H. Spliid, October 20, 2010

## Solution to exercise 10

The design is a Latin Square. The corresponding mathematical model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk};\ \ \varepsilon_{ijk} \in \text{NID}(0, \sigma^2),$$

where

$\alpha_i$ = effect (block effect) from town $i$; $i = 1, \ldots, 4$
$\beta_j$ = effect from type of packaging (emball.) $j$; $j$ = I,...,IV
$\tau_k$ = effect from type of advertising $k$; $k$ = A,...,D
$\varepsilon_{ijk}$ = the measurement errors

In order to facilitate the computations one can code the data with -40:

| | 1 | | 2 | | 3 | | 4 | | $T_{emb}$ | | $T_{rekl}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I | A | 12 | B | 11 | C | 15 | C | 16 | 54 | A | 21 |
| II | B | 10 | C | 5 | D | 9 | A | 11 | 35 | B | 22 |
| III | C | -1 | D | 1 | A | -3 | B | -1 | -4 | C | 21 |
| IV | D | 3 | A | 1 | B | 2 | C | 2 | 8 | D | 29 |
| $T_{by}$ | | 24 | | 18 | | 23 | | 28 | 93 | | 93 |

$$\begin{aligned}
\sum_i \sum_j \sum_k y_{ijk}^2 &= 12^2 + 11^2 + \cdots + 2^2 = 1103 \\
\sum_i \sum_j \sum_k (y_{ijk} - \overline{y}_{\ldots})^2 &= 1103 - 93^2/16 = 562.44 = \text{ssq}_{total} \\
\text{ssq}_{emb} &= \frac{1}{4}(54^2 + 35^2 + (-4)^2 + 8^2) - 93^2/16 = 514.69 \\
\text{ssq}_{town} &= \frac{1}{4}(24^2 + 18^2 + 23^2 + 28^2) - 93^2/16 = 12.69 \\
\text{ssq}_{advert} &= \frac{1}{4}(21^2 + 22^2 + 21^2 + 29^2) - 93^2/16 = 11.19 \\
\text{ssq}_{rest} &= \text{ssq}_{total} - (\text{ssq}_{emb} + \text{ssq}_{town} + \text{ssq}_{advert}) \\
&= 562.44 - 514.69 - 12.69 - 11.19 = 23.87
\end{aligned}$$

ANOVA concerning sales figures for cornflakes

| Source of variation | ssq | $d.f.$ | $s^2$ | F-value |
|---|---|---|---|---|
| Packaging | 514.69 | 3 | 171.56 | 43.11 |
| Town | 12.69 | 3 | 4.23 | 1.06 |
| Advertising | 11.19 | 3 | 3.73 | 0.99 |
| Residual | 23.87 | 6 | 3.98 | |
| Total | 562.44 | 15 | | |

$F(3,6)_{0.99} = 9.78$

We observe that the type of packaging has a significant influence on the sales figures while the other sources of variation do not seem to be of (significant) importance. Our model could therefore be reduced to:

$$Y_{ijk} = \mu + \beta_j + \varepsilon_{ijk};\ \varepsilon \in \text{NID}(0, \sigma^2),$$

An improved residual variance estimate is: $\hat{\sigma}^2 = (12.69 + 11.19 + 23.87)/12 = 3.98 \simeq 2^2$, which has $3 + 3 + 6 = 12$ degrees of freedom.

The mean sale for each of the four types of packaging is estimated giving:

$$\overline{y}_I = 53.50, \quad \overline{y}_{II} = 48.75, \quad \overline{y}_{III} = 39.00, \quad \overline{y}_{IV} = 42.00$$

Finally it could be of interest to identify a 'best' type of packaging, if any. For this purpose Duncans multiple range test or Newman-Keuls test is appropriate. The latter gives:

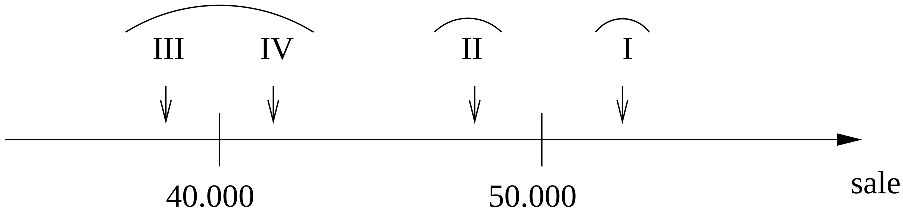
$$f = 12, s^2 = 3.98 \text{ and thus } s_{\overline{y}} = s/\sqrt{4} \simeq 1$$

Using fx level of significance $\alpha = 0.01$ the table over "studentized ranges" for p=2,3,4 and for $f = 12$ gives critical studentized ranges 4.32, 5.04 and 5.50, respectively. Multiplying these with $s_{\overline{y}} \simeq 1$, we get the 1% least significant ranges for the observed means. The result is LSR $= \simeq 4.32, 5.04, 5.50$, for p=2,3,4, respectively.

We now find:

| | | | | |
|---|---|---|---|---|
| $\overline{y}_I - \overline{y}_{III}$ | $=$ | $14.5 > 5.50$ | (p=4) | significant |
| $\overline{y}_I - \overline{y}_{IV}$ | $=$ | $11.5 > 5.04$ | (p=3) | significant |
| $\overline{y}_I - \overline{y}_{II}$ | $=$ | $4.75 > 4.32$ | (p=1) | significant |
| $\overline{y}_{II} - \overline{y}_{III}$ | $=$ | $9.75 > 5.04$ | (p=3) | significant |
| $\overline{y}_{II} - \overline{y}_{IV}$ | $=$ | $6.75 > 4.32$ | (p=2) | significant |
| $\overline{y}_{IV} - \overline{y}_{III}$ | $=$ | $3.00 < 4.32$ | (p=2) | not significant |

Type of packaging:




We conclude that packaging type I is significantly better than the three other types with respect to promoting the sale.

We can construct , for example, a 95% confidence interval for $\mu_I$:

$$I_{0.95}(\mu_I) = \overline{y}_I \pm s_{\overline{y}} \cdot t(12)_{1-0.025} \approx 53.5 \pm 1 \cdot 2.18 = [50.32, 55.68]$$

If one has access to the statistical computing system *SAS*, the exercise can be solved using the following small program

```
title Solution to exercise 10 using SAS;
data cornfl;
do emb=1 to 4;
   do town=1 to 4;
     input advert $ sale @ ;
     output;
   end;
end;
cards;
a 52 b 51 c 55 d 56
b 50 c 45 d 49 a 51
c 39 d 41 a 37 b 39
d 43 a 41 b 42 c 42
;
proc print;
proc glm; class emb town advert;
model     sale= emb town advert;
proc glm; class emb;
model sale=emb;
means emb / snk;
run;
```

A dataset called “cornfl” is defined. The observations are read using two loops. The beginning of the data is indicated by the statement “cards;”

The statement "input ..." reads the variable "advert" (where $ denotes a non-numerical value) and the variable "sale" (being a number).

The @-sign indicates that reading is to be continued on the same line until it is empty.

"output;" saves the data in the dataset "cornfl".

"proc print;" prints the data.

"proc glm;" starts a general analysis of variance and regressions analysis procedure.

"class emb town advert;" declares the three variables to categorical.

"model sale= emb town advert;" declares the proper model to be analysed (3-way additive model).

The last part is repeated, but now using the reduced model, where only packaging is in the model.

The statement "means emb / snk;" asks for a student-Newman-Keuls (snk) test, on the means of the data for the packaging types.