Stat 500 – Homework 12 answers – 2009

Notes are marked by •

1. Bottling machines

   (a) Yes, there are no blocks in the design. However, this **is not** a reasonable analysis. Days are subsamples; there are only 12 e.u.'s. A more appropriate analysis would be to either:
   average the subsamples to get one average per machine, then use ANOVA
   or include the e.u., machine(model), as a random effect in the ANOVA.

   (b) My SAS code, for the subsampling analysis:

```
data bottle;
  infile 'bottle.txt' firstobs=2 missover;
  input model machine @;
  do day = 1 to 5;
    input y @;
    output;
    end;
proc mixed method=type3;
  class model machine;
  model y = model;
  random machine(model);
  run;
```

   I got: estimated difference (model 1 - 2): -9.8 s.e.: 5.0 p-value: 0.084

2. Barley fungus experiment design

   All four studies have the same treatment structure (2 way factorial) but differ in their experimental design.

   (a) This is a CRD.

Source	d.f.
barley	2
fungus	1
b*f	2
error	12
total	17

   use MSE to test each effect

(b) This is an RCB, with repetition of the study being the block.

Source	d.f.
block	2
barley	2
fungus	1
b*f	2
error	10
total	17

use MSE to test each effect

• My answer uses a pooled block*treatment error. You could also use separate block*barley, block*fungus and block*b*f effects. Each would be random; each would be the error for the corresponding effect. I would do this if I believed the variances were not similar. However, each error has a very small d.f., so my default would be to pool.

(c) This is a split plot because there are two e.u.'s:
chambers are randomly assigned to fungal genotypes,
flats are randomly assigned to barley genotypes.
The main plot design is an RCBD because repetitions of the study are being treated as blocks.

Source	d.f.	
block	2	
fungus	1	
block*fungus	2	main plot error
barley	2	
b*f	2	
error	8	split plot error
total	17	

the main plot error is used to test the fungus main effect
the split plot error is used to test barley and b*f effects

(d) This is also a split plot, for the same reasons as in part c. However, there is no replication of the main plot effect. This causes a problem, which is apparent when you write out the skeleton ANOVA table:

Source	d.f.
fungus	1
main plot error	0
barley	2
b*f	2
split plot error	12
total	17

There is no estimate of variability between growth chambers (main plots).
The split plot error is used to test barley and b*f effects.

3. Barley fungus experiment analysis

   (a) Tests:

Hypothesis #	Test of	F	p
1	fungi	0.10	0.78
2	barley	42.49	< 0.0001
3	barley*fungi	8.58	0.010

   (b) Estimates:

Quantity	Estimate	s.e.
1	0.073	0.197
2	0.063	0.086
3	-0.287	0.197

   (c) This demonstrates the pattern I mentioned in class. There are two different standard errors for comparisons of cell means, even if the sample sizes are all equal. The smaller s.e. (for comparison 2) is the comparion between split plot effects **within a main plot**. The larger s.e. if for comparisons between different main plot treatments.

4. LINX study - multilocation analysis

   (a)

Source	d.f.
location	7
stream	2
location*stream	14
error	48
total	71

   (b) Location: F or R, Stream: F, L*S: R, error: always R.
   • to do broad sense inference, or equivalently to compare treatment differences to their consistency across locations, the interaction needs to be random.

   (c) My SAS code, not including the data step:

```
proc mixed method=type3;
  class location type;
  model dist = location type;
  random location*type;
  title 'Broad sense inference';
  run;
```

   I got F = 2.26, p = 0.14.

   (d) My SAS code, not including the data step:

```
proc mixed method=type3;
  class location type;
  model dist = location type location*type;
  title 'Narrow sense inference';
  run;
```

I got F = 7.74, p = 0.0012.

(e) I got F = 10.66, p = 0.0015. SAS code same as in 4c, except using logdist as the response.
• If you used REML estimates of the variance components, you got F=8.42, p = 0.0040. The difference is because the ANOVA and REML variance component estimates are different.

(f) Broad sense, log transformed responses. Since the intended population is streams throughout the US, you must use broad sense inference. The choice of response is more subjective. My argument for reporting the transformed responses is that the stream effect does seem to be consistent across locations when measured on the log scale. Since science looks for general patterns, this is the result to report. A reasonable argument for untransformed responses (e.g. previous experience suggests that responses are best measured on the untransformed scale) was accepted for full credit.