12 Simultaneous Confidence Intervals

12.1 Confidence Intervals for Estimable Functions

The goal is to compute a confidence interval for an estimable function $\mathbf{a}'\boldsymbol{\beta}$:

In the full rank $\mathbf{X}$ case we have

- $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$,
- $(n-p)S^2/\sigma^2 \sim \chi^2_{n-p}$,
- $\hat{\boldsymbol{\beta}}$ is independent of S^2.

Therefore,

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \sim N_1(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$$

and

$$T = \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta})/[\sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]^{1/2}}{\{[(n-p)S^2/\sigma^2]/(n-p)\}^{1/2}} = \frac{Z}{\{W/(n-p)\}^{1/2}} \sim t_{n-p},$$

by the definition of the t distribution, where $Z \sim N(0,1)$ and $W \sim \chi^2_{n-p}$ are independent. Note that

$$T = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{[S^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]^{1/2}} = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{\widehat{\text{se}}(\mathbf{a}'\hat{\boldsymbol{\beta}})}.$$

Now a $100(1-\alpha)\%$ confidence interval for $\mathbf{a}'\boldsymbol{\beta}$ is

$$I = [\ \mathbf{a}'\hat{\boldsymbol{\beta}} - t^{\alpha/2}_{n-p} \times \widehat{\text{se}}(\mathbf{a}'\hat{\boldsymbol{\beta}})\ ,\ \mathbf{a}'\hat{\boldsymbol{\beta}} + t^{\alpha/2}_{n-p} \times \widehat{\text{se}}(\mathbf{a}'\hat{\boldsymbol{\beta}})\ ],$$

where $t^{\alpha/2}_{n-p}$ is the $100(1-\alpha/2)$-percentile of the t_{n-p} distribution.

12.1 Note: The defining property of the confidence interval is $P(\mathbf{a}'\boldsymbol{\beta} \in I) = 1-\alpha$.

12.2 Note: In the less than full rank case our t statistics have the form

$$T = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{[S^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}]^{1/2}}$$

for a generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$. These statistics have a t_{n-r} distribution, where $r = \text{rank}(\mathbf{X})$, provided that $\mathbf{a} \neq \mathbf{0}$.

What if we wish to form confidence intervals for a set of estimable functions $\mathbf{a}_j'\boldsymbol{\beta}, j = 1, \dots, k$? As above we could find I_j such that

$$P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j) = 1-\alpha.$$

But we can expect that

$$P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j, \forall j = 1, \dots, k) < 1-\alpha,$$

i.e. there will be a greater than α chance of at least 1 of the confidence intervals not containing the true value. This is referred to as the multiple comparisons problem.

12.2 Multiple Comparisons: The Bonferroni Method

Suppose the jth confidence interval has coverage probability $1-\alpha_j$, i.e. $P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j) = 1-\alpha_j$. Then

$$P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j, \forall j \in \{1,\ldots,k\}) = 1 - P(\mathbf{a}_j'\boldsymbol{\beta} \notin I_j \text{ for some } j) \geq 1 - \sum_{j=1}^{k} P(\mathbf{a}_j'\boldsymbol{\beta} \notin I_j) = 1 - \sum_{j=1}^{k} \alpha_j.$$

The Bonferroni method proceeds by setting the α_j so that $\sum_{j=1}^{k} \alpha_j = \alpha$, (usually $\alpha_j = \alpha/k$). Then

$$P(\text{all confidence intervals contain their true values}) = P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j, \forall j \in \{1,\ldots,k\}) \geq 1-\alpha.$$

The main features of the Bonferroni method are that it is simple to use, and that it is conservative (i. e. the actual coverage probability is greater than claimed and the confidence intervals are wider than they have to be).

12.3 Example: One-way ANOVA $Y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ (i = 1,\ldots,k)$.
If we want confidence intervals for all pairwise comparisons $\{\tau_i - \tau_j, i \neq j\}$, there are $n_k = k \times (k-1)/2$ such comparisons. The Bonferroni method uses $\alpha_j = \alpha/n_k$. Hence with $\alpha = .05$ we have:

k	n_k	α_j
2	1	0.0500
3	3	0.0167
4	6	0.0083
5	10	0.0050
6	15	0.0033

12.4 Note: A better method for this situation is Tukey's studentized range method which we will see later when we take up ANOVA in more detail.

12.3 Multiple Comparisons: Maximum Modulus Method

What if $\{\mathbf{a}_j'\hat{\boldsymbol{\beta}}, j = 1,\ldots,k\}$ are independent? In this case, the t statistics

$$T_j = \frac{\mathbf{a}_j'\hat{\boldsymbol{\beta}} - \mathbf{a}_j'\boldsymbol{\beta}}{[S^2\mathbf{a}_j'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}_j]^{1/2}}$$

are conditionally independent given S^2. The t statistics are marginally uncorrelated:

$$\text{cov}(T_i, T_j) = E[T_iT_j] = E\{E[T_iT_j|S^2]\} = E\{E[T_i|S^2]E[T_j|S^2]\} = 0,$$

because $E[T_i|S^2] = E[\mathbf{a}_i'\hat{\boldsymbol{\beta}} - \mathbf{a}_i'\boldsymbol{\beta}|S^2]/[S^2\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}_i]^{1/2} = 0,$ by the independence of $\mathbf{a}'\hat{\boldsymbol{\beta}}$ and S^2.

12.5 Definition: Let $T_1,\ldots,T_k$ be a set of pairwise uncorrelated random variables with the distribution t_ν. Define $U \equiv \max\{|T_j|, j = 1,\ldots,k\}$. Then U has the Studentized Maximum Modulus Distribution, denoted by $u_{k,\nu}$.

12.6 Note: Let the $100(1-\alpha)$-percentile of the distribution be denoted $u_{k,\nu}^{\alpha}$. Then

$$P(|T_j| \leq u_{k,n-r}^{\alpha}, \forall j = 1,\ldots,k) = P(\max\{|T_j|, j = 1,\ldots,k\} \leq u_{k,n-r}^{\alpha}) = 1-\alpha,$$

and a set of simultaneous confidence intervals is

$$I_j = [\ \mathbf{a}_j'\hat{\boldsymbol{\beta}} - u_{k,n-r}^{\alpha}\hat{\text{se}}(\mathbf{a}_j'\hat{\boldsymbol{\beta}})\ ,\ \mathbf{a}_j'\hat{\boldsymbol{\beta}} + u_{k,n-r}^{\alpha}\hat{\text{se}}(\mathbf{a}_j'\hat{\boldsymbol{\beta}})\ ]$$

and hence

$$P(\mathbf{a}_j'\boldsymbol{\beta} \in I_j, \forall j = 1,\ldots,k) = 1-\alpha.$$

12.7 Example: Comparison with the Bonferroni method for 1-way ANOVA.
Suppose we want simultaneous confidence intervals for the group means $\mathbf{a}_j'\boldsymbol{\beta} = \mu + \tau_j$. Then $\mathbf{a}_j'\hat{\boldsymbol{\beta}} = \bar{Y}_j$ is the sample mean for the jth group. Thus the $\mathbf{a}_j'\hat{\boldsymbol{\beta}}$ are independent and we can apply the maximum modulus method. It can be shown that the confidence intervals for this method are narrower than the Bonferroni confidence intervals, i.e. $t_{n-r}^{\alpha/(2k)} \geq u_{k,n-r}^{\alpha}$.

12.4 Multiple Comparisons: Scheffé Method

The goal is to obtain simultaneous confidence intervals for estimable functions $\mathbf{a}_1'\boldsymbol{\beta},\ldots,\mathbf{a}_k'\boldsymbol{\beta}$.

Rearrange the vectors $\mathbf{a}_1,\ldots,\mathbf{a}_k$ so that $\mathbf{a}_1,\ldots,\mathbf{a}_d$ are linearly independent, and $\mathbf{a}_{d+1},\ldots,\mathbf{a}_k$ are linear combinations of $\mathbf{a}_1,\ldots,\mathbf{a}_d$ $(0 \leq d \leq k)$. Then a set of simultaneous confidence intervals for $\mathbf{a}_1,\ldots,\mathbf{a}_d$ automatically produce confidence intervals for the remaining ones $\mathbf{a}_{d+1},\ldots,\mathbf{a}_k$. So we work only with the d linearly independent vectors. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_d' \end{pmatrix}$$

be a $d \times p$ matrix of rank d (like $\mathbf{A}$ in $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{0}$). Then we have

$$\frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{A}\boldsymbol{\beta})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{A}\boldsymbol{\beta})}{dS^2} \sim F_{d,n-r}.$$

Now define $\boldsymbol{\phi} = \mathbf{A}\boldsymbol{\beta}$ and $\hat{\boldsymbol{\phi}} = \mathbf{A}\hat{\boldsymbol{\beta}}$, so that

$$P\left(\frac{(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})}{dS^2} \leq F_{d,n-r}^{\alpha}\right) = 1-\alpha.$$

This defines the following confidence region for $\boldsymbol{\phi}$:

$$\left\{\mathbf{u} : \frac{(\hat{\boldsymbol{\phi}} - \mathbf{u})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-1}(\hat{\boldsymbol{\phi}} - \mathbf{u})}{dS^2} \leq F_{d,n-r}^{\alpha}\right\},$$

which is a region enclosed by an ellipsoid centered at $\hat{\boldsymbol{\phi}}$. The interpretation is that this ellipsoidal confidence region contains the true value $\boldsymbol{\phi}$ with probability $1-\alpha$.

12.8 Theorem: If $\mathbf{L}$ is positive definite, then $\sup_{\mathbf{h}\neq\mathbf{0}} \frac{(\mathbf{h}'\mathbf{b})^2}{\mathbf{h}'\mathbf{L}\mathbf{h}} = \mathbf{b}'\mathbf{L}^{-1}\mathbf{b}$.

Apply Theorem 12.8 with $\mathbf{b} = \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}$ and $\mathbf{L} = \mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}'$:

$$
\begin{aligned}
& (\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \le dS^2 F^{\alpha}_{d,n-r}, \\
\iff \quad & \sup_{\mathbf{h}\neq\mathbf{0}} \frac{[\mathbf{h}'(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})]^2}{\mathbf{h}'\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}'\mathbf{h}} \le dS^2 F^{\alpha}_{d,n-r}, \\
\iff \quad & |\mathbf{h}'(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})| \le (dF^{\alpha}_{d,n-r})^{1/2} S\{\mathbf{h}'\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}'\mathbf{h}\}^{1/2} \ \ \forall \mathbf{h}.
\end{aligned}
$$

Therefore, the probability is $1 - \alpha$ that

$$
\mathbf{h}'\hat{\boldsymbol{\phi}} - (dF^{\alpha}_{d,n-r})^{1/2}\hat{\text{se}}(\mathbf{h}'\hat{\boldsymbol{\phi}}) \le \mathbf{h}'\boldsymbol{\phi} \le \mathbf{h}'\hat{\boldsymbol{\phi}} + (dF^{\alpha}_{d,n-r})^{1/2}\hat{\text{se}}(\mathbf{h}'\hat{\boldsymbol{\phi}}), \ \ \forall \mathbf{h},
$$

where

$$
\hat{\text{se}}(\mathbf{h}'\hat{\boldsymbol{\phi}}) = S\{\mathbf{h}'\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}'\mathbf{h}\}^{1/2}
$$

is the estimated standard error of $\mathbf{h}'\hat{\boldsymbol{\phi}}$. This defines the following set of simultaneous confidence intervals for all linear combinations $\mathbf{h}'\boldsymbol{\phi}$:

$$
\mathbf{h}'\boldsymbol{\phi} \in \Big[\ \mathbf{h}'\hat{\boldsymbol{\phi}} - (dF^{\alpha}_{d,n-r})^{1/2}\hat{\text{se}}(\mathbf{h}'\hat{\boldsymbol{\phi}})\ ,\ \mathbf{h}'\hat{\boldsymbol{\phi}} + (dF^{\alpha}_{d,n-r})^{1/2}\hat{\text{se}}(\mathbf{h}'\hat{\boldsymbol{\phi}})\ \Big], \ \ \forall \mathbf{h}.
$$

12.9 Note: We actually obtained simultaneous confidence intervals for all linear combinations of the estimable functions $\phi_j = \mathbf{a}_j'\boldsymbol{\beta}, \ \ j = 1, \ldots, k$.

12.10 Example: (Simultaneous confidence intervals for linear combinations of β's).

In the full rank case, we can choose the estimable functions $\beta_0, \beta_1, \ldots, \beta_{p-1}$ i.e. $\mathbf{A} = \mathbf{I}_{p\times p}$. Then we have the following confidence intervals for all linear combinations of the β's:

$$
\mathbf{h}'\boldsymbol{\beta} \in \Big[\ \mathbf{h}'\hat{\boldsymbol{\beta}} - \{pF^{\alpha}_{p,n-p}S^2\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}\}^{1/2}\ , \quad \mathbf{h}'\hat{\boldsymbol{\beta}} + \{pF^{\alpha}_{p,n-p}S^2\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}\}^{1/2}\ \Big], \ \ \forall \mathbf{h}.
$$

12.11 Example: (Confidence bands for a regression surface).

Suppose we want simultaneous confidence intervals for the mean of the response variable Y at a given set of predictor variables $\mathbf{x}' = (x_{i0}, x_{i1}, \ldots, x_{i,p-1})$, i.e. $E[Y] = \mathbf{x}'\boldsymbol{\beta}$. Assuming the full rank case, in the previous example set $\mathbf{h} = \mathbf{x}$:

$$
\mathbf{x}'\boldsymbol{\beta} \in \Big[\ \mathbf{x}'\hat{\boldsymbol{\beta}} - \{pF^{\alpha}_{p,n-p}S^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\}^{1/2}\ ,\ \ \mathbf{x}'\hat{\boldsymbol{\beta}} + \{pF^{\alpha}_{p,n-p}S^2\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\}^{1/2}\ \Big], \ \ \forall \mathbf{x}.
$$

This gives us simultaneous confidence intervals for the mean of Y at all values of the predictors. Plotted against the predictors, this yields a confidence band around the fitted model.

12.12 Example: (Simple linear regression).

If $p = 2$ we get the following confidence region for $\boldsymbol{\beta} = (\beta_0, \beta_1)'$:

$$
\{\boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le 2F^{\alpha}_{2,n-2}S^2\}.
$$

The simultaneous confidence intervals for $\beta_0 = (1,0)\boldsymbol{\beta}$ and $\beta_1 = (0,1)\boldsymbol{\beta}$ are

$$\begin{aligned}\beta_0 &\in \left[\hat{\beta}_0 - \sqrt{\frac{2F^{\alpha}_{2,n-2}S^2\sum x_i^2/n}{\sum(x_i-\bar{x})^2}}, \hat{\beta}_0 + \sqrt{\frac{2F^{\alpha}_{2,n-2}S^2\sum x_i^2/n}{\sum(x_i-\bar{x})^2}}\right],\\ \beta_1 &\in \left[\hat{\beta}_1 - \sqrt{\frac{2F^{\alpha}_{2,n-2}S^2}{\sum(x_i-\bar{x})^2}}, \hat{\beta}_1 + \sqrt{\frac{2F^{\alpha}_{2,n-2}S^2}{\sum(x_i-\bar{x})^2}}\right],\end{aligned}$$

where we have

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{\sum(x_i-\bar{x})^2}\begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

The confidence band for the regression line is

$$\beta_0 + \beta_1 x \in \left[\hat{\beta}_0 + \hat{\beta}_1 x - \left\{2F^{\alpha}_{2,n-2}S^2\frac{\sum(x_i-x)^2}{\sum(x_i-\bar{x})^2}\right\}^{1/2},\ \ \hat{\beta}_0 + \hat{\beta}_1 x + \left\{2F^{\alpha}_{2,n-2}S^2\frac{\sum(x_i-x)^2}{\sum(x_i-\bar{x})^2}\right\}^{1/2}\right].$$

Check that $\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} = \sum(x_i-x)^2/\sum(x_i-\bar{x})^2$. Note that the width of the confidence band depends on $\sum(x_i-x)^2/\sum(x_i-\bar{x})^2$, i.e. how far x is from $\bar{x}$.

12.13 Note: The confidence region for $\boldsymbol{\phi} = \mathbf{A}\boldsymbol{\beta}$ contains all points $\mathbf{u}$ which would not be rejected by the F test of $H : \boldsymbol{\phi} = \mathbf{u}$. This is the set of $\mathbf{u}$'s which satisfy

$$\frac{(\hat{\boldsymbol{\phi}}-\mathbf{u})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-}\mathbf{A}']^{-1}(\hat{\boldsymbol{\phi}}-\mathbf{u})}{dS^2} \leq F^{\alpha}_{d,n-r}.$$

12.5 Comparison of Methods

Bonferroni, Maximum Modulus and Scheffé confidence intervals for a set of estimable functions $\phi_j = \mathbf{a}_j'\boldsymbol{\beta},\ j = 1,\ldots,k$ all have the form:

$$[\hat{\phi}_j - c\,\hat{\text{se}}(\hat{\phi}_j), \hat{\phi}_j + c\,\hat{\text{se}}(\hat{\phi}_j)],$$

where $c = t^{\alpha/(2k)}_{n-r}$ (Bonferroni), $c = u^{\alpha}_{k,n-r}$ (Maximum Modulus), or $c = (dF^{\alpha}_{d,n-r})^{1/2}$, where d is the number of linearly independent $\mathbf{a}_j$ (Scheffé). Hence the relative widths of the intervals depend on the relative sizes of the values of c. Some general rules are:

(a) $u^{\alpha}_{k,n-r} \leq t^{\alpha/(2k)}_{n-r}$.

(b) If $k \approx d$, $u^{\alpha}_{k,n-r} < (dF^{\alpha}_{d,n-r})^{1/2}$.

(c) If $k >> d$, $(dF^{\alpha}_{d,n-r})^{1/2} < u^{\alpha}_{k,n-r}$.