L07: Models with normal distributions

1. Normal model

   (1) Normal model and likelihood function
   For $y = X\beta + e$ with $e \sim N(0, \sigma^2 I_n)$, the likelihood function

   $$\begin{aligned} L(\beta, \sigma^2) &= \tfrac{1}{(2\pi)^{n/2}|\sigma^2 I_n|^{1/2}} \exp\left[-\tfrac{1}{2}(y - X\beta)'(\sigma^2 I)^{-1}(y - X\beta)\right] \\ &= \tfrac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\tfrac{1}{2\sigma^2}\|y - X\beta\|^2\right). \end{aligned}$$

   (2) MLE of β and LSE of β
   For this particular $L(\beta, \sigma^2)$,

   $$L(\widehat{\beta}, \sigma^2) \le L(\beta, \sigma^2) \text{ for all } \beta \text{ and } \sigma^2 \iff \|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2 \text{ for all } \beta.$$

   Thus $\widehat{\beta}$ is MLE for $\beta \iff \|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2$ for all $\beta \iff \widehat{\beta}$ is LSE for β.
   So MLE(β) = LSE(β) = $X^+y + \mathcal{N}(X)$.

   (3) MLE of σ^2
   $\|y - X\widehat{\beta}\|^2 = \|y - \widehat{y}\|^2 = \sum_{i=1}^n (y_i - \widehat{y}_i)^2 =$ SSE, the Sum of Squared Errors.
   $L(\widehat{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{SSE}{-2\sigma^2}\right)$. $\ln[L(\widehat{\beta}, \sigma^2)] = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{SSE}{2\sigma^2}$
   $[\ln L(\widehat{\beta}, \sigma^2)]'_{\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{SSE}{2\sigma^4} = -\frac{n}{2\sigma^4}\left(\sigma^2 - \frac{SSE}{n}\right)$.
   By 1st derivative test $L(\widehat{\beta}, \sigma^2)$ is maximized at $\sigma^2 = \frac{SSE}{n}$. So MLE for σ^2 is $\frac{SSE}{n}$.
   Recall: UE for σ^2 is $|FracSSEn - \text{rank}(X) = MSE$, the mean of SSE.

   Comments: SSE$= y'(I - XX^+)y$ has DF = rank$(I - XX^+) = n - \text{rank}(X)$.
   $\frac{SSE}{DF} = \frac{SSE}{n - \text{rank}(X)} = MSE$, the mean of SSE.
   MSE is UE for σ^2.

2. Distributions of $\widehat{\beta}$ and SSE

   (1) Distribution of $\widehat{\beta}$: $\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

   $$\begin{aligned} \beta \text{ is estimable} &\iff I_p = LX \text{ for some } L \iff X \text{ has full column rank } p \\ &\Longrightarrow \text{BLUE of } \beta \text{ is } \widehat{\beta} = X^+y. \end{aligned}$$

   $$\widehat{\beta} = X^+y = (X'X)^{-1}X'y \sim (X'X)^{-1}X'\ N(X\beta, \sigma^2 I_n) = N(\beta, \sigma^2(X'X)^{-1})$$

   (2) Distribution of $\frac{SSE}{\sigma^2}$: $\frac{SSE}{\sigma^2} \sim \chi^2(n - p)$
   Tool: Suppose $\mathbf{x} \sim N(\mu, \Sigma)$ and $A' = A$. $A\Sigma A = A \Longrightarrow \mathbf{x}'A\mathbf{x} \sim \chi^2(\mu' A\mu, \text{tr}(A\Sigma))$.

   $\frac{SSE}{\sigma^2} = y'\frac{H}{\sigma^2}y$ where $H = I_n - XX^+$ is symmetric and idempotent and
   $y \sim N(X\beta, \sigma^2 I_n)$.
   $\frac{H}{\sigma^2}(\sigma^2 I_n)\frac{H}{\sigma^2} = \frac{H}{\sigma^2}$; $(X\beta)'\frac{H}{\sigma^2}(X\beta) = 0$ and $\text{tr}\left(\frac{H}{\sigma^2}\sigma^2 I_n\right) = n - \text{rank}(X) = n - p$.
   So $\frac{SSE}{\sigma^2} \sim \chi^2(n - p)$.

   (3) Independence of $\widehat{\beta}$ and SSE
   Tool: Suppose $\mathbf{x} \sim (\mu, \Sigma)$ and $B' = B$. Then $A\Sigma B = 0 \Longrightarrow A\mathbf{x}$ and $\mathbf{x}'B\mathbf{x}$ are independent

   With $\widehat{\beta} = X^+y$ and $SSE = y'Hy$ and $y \sim N(X\beta, \sigma^2 I_n)$, $X^+(\sigma^2 I)(H) = 0$.
   So $\widehat{\beta}$ and SSE are independent.

3. Confidence intervals

(1) A t-distribution: $\frac{\widehat{\beta}_i-\beta_i}{s_{\widehat{\beta}_i}} \sim t(n-p)$

$\widehat{\beta} \sim N(\beta,\, \sigma^2(X'X)^{-1}) \Longrightarrow \widehat{\beta} \sim N(\beta_i,\, \sigma^2[(X'X)^{-1}]_{ii})$ where $\text{var}(\widehat{\beta}_i) = \sigma^2[(X'X)^{-1}]_{ii}$ is estimated by $s^2_{\widehat{\beta}_i} = MSE\,[(X'X)^{-1}]_{ii}$ and $s_{\widehat{\beta}_i}$ is called the standard error for $\widehat{\beta}_i$. So $\frac{\widehat{\beta}_i-\beta_i}{\sigma_{\widehat{\beta}_i}} \sim N(0,\, 1^2)$.

$\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$ is independent to $\widehat{\beta}_i$. Therefore $\frac{\widehat{\beta}_i-\beta_i}{\sigma_{\widehat{\beta}_i}\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t(n-p)$.

With $\sigma_{\widehat{\beta}_i}\sqrt{\frac{SSE}{\sigma^2}/(n-p)} = \sqrt{\sigma^2[(X'X)^{-1}]_{ii}\frac{SSE}{\sigma^2}/(n-p)} = \sqrt{MSE\,[(X'X)^{-1}]_i i} = s_{\widehat{\beta}_i}$,

$$\frac{\widehat{\beta}_i-\beta_i}{s_{\widehat{\beta}_i}} \sim t(n-p).$$

(2) Confidence interval for β_i

$\widehat{\beta}_i \pm t_{\alpha/2}(n-p)s_{\widehat{\beta}_i}$ is a $1-\alpha$ confidence interval for β_i.

Proof
$$\begin{aligned} 1-\alpha &= P(-t_{\alpha/2}(n-p) \le t(n-p) \le t_{\alpha/2}(n-p)) \\ &= P\left(-t_{\alpha/2}(n-p) \le \tfrac{\widehat{\beta}_i-\beta_i}{s_{\widehat{\beta}_i}} \le t_{\alpha/2}(n-p)\right) \\ &= P\left(\widehat{\beta}_i - t_{\alpha/2}(n-p)s_{\widehat{\beta}_i} \le \beta_i \le \widehat{\beta}_i + t_{\alpha/2}(n-p)s_{\widehat{\beta}_i}\right). \end{aligned}$$

(3) Confidence interval for σ^2

$\left(\frac{SSE}{\chi^2_{\alpha/2}(n-p)},\, \frac{SSE}{\chi^2_{1-\alpha/2}(n-p)}\right)$ is a $1-\alpha$ confidence interval for σ^2.

Proof
$$\begin{aligned} 1-\alpha &= P(\chi^2_{1-\alpha/2}(n-p) \le \chi^2(n-p) \le \chi^2_{\alpha/2}(n-p)) \\ &= P\left(\chi^2_{1-\alpha/2}(n-p) \le \tfrac{SSE}{\sigma^2} \le \chi^2_{\alpha/2}(n-p)\right) \\ &= P\left(\tfrac{SSE}{\chi^2_{\alpha/2}(n-p)} \le \sigma^2 \le \tfrac{SSE}{\chi^2_{1-\alpha/2}(n-p)}\right). \end{aligned}$$

L08: Simultaneous CIs

1. Confidence intervals for $l'\beta$
   When β is estimable, $l'\beta$ is estimable for all $l \in R^p$. So we need to find CI for $l'\beta$.

   (1) A variable with t-distribution
   $\widehat{\beta} \sim N(\beta,\, \sigma^2(X'X)^{-1}) \Longrightarrow l'\widehat{\beta} \sim N(l'\beta,\, \sigma^2 l'(X'X)^{-1}l) = N(l'\beta,\, \sigma^2_{l'\widehat{\beta}})$.
   $\sigma^2_{l'\widehat{\beta}}$ is estimated by $s^2_{l'\widehat{\beta}} = \mathrm{MSE}\, l'(X'X)^{-1}l$ and $s_{l'\widehat{\beta}}$ is called the standard error for $l'\widehat{\beta}$.
   Since $\frac{l'\widehat{\beta}-l'\beta}{\sigma_{l'\widehat{\beta}}} \sim N(0,\, 1^2)$ is independent to $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$, $\quad \frac{l'\widehat{\beta}-l'\beta}{\sigma_{l'\widehat{\beta}}\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t(n-p)$.
   Therefore $\qquad \frac{l'\widehat{\beta}-l'\beta}{s_{l'\widehat{\beta}}} \sim t(n-p)$.

   (2) t-interval for $l'\beta$
   Based on the variable with t-distribution in (1), one can derive $1-\alpha$ CI for $l'\beta$:
   $$l'\widehat{\beta} \pm t_{\alpha/2}(n-p)s_{l'\widehat{\beta}}.$$

   (3) Bonferroni simultaneous CIs
   $l_i'\widehat{\beta} \pm t_{\alpha/(2k)}(n-p)s_{l_i'\widehat{\beta}}$, $i = 1, ..., k$, is a set of k simultaneous CIs for $l_i'\beta$, $i = 1, ..., k$, with overall confidence coefficient $1-\alpha$.

   Proof Let $I_i = [l_i'\widehat{\beta} - t_{\alpha/(2k)}(n-p)s_{l_i'\widehat{\beta}} \le l_i'\beta \le l_i'\widehat{\beta} + t_{\alpha/(2k)}(n-p)s_{l_i'\widehat{\beta}}]$.
   Then $P(I_i) \ge 1 - \frac{\alpha}{k}$. We show $P(I_1 \cap \cdots \cap I_k) \ge 1-\alpha$. By De Morgan law

   $$\begin{aligned} P(I_1 \cap \cdots \cap I_k) &= 1 - P((I_1 \cap \cdots \cap I_k)^c) = 1 - P(I_1^c \cup \cdots \cup I_k^c) \\ &\ge 1 - [P(I_1^c) + \cdots + P(I_k^c)] \\ &= 1 - [(1 - P(I_1)) + \cdots + (1 - P(I_k))] \\ &= 1 - k + P(I_1) + \cdots + P(I_k) \ge 1 - k + k\left(1 - \tfrac{\alpha}{k}\right) = 1 - \alpha. \end{aligned}$$

   .

   Comment: To construct k simultaneous CIs with overall CC $1-\alpha$, one can construct them one by one, each has CC $1-\frac{\alpha}{k}$. This method is called Bonferroni method.

2. F-confidence region for β

   (1) A variable with F-distribution
   $\widehat{\beta} \sim N(\beta,\, \sigma^2(X'X)^{-1}) \Longrightarrow y = \frac{1}{\sigma}(X'X)^{1/2}(\widehat{\beta}-\beta) \sim N(0,\, I_p)$
   So $y'y = \frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{\sigma^2} \sim \chi^2(p)$.
   But as a function of $\widehat{\beta}$, $y'y$ is independent to $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.
   Hence $\qquad \frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{p\cdot MSE} \sim F(p,\, n-p) \qquad$ can be derived.

   (2) F-confidence region for β
   $1-\alpha = P(F(p, n-p) \le F_\alpha(p, n-p)) = P\left(\frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{p\cdot MSE} \le F_\alpha(p,\, n-p)\right)$.
   Thus the random set $\left[\beta \in R^p : \frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{p\cdot MSE} \le F_\alpha(p,\, n-p)\right]$ is a $1-\alpha$ confidence region for β.

   Comment: This CR is an ellipsoid in R^p with center $\widehat{\beta}$.

3. Scheffe's simultaneous CIs for $l_i'\beta$

(1) A lemma

$x \in R^p$ and $A \in R^{p\times p}$ is a p.d matrix. With $c > 0$

$$0 \le x'Ax \le c \Longrightarrow -\sqrt{c\,(y'A^{-1}y)} \le x'y \le \sqrt{c\,(y'A^{-1}y)} \text{ for all } 0 \ne y \in R^p.$$

Proof: For x, $y \in R^p$, $(x'y)^2 \le (x'x)(y'y)$ is Cauchy-Schwartz inequality.

With $A > 0$ replace x by $A^{-1/2}x$ and y by $A^{1/2}y$. We have $(x'y)^2 \le (x'Ax)(y'A^{-1}y)$ called extended Cauchy-Schwartz inequality.

With $0 \ne y \in R^p$, $0 \le \frac{(x'y)^2}{y'A^{-1}y} \le x'Ax$. So if $0 \le x'Ax \le c$, then $0 \le \frac{(x'y)^2}{y'A^{-1}y} \le c$.

Consequently $-\sqrt{c\,(y'A^{-1}y)} \le x'y \le \sqrt{c\,(y'A^{-1}y)}$. $\Box$

(2) Scheffe's simultaneous CIs

$l_i'\widehat{\beta} \pm \sqrt{p\,F_\alpha(p,\, n-p)}\, s_{l_i'\widehat{\beta}}$, $i = 1, 2, \cdot$, are simultaneous CIs for $l_i'\beta$, $i = 1, 2, ...,$ with overall CC $1 - \alpha$.

Proof: Let $I_i = \left[l_i'\widehat{\beta} - \sqrt{pF_\alpha(p, n-p)}\, s_{l_i'\widehat{\beta}} \le l_i'\beta \le l_i'\widehat{\beta} + \sqrt{pF_\alpha(p, n-p)}\, s_{l_i'\widehat{\beta}}\right]$.

We need to show $P(\cap_i I_i) \ge 1 - \alpha$.

In the lemma in (1) let $x = \widehat{\beta} - \beta$, $A = \frac{X'X}{p\,MSE}$ and $c = F_\alpha(p,\, n-p)$.

Then $B = [0 \le x'Ax \le c] = \left[\frac{(\widehat{\beta}-\beta)'(X'X)(\widehat{\beta}-\beta)}{p\,MSE} \le F_\alpha(p,\, n-p)\right]$ is a random event with $P(B) = 1 - \alpha$.

Let $y = l_i$. Then $c(y'A^{-1}y) = [p\,F_\alpha(p,\, n-p)]\, s^2_{l_i'\widehat{\beta}}$. So

$$\begin{aligned}
&[-\sqrt{c\,(y'A^{-1}y)} \le x'y \le \sqrt{c\,(y'A^{-1}y)}\ ] \\
&\quad = [-\sqrt{p\,F_\alpha(p,\, n-p)} s_{l_i'\widehat{\beta}} \le l_i'(\widehat{\beta} - \beta) \le \sqrt{p\,F_\alpha(p,\, n-p)} s_{l_i'\widehat{\beta}}] \\
&\quad = [l_i'\widehat{\beta} - \sqrt{pF_\alpha(p, n-p)}\, s_{l_i'\widehat{\beta}} \le l_i'\beta \le l_i'\widehat{\beta} + \sqrt{pF_\alpha(p, n-p)}\, s_{l_i'\widehat{\beta}}] \\
&\quad = I_i
\end{aligned}$$

So $B \subset I_i \Longrightarrow B \subset \cap_i I_i \Longrightarrow 1 - \alpha = P(B) \le P(\cap_i I_i)$. $\Box$

(3) Comments

The intervals in (2) are called Scheffe's simultaneous CIs.

There is no limit on the number of intervals included. It could be countable many.

To get the overall confidence coefficient $1-\alpha$, the widthes of the intervals are wider than that of Bonferroni intervals.