

Stat 402 A - Spring 2008 - Exam 2 Answers

Additional explanations are included as Notes.

1. Fungicide response

- (a) Expt. design is an RCB; trt design is a 2 way factorial.
Note: Shelves are the blocks.
- (b) This is the F test for fungicide concentration: $F = 307.18$, $p < 0.0001$.
Note: if you got $F = 243$, you forgot to include shelf (the blocks) in the model.
- (c) Average growth, averaged over the isolates, declines with increasing concentration of the fungicide.

conc.	mean	s.e.
0	3.64	0.062
0.001	2.65	0.062
0.002	1.48	0.062

- (d) This is the F test for fungicide concentration * isolate. $F = 122.96$, $p < 0.0001$
- (e) Slicing gives you the 2 df F test for the difference between concentrations for CUE2: $F = 0.61$, $p = 0.55$
- (f) Because of the strong interaction, you need the simple effect. Most folks used the pooled error variance. That gives estimate = -0.215, s.e. = 0.195. $p = 0.27$.

Notes:

- 1) The easiest way to do this is to use the trt variable and write the contrast in terms of the treatments. If you use the two-way model, you need to include coefficients for both the main effect and the interaction.
 - 2) If you got estimate = -0.1067, your contrast coefficients were in the wrong places.
 - 3) Because isolates have very different error variances (the smallest is 0.023, the largest is 0.21), it would be better to split the data. That gives the same estimate but $se=0.097$ and $p = 0.051$
- (g) Test b is more powerful because the underlying contrast coefficients (remember our discussion of test as contrasts between cell means) are smaller. That means the s.e. of each contrast is smaller.

Notes: lots of folks missed this one. The issue is not a larger sample size, because all observations are used in the main effect test; all observations are used in the contrast test. Both use the same number of observations. Saying that that main effect tests are more powerful without explanation got partial credit.

2. Analysis of different designs for a microarray experiment

Note: the treatment structure is the same in all studies: a 2-way factorial. That means barley, fungus and their interaction should be in all models.

(a) This is a CRD. The e.u. is the chamber; treatments are randomly assigned to chambers.

Source	d.f.	F test denominator
barley	2	error
fungus	1	error
B*F	2	error
error	18	
c.total	23	

(b) Each repetition is a block (6 chambers, one for each treatment, in a repetition).

Source	d.f.	F test denominator
repetition	3	
barley	2	error
fungus	1	error
B*F	2	error
error	15	
c.total	23	

(c) This is a split plot. The e.u. for fungus is the chamber; the e.u. for barley is the flat. Both levels are CRD.

Source	d.f.	F test denominator
fungus	1	chamber(fungus)
chamber(fungus)	6	
barley	2	split error
B*F	2	split error
split error	12	
c.total	23	

(d) This is also a split plot. The e.u. for fungus is the chamber; the e.u. for barley is the flat. The main plots are arranged in a latin rectangle. The split plots are CRD within each main plot.

Source	d.f.	F test denominator
chamber	1	
time	3	
fungus	1	main error
main error	2	
barley	2	split error
B*F	2	split error
split error	12	
c.total	23	

Notes: Quite a few folks omitted chamber, which lost a point. Since the experimenters clearly made sure that each treatment occurred twice in each chamber, you should include chamber in the model.

- (e) This is also a split plot. The e.u. for fungus is the chamber; the e.u. for barley is the flat. Both are CRD; the problem is that there is no replication at the main plot level. This is only a problem if you want to test the main effect of fungus.

Source	d.f.	F test denominator
fungus	1	none
chamber(fungus)	0	
barley	2	error
B*F	2	error
error	12	
c.total	23	

Note: Quite a few folks wanted to substitute the split plot error for the missing main plot error, chamber(fungus). In some fields this is called pseudo-replication and is considered a statistical sin. The best approach is to realize that you have valid tests of some effects but not others.

3. Bison bone beds

- (a) The key here was to realize that age and geog should be treated as if they were assigned to sites but bone was assigned to observations. The sentence 'Age and Geog are characteristic of site ...' is the key. There are two sizes of "eu"s. The treatment structure for the main plots is a 2way factorial. This means that the ANOVA table is:

Source	d.f.	F/R
age	2	F
geog	2	F
age*geog	4	F
site(A*G)	13	R
bone	3	F
bone*age	6	F
bone*geog	6	F
bone*A*G	12	F
split error	1040	R

- (b) This is the test of bone*age: $F = 1.59$, $p = 0.14$
 Since the geog*bone*age interaction is not significant, the bone*age interaction can be assumed similar at each geog. that means you can don't have to look separately at each geog area; you can use the 2 way interaction to answer the question.
- (c) This is the test of geog*age: $F = 4.20$, $p = 0.021$
 Similar reasoning as above, the geog*age*bone is far from significant, so the 'average' answer is sufficient. This markedly simplifies the table of means:

Geog	Age		
	E	T	L
N	3.65	2.58	3.07
M	3.72	3.55	3.16
S	3.54	3.29	3.24

- (d) no missing cells. The 3 way interaction (bone*age*geog) has the correct d.f.
- (e) Age*geog is a comparison among main plots; the difference between bones a and b is a comparison between two split levels in the same main plot (site). The split plot comparison (bones a - b) can take advantage of the 'mini-blocks' provided by the main plots. The variance of this difference includes on the split plot variance. The comparions between main plots includes both the main plot variance and the split plot variance. This is larger.

Notes: Most folks missed this. In fact, the question wouldn't make sense if you didn't see the split plot. I didn't worry about this because even those who saw the split plot didn't get this correct.

Effectively the midterm was a 95 point exam; this question became a bonus. I took that into account when I made up my grade ranges.