MAS programmes - Statistical Analysis (Autumn Term)

Exercises 10 - SOLUTIONS

1.

(a) There are two factors: *machine* and *outlet* (nested within a machine), both of which should be considered **fixed**. (After all, they have not arisen as a result of sampling from larger populations).

(b) We fit the model

$$y_{ijk} = \mu + \tau_i + \beta_{(i)j} + \epsilon_{(ij)k}$$

$\sum_{i=1}^{a} \tau_i = 0$, $\sum_{j=1}^{b} \beta_{(i)j} = 0$, $\{\epsilon_{(ij)k}\} \sim \text{NID}(0, \sigma^2)$,
$i=1,2,3$, $j=1,2$, $k=1,2,3,4$.

R analysis runs as follows:

```
> reading <- c(40, 45, 30, 50, 35, 35, 55, 60, 40, 40, 50, 45, 55, 50, 60,
    75, 60, 50, 90, 65, 70, 75, 75, 75)/10
> m <- rep(rep(1:3, rep(2, 3)), 4)
> o <- rep(rep(c("L", "R"), 3), 4)
> machine <- factor(m)
> outlet <- factor(o)
> resistance <- data.frame(reading, machine, outlet)
> resistance.aov <- aov(reading ~ machine/outlet, data = resistance)
> summary(resistance.aov)
               Df  Sum Sq  Mean Sq  F value  Pr(>F)
machine         2    0.77   0.3854    0.132   0.877
machine:outlet  3    2.78   0.9271    0.317   0.813
Residuals      18   52.69   2.9271
```

(c) The F-ratio for differences due to machines (derived from lines 1 and 3 of ANOVA table), and the one for differences in outlets, nested within machines (derived from lines 2 and 3), are associated with rather high p-values. Thus, we conclude that the resistances do not vary according to the machine, or the outlet (nested within a machine) used in the production. Nothing from the various residual plots to suggest that the model fits inadequately.

2.

(a) Responses are classified according to the *region* which is **fixed**, the *council* which is **random**, and the *ward* which is also **random**.

(b)

$$y_{ijkl} = \mu + \tau_i + \beta_{(i)j} + \gamma_{(ij)k} + \epsilon_{(ijk)l}$$

where y_{ijkl} is the number of hours of television watched by the l-th individual sampled from the k-th ward nested within the j-th council nested within the i-th region,

μ is the 'overall' mean,
$\{\tau_i\}$ are the *region* effects, with $\sum_{i=1}^{a} \tau_i = 0$,
$\{\beta_{(i)j}\} \sim \text{NID}(0, \sigma_B^2)$ are the council effects, nested within regions
$\{\gamma_{(ij)k}\} \sim \text{NID}(0, \sigma_C^2)$ are the *ward* effects, nested within councils, and
$\{\epsilon_{(ijk)l}\} \sim \text{NID}(0, \sigma^2)$ are the random error terms.

(c)
```
> hours <- c(20, 13.5, 47.5, 34.5, 32.5, 39, 11, 11.5, 41.5, 37.5,
30.5, 17.5,
    31.5, 15, 33.5, 23.5, 27, 22.5, 5, 19.5, 31, 32, 38.5, 31)
> r <- rep(rep(c("SE", "SW"), rep(6, 2)), 2)
> c <- rep(rep(1:3, rep(2, 3)), 4)
> w <- rep(rep(1:2, 6), 2)
> region <- factor(r)
> council <- factor(c)
> ward <- factor(w)
> viewing <- data.frame(hours, region, council, ward)
> viewing.raov <- aov(hours ~ region/council/ward, data = viewing)
> summary(viewing.raov)
                   Df Sum Sq Mean Sq F value Pr(>F)
region              1   46.8    46.8   0.904 0.3603
region:council      4 1673.5   418.4   8.092 0.0021 **
region:council:ward 6  429.1    71.5   1.383 0.2972
Residuals          12  620.4    51.7
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

To test for non-zero region effects, construct an F-ratio from the mean squares of lines 1 and 2 of the ANOVA table.

To test for a non-zero variance for the council effects, compare lines 2 and 3.

And to test for a non-zero variance for the ward effects, compare lines 3 and 4.

The relevant p-values are computed below:

```
> p.value.region <- pf(46.8/418.4, 1, 4, lower.tail = F)
> p.value.council <- pf(418.4/71.5, 4, 6, lower.tail = F)
> p.value.ward <- pf(71.5/51.7, 6, 12, lower.tail = F)
> p.value.region
[1] 0.7548435
>  p.value.council
```

```
[1] 0.02877614
> p.value.ward
[1] 0.297264
```

Councils nested within regions effects appear to have effects with a non-zero variance. Neither regions nor wards seems to play a part in the overall variation in the responses.