

Notes are marked by ●

1. Average Daily Weight Gain in pigs

(a) The d.f. are litters = (10-1) = 9, pigs(litters) = (2-1)\*10 = 10. Total = 20 - 1 = 19

●If you need a reminder about computing SS, here's how I did it:

$$\begin{aligned} \text{SS(litters)} &= \text{sum of } 2 \times (\text{litter mean} - \text{overall mean})^2 \\ &= 2 \times (2.57 - 2.574)^2 + 2 \times (2.76 - 2.574)^2 + \dots \\ &= 0.6705 \end{aligned}$$

$$\begin{aligned} \text{SS(pigs in litters)} &= \text{SS(total)} - \text{SS(litters)} \\ &= \text{pooling } (n-1) \times \text{s.d.}^2 \text{ for each litter} \\ &= .3834 \end{aligned}$$

(b) If  $Y_{ij}$  is the ADWG from the  $j$ 'th pig in litter  $i$ , then one way to write the model is:

$$\begin{aligned} Y_{ij} &= \mu_i + \epsilon_{ij} \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \mu_i &\sim N(0, \sigma_\mu^2). \end{aligned}$$

You could also write an effects model:

$$\begin{aligned} Y_{ij} &= \mu + \tau_i + \epsilon_{ij} \\ \epsilon_{ij} &\sim N(0, \sigma^2) \\ \tau_i &\sim N(0, \sigma_\mu^2). \end{aligned}$$

The only difference is that the mean of the pig effects is separated out as a distinct parameter.

(c) The expected mean squares are:

$$\text{Litters: } \sigma^2 + 2\sigma_{\mu}^2$$

$$\text{Error: } \sigma^2$$

Hence, the method of moments estimate of  $\sigma^2$  is the  $\text{MSE} = .03834$ . Litter-to-litter variation,  $\sigma_\mu^2$ , can be estimated using method of moments  $\sigma_\mu^2 = (\text{MS}_{\text{litters}} - \text{MS}_{\text{pigs in litters}}) / (\# \text{pigs in litter}) = (.0745 - .0383) / 2 = .0181$

(d) Under the model in b, the variance of  $Y_{ij}$  is the sum of the two variance components 0.0564. The fraction of this that is due to genetic factors is  $.0181 / (.0181 + .0383) = .32$ . You may notice that this fraction is also the intraclass correlation, also known as the heritability when the classes are genetic groups.

(e) The variance of the mean for  $r$  litters and  $n$  pigs per litter is:  $\sigma_\mu^2 / r + \sigma^2 / (rn)$ .

For the sample sizes given this is:

r	n	Var mean
2	4	0.01385
4	2	0.00932
8	1	0.00706

$r=8, n=1$  is the most precise estimate.

•While  $r=8$ ,  $n=1$  gives the most precise estimate of the mean, it does not allow you to estimate the variance components. If you want both, you need to compromise, e.g. use  $r=4$ ,  $n=2$ .

## 2. Bottling machines

- (a) Operator is nested within machine. 12 DIFFERENT operators were used in the study.
- Operators would be crossed if the same four were used with each of the four machines.
- (b) If you want to make conclusions (narrow sense) about these specific machines and operators, you should consider both machines and operators as fixed effects. The only random effect would be the observational variance (day).
- (c) Machines: Random, Operators: Random. To estimate the variance components, machines and operators should be random.
- (d) An effects model is:

$$\begin{aligned}
 Y_{ijk} &= \mu + \alpha_i + \beta_{ji} + \epsilon_{ijk} \\
 \mu &= \text{overall mean} \\
 \alpha_i &\sim N(0, \sigma_\alpha^2) \quad \text{machine } i \text{ effect} \\
 \beta_{ji} &\sim N(0, \sigma_\beta^2) \quad \text{operator } j \text{ effect (nested in machine } i) \\
 \epsilon_{ijk} &\sim N(0, \sigma^2) \quad \text{day - day variability} = \text{error}
 \end{aligned}$$

- (e) My SAS code:

```

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

```

The estimates are:

$$\begin{aligned}
 \sigma_{\text{machine}}^2 &= 29.7 \\
 \sigma_{\text{machine}}^2 &= 45.8 \\
 \sigma^2 = \sigma_{\text{days}}^2 &= 23.6
 \end{aligned}$$

(f) The ANOVA table is:

Source	df	MS	F	p-value
Machine	2	847.8	3.36	$0.05 < p < 0.10$
Op(Mach)	9	252.5	10.6	$< 0.0001$
Error	48	23.6		

•Remember that Machine is tested using Operator(Machine).

(g) Var(machines) is poorly estimated (sample size = 3 machines).

Sample size for Var(operator(machine)) is better (9) but still on the small side.

•The model also assumes constant variance. That is Var(operator(machine)) is the same for each machine and Var(error) is the same for each machine and each combination of operator and machine. This can be assessed using methods from earlier in the semester (residual plots, calculation of s.d.'s and tests of variances).

### 3. 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 RCBD, 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.

#### 4. 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 comparison between split plot effects **within a main plot**. The larger s.e. is for comparisons between different main plot treatments.

#### 5. Freeway rods, as a split plot

- (a) There are 54 ( $9 * 6$ ) experimental units in this study. To design this as a completely randomized design, you would to indentify 54 “places” and randomly assign one combination of rod, location, and lane to that place.

- This is a bit contrived since location and lane are observational factors, not randomly assigned. A way to do a CRD when two factors are not randomized is to find 9 'slow lane/outside' places in the freeway and randomly assign rod types to those 9 places. Then, do another randomization for the 9 'slow lane / center' places, ditto for the rest of the six places.

(b) There are two sizes of e.u.: rod types are assigned to road segments, combinations of lane and location) are “randomly” assigned to place.

- Randomly is in quotes because lane and location are observational factors and are not randomly assigned. The difference between observational or randomized treatment doesn't change the ANOVA. It definitely changes the interpretation.

(c) No, last week's analysis only considered one size of e.u. A more appropriate ANOVA table is:

rod	2	
segment(rod)	6	main plot error
lane	1	
location	2	
lane*loc	2	
rod*lane	2	
rod*loc	4	
rod*lane*loc	4	
split plot error	30	
total	53	

- Including segment in the model and ANOVA table is not appropriate because the main plot design is a CRD; segment is not a crossed factor. If the main plot design used blocking, BLOCK would be in the model and BLOCK\*MAINtrt would be the main plot error.

(d) My SAS code:

```
data freeway;
  infile 'c:/philip/stat 500/data/freeway.txt' firstobs=2;
  input rod segment lane location width;
  logwidth = log(width);
proc mixed method=type3;
  class rod segment lane location;
  model logwidth = rod | lane | location;
  random segment(rod);
run;
```

Which gives you:

Source	F	p-value
rod	15.85	0.0040
lane	324.61	<.0001
rod*lane	0.08	0.92
location	294.72	<.0001
rod*location	0.44	0.7769
lane*location	0.65	0.5291
rod*lane*location	0.31	0.8720
segment(rod)	46.48	<.0001