Solutions

TAMS38 – Experimental Design and Biostatistics, 6 hp
August 20 2019, 14–18

1. (a) Fertilizer $i = 1, 2, 3$ and observation $j = 1, ..., n_i$ give y_{ij} where $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $\sum_{i=1}^{3} n_i \tau_i = 0$ and $\varepsilon_{ij} \sim N(0, \sigma)$ independently.
Test the hypotesis

$$H_0: \ \mu_1 = \mu_2 = \mu_3 \ \Leftrightarrow \ \tau_1 = \tau_2 = \tau_3 = 0$$
$$\text{versus}$$
$$H_1: \ \mu_i \neq \mu_j \text{ at least one } (i, j) \ \Leftrightarrow \ \text{ at least one } \tau_i \neq 0$$

Reject the hypothesis, since $v = \dfrac{SS_F/2}{SS_E/15} = 21.48 > 3.685 = F_{0.95}(2, 15)$, where

$$SS_F = \sum_{i=1}^{3} n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = 171.137,$$
$$SS_E = \sum_{i=1}^{3} \sum_{l=1}^{n_i} (y_{il} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{3} (n_i - 1) s_i^2 = 59.754.$$

(b) Test the hypothesis $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 \neq \sigma_2$ at level 5%. Use the test statistic

$$v = \frac{s_1^2}{s_2^2} = 9.62.$$

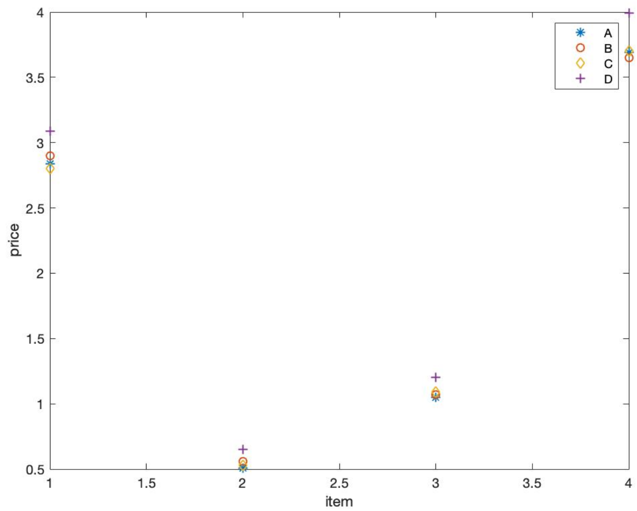
Reject the hypothesis if $v < F_{0.025}(6, 4) < 1$ or $v > F_{0.975}(6, 4) > 1$. Since, $F_{0.975}(6, 4) = 9.20$ we can reject H_0, i.e., the standard deviations seems to differ.

2. The Kruskal-Wallis test. Use the test statistic

$$T = \frac{12 S_a}{N(N+1)} - 3(N+1) = 13.09$$

where $N = \sum_{i=1}^{a} n_i$, $S_a = \sum_{i=1}^{a} \frac{s_i^2}{n_i}$, $s_i = \sum_{j=1}^{n_i} r_{ij}$ and r_{ij} is the rank for observation y_{ij}. Reject the hypothesis, that the effect by the fertilizers are equal, if $T > \chi^2_{0.95}(2) = 5.99$ (we have *large* $n_i > 5$), i.e., reject the hypothesis. The fertilizers seems to be diffferent.

3. (a) To show prices differences among the four supermarkets, we can graph the prices on the y-axis and graph one of the two background variables, say item, on the x-axis. Use a different plotting symbol for the four supermarkets. One can see that the prices at supermarket D are consistently the highest.



(b) $y_{ij(k)}$ are observations from $Y_{ij(k)} = \mu + \tau_i + \beta_j + \gamma_k + \varepsilon_{ij(k)}$, where $\sum_i \tau_i = 0, \sum_j \beta_j = 0, \sum_k \gamma_k = 0$ and $\varepsilon_{ij(k)} \sim N(0, \sigma)$ independently.

We test the hypothesis $H_0 : \gamma_k = 0$ for all k using the test statistic

$$v_{Market} = \frac{SS_{Market}/(p-1)}{SS_E/[(p-2)(p-1)]} = \frac{0.1210/3}{0.0127/6} = 19.1 > F_{0.95}(3,6) = 4.76.$$

Hence, we reject H_0, i.e., the market seems to have different prices.

(c) The Tukey's simultaneous confidence intervals are given by

$$I_{\gamma_i - \gamma_j} = \left(\bar{y}_{\cdot\cdot i} - \bar{y}_{\cdot\cdot j} \mp \underbrace{q_{0.05}(4,6)}_{=4.90} \frac{s}{\sqrt{4}}\right) = \left(\bar{y}_{\cdot\cdot i} - \bar{y}_{\cdot\cdot j} \mp 0.1127\right),$$

where $s^2 = \dfrac{SS_E}{(p-2)(p-1)} = 0.0021$ and $\bar{y}_{\cdot\cdot i} = 2.0225,\ 2.0450,\ 2.0300,\ 2.2325$ for supermarket A, B, C and D, respectively. These intervals indicate that supermarket A, B and C appear to be the same but all of them less price than supermarket D.

4. (a) It seems like it effects number 5 (factor C) and number 9 (factor D) are the most important.

The E-effect are together with ABCD, i.e., number 16.

(b) Model: $Y_{ijk} = \mu + \gamma_i + \delta_j + \varepsilon_{ijk}$, where $\varepsilon_{ijk} \sim N(0, \sigma)$ independently. γ_i is for the C-level i and δ_j for the D-level j. Derive $I_{\gamma_1 - \gamma_{-1}}$ and $I_{\delta_1 - \delta_{-1}}$ with confidence level 95% each.

$\gamma_1 - \gamma_{-1}$ is estimated by $\hat{\gamma}_1 - \hat{\gamma}_{-1} = \bar{y}_{1\cdot\cdot} - \bar{y}_{-1\cdot\cdot}$ which is an observation from $\bar{Y}_{1\cdot\cdot} - \bar{Y}_{-1\cdot\cdot} \sim N\left(\gamma_1 - \gamma_{-1}, \sigma\sqrt{\dfrac{1}{8} + \dfrac{1}{8}}\right)$. Use the statistic

$$\frac{\bar{Y}_{1\cdot\cdot} - \bar{Y}_{-1\cdot\cdot} - (\gamma_1 - \gamma_{-1})}{S/2} \sim t(13),$$

which gives the interval

$I_{\gamma_1-\gamma_{-1}} = \left(\bar{y}_{1\cdot\cdot} - \bar{y}_{-1\cdot\cdot} \mp t\,\frac{s}{2}\right)$, where $t = t_{0.975}(13) = 2.16$ and $s = \sqrt{\frac{SS_E}{13}} = 1.83$, i.e.,

$I_{\gamma_1-\gamma_{-1}} = \left(21.325 - 14.575 \mp 2.16\frac{1.83}{2}\right) = (4.77\ ,\ 8.73)$.

In the same way $I_{\delta_1-\delta_{-1}} = (2.75\ ,\ 6.70)$.

Hence, choose C=D=1.

5. Derive the interval $I_{\sigma_\tau^2/(\sigma_\tau^2+\sigma^2)}$. Use the statistic

$$\frac{SS_{TREAT}/((n\sigma_\tau^2+\sigma^2)(a-1))}{SS_E/(\sigma^2(N-a))} = \frac{MS_{TREAT}}{MS_E}\frac{\sigma^2}{(n\sigma_\tau^2+\sigma^2)} \sim F(a-1, N-a),$$

with $N = a \cdot n$. Hence, we have

$$P\left(b_0 \le \frac{MS_{TREAT}}{MS_E}\frac{\sigma^2}{(n\sigma_\tau^2+\sigma^2)} \le b_1\right) = 95\%,$$

where $b_0 = F_{0.025}(a-1, N-a) = 1/F_{0.975}(N-a, a-1)$ and $b_1 = F_{0.975}(a-1, N-a)$. If we solve this for $\sigma_\tau^2/(\sigma_\tau^2+\sigma^2)$ we get

$$P\left(\frac{L}{1+L} \le \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \le \frac{U}{1+U}\right) = 95\%,$$

where

$$L = \frac{1}{n}\left(\frac{MS_{TREAT}}{MS_E}\frac{1}{b_1} - 1\right) \quad \text{och} \quad U = \frac{1}{n}\left(\frac{MS_{TREAT}}{MS_E}\frac{1}{b_0} - 1\right).$$