Tukey's Method

These slides cover Tukey's method for all pairwise comparisons of means in a balanced single-factor ANOVA.

Suppose $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1)$.

Suppose $U \sim \chi^2_m$ and U is independent of $Z_1, \ldots, Z_k$.

Let

$$R = \frac{\max_{i=1,\ldots,k} Z_i - \min_{i=1,\ldots,k} Z_i}{\sqrt{U/m}}.$$

The distribution of R is known as the distribution of the studentized range.

Let $R_{k,m,\alpha}$ denote the upper α quantile of the distribution of the studentized range so that

$$\mathbb{P}(R \leq R_{k,m,\alpha}) = 1 - \alpha.$$

Now suppose

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

where the ε_{ij} terms are iid $N(0, \sigma^2)$ and $i = 1, \ldots, t$ and $j = 1, \ldots, n$.

Prove that

$$\mathbb{P}\left[(\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) - \frac{\hat{\sigma}}{\sqrt{n}} R_{t,t(n-1),\alpha} \leq \mu_i - \mu_{i^*} \right.$$
$$\left. \leq (\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) + \frac{\hat{\sigma}}{\sqrt{n}} R_{t,t(n-1),\alpha} \quad \forall\, i \neq i^* \right] = 1 - \alpha.$$

Proof:

$$(\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) - \frac{\hat{\sigma}}{\sqrt{n}} R_{t,t(n-1),\alpha} \le \mu_i - \mu_{i^*} \le (\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) + \frac{\hat{\sigma}}{\sqrt{n}} R_{t,t(n-1),\alpha}$$

$$\forall\ i \neq i^*$$

$$\iff -R_{t,t(n-1),\alpha} \le \frac{(\mu_i - \mu_{i^*}) - (\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot})}{\hat{\sigma}/\sqrt{n}} \le R_{t,t(n-1),\alpha} \quad \forall\ i \neq i^*$$

$$\iff \left| \frac{(\mu_i - \mu_{i^*}) - (\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot})}{\hat{\sigma}/\sqrt{n}} \right| \le R_{t,t(n-1),\alpha} \quad \forall\ i \neq i^*$$

$$\iff \left| \frac{(\bar{y}_{i\cdot} - \mu_i) - (\bar{y}_{i^*\cdot} - \mu_{i^*})}{\hat{\sigma}/\sqrt{n}} \right| \leq R_{t,t(n-1),\alpha} \quad \forall\ i \neq i^*$$

$$\iff \left| \frac{\frac{\sqrt{n}(\bar{y}_{i\cdot} - \mu_i)}{\sigma} - \frac{\sqrt{n}(\bar{y}_{i^*\cdot} - \mu_{i^*})}{\sigma}}{\sqrt{\hat{\sigma}^2/\sigma^2}} \right| \leq R_{t,t(n-1),\alpha} \quad \forall\ i \neq i^*$$

$$\iff \frac{\max_{i=1,\ldots,t} \frac{\sqrt{n}(\bar{y}_{i\cdot} - \mu_i)}{\sigma} - \min_{i=1,\ldots,t} \frac{\sqrt{n}(\bar{y}_{i\cdot} - \mu_i)}{\sigma}}{\sqrt{\hat{\sigma}^2/\sigma^2}} \leq R_{t,t(n-1),\alpha}.$$

$$\frac{\sqrt{n}(\bar{y}_{1\cdot} - \mu_1)}{\sigma}, \ldots, \frac{\sqrt{n}(\bar{y}_{t\cdot} - \mu_t)}{\sigma} \overset{iid}{\sim} N(0,1)$$

independent of

$$\hat{\sigma}^2/\sigma^2 \sim \chi^2_{t(n-1)}/[t(n-1)].$$

Thus, the result follows. $\square$

The intervals

$$(\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) \pm \frac{\hat{\sigma}}{\sqrt{n}} R_{t,t(n-1),\alpha} \quad i \neq i^*$$

are Tukey's $100(1-\alpha)\%$ simultaneous confidence intervals for all possible pairwise differences $\mu_i - \mu_{i^*},\ i \neq i^*$.

The previous result shows that these intervals have simultaneous coverage $1 - \alpha$.

For the unbalanced case

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, t;\ j = 1, \ldots, n_i,$$

the Tukey-Kramer intervals

$$(\bar{y}_{i\cdot} - \bar{y}_{i^*\cdot}) \pm \hat{\sigma}\sqrt{\frac{1/n_i + 1/n_{i^*}}{2}} R_{t,t(n-1),\alpha} \quad i \neq i^*$$

have simultaneous coverage at least $1 - \alpha$.

(See Hayter, A.J. (1984). Annals of Statistics.12, 61-75.)