

Anova exercise class

Sylvain

17 Nov 2014

Series 3

In general it was good, no big issue.

I didn't correct it myself, Gian (the new assisting assistant) did. But ask me if anything is unclear.

Some people seemed to have trouble with interactions in exercise 1.

If estimated effect are 0, it doesn't automatically imply that interaction is 0 too (see blackboard).

Series 4, exercise 1

Experiment to study the effect of three feed compositions on the concentration of an hormone in cattle.



Series 4, exercise 1

Load the data:

```

feed <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/feed.txt",header=TRUE)
feed$Feeding <- as.factor(feed$Feeding)
str(feed)

```

```

## 'data.frame': 32 obs. of 3 variables:
## $ Initial: int 207 196 217 210 202 201 214 223 190 220 ...
## $ Final : int 216 199 256 234 203 214 225 255 182 225 ...
## $ Feeding: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 2 ...

```

Series 4, exercise 1

We test three different models:

- 1) model a: $Final_{ij} = \mu + Feeding_i + \epsilon_{ij}$
- 2) model b: $(Final - Initial)_{ij} = \mu + Feeding_i + \epsilon_{ij}$
- 3) model c: $Final_{ij} = \mu + Initial_{ij} + Feeding_i + \epsilon_{ij}$

Do you expect different results?

Series 4, exercise 1

To understand differences between model b and c, notice that model b is equivalent to:

$$Final_{ij} = \mu + 1 \cdot Initial_{ij} + A_i + \epsilon_{ij}$$

There is no *flexible* coefficient $\hat{\beta}$ for the *Initial* covariate. Try to plot *Final* vs. *Initial* and see if you think it makes sense and which model should work best.

To gain additional insight, try to plot the residuals of model b versus the covariate *Initial*. Are the residuals uncorrelated?

Series 4, exercise 2

Back to exercise 3 from series 3, where we looked at the effect of four different factors on the quality of a soft drink. There were 2 replicates of the full factorial design. Let's assume now that we have only the first replicate, so only $2^4 = 16$ observations, and that we have to divide them in two blocks of 8 observations (for example between two different days).

Question: Design the splitting between block such that ABCD is confounded.



Series 4, exercise 2

Read in the data and get first replicate.

```
soft<-read.table("http://stat.ethz.ch/Teaching/Datasets/softdrinkANOVA.txt",header=TRUE)
soft$sugar <- as.factor(soft$sugar)
soft$soda <- as.factor(soft$soda)
soft$water <- as.factor(soft$water)
soft$temp <- as.factor(soft$temp)
soft <- soft[1:16*2-1,] #get first replicate
soft
```

```
##      score sugar soda water temp
## 1      159     1    1     1     1
## 3      168     2    1     1     1
## 5      158     1    1     2     1
## 7      166     2    1     2     1
## 9      175     1    2     1     1
## 11     179     2    2     1     1
## 13     173     1    2     2     1
## 15     179     2    2     2     1
## 17     164     1    1     1     2
```

```
## 19 187 2 1 1 2
## 21 163 1 1 2 2
## 23 185 2 1 2 2
## 25 168 1 2 1 2
## 27 197 2 2 1 2
## 29 170 1 2 2 2
## 31 194 2 2 2 2
```

Series 4, exercise 2

See lecture notes p.53-56 for details on blocking for factorial designs.

Set to -1 the low level and to +1 the high level of each factor.

```
A <- as.numeric(soft[,2])*2-3
B <- as.numeric(soft[,3])*2-3
C <- as.numeric(soft[,4])*2-3
D <- as.numeric(soft[,5])*2-3
```

From this construct the factor BLOCK with value indicating in which block an observation should belong.

Once you have that fit the model $Y \sim \dots + \text{BLOCK}$ and analyse the results.

Series 4, exercise 3

Construct a design to test 5 two-level factors in 8 runs.

A full factorial design means that a replicate would have size $2^5 = 32$. We want to stick to 8, which is a 2^3 design.

We are thus doing a *fractional factorial* design 2^{k-l} with $l = 2$.

How does it work?

Series 4, exercise 3

fractional factorial design 2^{k-l} with $l = 2$, implies that we need 2 confounding relations.

Let's call our 5 factors A , B , C and D .

Now we choose (arbitrarily) $D = AB$ and $E = AC$

Series 4, exercise 3

From this you can construct a fractional factorial design with the construction method II (see lecture notes p.60).

- 1) construct a 2^3 design (with + and - signs).
- 2) compute AB and AC (simple sign multiplication: $++=+$, $+--$, etc).
- 3) identify D with AB .
- 4) identify E with AC .

Series 4, exercise 3

Which effects are confounded with each other? (or what is the aliasing structure?)

From the confounding relations, you know that $D \cong AB$ and $E \cong AC$.

Now because $D \cong AB$, you know that ABD will always be a (+) sign.

A factor with all (+) is the intercept, called I .

So you can write $I \cong ABD \cong ACE$.

From this, you can also see that $I \cong BDCE$.

Series 4, exercise 3

From the confounding relation: $I \cong ABD \cong ACE \cong BCDE$ that we found before, we can find the whole aliasing structure. To do that, multiply by A, B, C , etc.

Remember that $AA = I$.

So, for example, by multiplying the relation by A , you obtain:

$$A \cong BD \cong CE \cong ABCDE$$

Series 4, exercise 3

In a 2^{k-l} design, you know that the relation holds $2^l - 1$ terms (in our case 3).

The *resolution* of the design is the length of the shortest word among these terms.

So what is it in our case?

Series 4, exercise 4

In exercise 4 we study the effect of four factors on the fabric strength in the field of high-speed weaving.

I assume everybody is familiar with *high-speed weaving*.

Series 4, exercise 4

High-speed weaving machine:



Series 4, exercise 4

You are given the following data:

Side-to-side	Yarn type	Pick density	Air pressure	Strength
-	-	-	-	24.50
+	-	-	+	22.05
-	+	-	+	24.52
+	+	-	-	25.00
-	-	+	+	25.68
+	-	+	-	24.51
-	+	+	-	24.68
+	+	+	+	24.23

Series 4, exercise 4

Find k and l for this 2^{k-l} fractional factorial design.

So, what is k ? how many observations are there?

Series 4, exercise 4

Determine the alias structure of this design.

1. The defining relation should have $2^l - 1$ words.
2. Find the relation.
3. Multiply by every possible term to find with what they are aliased.

Series 4, exercise 4

Calculate estimates of the effects. How do we do that? very simple

$$\hat{A} = \bar{y}_{A+} - \bar{y}_{A-}$$

For A , add up all the values with a corresponding $+$ and subtract all the values with a corresponding $-$. Then divide by $4n$, where n is the number of replicates.

Easy, no?

e.g. $\hat{A} = \frac{1}{4}(-24.5 + 22.05 - 24.52 + 25 - 25.68 + 24.51 - 24.68 + 24.23) = \dots$

Series 4, exercise 4

Suppose that experimentation shows that only effect whose magnitude exceed 0.35 in absolute value are important. Which factors or interactions have a **practically** significant effect on fabric strength?

1. With enough sample size you can find any effect to be significant (even $\beta = 0.000001$).
2. Always consider what is **practically** significant (depends on the application).

see blackboard