

1. LINX experiment in multiple years

(a) 2 pts.	Source	d.f.
	location*year	23
	treatment	2
	treatment*location*year	46
	error	144
	<hr/>	
	c.total	215

Notes:

1) Do not want main effect for year in the model, because want to pool location(year) and year effects into a single source of variability, that between all combinations of year and location.

2) Could write location\*year as location(year) just as long as year is not in the model.

3) error d.f. determined by difference or by going forward: There are 3 streams for each stream type at a location/year. That means each gives you 2 d.f. for the error for that "cell". There are a total of 3 types \* 8 locations \* 3 years = 72 cells, so  $2*72 = 144$  error d.f.

(b) 2 pts. This is intermediate sense inference so trt\*location\*year is random. Treatment is fixed. If you do intermediate sense inference, location\*year is fixed.

If you wanted to do broad sense inference, you would make location\*year random. I prefer intermediate sense inference when the focus is on differences among treatments; many other folks prefer broad sense.

Either way, treatment\*location\*year is the appropriate error term to test treatment effects using intermediate sense inference.

Note: If wanted to make inferences to means of a stream type in a location/year, then you would need broad sense inference.

(c) 2 pts. Now, can not pool year and location(year). This gives:

	Source	d.f.
	year	2
	location(year)	21
	treatment	2
	treatment*year	4
	treatment*location(year)	42
	error	144
	<hr/>	
	c.total	215

If I had data, my SAS code for 1a, i.e. pooling, would be:

```
proc mixed method=type3;
class location year treatment;
model dist = location*year treatment;
random location*year*treatment;\
run;
```

Of course, this core code would be followed up by the appropriate set of lsmeans, estimates, and/or contrasts. You could get the same thing by defining a new variable: env, with values from 1 to 24, indicating each unique combination of year and location, then replacing location\*year by env.

My code for 1c, i.e. not pooling, would be:

```
proc mixed method=type3;
class location year treatment;
model dist = year location(year) treatment;
random treatment*year treatment*location(year);
run;
```

2. Three drugs and heart function. My SAS code for all parts was:

```
data heart;
  infile 'heart.txt';
  input group $ pre post;
```

```
proc plot;
  plot post*pre = group;
run;
```

```
proc glm;
  class group;
  model post = group;
  lsmeans group /stderr;
  estimate 'diff btwn A and C' group 1 0 -1;
  title 'Post only analysis';
```

```
proc glm;
  class group;
  model post = group pre /solution;
  lsmeans group /stderr;
  estimate 'slope' pre 1;
  estimate 'diff btwn A and C' group 1 0 -1;
  title "ANCOVA";
```

```
proc glm;
  class group;
  model post = group pre pre*group;
  title "Heterogenous Regressions, to test equal slopes";
run;
```

(a) 2 pts. No evidence. The p-value for drug is 0.496.

(b) 2 pts. The estimate (drug A - drug C) is -2.17 with a s.e. of 2.28.

- (c) 2 pts. Yes, there is evidence that at least one drug has a different mean at the same pre-treatment value.  $p=0.0409$ .
- (d) 2 pts. The estimate (drug A - drug C) is -4.76 with a s.e. of 1.80.
- (e) 2 pts. Yes, the assumption is reasonable. The p-value for the test of unequal slopes (pre\*group) in the heterogeneous regression model is 0.58.
- (f) 2 pts. The estimates in b) and d) are not the same number because the estimates in d) are adjusted for the difference in mean X. If you calculate the means for each drug, you find that the subjects who got drug A had a slightly higher pre-treatment heart rate (76.0) than those who got drug C (72.9). There is a strong relationship between pre- and post- treatment heart rates (estimated regression slope is 0.834). So, you expect the adjusted difference (at the same pre-treatment value) to be different from the unadjusted value.

Note: These numbers illustrate the relationship given in class:

$$\text{adjusted diff } \mu_A - \mu_C = \bar{Y}_A - \bar{Y}_C - \hat{\beta}_1(\bar{X}_A - \bar{X}_C) = -2.17 - 0.834(76.0 - 72.9) = -4.76.$$

The tests in a) and c) give different results because the differences are more precisely estimated (smaller se) in the ANCOVA.

- (g) 2 pts. The ANCOVA is more appropriate because the tremendous between-subjects variability in pre-treatment heart rate masks the effect of the drugs. The equal slope assumption of ANCOVA is appropriate.