ***
Age          0.023387   0.003348   6.984  7.1e-12 ***
Ht           0.042796   0.001679  25.489  < 2e-16 ***
GenderM      0.029319   0.011719   2.502   0.0126 *  
SmokeSmoker -0.046068   0.020910  -2.203   0.0279 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

X_1 Age, X_2 Ht, X_3 GenderM, X_4 SmokeSmoker

$X_3 = \begin{cases} 1 & M \\ 0 & F \end{cases}$

$X_4 = \begin{cases} 1 & \text{smoker} \\ 0 & \text{non-smoker} \end{cases}$

```
Residual standard error: 0.1455 on 649 degrees of freedom
Multiple R-squared:  0.8106,	Adjusted R-squared:  0.8095 
F-statistic: 694.6 on 4 and 649 DF,  p-value: < 2.2e-16
```

比R^2小一些

$H_0: \beta_1 = \beta_2 = \cdots = \beta_4 = 0$

H_1: at least one $\beta_j \neq 0$ for $j = 1, 2, 3, 4$

```
> anova(reg1)
Analysis of Variance Table

Response: log(FEV)
           Df Sum Sq Mean Sq   F value    Pr(>F)    
Age         1 43.210  43.210 2041.9564 < 2.2e-16 ***
Ht          1 15.326  15.326  724.2665 < 2.2e-16 ***
Gender      1  0.153   0.153    7.2451  0.007293 ** 
Smoke       1  0.103   0.103    4.8537  0.027937 *  
Residuals 649 13.734   0.021                        
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Overall utility of the model?
- What are the values of R squares and Adj R squares? Interpretation?
- Tests of usefulness of individual predictor variables?

(Model A) reg1: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$

Age Ht gender Smoke

↓ R^2 will decrease (能解释的部分↓)

(model B) reg2: $H_0: \beta_2 = \beta_3 = 0$

H_1: model A

PT likelihood ↑ AIC ↓↑ ⇒ balance

Remark 8.1

Model B: with independent variables Age and Smoke (reduced model)

```
> reg2 <- lm(log(FEV)~Age+Smoke,data=lungcap)
> summary(reg2)

Call:
lm(formula = log(FEV) ~ Age + Smoke, data = lungcap)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.71124 -0.13458  0.00104  0.14909  0.60261 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.022939   0.030376   0.755  0.45041    
Age          0.090768   0.003053  29.733  < 2e-16 ***
SmokeSmoker -0.089927   0.030118  -2.986  0.00293 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2108 on 651 degrees of freedom
Multiple R-squared:  0.6012,Adjusted R-squared:    0.6 
F-statistic: 490.8 on 2 and 651 DF,  p-value: < 2.2e-16

> anova(reg2)
Analysis of Variance Table

Response: log(FEV)
           Df Sum Sq Mean Sq  F value    Pr(>F)    
Age         1 43.210  43.210 972.6805 < 2.2e-16 ***
Smoke       1  0.396   0.396   8.9151  0.002934 ** 
Residuals 651 28.920   0.044                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question: Is the full model better than the reduced model? (partial F test)

$n-p-1 = 654-2-1 = 651$

```
> anova(reg2,reg1)
Analysis of Variance Table

Model 1: log(FEV) ~ Age + Smoke
Model 2: log(FEV) ~ Age + Ht + Gender + Smoke
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1    651 28.920                                  
2    649 13.734  2    15.186 358.82 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual sum of square: $y'(I-H)y$

$\dfrac{Q/2 \text{ (df of } Q)}{SSE/n-k-1} = \dfrac{15.186/2}{13.734/649} = 358.82 \overset{H_0}{\sim} F_{2,649}$

P value $= P(F_{2,649} \geq 358.82)$

$n-k-1 = 654-4-1 = 649$

$Q = y'(H-H_0)y$

$= SSE_{H_0} - SSE$

Conclusion:

P is too small ⇒ reject H_0

Example Ch8: on confidence interval

R Package: GLMsData   Dataset: lungcap

A study of 654 youths in East Boston investigate the relationships between lung capacity (measured by forced expiratory volume in litres (FEV)) and several factors.

Dependent variable: FEV

Independent variables:

- Smoke status: 1 = smokers; 0 = non-smokers
- age
- height
- gender

Multiple regression

Model reg1: with independent variables Age, Ht, Gender, and Smoke (full model)

```
> reg1 <- lm(log(FEV)~Age+Ht+Gender+Smoke,data=lungcap)
```

Confidence interval for betas

```
> confint(reg1, level=0.95)
                    2.5 %       97.5 %
(Intercept) -2.098414941 -1.789581413
Age          0.016812109  0.029962319
Ht           0.039498923  0.046092655
GenderM      0.006308481  0.052330236
SmokeSmoker -0.087127344 -0.005007728
```

$\beta_j,\ j=0,1,2,3,4$

Confidence interval for expected value of the dependent variable for given **x**

```
> predict(reg1,level=0.95,newdata=data.frame(Age=18, Ht=66, Gender="F",
Smoke="Smoker"), interval="confidence")
       fit      lwr      upr
1 1.255426 1.209268 1.301584
```

at new point $\underset{\sim}{x}^*$

CI of $E(y^*)$

prediction interval for expected value of the dependent variable for given **x**

```
> predict(reg1,level=0.95,newdata=data.frame(Age=18, Ht=66, Gender="F",
Smoke="Smoker"), interval="prediction")
       fit      lwr      upr
1 1.255426 0.966075 1.544777
```

C.I. (predictive interval)

Confidence Ellipse for betas related to Age and Ht

```
> install.packages("ellipse")
Installing package into ‘/Users/siuhungcheung/Library/R/3.6/library’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.6/ellipse_0.4.1.tgz'
Content type 'application/x-gzip' length 71606 bytes (69 KB)
==================================================
downloaded 69 KB


The downloaded binary packages are in
	/var/folders/0m/hwl7r7_94f9ddhyn1sjywmg40000gn/T//Rtmp4PkszD/downloaded_packages
> library(ellipse)
```

β_1: Age β_2: Ht

```
> plot(ellipse(reg1, c(2,3)), type = "l")
```

joint CI of (β_1, β_2)

```
> points(coef(reg1)[2],coef(reg1)[3],pch=18)
> abline(v=confint(reg1)[2,],lty=2)
> abline(h=confint(reg1)[3,],lty=2)
```