

Stat 402 A - Spring 2009 - Exam 2 Answers

Additional explanations are included as Notes.

1. Swamp tree growth. My SAS code for parts d-g:

```
data nyssa;
  infile 'nyssa.txt';
  input block pool light species $ water growth;
  run;
proc mixed method=type3;
  class block light species water;
  model growth = block light|species|water / ddfm=satterth;
  random block*light;
  estimate 'light 100-35' light -1 0 1;
  estimate 'water 0 - 10' water 1 0 -1;
  run;
```

- (a) 2 pts. Two e.u.'s: Pool and pot.

Note: Light levels are assigned to pools; water level and species are assigned to pots

- (b) 3 pts. Trt design: 3 way factorial (or 3x3x2 complete factorial)

Expt design: split plot with RCBD at the main plot level and CRD at the split plot level

(c) 10 pts.	Source	d.f.	F/R	
	Block	4	F or R	
	Light	2	F	main plot trt
	B*L	8	R	Main plot error
	Species	1	F	split plot trts
	Water	2	F	"
	S*W	2	F	"
	L*S	2	F	main*split interactions
	L*W	4	F	"
	L*S*W	4	F	
	error	60	R	split plot error
	c.total	89		

Note: There are 15 pools (main plot eu), so main plot d.f. add up to 14. There are 90 observations so all df add up to 89.

- (d) 5 pts.

Strong evidence that species respond differently to water: F for water*species = 46.28, $p < 0.0001$

No evidence that species respond differently to light: F for light*species = 1.09, $p = 0.34$

No evidence that species have different 2way interactions between light and water: F for species*light*water = 0.70, $p = 0.59$

Note: These questions all ask whether species have the same pattern of response, i.e. whether the curves (e.g. height vs water level) for each species are parallel. These are questions about the interaction between species and another factor.

- (e) 5 pts. F for light = 26.0, $p = 0.0003$. Strong evidence that light affects height.
- (f) 5 pts. The difference in mean height between 100% and 35% light, averaged over species and water levels, is estimated to be 15.5 with a s.e. of 2.3.
 Note: I deducted a point if the difference was taken the other way around. That really changes the conclusion!
 Note: A few people reported means, not the mean difference. Please answer the question that I ask.
- (g) 5 pts. The difference in mean height between 10cm and 0cm water depth, averaged over species and water levels, is estimated to be 31.0 with a s.e. of 1.5.
- (h) 5 pts. No they are not. The difference between water levels is more precisely estimated because water is a split plot factor. "Miniblocking" increases the precision.
 Note: A lot of folks answered that the split plot was more precise because the split plot error had more d.f. D.f. have nothing to do with the magnitude of the s.e. D.f. tell you much information was used to calculate that variance. So, they tell you how precisely you have estimated a variance. That has nothing to do with the size of the s.e. When you test a hypothesis or calculate a confidence interval, the d.f. influence the power of the test and width of the ci because d.f. determine the t or F quantiles. However, s.e. and d.f. are unrelated.
 Note: The other common explanation of unequal s.e.'s, different sample sizes, doesn't apply here. Water means are based on 6 cells (3 light x 2 spp); light means are based on 6 cells (3 water x 2 spp). Each cell has 5 observations.

2. Potato scab. My approach combined a 1 way and 2 way ANOVA. SAS code for all parts using this approach :

```
data scab;
  infile 'c:/philip/stat 402/scab.txt';
  input block trtcode1 trtcode2 amount season $ disease;
  run;
proc glm;
  class trtcode2 block;
  model disease = block trtcode2;
  estimate 'fall - spring, when applied' trtcode2 0 1 -1 1 -1 1 -1 / divisor=3;
  contrast 'interaction when applied' trtcode2 0 1 -1 -1 1 0 0,
      trtcode2 0 1 -1 0 0 -1 1;
  contrast 'amount in fall incl 0' trtcode2 1 -1 0 0 0 0 0,
      trtcode2 1 0 0 -1 0 0 0,
      trtcode2 1 0 0 0 0 -1 0;
  contrast 'amount in spring incl 0' trtcode2 1 0 -1 0 0 0 0,
      trtcode2 1 0 0 0 -1 0 0,
      trtcode2 1 0 0 0 0 0 -1 ;
  contrast 'amount, ave over season' trtcode2 2 -1 -1 0 0 0 0,
      trtcode2 2 0 0 -1 -1 0 0,
      trtcode2 2 0 0 0 0 -1 -1;
  estimate 'control - fungicide' trtcode2 6 -1 -1 -1 -1 -1 -1 / divisor=6;
  title 'Analysis using trtcode2 ';
  run;
```

```

proc glm;
  class block season amount;
  model disease = block season amount season*amount;
  lsmeans season*amount /slice = season;
  title 'analysis using 2 way factorial';
run;

```

Note that I approached this using a combination of 1way ANOVA followed by contrasts and 2way ANOVA. This is because of the two “treatments” with 0 fungicide. Some questions are appropriate to answer using the 0 fungicide; others are easiest answered by constructing contrasts. Problems with the standard 2 way factorial are noted in appropriate places.

- (a) 8 pts. My preferred approach starts with
- | Source | d.f. | F/R |
|----------|------|--------|
| Block | 3 | F or R |
| Trtcode2 | 6 | F |
| Error | 22 | R |

Note: The choice between trtcode1 and trtcode2 is determined by whether the 0/f and 0/s treatments are considered to have the same population mean or not. If they have the same mean, then there are only 7 means to estimate and you use trtcode2. If they do not, then you estimate 8 means using trtcode1.

Everyone in the class started with a 2 way factorial, which is appropriate so long as you think carefully about some of the subsequent questions. That gives you:

Source	d.f.	F/R
Block	3	F or R
Season	1	F
Amount	3	F
S*A	3	F
Error	21	R

Note: Throughout the rest of the answers, the s.e.’s you get may differ from those given here. That’s because the two approaches have slightly different errors. I only deducted points when a s.e. was clearly incorrect.

Note: Most of my questions focused on estimates and their s.e. Various folks forgot to report the s.e. Again, please answer the question that was asked.

- (b) 5 pts. The estimated difference is -6.2 scab units with a s.e. of 2.8.

Note: Here is one place where care is required with the 2 way factorial approach. The question asks for the difference when fungicide is applied. That excludes the 0 amounts! The standard season main effects include the amount = 0 cells, so your answers are quite different. Using a 2 way factorial gives a difference of -2.3 with p-value of 0.33. The conclusion about season effects is quite different when the ”difference” at amount = 0 is ignored; that p-value is 0.040.

- (c) 5 pts. My approach is to decide this by looking at the interaction. Since the question includes the 0 amount, this is easiest done using the 2 way factorial model. The interaction is just significant, $p = 0.049$. You should examine amounts in each season separately.

Note: Why is amount = 0 relevant here but not in the previous part? Because the ”effect of amount” includes the change in reponse when fungicide is added (0 vs 300 or 600 or 1200).

- (d) 5 pts. This could be reasonably answered either using the 2 way factorial and slicing, or by constructing contrasts among 7 treatments. The answers differ slightly because the 0 group is defined differently in the two approaches. If you used contrasts with trtcode1 (8 treatments), you get the same answer as the 2 way factorial.

Using 2 way factorial: fall, $p = 0.0008$; spring, $p = 0.81$

Using contrasts: fall, $p = 0.0027$; spring $p = 0.24$

There is strong evidence that amount of fungicide applied in the fall affects the amount of disease. There is no evidence that amount applied in the spring affects the amount of disease.

- (e) 5 pts. Yes, it does. The estimated difference between control (0 fungicide) and the average of all fungicide treatments is 9.29 units with a s.e. of 2.83.

3. Barley-fungus microarray studies

All studies have the same treatment design. They differ in the experimental design.

- (a) 8 pts. This is a CRD. There are 18 chambers, each with 1 flat. The e.u. is the combination of chamber and flat. Each combination of barley and fungus is randomly assigned to an e.u.

Source	d.f.	error term
Barley	2	error
Fungus	1	error
B*F	2	error
Error	12	
c.total	17	

Note: Some folks thought this was a split plot because fungus went to chamber and barley to flat. It's not a split plot because 1 chamber is 1 flat. If you did this please bring your exam and see me.

- (b) 8 pts. This is a RCBD with 3 blocks, one for each repetition of the study. It is identical to the previous problem, except the 18 chambers are grouped into 3 repetitions.

Source	d.f.	error term
Block	2	
Barley	2	error
Fungus	1	error
B*F	2	error
Error	10	
c.total	17	

- (c) 8 pts. This is a split plot. The two eu's are chamber (fungus, in blocks) and flat (barley). There are three flats in this chamber. Chamber is the main plot to which fungus is assigned. Flat is the split plot to which barley is assigned.

Source	d.f.	error term
Block	2	
Fungus	1	Block*F
Block*F	2	(= main error)
Barley	2	split error
B*F	2	split error
split error	8	
c.total	17	

- (d) 8 pts. This is also a split plot, with same same eu's as before, except only one repetition of the main plot trt. There are only 2 e.u.'s, so only one d.f., at the main plot level. There is no main plot error.

Source	d.f.	error term
Fungus	1	none
main error	0	
Barley	2	split error
B*F	2	split error
split error	12	
<hr/>		
c.total	17	

Note: I talked about pseudoreplication after the cutoff for what was covered on this exam. Everything above can be figured out very easily from the same principles used in the previous part. I gave full credit if you claimed that the split error could be used to test fungus AND explicitly said you were committing pseudoreplication.