ANSWERS

TAMS38 – Experimental Design and Biostatistics, 4 p / 6 hp
Examination on 10 January 2017 kl 14–18.

1 a)

$$I_{\mu_1-\mu_2} = \left(\bar{y}_{1\cdot} - \bar{y}_{2\cdot} \pm z_{0.975} \cdot \sigma\sqrt{\frac{1}{n}+\frac{1}{n}}\right)$$

Length of interval is

$$2 \cdot \underbrace{z_{0.975}}_{1.96} \cdot \sigma\sqrt{\frac{1}{n}+\frac{1}{n}} \leq 3$$

Then

$$\sqrt{\frac{2}{n}} \leq 0.638$$
$$n \geq 4.92$$

Answer: $n \geq 5$.

b) $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$

$$s^2 = \frac{7 \cdot s_1^2 + 7 \cdot s_2^2}{14} = 0.1284$$

$$t = \frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{9.925 - 10.3625}{\sqrt{0.1284}\sqrt{\frac{1}{8}+\frac{1}{8}}} = \frac{-0.4375}{0.179} = -2.44$$

Test statistic is $t(14)-$distributed, so critical region is $C = (-\infty, -t_{0.975}) \cup (t_{0.975}, \infty) = (-\infty, -2.14) \cup (2.14, \infty)$. As $t \in C$ we can reject H_0 on $\alpha = 0.05$. There is significant difference between Solution 1 and 2 on $\alpha = 0.05$.

2 a) factor1=bolt, factor2=chemical.
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ against $H_0 : \beta_i \neq 0$ for some i.
F-test statistic for chemical = 2.92 and $F_{0.9}(3, 12) = 2.61$. We should reject H_0 on $\alpha = 0.10$. With high probability there is difference between chemical agents.

```
Analysis of Variance for strength
Source    DF       SS      MS      F      P
bolt       4  153.700  38.425  21.85  0.000
chemical   3   15.400   5.133   2.92  0.078
Error     12   21.100   1.758
Total     19  190.200

S = 1.32602   R-Sq = 88.91%   R-Sq(adj) = 82.44%

Means
bolt  N  strength
```

```
1     4    73.250
2     4    68.500
3     4    75.500
4     4    72.750
5     4    68.500

chemical  N  strength
1         5    70.400
2         5    71.400
3         5    72.400
4         5    72.600
```

b) H_0 : No difference between chemical agents.
against
H_1 : There is difference between chemical agents.
Test statistic =7.34. $\chi^2(3) = 6.25$. We should reject H_0 on $\alpha = 0.10$. With high probability there is difference between chemical agents.

```
Friedman Test: strength versus chemical blocked by bolt
S = 6.90  DF = 3  P = 0.075
S = 7.34  DF = 3  P = 0.062 (adjusted for ties)

                          Sum of
chemical  N  Est Median   Ranks
1         5      71.125     6.5
2         5      72.375    12.0
3         5      72.875    15.5
4         5      73.125    16.0

Grand median = 72.375
```

c) $I_{\theta_1-\theta_4} = (A_{(k)}, A_{(15-k+1)}) = (A_{(2)}, A_{(14)}) = (-3.5, -1)$ as $k = 2$ using Wilcoxons sign rank test on $\alpha = 14\%$.

```
Wilcoxon Signed Rank CI: C24

                                 Confidence
        Estimated    Achieved    Interval
     N     Median  Confidence  Lower  Upper
C24  5      -2.00        89.4  -3.50  -1.00
```

d) We need 3 dummy variables, f.ex.

$$z_1 = \begin{cases} 1 & \text{for Chemical agent 1} \\ 0 & \text{for otherwise} \end{cases}$$

$$z_2 = \begin{cases} 1 & \text{for Chemical agent 2} \\ 0 & \text{for otherwise} \end{cases}$$

$$z_3 = \begin{cases} 1 & \text{for Chemical agent 3} \\ 0 & \text{for otherwise} \end{cases}$$

3 a) We test $H_0 : \Delta = 0$ (no curvature) against $H_1 : \Delta \neq 0$ (curvature). Test statistics $v_{PQ} = \dfrac{SS_{PQ}/1}{SS_E/4} = 17.62$. Den s.v. V_{PQ} är $F(1,4)$ is there is no curvature. Critical region is given by $c = F_{0.95}(1,4) = 7.71$. $v_{PQ} > c$, i.e. there is curvature.

$SS_{PQ} = \dfrac{(\bar{y}_F - \bar{y}_C)^2}{\frac{1}{8} + \frac{1}{5}} = 2486.1,$

where $\bar{y}_F = \dfrac{1}{8}(66+70+78+64+80+70+100+75) = 75.375$, $\bar{y}_C = \dfrac{1}{5}(113+100+118+88+100) = 103.8$

and $SS_E = (5-1)s_C^2 = 564.5$ where s_C^2 is sample variance from the points in point (0,0,0).

b) We should do additional measurements around central point, for example in $(\sqrt{3},0,0)$, $(-\sqrt{3},0,0)$, $(0,0,\sqrt{3})$, $(0,0,-\sqrt{3})$, $(0,\sqrt{3},0)$ and $(0,-\sqrt{3},0)$.

4 a) ii), generators are $I = ABCE$, $I = BCDF$ and $I = ADEF$.

b)

number 2 that is $\hat{\tau}_1 = 25.375$, aliases: A+BCE+ABCDF+DEF

number 9 that is $\hat{\delta}_1 = 18.375$, aliases: D+ABCDE+BCF+AEF

number 8 that is $\widehat{(\tau\beta\gamma)}_{111} = 17.375$, aliases: ABC+E+ADF+BCDEF

number 16 that is $\widehat{(\tau\beta\gamma\delta)}_{1111} = -10.875$, aliases: ABCD+DE+AF+BCEF

c) Model:

$$Y_{ijkl} = \mu_{ijk} + \delta_l + \varepsilon_{ijkl} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \delta_l + \varepsilon_{ijkl},$$

where $\varepsilon_{ijkl} \sim N(0,\sigma)$ (independent) and $\sum_{i=1}^2 \tau_i = 0$, $\sum_{j=1}^2 \beta_j = 0$, $\sum_{k=1}^2 \gamma_k = 0$, $\sum_{l=1}^2 \delta_l = 0$, $\sum_{i=1}^2 (\tau\beta)_{ij} = 0$ for each $j = 1, 2$, $\sum_{j=1}^3 (\tau\beta)_{ij} = 0$ for each $i = 1, 2$, ...

$s = \sqrt{SS_E/df_E} = 21.14$

Tukey-interval gives

$$I_{\mu_{ijk} - \mu_{lmn}} = \left(\bar{y}_{ijk\cdot} - \bar{y}_{lmn\cdot} \mp \underbrace{q_{0.05}(8,7)}_{=5.82} s\frac{1}{\sqrt{2}} \right) = \left(\bar{y}_{ijk\cdot} - \bar{y}_{lmn\cdot} \mp 87 \right).$$

The highest average obtained for $A = 1, B = -1, C = -1$ is significantly better than $A = -1, B = -1, C = -1$ and $A = -1, B = 1, C = -1$. It does not give

significantly better value than the remaining level combinations.
t-interval gives

$$I_{\delta_{-1}-\delta_1} = \left(\bar{y}_{\cdots -1} - \bar{y}_{\cdots 1} \mp \underbrace{t_{0.975}(7)}_{=2.37} s\sqrt{\frac{1}{8}+\frac{1}{8}}\right) = \left(117.25 - 154 \mp 25.1\right) < 0$$

Best combination is given by $D = 1$.

d) $y_{ijkl} = \beta_0 + \beta_1 x_A + \beta_2 x_B + \beta_3 x_C + \beta_4 x_D + \beta_5 x_A x_B + \beta_6 x_A x_C + \beta_7 x_B x_C + \beta_8 x_A x_B x_C + \epsilon_{ijkl}$
One test significance of factors with t-test or do confidence interval for the interesting coefficient. $df_E = 7$ as regression is equivalent to the previous anova analysis. One can also calculate $df_E = 16 - 9 = 7$ as we have 16 observations and 9 coefficients estimated.

```
Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant  135.63     5.29    25.66    0.000
A          25.38     5.29     4.80    0.002  1.00
B          -0.37     5.29    -0.07    0.945  1.00
C          -6.38     5.29    -1.21    0.267  1.00
D          18.38     5.29     3.48    0.010  1.00
A*B        -2.13     5.29    -0.40    0.700  1.00
A*C        -1.37     5.29    -0.26    0.802  1.00
B*C         0.38     5.29     0.07    0.945  1.00
A*B*C      17.38     5.29     3.29    0.013  1.00

Regression Equation
y = 135.63 + 25.38 A - 0.37 B - 6.38 C + 18.38 D - 2.13 A*B - 1.37 A*C
    + 0.38 B*C + 17.38 A*B*C
```