University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Bonferroni confidence intervals

Suppose we want to find confidence intervals $I_1, I_2, \ldots, I_m$ for parameters $\theta_1, \theta_2, \ldots, \theta_m$. In the analysis of variance problem $\theta_1 = \mu_1 - \mu_2, \theta_2 = \mu_1 - \mu_3, \ldots, \theta_m = \mu_{k-1} - \mu_k$. We want

$$P\left(\theta_j \in I_j, j = 1, 2, \ldots, m\right) \geq 1 - \alpha,$$

i.e. we want a simultaneous confidence level $1 - \alpha$.

Aside note:
De Morgan's law states that:

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) = 1 - P(A_1' \cup A_2' \cup \ldots \cup A_m')$$

Bonferroni inequality:

$$P(A_1' \cup A_2' \cup \ldots \cup A_m') \leq \sum_{i=1}^{m} P(A_i')$$

Therefore,

$$P(A_1 \cap A_2 \cap \ldots \cap A_m) \geq 1 - \sum_{i=1}^{m} P(A_i') \tag{1}$$

Suppose $P(\theta_j \in I_j) = 1 - \alpha_j$. Let's denote with A_j the event that $\theta_j \in I_j$. Then from Bonferroni inequality (see expression (1) above) we get:

$$\begin{aligned} P(\theta_1 \in I_1, \theta_2 \in I_2, \ldots \theta_m \in I_m) &\geq 1 - \sum_{i=1}^{m} P(\theta_j \notin I_j) \\ &\geq 1 - \sum_{i=1}^{m} \alpha_j. \end{aligned}$$

If all $\alpha_j, j = 1, 2, \ldots, m$ are chosen equal to α we observe that the simultaneous confidence level is only $\geq 1 - m\alpha$ which is smaller than $1 - \alpha$ because $m \geq 1$. In order to have a simultaneous confidence interval $1-\alpha$ we will need to use confidence level $1 - \frac{\alpha}{m}$ for each confidence interval for $\mu_1 - \mu_2, \mu_1 - \mu_3, \ldots, \mu_{k-1} - \mu_k$.

Bonferroni confidence interval for $\mu_i - \mu_j$:

$$\bar{x}_i - \bar{x}_j \pm t_{\frac{\alpha}{2m};N-k} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}},$$

where s_p is the estimate of the common but unknown standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{N - k}}$$

Example (see ANOVA handout):
Construct a Bonferroni confidence interval for $\mu_1 - \mu_4$ using 95% simultaneous confidence level. The data are $\bar{x}_1 = 4.062, \bar{x}_4 = 3.920, n_1 = 10, n_2 = 10, s_p = 0.06061$. We also need $t_{\frac{0.05}{2m};63} = t_{\frac{0.05}{2\times 21};63} = t_{0.00119;63} = 3.1663$.

$$\mu_1 - \mu_4 \in 4.062 - 3.920 \pm 3.1663(0.06061)\sqrt{\frac{1}{10} + \frac{1}{10}},$$

or

$$\mu_1 - \mu_4 \in 0.142 \pm 0.086,$$

or

$$0.056 \leq \mu_1 - \mu_4 \leq 0.228.$$