

Lathyrus ms2: Comparison datafiles old period

Contents

Problems with fcode variable	2
Cases where fcode=0 or NA, but there is data on FFD, or n_fl	2
Cases where fcode=0 or NA, but there are fruits or seeds	3
Cases where fcode is NA, but there is data on shoot volume	3
Differences in variables among the three datasets	3
FFD	3
Number of flowers	4
Number of fruits	5
Number of total seeds	6
Number of intact seeds	7
Shoot volume	8
Definitive dataset	9

data1: Data derived from Excel file Johan sent for ms1 in the beginning (before imputations by Alicia)

data2: Data derived from Excel file Johan sent for ms1 in the beginning (after imputations by Alicia)

data3: Data derived from Excel file Johan sent recently (88-95, for 87 and 96 the data should be the same as in data2)

The three datasets merged: FFD1 comes from data1, FFD2 from data2, FFD3 from data3, same for n_fl1, n_fl2, n_fl3... This is how the data looks:

```
##   year id_nr      id fcode FFD1      FFD2 FFD3 intactseed1 intactseed2
## 1 1987   260 old_260      1  144 63.88194  144              0              0
## 2 1987   235 old_235      1  145 64.88194  145             13             13
## 3 1987   269 old_269      1  143 62.88194  143             17             17
## 4 1987   833 old_833      1  143 62.88194  143             14             14
## 5 1987   277 old_277      1  147 66.88194  147             11             11
## 6 1987   251 old_251      1  143 62.88194  143             22             22
##   intactseed3 n_fl1 n_fl2 n_fl3 n_fr1 n_fr2 n_fr3 shoot_vol1 shoot_vol2
## 1              0   25   25   25      0      0      0          NA          NA
## 2              13   24   24   24      2      2      2          NA          NA
## 3              17   25   25   25      3      3      3          NA          NA
## 4              14   11   11   11      3      3      3          NA          NA
## 5              11   16   16   16      2      2      2          NA          NA
## 6              22   21   21   21      5      5      5          NA          NA
##   shoot_vol3 totseed1 totseed2 totseed3
## 1          NA        0         0         0
## 2          NA       13        13        13
## 3          NA       17        17        17
## 4          NA       14        14        14
## 5          NA       11        11        11
## 6          NA       22        22        22
[...]
```

Problems with fcode variable

Cases where fcode=0 or NA, but there is data on FFD, or n_fl

```
##   year id_nr      id fcode FFD1      FFD2 FFD3 n_fl1      n_fl2 n_fl3
## 1 1988   470 old_470      0   NA        NA   NA      5 5.000000      0
## 2 1989   416 old_416      NA  135 55.39653   NA      3 3.000000     NA
## 3 1989   100 old_100      0   NA        NA      1      0 0.000000      0
## 4 1990   103 old_103      0   NA        NA      5      0 0.000000      0
## 5 1993    27 old_27       0  125 45.42639      1      0 7.675872      0
## 6 1993   415 old_415      0  140 60.42639      4      0 8.231816      0
## 7 1993   258 old_258      0  135 55.42639      3      0 7.779139      0
## 8 1993   444 old_444      0  140 60.42639      4      0 10.556234      0
```

The first row in the table above has a value of n_fl in the previous data I got (n_fl1, n_fl2), but not in the last data sent by Johan (n_fl3=NA). **I changed fcode to 1.**

The second row in the table above has values of FFD and n_fl in the previous data I got (FFD1, FFD2, n_fl1, n_fl2), but not in the last data sent by Johan (FFD3=NA, n_fl3=NA). **I changed fcode to 1.**

The next two rows in the table above have a value of phenology in the last data sent by Johan (FFD3), but not in the previous data I got (FFD1=FFD2=NA). **I changed fcode to 1,** because apparently they

flowered. Can we get a FFD value from those values in FFD3 (What do 1 and 5 mean?).

The last four rows in the table above have values of FFD in all three datasets (FFD1, FFD2 and FFD3) but fcode=0 -> **I changed to fcode=1**. For these rows, values of n_fl were missing in the previous data I got (n_fl1=NA), but I imputed them from shoot volume (that is why there are values of n_fl2).

Cases where fcode=0 or NA, but there are fruits or seeds

```
## [1] year      id_nr      id      fcode      n_fr1      n_fr2
## [7] n_fr3      totseed1  totseed2  totseed3  intactseed1 intactseed2
## [13] intactseed3
## <0 rows> (or 0-length row.names)
```

No cases. Good!

Cases where fcode is NA, but there is data on shoot volume

```
##   year id_nr      id fcode shoot_vol1 shoot_vol2 shoot_vol3
## 1  1988  384 old_384   NA    732.5000    732.5000    733.0000
## 2  1988  270 old_270   NA    732.5000    732.5000    733.0000
## 3  1988   88 old_88    NA    439.5000    439.5000    440.0000
## 4  1989   8  old_8     NA   1038.0000   1038.0000   1038.0000
## 5  1990  389 old_389   NA    635.0000    635.0000    635.0000
## 6  1992  191 old_191   NA    659.9552    659.9552    659.9552
## 7  1992  119 old_119   NA    706.8611    706.8611    706.8611
## 8  1992  247 old_247   NA   1032.3662   1032.3662   1032.3662
## 9  1993  827 old_827   NA    659.9552    659.9552    660.0000
## 10 1993   50 old_50    NA     78.5400     78.5400     79.0000
## 11 1994  385 old_385   NA    388.7736    388.7736    388.7736
## 12 1994   76 old_76    NA   2993.2512   2993.2512   2993.2512
## 13 1988   1  old_1     NA         NA         NA     879.0000
## 14 1988  376 old_376   NA         NA         NA    1465.0000
## 15 1988  381 old_381   NA         NA         NA    1026.0000
```

I changed fcode to 0.

Differences in variables among the three datasets

FFD

For the rows below, there are no values in FFD3 (last data sent by Johan), but there are values in FFD1 and FFD2 -> **I will use values of FFD2** (which are days from the vernal equinox).

```
##   year id_nr      id fcode FFD1      FFD2 FFD3
## 643 1989  800 old_800     1  142 62.39653  NA
## 737 1989  249 old_249     1  131 51.39653  NA
## 741 1989  694 old_694     1  135 55.39653  NA
## 900 1989  416 old_416     1  135 55.39653  NA
## 1742 1990  752 old_752     1  141 61.15417  NA
## 2555 1991  251 old_251     1  140 59.91181  NA
## 5142 1993  259 old_259     1  140 60.42639  NA
## 5665 1993  428 old_428     1  140 60.42639  NA
## 6433 1994   1  old_1      1  139 59.18403  NA
```

For the rows below: there are no values in FFD1 and FFD2, but there are values in FFD3 (last data sent by Johan) -> **What to do? Can we get a FFD value from those values in FFD3 (What do those values mean?).**

```
##      year id_nr      id fcode FFD1 FFD2 FFD3
## 898  1989   199 old_199      1   NA   NA    3
## 1001 1989   100 old_100      1   NA   NA    1
## 1145 1989   739 old_739      1   NA   NA    1
## 1908 1990   103 old_103      1   NA   NA    5
## 2285 1990   407 old_407      1   NA   NA    8
```

Number of flowers

778 rows where there is difference in the number of flowers among the three datasets

```
##      year id_nr      id fcode n_fl1      n_fl2 n_fl3
## 327  1988   191 old_191      1      3 3.000000      2
## 422  1988   470 old_470      1      5 5.000000      0
## 898  1989   199 old_199      1   NA      NA      3
## 900  1989   416 old_416      1      3 3.000000     NA
## 1254 1990   810 old_810      1   NA 6.974556     NA
## 2284 1990   409 old_409      1   NA 10.965545     NA
```

[...]

601 cases where fcode=0 and n_fl3=0 (NA values for n_fl1 and n_fl2) -> **Assign 0 flowers**

```
##      year id_nr      id fcode n_fl1 n_fl2 n_fl3
## 10168 1988      5 old_5      0   NA   NA      0
## 10171 1988     12 old_12      0   NA   NA      0
## 10173 1989     13 old_13      0   NA   NA      0
## 10174 1988     14 old_14      0   NA   NA      0
## 10175 1989     14 old_14      0   NA   NA      0
## 10183 1988     20 old_20      0   NA   NA      0
```

[...]

Cases where fcode=0 and n_fl3<0 -> must be errors -> **Assign 0 flowers**

```
##      year id_nr      id fcode n_fl1 n_fl2      n_fl3
## 2864 1991   704 old_704      0      0      0 -4.15e-05
## 3192 1991   258 old_258      0      0      0 -4.15e-05
## 3475 1991   113 old_113      0      0      0 -4.15e-05
```

47 cases where fcode=1 and there is only data for n_fl2 -> These are data I have imputed from shoot volume
-> **Use n_fl2**

```
##      year id_nr      id fcode n_fl1      n_fl2 n_fl3
## 1254 1990   810 old_810      1   NA 6.974556     NA
## 2284 1990   409 old_409      1   NA 10.965545     NA
## 2541 1991   384 old_384      1   NA 8.288818     NA
## 2542 1991      8 old_8      1   NA 8.731097     NA
## 2546 1991   284 old_284      1   NA 9.687712     NA
## 2547 1991   262 old_262      1   NA 20.058023     NA
```

[...]

119 cases where fcode=1 and there is only data for n_fl3 -> **Use n_fl3**

```
##      year id_nr      id fcode n_fl1 n_fl2 n_fl3
## 898  1989   199 old_199     1    NA    NA     3
## 10186 1988    23 old_23     1    NA    NA     4
## 10188 1988    25 old_25     1    NA    NA     7
## 10189 1989    26 old_26     1    NA    NA     3
## 10196 1989    30 old_30     1    NA    NA     3
## 10197 1989    31 old_31     1    NA    NA    75
```

[...]

Cases where fcode=1 and there is data for nfl1 and nfl2 but not for nfl3 -> **Use n_fl2**

```
##      year id_nr      id fcode n_fl1 n_fl2 n_fl3
## 900  1989   416 old_416     1     3     3    NA
## 3533 1991   376 old_376     1     4     4    NA
```

Cases where fcode=1 and there is data for the three years but the values are different -> **Use n_fl2**

```
##      year id_nr      id fcode n_fl1 n_fl2 n_fl3
## 327  1988   191 old_191     1     3     3     2
```

Cases where n_fl3=0 and there is info on number of flowers from data2 -> **Use n_fl2**

```
##      year id_nr      id fcode n_fl1      n_fl2 n_fl3
## 422  1988   470 old_470     1     5 5.000000     0
## 5150 1993    27 old_27     1     0 7.675872     0
## 5414 1993   415 old_415     1     0 8.231816     0
## 5720 1993   258 old_258     1     0 7.779139     0
## 6053 1993   444 old_444     1     0 10.556234     0
```

Number of fruits

1032 rows where there is difference in the number of fruits among the three datasets

```
##      year id_nr      id fcode n_fr1 n_fr2 n_fr3
## 248  1988   211 old_211     1     0     0    NA
## 251  1988   279 old_279     1     0     0    NA
## 261  1988   223 old_223     1     0     0    NA
## 263  1988   229 old_229     1     6     6     5
## 264  1988   139 old_139     1     0     0    NA
## 269  1988   202 old_202     1     0     0    NA
```

[...]

246 cases where n_fr1=n_fr2 and n_fr3 is NA -> **Use n_fr2**

```
##      year id_nr      id fcode n_fr1 n_fr2 n_fr3
## 248  1988   211 old_211     1     0     0    NA
## 251  1988   279 old_279     1     0     0    NA
## 261  1988   223 old_223     1     0     0    NA
## 264  1988   139 old_139     1     0     0    NA
## 269  1988   202 old_202     1     0     0    NA
## 270  1988   841 old_841     1     0     0    NA
```

[...]

Cases where n_fr1=n_fr2 and n_fr3 is not NA

```
##      year id_nr      id fcode n_fr1 n_fr2 n_fr3
## 263  1988   229 old_229     1     6     6     5
```

```
## 484 1988 355 old_355 1 0 0 7
## 4199 1992 464 old_464 1 3 3 0
```

First and third case -> **Use n_fr2**

Second case -> **Use n_fr3**

695 cases when n_fr1 and n_fr2 are NA, and n_fr3 is 0 -> **Use n_fr3**

```
##      year id_nr      id fcode n_fr1 n_fr2 n_fr3
## 275 1988  810 old_810    1  NA   NA    0
## 328 1988  119 old_119    1  NA   NA    0
## 334 1988   19 old_19    1  NA   NA    0
## 368 1988  415 old_415    1  NA   NA    0
## 556 1988  188 old_188    1  NA   NA    0
## 905 1989 1034 old_1034    1  NA   NA    0
```

[...]

88 cases when n_fr1 and n_fr2 are NA, and n_fr3 is not 0 -> **Use n_fr3**

```
##      year id_nr      id fcode n_fr1 n_fr2 n_fr3
## 540  1988  315 old_315    1  NA   NA    2
## 3554 1991  448 old_448    1  NA   NA    1
## 10186 1988   23 old_23    1  NA   NA    1
## 10188 1988   25 old_25    1  NA   NA    1
## 10197 1989   31 old_31    1  NA   NA    5
## 10215 1989   44 old_44    1  NA   NA    1
```

[...]

Number of total seeds

965 rows where there is difference in the number of total seeds among the three datasets

```
##      year id_nr      id fcode totseed1 totseed2 totseed3
## 249 1988   64 old_64    1 14.90700 14.90700 0.00000
## 252 1988  419 old_419    1 19.87600 19.87600 0.00000
## 256 1988  426 old_426    1  9.93800  9.93800 0.00000
## 259 1988  216 old_216    1 37.333333 37.333333 37.33333
## 263 1988  229 old_229    1 26.40000 26.40000 22.00000
## 268 1988  212 old_212    1  4.96900  4.96900 0.00000
```

[...]

91 cases where totseed1=totseed2 and totseed3 is NA or 0 -> **Use totseed2**

```
##      year id_nr      id fcode totseed1 totseed2 totseed3
## 249 1988   64 old_64    1  14.907  14.907    0
## 252 1988  419 old_419    1  19.876  19.876    0
## 256 1988  426 old_426    1   9.938   9.938    0
## 268 1988  212 old_212    1   4.969   4.969    0
## 303 1988  284 old_284    1   4.969   4.969    0
## 307 1988  213 old_213    1  19.876  19.876    0
```

[...]

Cases where totseed1=totseed2 and totseed3 is > 0

```
##      year id_nr      id fcode totseed1 totseed2 totseed3
## 259 1988  216 old_216    1 37.333333 37.333333 37.333333
```

```
## 263 1988 229 old_229 1 26.400000 26.400000 22.000000
## 302 1988 252 old_252 1 9.333333 9.333333 9.333333
## 484 1988 355 old_355 1 0.000000 0.000000 19.000000
## 2619 1991 278 old_278 1 36.000000 36.000000 12.000000
```

First, second, third and fifth case -> **Use totseed2**

Fourth case -> **Use totseed3**

803 cases when totseed1 and totseed2 are NA, and totseed3 is not NA -> **Use totseed3**

```
##      year id_nr      id fcode totseed1 totseed2 totseed3
## 275 1988  810 old_810    1      NA      NA      0
## 328 1988  119 old_119    1      NA      NA      0
## 334 1988   19 old_19     1      NA      NA      0
## 368 1988  415 old_415    1      NA      NA      0
## 540 1988  315 old_315    1      NA      NA     10
## 556 1988  188 old_188    1      NA      NA      0
```

[...]

127 remaining cases when totseed2 is not NA, and totseed3 is 0 or NA -> **Use totseed2**

```
##      year id_nr      id fcode totseed1 totseed2 totseed3
## 900 1989  416 old_416    1      0 0.000000      NA
## 1232 1990  133 old_133    1      NA 28.829268      0
## 1237 1990    6 old_6     1      NA 4.804878      0
## 1239 1990   58 old_58    1      NA 9.609756      0
## 1267 1990  255 old_255    1      NA 4.804878      0
## 1282 1990  284 old_284    1      NA 24.024390      0
```

[...]

Number of intact seeds

965 rows where there is difference in the number of intact seeds among the three datasets

```
##      year id_nr      id fcode intactseed1 intactseed2 intactseed3
## 249 1988   64 old_64     1   6.901690   6.901690   0.000000
## 252 1988  419 old_419    1   9.171920   9.171920   0.000000
## 253 1988  133 old_133    1   3.818182   3.818182   3.818182
## 256 1988  426 old_426    1   4.631460   4.631460   0.000000
## 259 1988  216 old_216    1  23.333333  23.333333  23.333333
## 263 1988  229 old_229    1  18.000000  18.000000  15.000000
```

[...]

90 cases where intactseed1=intactseed2 and intactseed3 is NA or 0 -> **Use intactseed2**

```
##      year id_nr      id fcode intactseed1 intactseed2 intactseed3
## 249 1988   64 old_64     1   6.90169   6.90169      0
## 252 1988  419 old_419    1   9.17192   9.17192      0
## 256 1988  426 old_426    1   4.63146   4.63146      0
## 268 1988  212 old_212    1   2.36123   2.36123      0
## 303 1988  284 old_284    1   2.36123   2.36123      0
## 307 1988  213 old_213    1   9.17192   9.17192      0
```

[...]

Cases where intactseed1=intactseed2 and intactseed3 is > 0

##	year	id_nr	id	fcode	intactseed1	intactseed2	intactseed3
## 253	1988	133	old_133	1	3.818182	3.818182	3.818182
## 259	1988	216	old_216	1	23.333333	23.333333	23.333333
## 263	1988	229	old_229	1	18.000000	18.000000	15.000000
## 302	1988	252	old_252	1	5.333333	5.333333	5.333333
## 484	1988	355	old_355	1	0.000000	0.000000	2.000000
## 2619	1991	278	old_278	1	36.000000	36.000000	12.000000

First, second, third, fourth and sixth case -> **Use intactseed2**

Fifth case -> **Use intactseed3**

803 cases when intactseed1 and intactseed2 are NA, and intactseed3 is not NA -> **Use intactseed3**

##	year	id_nr	id	fcode	intactseed1	intactseed2	intactseed3
## 275	1988	810	old_810	1	NA	NA	0
## 328	1988	119	old_119	1	NA	NA	0
## 334	1988	19	old_19	1	NA	NA	0
## 368	1988	415	old_415	1	NA	NA	0
## 540	1988	315	old_315	1	NA	NA	10
## 556	1988	188	old_188	1	NA	NA	0

[...]

127 remaining cases when intactseed2 is not NA, and intactseed3 is 0 or NA -> **Use intactseed2**

##	year	id_nr	id	fcode	intactseed1	intactseed2	intactseed3
## 900	1989	416	old_416	1	0	0.000000	NA
## 1232	1990	133	old_133	1	NA	22.035366	0
## 1237	1990	6	old_6	1	NA	3.672561	0
## 1239	1990	58	old_58	1	NA	7.345122	0
## 1267	1990	255	old_255	1	NA	3.672561	0
## 1282	1990	284	old_284	1	NA	18.362805	0

[...]

Shoot volume

2108 rows where there is difference in shoot volume among the three datasets

##	year	id_nr	id	fcode	shoot_vol1r	shoot_vol2r	shoot_vol3r
## 248	1988	211	old_211	1	732	732	733
## 250	1988	791	old_791	0	732	732	733
## 257	1988	265	old_265	0	732	732	733
## 258	1988	6	old_6	1	2490	2490	2491
## 259	1988	216	old_216	1	2490	2490	2491
## 260	1988	58	old_58	0	732	732	733

[...]

123 cases where the difference is very small, due to rounding of values -> **Use shoot_vol2**

##	year	id_nr	id	fcode	shoot_vol1r	shoot_vol2r	shoot_vol3r
## 248	1988	211	old_211	1	732	732	733
## 250	1988	791	old_791	0	732	732	733
## 257	1988	265	old_265	0	732	732	733
## 258	1988	6	old_6	1	2490	2490	2491
## 259	1988	216	old_216	1	2490	2490	2491
## 260	1988	58	old_58	0	732	732	733

[...]

1230 cases where shoot_vol3 is NA -> **Assign shoot_vol as NA**

##	year	id_nr	id	fcode	shoot_vol1r	shoot_vol2r	shoot_vol3r
##	8813	1996	791 old_791	NA	0	0	NA
##	8814	1996	279 old_279	NA	0	0	NA
##	8815	1996	419 old_419	NA	0	0	NA
##	8817	1996	264 old_264	NA	0	0	NA
##	8818	1996	195 old_195	NA	0	0	NA
##	8820	1996	265 old_265	NA	0	0	NA

[...]

755 cases where shoot_vol1 and shoot_vol2 are NA, but there is a value of shoot_vol3 -> **Use shoot_vol3**

##	year	id_nr	id	fcode	shoot_vol1r	shoot_vol2r	shoot_vol3r
##	10165	1988	1 old_1	0	NA	NA	879
##	10168	1988	5 old_5	0	NA	NA	192
##	10171	1988	12 old_12	0	NA	NA	733
##	10173	1989	13 old_13	0	NA	NA	48
##	10174	1988	14 old_14	0	NA	NA	170
##	10175	1989	14 old_14	0	NA	NA	82

[...]

Definitive dataset

I created a “defintive” dataset with the values chosen in each case

##	year	id_nr	id	fcode	FFD	n_fl	n_fr	totseed	intactseed	shoot_vol
##	1	1989	1 old_1	1	NA	6	3	8	6	1418.6000
##	2	1990	1 old_1	0	NA	0	0	0	0	523.2000
##	3	1991	1 old_1	1	59.91181	23	3	12	12	1915.4000
##	4	1992	1 old_1	1	55.66944	19	2	6	1	1460.1917
##	5	1993	1 old_1	1	NA	NA	0	0	0	879.6493
##	6	1994	1 old_1	1	59.18403	14	1	3	3	1338.6727

[...]