

PARTITIONING VARIANCE WITH S-PLUS

The first problem you face is getting your data into S-Plus. I've made it easy, because I've provided two data sets that are ready for you:

1. `full-sib-data.txt` and
2. `full-sib-data-unbalanced.txt`.

To get the data into S-Plus, click on the **File** menu, and select **Import Data From File...** Now point to the class directory on the network drive, and click on `full-sib-data.txt`, but don't double click on it. At the bottom of the dialog box, you'll see a little window labeled *Import to Data Frame*. Change the name here to `full-sib`.¹ Now hit the **Open** button. You'll see a little message saying "This object will be imported as `full-sib-data.txt`". Just press **OK**, and you're done.

Well, at least you're done getting the data into S-Plus. Now, we need to prepare it for analysis. You'll see a spreadsheet on your screen. The columns correspond to variables; the rows correspond to individual observations. We now need to tell S-Plus to treat `sire`, `dam`, and *individual* as treatment factors instead of as numbers. You do this by typing the following at the S-Plus command prompt:²

```
> full.sib$sire <- factor(full.sib$sire)
> full.sib$dam <- factor(full.sib$dam)
> full.sib$individual <- factor(full.sib$individual)
```

We also need to tell S-Plus that these are random factors, i.e., we regard each sire and dam as a randomly sampled from the population and each individual as a randomly chosen offspring from that mating. You do this simply by typing:

```
> is.random(full.sib) <- T
```

Now we're ready to go. Since this data set is completely balanced (all sires are mated to the same number of dams and we examine the same number of progeny from every dam), we can use `raov()` to do a random-effects analysis of variance.

```
> raov(weight ~ sire/dam, data = full.sib)
Call:
  raov(formula = weight ~ sire/dam, data = full.sib)
```

Terms:

	sire	dam	%in%	sire	Residuals
Sum of Squares	36.1887			44.8167	122.9101
Deg. of Freedom	5			30	180

Residual standard error: 0.8263378

¹ This will save you some typing later.

² Don't put in the `>` character. That's the command prompt.

Estimated effects are balanced

You can calculate the mean squares associated with sires, dams, and sibships by dividing the mean squares above by their appropriate degrees of freedom. You can then calculate the variance components for dams, sires, and within sibships using the formulas from the last handout. That's a bit tedious, though, and you can't use `raov` with an unbalanced design. Fortunately, there is an alternative:

```
> vc.ml <- varcomp(weight ~ sire/dam, data = full.sib,
  method = "ml")
> summary(vc.ml)
Call:
varcomp(formula = weight ~ sire/dam, data = full.sib,
  method = "ml")
Variance Estimates:
              Variance
      sire 0.1260435
dam %in% sire 0.1351760
  Residuals 0.6828342
Method: ml
Approximate Covariance Matrix of Variance Estimates:
              sire dam %in% sire Residuals (Intercept)
      sire 0.00945262 -0.00041082      0 0.00000008
dam %in% sire -0.00041082 0.00001785      0 0.00000000
  Residuals 0.00000000 0.00000000      0 0.00000000
(Intercept) 0.00000008 0.00000000      0 0.02792341

Coefficients:
(Intercept)
1.224955
Approximate Covariance Matrix of Coefficients:
(Intercept)
(Intercept) 0.01906706
```

The variance estimates are what we're interested in here. If you do the hand calculation based on mean squares above, you'll find the numbers in close agreement.

Once you've managed to duplicate this, try the whole procedure again using `full-sib-data-unbalanced.txt` and compare the results. You may be interested to know that the *only* difference between these data sets is that the first is completely balanced and the second is very unbalanced. I used exactly the same underlying genetic parameters to produce the data in both cases.