Solutions

TAMS38 – Experimental Design and Biostatistics, 6 hp
January 17 2019, 14–18

1. (a) y_{ij} is an observation from $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, for $i = 1, ..., 4$ and $j = 1, ..., n_i$ with $n_1 = 2, n_2 = 2, n_3 = 3, n_4 = 4$ and where $\varepsilon_{ij} \sim N(0, \sigma)$ independently and $\sum_{i=1}^{4} \tau_i = 0$. Test the hypotesis $H_0 : \tau_1 = ... = \tau_4 = 0$ vs. H_1 : not H_0 at level 1%. Test statistic

$$v = \frac{SS_A/(4-1)}{SS_E/(n_1 + ... + n_4 - 4)} = \frac{SS_A/3}{SS_E/7} \sim F(3, 7) \quad \text{under } H_0.$$

Reject H_0 if $v = 8.90 > c = F_{0.01}(3, 7) = 8.45 \quad \Rightarrow \quad$ Reject H_0.

(b) Kruskal-Wallis test. Test the hypothesis using the statistic

$$T = \frac{12 S_4}{N(N+1)} - 3(N+1),$$

where $s_i = \sum_{j=1}^{n_i} r_{ij}$, $S_4 = \sum_{i=1}^{4} \frac{s_i^2}{n_i} = 498$, and $N = \sum_{i=1}^{4} n_i = 11$. The statistic is $T \approx \chi^2(4-1) = \chi^2(3)$ under the hypothesis.

Reject the hypothesis if the test statistic $T > \chi^2_{0.99}(4-1) = 11.32$. But since $T = 9.27$, we can not reject the hypothesis. Compare with a)

2. (a) y_{ij} is an observation from $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, for $i = 1, ..., 4$ and $j = 1, ..., 7$, where$\sum_{i=1}^{4} \tau_i = 0$, $\sum_{j=1}^{7} \beta_j = 0$, $\varepsilon_{ij} \sim N(0, \sigma)$ independently. The model is analysed with analysis (ii).

(b) Test the hypotesis $H_0 : \tau_1 = ... = \tau_4 = 0$ vs. H_1 : not H_0 at level 5%. Test statistic

$$v = \frac{SS_A/3}{SS_E/18} \sim F(3, 18) \quad \text{under } H_0.$$

Reject H_0 if $v = 16.5 > c = F_{0.05}(3, 18) = 3.16 \quad \Rightarrow \quad$ Reject H_0, the gas mixture seems to effect the walking distance.

(c) Construct Tukey-intervals

$$I_{\beta_j - \beta_k} = \left(\bar{y}_{\cdot j} - \bar{y}_{\cdot k} \mp q_{0.05} \frac{s}{\sqrt{7}} \right) = \left(\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \mp 45.52 \right),$$

where $q_{0.05}(4, 18) = 4.00$ and $s = 30.11$. Hence, $\beta_3 < \beta_1, \beta_2, \beta_4$, i.e., air + CO seems to give the shortest distance.

3. (a) Test the hypothesis $H_0 : \sigma_\tau^2 = 0$ vs. $H_1 : \sigma_\tau^2 \neq 0$ using the test statistic $v = \dfrac{SS_M/3}{SS_E/16} = 20.87$. Under H_0 we have $V \sim F(3, 16)$, i.e., reject H_0 if $v > c = F_{0.95}(3, 16) = 3.24$. Hence, reject H_0. The machines are probably **not** homogenous.

(b) Let $z_i = \bar{y}_{i\cdot}$ which are observations from $Z_i = \bar{Y}_{i\cdot} \sim N(\mu, \sigma_z)$. Estimate μ with $\hat{\mu} = \bar{z} = 1.2045$ and σ_z^2 with $s_z^2 = \dfrac{1}{4-1} \sum_{i=1}^{4} (z_i - \bar{z})^2 = 0.000313$.

Hence, we have $\dfrac{\bar{Z} - \mu}{S/\sqrt{4}} \sim t(3)$ which gives

$I_\mu = \left(\bar{z} \mp t \dfrac{s}{\sqrt{4}} \right)$, where $t = t_{0.975}(3) = 3.18$, i.e., $I_\mu = (1.18\ ,\ 1.23)$.

The overall mean machine fills the cans between 1.18 and 1.23. However, there is a large variations among the machines since they are not homogeneous.

4. (a) The most important effects seems to be nr 9, 10 and 2, i.e., D, AD and A.

(b) The model is given $Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \varepsilon_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ with the normal parameter restrictions and $\varepsilon_{ijk} \sim N(0, \sigma)$ independently.

Since the interaction effect between A and D is significant we compare μ_{ij} pairvise. Use Tukeys-method.

$$I_{\mu_{ij} - \mu_{pq}} = \left(\bar{y}_{ij\cdot} - \bar{y}_{pq\cdot} \mp \underbrace{q_{0.05}(4, 12)}_{=4.20} \frac{s}{\sqrt{4}} \right) = (\bar{y}_{ij\cdot} - \bar{y}_{pq\cdot} \mp 87.55),$$

where $s = 41.7$.

The level combination $A = -1$ and $D = 1$ gives significant higher speeds the the others, i.e., choose a small gap and high effect.

5. (a) $\hat{\tau}_2 - \hat{\tau}_1 = \bar{y}_{2\cdot} - \bar{y}_{1\cdot} = 0.0057$ which is an observation form the random variable $\bar{Y}_{2\cdot} - \bar{Y}_{1\cdot} = \mu + \tau_2 + \bar{\beta}_\cdot + \bar{\varepsilon}_{2\cdot} - (\mu + \tau_1 + \bar{\beta}_\cdot + \bar{\varepsilon}_{1\cdot}) = \tau_2 - \tau_1 + \bar{\varepsilon}_{2\cdot} - \bar{\varepsilon}_{1\cdot}$.
$\mathrm{E}(\bar{Y}_{2\cdot} - \bar{Y}_{1\cdot}) = \tau_2 - \tau_1 + \mathrm{E}(\bar{\varepsilon}_{2\cdot} - \bar{\varepsilon}_{1\cdot}) = \tau_2 - \tau_1$ which is unbiased since $\mathrm{E}(\varepsilon_{ij}) = 0$ gives $\mathrm{E}(\bar{\varepsilon}_{i\cdot}) = 0$.

(b) We have $SS_B = \sum_{i=1}^{2} \sum_{j=1}^{7} (Y_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 = 2 \sum_{j=1}^{7} (\mu + 0 + \beta_j + \bar{\varepsilon}_{\cdot j} - (\mu + 0 + \bar{\beta}_\cdot + \bar{\varepsilon}_{\cdot\cdot}))^2 = 2 \sum_{j=1}^{7} (\nu_j - \bar{\nu})^2$, where $\nu_j = \beta_j + \bar{\varepsilon}_{\cdot j} \sim N\left(0, \sqrt{\sigma_\beta^2 + \dfrac{\sigma^2}{2}}\right)$.

Let $\sigma_\nu^2 = \sigma_\beta^2 + \dfrac{\sigma^2}{2}$.

Then we have $\mathrm{E}(SS_B) = 2(7-1)\sigma_\nu^2 = 12\left(\sigma_\beta^2 + \dfrac{\sigma^2}{2}\right)$ and $\mathrm{E}(SS_E) = 6\sigma^2$.

Hence, an unbiased estimator of σ_β^2 is given by $\hat{\sigma}_\beta^2 = \dfrac{SS_B}{12} - \dfrac{SS_E}{12} = 0.000317$ since $\mathrm{E}\left(\hat{\sigma}_\beta^2\right) = \dfrac{\mathrm{E}(SS_B)}{12} - \dfrac{\mathrm{E}(SS_E)}{12} = \sigma_\beta^2 + \dfrac{\sigma^2}{2} - \dfrac{\sigma^2}{2} = \sigma_\beta^2$.