Advanced Methods in Biostatistics II

Lecture 3

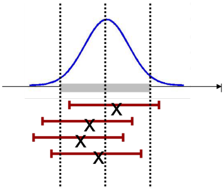
October 31, 2017

Simultaneous Inference

- In many situations we are interested in constructing a collection, or family, of confidence intervals each with a specific confidence level.
- In these situations we need to determine our level of confidence that all of the intervals simultaneously contain the true parameter value.
- Throughout we assume the standard normal error model.

Illustration

- In a family of four independent 95% confidence intervals, the probability that all intervals in the family simultaneously capture the true parameter value will be less than 0.95.



- In fact, the probability is $(0.95)^4 = 0.8145$.

Simultaneous Inference

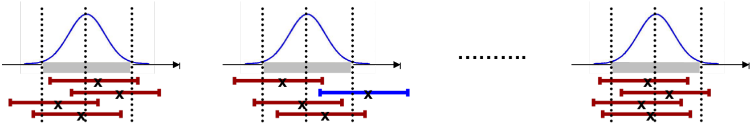
- In the context of a family of confidence intervals, we distinguish between an individual confidence level and family-wise confidence level.
- The individual confidence level is the confidence we have that any particular confidence interval contains the true parameter value.
- The family-wise confidence level is the confidence we have that all the confidence intervals in a family of intervals simultaneously contain the true parameter value.

Simultaneous Inference

- Simultaneously making a large number of comparisons compounds the statistical uncertainty and introduces the need to adjust the individual confidence levels for multiple comparisons.

Illustration

- To control the family confidence level at 95% we must widen the confidence intervals so that the probability that all of the intervals in the family simultaneously capture the true parameter value is at least 0.95.



- This means that at most 1 out of 20 of such families of confidence intervals may contain an individual interval that does not contain the true parameter value.

Bonferroni Method

- Perhaps the most well-known and simplest procedure for controlling for multiple comparisons is the Bonferroni method.
- It provides control for multiple comparisons by adjusting the width of the intervals.

Bonferroni Method

- To illustrate, let us assume we are interested in creating k confidence intervals for the parameters $\beta_1, \beta_2, \ldots \beta_k$.
- Suppose the j^{th} confidence interval has coverage probability $1 - \alpha$ and we want the family-wise confidence level to be $1 - \alpha_f$.
- Let E_j be the event that the j^{th} confidence interval includes β_j, and E_j^c be the complement.

Bonferroni's Inequality

Theorem

Let $P(E_i)$ be the probability that E_i is true, and $P(\cup_{i=1}^n E_i)$ the probability that at least one of $E_1, E_2, \ldots, E_n$ is true. Then the Bonferroni inequality, also known as Boole's inequality, states:

$$P(\cup_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i).$$

Bonferroni Method

- Using this result, we can write:

$$
\begin{aligned}
1-\alpha_f &= P(E_1 \cap E_2 \cap \ldots \cap E_k) \\
&= 1 - P(E_1^c \cup E_2^c \cup \ldots \cup E_k^c) \\
&\leq 1 - \sum_{j=1}^{k} P(E_j^c) \\
&= 1 - k\alpha.
\end{aligned}
$$

Bonferroni Method

- Thus, if the individual confidence level for two intervals is $1 - \alpha = 0.95$, then we have a family confidence level of at least $1 - 2\alpha = 0.90$.
- To guarantee a family confidence level of at least 0.95 we instead need the individual confidence level to be $(1 - 0.05/2) = 0.975$.
- In general, we can ensure appropriate control by setting $\alpha = \alpha_f / k$.

Bonferroni Confidence Intervals

- Using this approach, Bonferroni confidence intervals for $\beta_1, \beta_2, \ldots \beta_k$ are given by

$$\hat{\beta}_j \pm t_{n-p,1-\alpha_f/2k} s \sqrt{g_{jj}}$$

for $j = 1, 2, \ldots k$, where g_{jj} is the j^{th} diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$.

Bonferroni Confidence Intervals

- More generally, for k linear functions $\mathbf{c}_1'\beta$, $\mathbf{c}_2'\beta, \ldots$ $\mathbf{c}_k'\beta$, Bonferroni confidence intervals are given by

$$\mathbf{c}_i'\hat{\beta} \pm t_{n-p,1-\alpha_f/2k} s \sqrt{\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_i}$$

  for $i = 1, 2, \ldots k$.

Bonferroni Prediction Intervals

- For simultaneous prediction of k new observations $y_{01}, y_{02}, \ldots y_{0k}$ at k values of the explanatory variables $\mathbf{x}_{01}, \mathbf{x}_{02}, \ldots \mathbf{x}_{0k}$, we can use Bonferroni prediction intervals:

$$\mathbf{x}_{0i}'\hat{\beta} \pm t_{n-p,1-\alpha_f/2k} s \sqrt{1 + \mathbf{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{0i}}$$

for $i = 1, 2, \ldots k$.

Properties

- The Bonferroni procedure provides a lower bound on the true family-wise confidence level, as it is guaranteed to control the family-wise confidence level.
- However, it tends to be overly conservative.
- It gives rise to confidence intervals that are unnecessarily wide and significance tests with too many false negatives.
- It becomes increasingly conservative as the number of comparisons increase.

Example

- Consider one-way ANOVA

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

  for $i = 1, \ldots, k$.

- Suppose we want confidence intervals for all pairwise comparisons $\{\tau_i - \tau_j, i \neq j\}$.
- In total there will be $n_k = k \times (k-1)/2$ such comparisons.
- The Bonferroni method would use $\alpha = \alpha_f / n_k$ to control the family-wise confidence level at α_f.

Example

- Hence with $\alpha_f = 0.05$ we have:

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

Scheffé Method

- Next, we turn our attention to an alternative approach towards handling multiple comparisons, namely Scheffé's method.
- It is based on the following theorem.

Scheffé Method

Theorem

If $\mathbf{L}$ is positive definite, then

$$\max_{\mathbf{h}\neq\mathbf{0}} \left(\frac{(\mathbf{h}'\mathbf{b})^2}{\mathbf{h}'\mathbf{L}\mathbf{h}} \right) = \mathbf{b}'\mathbf{L}^{-1}\mathbf{b}.$$

Scheffé Method

- Recall that we can write:

$$\frac{(\hat{\beta}-\beta)'\mathbf{X}'\mathbf{X}(\hat{\beta}-\beta)}{ps^2} \sim F_{p,n-p}.$$

- Applying the theorem with $\mathbf{b} = \hat{\beta} - \beta$ and $\mathbf{L} = (\mathbf{X}'\mathbf{X})^{-1}$, we find that

$$(\hat{\beta}-\beta)'\mathbf{X}'\mathbf{X}(\hat{\beta}-\beta) = \max_{\mathbf{h}\neq\mathbf{0}} \left(\frac{(\mathbf{h}'(\hat{\beta}-\beta))^2}{\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}} \right).$$

Scheffé Method

- Therefore,

$$P\left(\frac{1}{ps^2}\max_{\mathbf{h}\neq\mathbf{0}}\left(\frac{(\mathbf{h}'(\hat{\beta}-\beta))^2}{\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}}\right)\leq F_{p,n-p,1-\alpha}\right)=1-\alpha.$$

- Equivalently, we can write:

$$P\left(\frac{|\mathbf{h}'\hat{\beta}-\mathbf{h}'\beta|}{\sqrt{\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}}}\leq\sqrt{ps^2F_{p,n-p,1-\alpha}}\ \forall\,\mathbf{h}\in\mathbb{R}^p\right)=1-\alpha.$$

Scheffé Method

- As a special case, for β_j we can write:

$$P\left(\frac{|\hat{\beta}_j - \beta_j|}{\sqrt{g_{jj}}} \leq \sqrt{ps^2 F_{p,n-p,1-\alpha}} \ \forall\, 1 \leq j \leq p\right) \geq 1 - \alpha.$$

- Hence, simultaneous confidence intervals for $\beta_1, \beta_2, \ldots \beta_k$ are given by

$$\hat{\beta}_j \pm \sqrt{ps^2 g_{jj} F_{p,n-p,1-\alpha}} \ \text{ for } \ j = 1, 2, \ldots k$$

Scheffé Method

- More generally, we can express simultaneous confidence intervals for any linear combination of β as follows:

$$\mathbf{h}'\hat{\beta} \pm \sqrt{pF_{p,n-p,1-\alpha}s^2\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{h}}.$$

- Note this does not depend upon the number of comparisons made.

Example

- As an example, let us consider confidence bands for a regression surface.
- Suppose we want simultaneous confidence intervals for the mean of the response variable y at different values of $\mathbf{x}_0$, i.e. $E[y] = \mathbf{x}_0'\beta$.

Example

- Set $\mathbf{h} = \mathbf{x}_0$ and we obtain:

$$\mathbf{x}_0'\hat{\beta} \pm \sqrt{pF_{p,n-p,1-\alpha} s^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0} \quad \forall \mathbf{x}_0.$$

- 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.

Example

- Now, consider the case of simple linear regression.
- If $p = 2$ we get the following confidence region for the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)'$:

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

Example

- Recall that

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

- The simultaneous confidence intervals for β_0 and β_1 are given by

$$\hat{\beta}_0 \pm \sqrt{\frac{2F_{2,n-2,1-\alpha}s^2 \sum x_i^2/n}{\sum(x_i - \bar{x})^2}}$$

$$\hat{\beta}_1 \pm \sqrt{\frac{2F_{2,n-2,1-\alpha}s^2}{\sum(x_i - \bar{x})^2}}$$

Example

- The confidence band for the regression line $\beta_0 + \beta_1 x$ is given by

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm \sqrt{2 F^{\alpha}_{2,n-2} s^2 \frac{\sum (x_i - x)^2}{\sum (x_i - \bar{x})^2}}.$$

- 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 $\bar{x}$.

Bonferroni vs. Scheffé Method

- Suppose we are studying k different linear functions.
- To choose between Bonferroni and Scheffé, one can compare:
$$\sqrt{pF_{p,n-p,1-\alpha}}$$
and
$$t_{n-p,1-\alpha/k}.$$
- In particular for large values of k, Scheffé tends to provide narrower intervals.

Hypothesis Testing

- In many modern applications (e.g., genomics and imaging) we seek to perform multiple hypothesis tests at the same time, rather than a single joint test.
- In the context of hypothesis tests we differentiate between the overall or family-wise α-level and the individual or comparison-wise α-level.
- The methods described above (e.g., Bonferroni) carry over to the hypothesis testing setting.

Hypothesis Testing

- Recall that there are two types of errors one can make when performing hypothesis tests: Type I and Type II errors.

- A Type I error occurs when H_0 is true, but we mistakenly reject it (i.e., a false positive). This is controlled by the significance level α.

- A Type II error occurs when H_0 is false, but we fail to reject it (i.e., a false negative).

Hypothesis Testing

- If more than one hypothesis test is performed, the risk of making at least one Type I error is greater than the α value for a single test.
- The more tests one performs, the greater the likelihood of getting at least one false positive.

False Positives

- Controlling for multiple comparisons involves controlling the Type I error rate.
- There exist several ways of quantifying the likelihood of obtaining false positives.
    - The family-wise error rate (FWER) is the probability of any false positives.
    - The false discovery rate (FDR) is the proportion of false positives among rejected tests.

Notation

- Suppose we seek to perform m hypothesis tests.
- Let H_{0i} be the null hypothesis for the i^{th} test.
- Let T_i be the value of the corresponding test statistic and p_i its p-value.

Family-wise Error Rate

- The family-wise null hypothesis,

$$H_0 = \cap_{i=1}^{m} H_{0i},$$

  states that all m individual null hypotheses are true.

- If we reject a single voxel null hypothesis, H_{0i}, we will reject the family-wise null hypothesis.

- Assuming H_0 is true, we want the probability of falsely rejecting H_0 to be controlled by α, i.e.,

$$P\left(\cup_{i=1}^{m} T_i \leq u | H_0\right) \leq \alpha$$

Bonferroni Correction

- Using similar reasoning as before, it is easy to show that the Bonferroni method controls the FWER by choosing:

$$P\left(T_i \leq u | H_0\right) \leq \frac{\alpha}{m}$$

False Discovery Rate

- A more recent approach towards dealing with multiple comparisons is the false discovery rate (FDR).
- It is due to Benjamini and Hochberg (1995).

False Positives

	Not Declared Significant	Declared Significant	
Null True	U	V	m_0
Alternative True	T	S	$m-m_0$
	m-R	R	m

- Here U, V, T and S are unobservable random variables.
- R is an observable random variable.

Notation

- Using this notation we can define the FWER as follows:

$$FWER = P(V \geq 1).$$

- We can define the FDR as follows:

$$FDR = E\left(\frac{V}{R}\right)$$

- Note $FDR = 0$ if $R = 0$.

Properties

- A procedure controlling the FDR ensures that on average the FDR is no bigger than a pre-specified rate q which lies between 0 and 1.
- However, for any given data set the FDR need not be below the bound.
- An FDR-controlling technique guarantee controls of the FDR in the sense that $FDR \leq q$.

Benjamini-Hochberg Procedure

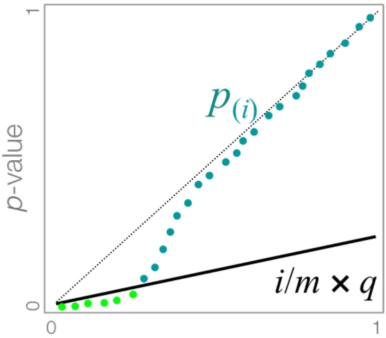
1. Select the desired limit q on the FDR (e.g., 0.05)
2. Rank the p-values, $p_{(1)} \leq p_{(2)} \leq \ldots p_{(m)}$.
3. Let r be largest i such that

$$p_{(i)} \leq \frac{iq}{m}$$

4. Reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$.

Benjamini-Hochberg Procedure

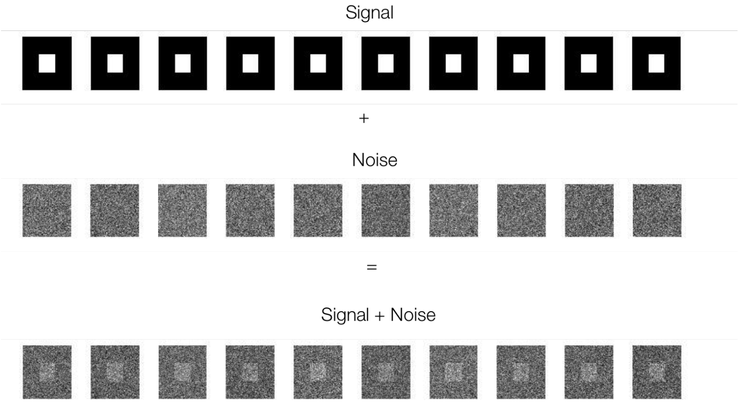
- Illustration of the Benjamini-Hochberg procedure procedure.



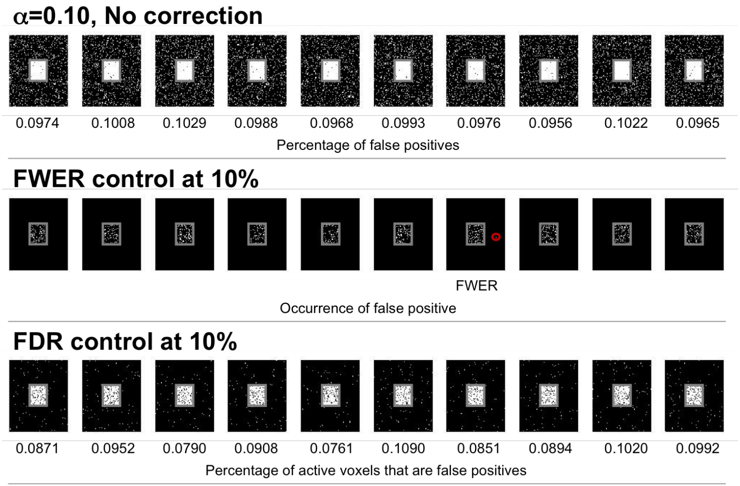
Properties

- If all null hypothesis are true, the FDR is equivalent to the FWER.
- Any procedure that controls the FWER also controls the FDR. A procedure that controls the FDR only can be less stringent and lead to a gain in power.
- Since FDR controlling procedures work only on the p-values and not on the actual test statistics, it can be applied to any valid statistical test.

Example



Example


α=0.10, No correction
0.0974 0.1008 0.1029 0.0988 0.0968 0.0993 0.0976 0.0956 0.1022 0.0965
Percentage of false positives
FWER control at 10%
FWER
Occurrence of false positive
FDR control at 10%
0.0871 0.0952 0.0790 0.0908 0.0761 0.1090 0.0851 0.0894 0.1020 0.0992
Percentage of active voxels that are false positives