ANSWERS

TAMS38 – Experimental Design and Biostatistics, 4 p / 6 hp Examination on 19 April 2017 kl 8–12.

1 a) Model: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ where $\varepsilon_{ij} \sim N(0, \sigma)$ (independent) and $\tau_i \sim N(0, \sigma_\tau)$ (independent) for each $i = 1, ..., 5$.

$SS_E = 4 \cdot (0.0687^2 + 0.0630^2 + 0.0688^2 + 0.0652^2 + 0.0650^2) = 0.087600$
$SS_{TR} = 5 \cdot ((23.458 - 23.444)^2 + (23.492 - 23.444)^2 + (23.524 - 23.444)^2 + (23.380 - 23.444)^2 + (23.364 - 23.444)^2) = 0.0970$

$\hat{\sigma}^2 = s^2 = 0.087600/(25 - 5) = 0.004380$
$E(SS_{TR}) = (a-1)(n\sigma_\tau^2 + \sigma^2)$, so $(E(SS_{TR}) - (a-1)\sigma^2)/(N-a) = \sigma_\tau^2$
$\hat{\sigma}_\tau^2 = (0.0970 - 4 \cdot 0.004380)/(25 - 5) = 0.0040$

b) $H_0 : \sigma_\tau = 0$ mot $H_1 : \sigma_\tau \neq 0$

Test statistic:

$$v = \frac{0.0970/(5-1)}{0.087600/(25-5)} = 5.54$$

$F_{0.95}(4, 20) \approx 2.87$. $v > 2.87$ so we can reject H_0 on $\alpha = 5\%$. There is a difference between calcium content for batches.

c) We want to test equality of variances for batch 2 and 3 i.e., $H_0 : \sigma_2 = \sigma_3$ mot $H_1 : \sigma_2 \neq \sigma_3$ on level 5%.
Test statistic

$$v = \frac{s_2^2}{s_3^2} = 0.063^2/0.0688^2 = 0.8385.$$

Underlying random variable $V \sim F(4, 4)$ under H_0, so we reject H_0 if $v < a$ or if $v > b$ where a and b come from F-table.
$b = F_{0.975}(4, 4) = 9.61$ and $a = \dfrac{1}{F_{0.975}(4,4)} = \dfrac{1}{9.61} = 0.1041$, i.e., we cannot reject H_0. Batch 2 and 3 have the same variance with probability 95%.

2 a) Model $Y_{ijk} = \mu + \tau_i + \beta_j + \gamma_j + \varepsilon_{ij(k)}$, where $\varepsilon_{ij(k)} \sim N(0, \sigma)$ (independent), τ_i denotes method i, β_i denotes operator j, γ_k denotes order of assembly k. Moreover $\sum_{i=1}^4 \tau_i = 0$, $\sum_{j=1}^4 \beta_j = 0$, $\sum_{k=1}^4 \gamma_k = 0$.

b) $H_0 : \tau_1 = ... = \tau_4 = 0$ against $H_0 : \tau_i \neq 0$ for some i.
F-test statistic for method = (72.5/3)/(10.5/6)=13.81 and $F_{0.95}(3, 6) = 4.76$. We should reject H_0 on $\alpha = 5\%$. With high probability there is difference between methods regarding assembly time.

c) One can do for example two Tukey-intervals each at level of 5% as follows

$$I_{\tau_i-\tau_j} = \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{j\cdot\cdot} \mp \underbrace{q_{0.95}(4,6)}_{=4.896}\frac{s}{\sqrt{4}}\right) = \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{j\cdot\cdot} \mp 3.2384\right)$$

where $s^2 = 10.5/6 = 1.75$. Method 1 is better than method 3 and 4. No significant difference between method 1 and 2.

$$I_{\beta_i-\beta_j} = \left(\bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot j\cdot} \mp \underbrace{q_{0.95}(4,6)}_{=4.896}\frac{s}{\sqrt{4}}\right) = \left(\bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot j\cdot} \mp 3.2384\right)$$

Operator 1 and 4 are better than operator nr 2. No significant difference between operators 1, 4 and 3.

3 a) We test H_0 : maximal model and logistic regression model are equally good against
H_1 : maximal model is better.
Test statistics *deviance* = 6.39 and the corresponding r.v. is $\chi^2(6)$-distributed. Critical region is given by $c = \chi^2_{0.95}(6) = 12.6$. As $6.39 < 12.6$, we cannot reject H_0 i.e., logistic regression model is ok.

b) $I_{\beta_1} = (\hat{\beta}_1 \mp \underbrace{z_{0.975}}_{=1.96} d(\hat{\beta}_1)) = (-0.2694 \mp 1.96 \cdot 0.055) = (-0.2694 \mp 0.1078) < 0$

c) $H_0 : \beta_2 = 0$ against $H_0 : \beta_2 \neq 0$.
Test statistic is $z = 0.831/0.134 = 6.2$ and critical region is $(-\infty, -1.96) \cup (1.96, \infty)$. We should reject H_0 on $\alpha = 5\%$. With high probability the new variable is significant.

4 a) nr 3 -B, nr 5 -C, nr 7- BC.

b) No. We would have $df_E = 0$.

c) $\widehat{E(Y_{111})} = \hat{\mu} + \hat{\tau}_1 + \hat{\beta}_1 + \hat{\gamma}_1 = 24.125 + 0.375 + 11.375 + 10.750 = \underline{\underline{46.625}}$.

d) $s = \sqrt{SS_E/df_E} = \sqrt{152.13/3} = 7.12$

t-interval gives

$$I_{\tau_1-\tau_{-1}} = \left(\bar{y}_{1\cdot\cdot} - \bar{y}_{-1\cdot\cdot} \mp \underbrace{t_{0.995}(3)}_{=5.84} s\sqrt{\frac{1}{4}+\frac{1}{4}}\right) = \left(23.75 - 24.5 \mp 29.4\right) \ni 0$$

No significance difference between results for $A = 1$ and $A = -1$.

e) Tukey-interval gives

$$I_{\mu_{jk}-\mu_{mn}} = \left(\bar{y}_{\cdot jk} - \bar{y}_{\cdot mn} \mp \underbrace{q_{0.05}(4,3)}_{=6.825} s\frac{1}{\sqrt{2}}\right) = \left(\bar{y}_{\cdot jk} - \bar{y}_{\cdot mn} \mp 34.36\right).$$

The highest value is obtained for $B = 1, C = 1$.

f) $1 - \alpha_{sim} \geq 1 - 5\% - 1\% = 94\%$