Chemometrics

MA4605

Week 10. Lecture 18. Two-way ANOVA with interactions

November 7, 2011

Two-way ANOVA

- In the previous approach, two factors possibly affecting the results of observed variable have been assumed to enter the model in a linear fashion.
- In particular, an increase in the observed values due to a change of levels of one of the factors, has been assumed independent of the levels set for the other factor.
- In reality, there maybe more complicated relations between factors and their levels - there can be interaction between factors.

Interactions

We can extend two-way ANOVA to include an interaction term.

- Two factors.
- One factor **A** with k levels (called treatments), another factor **B** with b levels (called blocks).
- One factor **AxB** called the interaction term.
- Four sources of variation: treatments, blocks, interaction and experimental variation.

Two-way ANOVA with interactions sum of squares

The total variability is partitioned into four components:

- the variability due to the different treatments (k)
- the variability due to the different blocks(b)
- the variability due to the interaction between treatments and blocks
- the error variability (residuals)

Adding the interacting terms requires replicating the measurements in each cell. Instead of one single measurement in a cell we will need **n**=2,3,... .
Two-way ANOVA is most powerful when the experiment has the same number of replicates in each group defined by the pair of factors. This is called a **balanced design**.

$$
\begin{aligned}
SS_{Total} &= SS_A + SS_B + SS_{AxB} + SS_{Residuals} \\
SS_{Total} &= \sum_{i=1}^{b}\sum_{j=1}^{k}\sum_{l=1}^{n}(y_{ijl} - \overline{\overline{y}})^2 \\
SS_A &= nb\sum_{j=1}^{k}(\overline{y}_{\cdot j} - \overline{\overline{y}})^2 \\
SS_B &= nk\sum_{i=1}^{b}(\overline{y}_{i\cdot} - \overline{\overline{y}})^2 \\
SS_{AxB} &= n\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} - \overline{\overline{y}})^2 \\
SS_{Residual} &= SS_{Total} - SS_A - SS_B - SS_{AxB}
\end{aligned}
$$

Degrees of freedom

Source of variation	Degrees of freedom
Total	n*k*b-1
Between treatments	b-1
Between blocks	k-1
Interactions	(k-1)*(b-1)
Residuals	(n-1)*k*b

Example

In an inter-laboratory collaborative experiment on the determination of arsenic in coal, samples of coal from three different regions were sent to each of three laboratories. Each lab performed a duplicate analysis on each sample.

Sample	1	2	3
A	5.1,5.1	5.3,5.4	5.3,5.1
B	5.8,5.4	5.4,5.9	5.2,5.5
C	6.5,6.1	6.6,6.7	6.5,6.4

Two-way ANOVA with interactions in R

To perform a two-way analysis of variance in *R* we need to code the new table structure.

```
y < - c(5.1,5.1, 5.3,5.4, 5.3,5.1,
5.8,5.4, 5.4,5.9, 5.2,5.5,
6.5,6.1, 6.6,6.7, 6.5,6.4)
A < - factor(rep(1:3,3,each=2))
A
1 1 2 2 3 3 1 1 2 2 3 3 1 1 2 2 3 3
B < - factor(rep(1:3,each=6))
B
1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3
model <- lm(y~A+B+A:B)
anova(model)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	2	0.1878	0.09389	2.3151	0.1545
B	2	5.0678	2.53389	62.4795	5.281e-06
A:B	4	0.1022	0.02556	0.6301	0.6533
Residuals	9	0.3650	0.04056		

— — —

Three separate statistical tests are performed (based on the F statistic),comparing

- the variability due to first factor,
- the variability due to second factor
- the variability due to the interaction between factors

to the error variability.

The error variability is estimated from the within-cell variation and it equals 0.365 with $(n-1)bk = 1 * 3 * 3 = 9$ degrees of freedom. The residual mean square is obtained as the ratio between the residual sum of square and its corresponding degrees of freedom: $MS_{Residuals} = \frac{SS_{Residuals}}{df} = \frac{0.3650}{9} = 0.04056$ Each source of variation is compared with the residual mean square to test whether it is significant.

Interactions effect. The interaction effect is not significant as the *p*-value 0.6533. Since the interaction term is not significant, we can interpret the separate effects of factor A and B.
Between-column effect The effect of factor A is not significant since the p-value $0.1545 > 0.05$.
Between-row effect The effect of factor B is significant since the p-value 5.281e-06 is less than 0.05 making this source of variation very significant.

Interaction plot

We can visually inspect the sample-laboratory interactions from the following plot in which the results are grouped by the two factors: laboratory and sample.