Solutions

TAMS38 – Experimental Design and Biostatistics, 6 hp
April 24 2019, 8–12

1. (a) Oat variety i and area j give observation y_{ij} where $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, $\sum_{i=1}^{6} \tau_i = 0$, $\sum_{j=1}^{6} \beta_j = 0$ and $\varepsilon_{ij} \sim N(0, \sigma)$ independently.

   (b) Pairwise comparisons of τ_i's with Tukeys method gives

   $$I_{\tau_i - \tau_k} = \left(\bar{y}_{i\cdot} - \bar{y}_{k\cdot} \mp q_{0.05}(6, 25) \frac{s}{\sqrt{6}} \right) = (\bar{y}_{i\cdot} - \bar{y}_{k\cdot} \mp 0.796),$$

   where $s^2 = \dfrac{SS_E}{25} = 0.2000$, $s = 0.4472$ and $q_{0.05}(6, 25) = 4.36$. One can see that 5 is not as good as the others, 2, 3 and 4 seems to be quite similar, but not as good as 1 which is not as good as 6.

   (c) Estimate the expected value

   $$\hat{\mu} + \hat{\tau}_6 + \hat{\beta}_1 = \bar{y}_{\cdot\cdot} + (\bar{y}_{6\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot}) = 21.172 + 17.753 - 18.166 = 20.759.$$

2. (a) One has randomly selected four batches from a large number. The purity from batch i is $\mu + \tau_i$, where μ is the overall mean for all possible batches and τ_i is a random variable which defines the variation between the batches. In this case a random effects model is suitable.

   (b) Test the hypothesis $H_0 : \sigma_\tau = 0$ versus $H_1 : \sigma_\tau \neq 0$, with the statistic $v = \dfrac{SS_{TREAT}/(4-1)}{SS_E/(12-4)} = 38.4$. Reject H_0 if $v > c = F_{0.95}(3, 8) = 4.07$. Hence, reject H_0, and it seems to be variations between the batches.

3. (a) We test curvature using the statistic $v_{PQ} = \dfrac{SS_{PQ}/1}{SS_E/8} = 7.53$. If no curvature the random variable $V_{PQ} \sim F_(1, 8)$ and large values indicates curvature. The critical value from the table is $c = 5.32$, i.e., there seems to exist curvature with high probability.

   (b) Since we have curvature we would like to fit a second order polynomial to the data. For this we need more data and these could be taken in the center point and in $(x_1, x_2, x_3) = ((\pm\sqrt{3}, 0, 0), (0, \pm\sqrt{3}, 0), (0, 0, \pm\sqrt{3}))$.

4. (a) Nr. 5, i.e., factor C, nr. 9, i.e., factor D, nr. 29, i.e., factor CDE and nr. 4, i.e., AB. AB seems not to be so important.

   (b) $Y_{ijkl} = \underbrace{\mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk}}_{=\mu_{ijk}} + \varepsilon_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}$, where $\varepsilon_{ijkl} \sim N(0, \sigma)$ (independently).
   Construct Tukey intervals $I_{\mu_{ijk} - \mu_{lmr}}$ with simultaneous confidence level 95%.

Estimate μ_{ijk} with $\hat{\mu}_{ijk} = \bar{y}_{ijk\cdot}$ which are independent observations from $\bar{Y}_{ijk\cdot} \sim N\left(\mu_{ijk}, \dfrac{\sigma}{\sqrt{4}}\right)$.

Estimate σ^2 with $s^2 = \dfrac{SS_E}{24} = 0.02458$, i.e., $s = 0.1568$ with $df = 24$ degrees of freedom.

Hence, the confidence intervals are given by

$$I_{\mu_{ijk}-\mu_{lmr}} = \left(\bar{y}_{ijk\cdot} - \bar{y}_{lmr\cdot} \mp \underbrace{q_{0.05}(8,24)}_{=4.68} \frac{s}{\sqrt{4}}\right) = \Big(\bar{y}_{ijk\cdot} - \bar{y}_{lmr\cdot} \mp 0.3669\Big).$$

No significant difference between $(-1,-1,-1)$, $(-1,-1,1)$ and $(-1,1,1)$, but they are all smaller than $(1,i,j)$ when $i = \pm 1$ and $j = \pm 1$.

(c) $de, ae, b, abd, cd, ac, bce, abcde$

5. (a) We have $\hat{\mu} = 38.9812$, $\hat{\tau}_1 = 5.5563$, $\hat{\beta}_1 = -20.0062 = -\hat{\beta}_{-1}$, i.e., $\hat{\beta}_{-1} = 20.0062$, $\hat{\gamma}_1 = 15.7312$ and $\hat{\delta}_1 = 0.2187$.

This gives the largest estimated mean as

$$\hat{\theta} = \widehat{\mathrm{E}(Y_{1,-1,1,1})} = \hat{\mu} + \hat{\tau}_1 + \hat{\beta}_{-1} + \hat{\gamma}_1 + \hat{\delta}_1 = 80.4936$$

(b) The random variables $\hat{\mu}$, $\hat{\tau}_1$, $\hat{\beta}_{-1}$, $\hat{\gamma}_1$ and $\hat{\delta}_1$ are independent and normal distributed each with variance $\dfrac{\sigma^2}{2^4}$ which gives $\mathrm{var}(\hat{\theta}) = \dfrac{5}{16}\sigma^2$.

Estimate σ^2 with $s^2 = \dfrac{SS_E}{11} = 107.87$, i.e., $s = 10.386$ with $df = 11$ degrees of freedom.

Let $\theta = \mathrm{E}(\hat{\theta}) = \mu + \tau_1 + \beta_{-1} + \gamma_1 + \delta_1$. Then,

$$\frac{\hat{\theta} - \theta}{s\sqrt{5/16}} \sim t(11)$$

gives the interval

$$I_\theta = \left(\hat{\theta} \mp \underbrace{t_{0.975}(11)}_{=2.20}\, s\sqrt{\frac{5}{16}}\right) = (80.4936 \mp 12.773) = (67.72, 93.27)\,.$$