Applied Statistical Methods - Solution 9

Peter von Rohr

2022-05-02

Problem 1: Milk Dataset

Use the dataset `milk` from package `pedigreemm` and fit a sire model to each of the response variables (`milk`, `fat`, `prot` and `scs`) in the data. The dataset can be loaded using the command `pedigreemm::milk`. The other variables like `lact` and `herd` can be used as fixed effects. The `sire` column is used as a random effect. For this analysis, we assume that sires are unrelated.

Your Tasks

- Analyse the `milk` dataset from package `pedigreemm` using the function `lme4::lmer()` for all given response variables. You can use the same model for each of the responses.
- Compute the estimated heritability for each response variable, using the fact that the heritability h^2 can be computed from the variance σ_s^2 of the sire effects and the phenotypic variance σ_p^2 with the formula

$$h^2 = \frac{4 * \sigma_s^2}{\sigma_p^2}$$

- Compute the summary statistic using the function `summary()` of all the predicted sire breeding values. Solutions for the sire breeding values are obtained using the function `ranef()`

Solution

- Assign milk dataset to a tibble

```
tbl_milk <- tibble::as_tibble(pedigreemm::milk)
```

- Analyse the data for each of the responses. The responses are pre-defined and must appear as column names

```
vec_resp_milk <- c("milk", "fat", "prot", "scs")
# check
if (!all(is.element(vec_resp_milk, colnames(tbl_milk)))){
  stop(" *** ERROR: Response not in dataset: ",
       vec_resp_milk[!is.element(vec_resp_milk, colnames(tbl_milk))])
}
```

For each of the responses run an analysis

```
vec_fix_fact_milk <- c("lact", "herd")
vec_fix_cov_milk <- c("dim")
s_rand_eff_milk <- "sire"
s_rhs <- paste0(c(paste0(vec_fix_fact_milk, collapse = " + "),
              paste0(vec_fix_cov_milk, collapse = " + "),
              paste0("(1|", s_rand_eff_milk, ")", collapse = "")),
              collapse = " + ")
```

Convert fixed effects to factors

```
# convert fixed effects to factors
for (f in vec_fix_fact_milk){
  tbl_milk[[f]] <- as.factor(tbl_milk[[f]])
}
```

Construct a function that runs an analysis for a given response

```
run_lmer_single_response <- function(ps_resp, ps_rhs, ptbl_data){
  # check
  # formula
  s_formula <- paste(ps_resp, ps_rhs, sep = " ~ ")
  frma_lmem <- as.formula(s_formula)
  return(lme4::lmer(frma_lmem, data = ptbl_data))
}
```

Let the analysis be done for all responses

```
l_lmer_results <- lapply(vec_resp_milk,
                         run_lmer_single_response,
                         ps_rhs = s_rhs,
                         ptbl_data = tbl_milk)
```

Print all the summaries as results

```
for (l in l_lmer_results){
  print(summary(l))
}
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: milk ~ lact + herd + dim + (1 | sire)
##    Data: ptbl_data
##
## REML criterion at convergence: 64682.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3515 -0.5939  0.0471  0.6391  3.0229
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sire     (Intercept)   436399  660.6
##  Residual             14323717 3784.7
## Number of obs: 3397, groups:  sire, 38
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 22179.3013   572.6886  38.728
## lact2        -567.0717   160.5548  -3.532
## lact3       -1027.2394   186.3072  -5.514
## lact4        -938.8171   239.1321  -3.926
## lact5       -1149.8552   395.5424  -2.907
## herd2        -405.1815   546.4401  -0.741
## herd4        1795.5114   873.3207   2.056
## herd5        3545.1581   753.3613   4.706
## herd6        -557.2013   706.5353  -0.789
```

```
## herd7        2050.2860   907.0908   2.260
## herd8        -349.0441   681.3300  -0.512
## herd13       2574.4707   735.7602   3.499
## herd14        556.1105   524.9750   1.059
## herd17       4020.2199   782.3927   5.138
## herd18      -1653.5144   703.7817  -2.349
## herd22       -765.3971   731.4518  -1.046
## herd23       -319.5227   605.6600  -0.528
## herd26       3815.7866  1389.9893   2.745
## herd27      -1467.1500  1282.3213  -1.144
## herd29         99.4121   776.3642   0.128
## herd30       4786.8342   831.7947   5.755
## herd34        299.2986   670.9064   0.446
## herd36       -992.0130   795.9012  -1.246
## herd37       3052.8175   707.6809   4.314
## herd38       2486.6492   638.8885   3.892
## herd41       1845.6266   721.4234   2.558
## herd45        595.2089   635.5650   0.937
## herd46       1957.0566   728.7574   2.685
## herd48       5747.5087   746.2061   7.702
## herd49       -533.7808  1034.2567  -0.516
## herd52       3836.7977   657.4239   5.836
## herd53       1155.4060   861.6142   1.341
## herd54        306.0101   803.6259   0.381
## herd55      -3168.1618   854.3604  -3.708
## herd59       1500.7075   605.7404   2.477
## herd60        -93.1071   620.9171  -0.150
## herd62       2867.7655   889.6308   3.224
## herd64       4488.2422   666.6853   6.732
## herd66       1568.2347   745.6632   2.103
## herd68       -318.6871   629.6671  -0.506
## herd69       -720.5528   610.7063  -1.180
## herd70      -2762.9034   673.9141  -4.100
## herd75       2527.1950   638.5451   3.958
## herd77      -1456.3385  1088.3172  -1.338
## herd89      -4164.5413   906.5904  -4.594
## herd90        835.3527   704.7541   1.185
## herd95       1898.5536  2749.3678   0.691
## herd96       2840.0896  2751.8931   1.032
## herd97       6006.9118  3836.3784   1.566
## herd98      -1589.8130  1550.0904  -1.026
## herd99      -2145.5194  2756.7237  -0.778
## herd100      1125.4711  2754.7748   0.409
## herd101       476.9742  2751.7144   0.173
## herd103     -6296.3744  2751.0073  -2.289
## herd104     -3698.2979  2273.0843  -1.627
## herd105     -3281.2862  3839.2518  -0.855
## herd106       593.8344  2785.4041   0.213
## herd107     -4731.4899  3840.2745  -1.232
## herd108      3288.2457  3841.1488   0.856
## herd109     -1315.0161  1027.7215  -1.280
## herd110      3993.4835  2749.6515   1.452
## dim             9.1286     0.6194  14.737
```

```
##
## Correlation matrix not shown by default, as p = 62 > 12.
## Use print(summary(l), correlation=TRUE)  or
##     vcov(summary(l))        if you need it

## Linear mixed model fit by REML ['lmerMod']
## Formula: fat ~ lact + herd + dim + (1 | sire)
##    Data: ptbl_data
##
## REML criterion at convergence: 42711.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4455 -0.6047 -0.0047  0.6297  5.0727
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sire     (Intercept)   675.7   25.99
##  Residual             19705.0  140.37
## Number of obs: 3397, groups:  sire, 38
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  913.64827   21.34689  42.800
## lact2         -3.12756    5.95520  -0.525
## lact3        -17.84204    6.91054  -2.582
## lact4        -18.58159    8.87041  -2.095
## lact5        -45.92505   14.67266  -3.130
## herd2       -123.27145   20.29486  -6.074
## herd4         67.40094   33.03755   2.040
## herd5         19.32710   28.09209   0.688
## herd6        -93.34786   26.27480  -3.553
## herd7        -92.45223   33.70998  -2.743
## herd8       -107.52354   25.35701  -4.240
## herd13        -0.98249   27.37090  -0.036
## herd14        25.97963   19.48136   1.334
## herd17       -29.33251   29.10324  -1.008
## herd18      -173.18362   26.17006  -6.618
## herd22      -155.14524   27.20152  -5.704
## herd23      -100.03718   22.54883  -4.436
## herd26       -44.84622   51.63849  -0.868
## herd27      -143.50703   47.65044  -3.012
## herd29       -28.14249   28.89087  -0.974
## herd30        19.73302   31.17323   0.633
## herd34       -53.04222   24.97092  -2.124
## herd36       -69.43396   29.60786  -2.345
## herd37       -38.01074   26.33616  -1.443
## herd38       -26.91082   23.74423  -1.133
## herd41       -11.92557   26.88505  -0.444
## herd45       -44.91041   23.66587  -1.898
## herd46        -6.06832   27.10444  -0.224
## herd48        68.73228   27.74948   2.477
## herd49       -54.19039   38.39780  -1.411
## herd52        83.47301   24.45448   3.413
```

```
## herd53       -34.28559   32.09108  -1.068
## herd54       -56.27583   29.88240  -1.883
## herd55      -197.87715   31.77086  -6.228
## herd59       -21.23528   22.56866  -0.941
## herd60      -105.62412   23.08122  -4.576
## herd62       -34.51391   33.05735  -1.044
## herd64        36.61065   24.83194   1.474
## herd66       -90.63616   27.73316  -3.268
## herd68       -53.87295   23.42433  -2.300
## herd69      -128.47790   22.72433  -5.654
## herd70      -245.93829   25.11742  -9.792
## herd75       -76.06256   23.75975  -3.201
## herd77      -140.65586   40.41529  -3.480
## herd89      -219.06039   34.79154  -6.296
## herd90      -122.09286   26.20182  -4.660
## herd95       160.31707  102.01316   1.572
## herd96       121.54513  102.10996   1.190
## herd97       228.93906  142.31594   1.609
## herd98       -87.30903   57.56693  -1.517
## herd99      -108.06358  102.30015  -1.056
## herd100       94.16122  102.23388   0.921
## herd101      -45.31427  102.10462  -0.444
## herd103     -351.44585  102.08486  -3.443
## herd104     -132.35824   84.36061  -1.569
## herd105     -194.04205  142.43226  -1.362
## herd106     -110.85591  103.52642  -1.071
## herd107      -95.48285  142.47691  -0.670
## herd108     -111.05764  142.50439  -0.779
## herd109     -106.74405   38.17333  -2.796
## herd110       97.05187  102.02807   0.951
## dim            0.27360    0.02298  11.907
##
## Correlation matrix not shown by default, as p = 62 > 12.
## Use print(summary(l), correlation=TRUE)  or
##     vcov(summary(l))        if you need it
## Linear mixed model fit by REML ['lmerMod']
## Formula: prot ~ lact + herd + dim + (1 | sire)
##    Data: ptbl_data
##
## REML criterion at convergence: 40474.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6737 -0.6009  0.0550  0.6335  3.0182
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sire     (Intercept)   239.4   15.47
##  Residual             10101.1  100.50
## Number of obs: 3397, groups:  sire, 38
##
## Fixed effects:
##               Estimate Std. Error t value
```

```
## (Intercept)  687.84915   15.05458  45.690
## lact2         -5.22348    4.26338  -1.225
## lact3        -13.56882    4.94697  -2.743
## lact4        -14.32312    6.34887  -2.256
## lact5        -21.14774   10.50091  -2.014
## herd2        -15.49471   14.46642  -1.071
## herd4         80.44214   22.26248   3.613
## herd5        101.42674   19.77209   5.130
## herd6        -49.97128   18.65285  -2.679
## herd7         53.85117   23.98051   2.246
## herd8        -24.85331   17.95591  -1.384
## herd13        66.03249   19.41086   3.402
## herd14         9.70295   13.92509   0.697
## herd17       109.89813   20.64402   5.323
## herd18       -74.94346   18.58298  -4.033
## herd22       -43.73434   19.31041  -2.265
## herd23       -11.58183   15.94946  -0.726
## herd26        20.61570   36.78421   0.560
## herd27       -48.33932   33.91609  -1.425
## herd29       -32.98307   20.46449  -1.612
## herd30       129.62836   21.60393   6.000
## herd34         1.63952   17.67893   0.093
## herd36       -19.36150   20.99663  -0.922
## herd37        74.19203   18.65291   3.978
## herd38        55.01220   16.88826   3.257
## herd41        20.74571   18.95723   1.094
## herd45       -10.91260   16.73092  -0.652
## herd46        48.82779   19.23558   2.538
## herd48       146.24036   19.70026   7.423
## herd49        -4.67361   27.40547  -0.171
## herd52        92.37518   17.34591   5.325
## herd53        25.79652   22.66968   1.138
## herd54       -18.60610   21.21973  -0.877
## herd55      -109.31983   22.55610  -4.847
## herd59        -3.59154   15.92558  -0.226
## herd60       -13.91199   16.40581  -0.848
## herd62        67.84174   23.52892   2.883
## herd64       112.02522   17.53687   6.388
## herd66        30.25665   19.68037   1.537
## herd68       -10.80834   16.61041  -0.651
## herd69        -7.17148   16.10200  -0.445
## herd70       -81.03427   17.70222  -4.578
## herd75        51.84225   16.83448   3.080
## herd77       -62.40786   28.82230  -2.165
## herd89      -136.54329   22.51005  -6.066
## herd90         9.09527   18.61498   0.489
## herd95        57.54430   72.94915   0.789
## herd96        18.32784   73.01133   0.251
## herd97       155.07864  101.83915   1.523
## herd98       -40.59226   41.04723  -0.989
## herd99       -79.08308   73.12301  -1.082
## herd100       -0.36401   73.06205  -0.005
## herd101        7.31608   73.00460   0.100
## herd103     -192.35676   72.97616  -2.636
```

```
## herd104     -144.28294   60.28039  -2.394
## herd105     -102.10287  101.90064  -1.002
## herd106        5.44255   73.66756   0.074
## herd107     -132.35566  101.91766  -1.299
## herd108      115.50101  101.94846   1.133
## herd109      -48.20625   27.20305  -1.772
## herd110      -57.81409   72.94988  -0.793
## dim            0.24386    0.01644  14.831
##
## Correlation matrix not shown by default, as p = 62 > 12.
## Use print(summary(l), correlation=TRUE)  or
##     vcov(summary(l))        if you need it
## Linear mixed model fit by REML ['lmerMod']
## Formula: scs ~ lact + herd + dim + (1 | sire)
##    Data: ptbl_data
##
## REML criterion at convergence: 11023.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0849 -0.7031 -0.1689  0.5332  5.1888
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sire     (Intercept) 0.02892  0.1701
##  Residual             1.47787  1.2157
## Number of obs: 3397, groups:  sire, 38
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  2.9028929  0.1807507  16.060
## lact2       -0.0068241  0.0515663  -0.132
## lact3        0.1015210  0.0598323   1.697
## lact4       -0.0161535  0.0767806  -0.210
## lact5        0.1372396  0.1269878   1.081
## herd2        0.2219766  0.1745385   1.272
## herd4       -0.5058221  0.2613789  -1.935
## herd5       -0.4795203  0.2369695  -2.024
## herd6       -0.4903804  0.2245745  -2.184
## herd7       -0.0414502  0.2889896  -0.143
## herd8       -0.2019707  0.2158770  -0.936
## herd13      -0.6138834  0.2335827  -2.628
## herd14       0.0024948  0.1682780   0.015
## herd17      -0.1710233  0.2484421  -0.688
## herd18       0.3150843  0.2237503   1.408
## herd22       0.2507096  0.2324894   1.078
## herd23      -0.1154279  0.1916420  -0.602
## herd26      -0.7957806  0.4437595  -1.793
## herd27      -0.2174592  0.4089785  -0.532
## herd29      -0.1476948  0.2460665  -0.600
## herd30       1.3440147  0.2569701   5.230
## herd34       0.5248772  0.2125322   2.470
## herd36       0.4901778  0.2526440   1.940
```

```
## herd37       0.6412569  0.2242814   2.859
## herd38      -0.5857264  0.2035064  -2.878
## herd41      -0.0321614  0.2274032  -0.141
## herd45       0.1408516  0.2009714   0.701
## herd46      -0.4831023  0.2315651  -2.086
## herd48      -0.4692556  0.2371760  -1.979
## herd49       0.4456772  0.3309014   1.347
## herd52       0.5490433  0.2087311   2.630
## herd53       0.0028894  0.2721992   0.011
## herd54      -0.6486718  0.2555015  -2.539
## herd55       1.1224558  0.2715587   4.133
## herd59      -0.0913609  0.1911135  -0.478
## herd60      -0.3185984  0.1976264  -1.612
## herd62       0.2421185  0.2836912   0.853
## herd64       0.0290572  0.2105039   0.138
## herd66       0.1926726  0.2368906   0.813
## herd68       0.0980277  0.1998564   0.490
## herd69       0.1578356  0.1936641   0.815
## herd70       0.9185132  0.2122622   4.327
## herd75       0.4677678  0.2024318   2.311
## herd77       0.6308696  0.3478703   1.814
## herd89      -0.0945232  0.2599028  -0.364
## herd90       0.6427054  0.2241886   2.867
## herd95       0.6639766  0.8817645   0.753
## herd96       0.2651099  0.8824714   0.300
## herd97       1.4345314  1.2314394   1.165
## herd98      -0.7483216  0.4953909  -1.511
## herd99       1.0221822  0.8836747   1.157
## herd100     -1.3204170  0.8828543  -1.496
## herd101      0.4115713  0.8823720   0.466
## herd103     -0.0370665  0.8819429  -0.042
## herd104      0.6027929  0.7283269   0.828
## herd105      1.8565615  1.2320502   1.507
## herd106     -0.4571387  0.8885900  -0.514
## herd107      1.6780828  1.2321659   1.362
## herd108      0.0589352  1.2326072   0.048
## herd109      0.6751412  0.3281733   2.057
## herd110     -0.7010447  0.8817106  -0.795
## dim         -0.0005187  0.0001988  -2.609

##
## Correlation matrix not shown by default, as p = 62 > 12.
## Use print(summary(l), correlation=TRUE)  or
##     vcov(summary(l))        if you need it
```

- Compute estimated heritability. For the sire model, the heritability h^2 is defined as

$$h^2 = \frac{4 * \sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

The estimates for the sire variance σ_s^2 and σ_e^2 can be extracted from the summary of the `lmer` result.

```
for (l in l_lmer_results){
  smry_l <- summary(l)
  # getting sire std. dev
  n_sire_var <- attr(smry_l$varcor[[1]], "stddev")[[1]]
```

```
  n_res_var <- smry_l$sigma
  n_h2 <- 4*n_sire_var / (n_sire_var+n_res_var)
  s_trait <- unlist(strsplit(as.character(smry_l$call)[2], split = " ~ ", fixed = TRUE))[1]
  cat("Heritability for trait: ", s_trait, ": ", n_h2, "\n")
}
```

```
## Heritability for trait:  milk :  0.5944334
## Heritability for trait:  fat :  0.624975
## Heritability for trait:  prot :  0.533636
## Heritability for trait:  scs :  0.4908639
```

- Obtain summary statistics for predicted sire breeding values. The breeding values can be obtained by the function `raneff()`

```
for (l in l_lmer_results){
  smry_l <- summary(l)
  s_trait <- unlist(strsplit(as.character(smry_l$call)[2], split = " ~ ", fixed = TRUE))[1]
  cat("Summary statistics of breeding values for trait: ", s_trait, "\n")
  print(summary(lme4::ranef(l)$sire[[1]]))
}
```

```
## Summary statistics of breeding values for trait:  milk
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1168.09  -304.68    25.59     0.00   275.90  1218.04
## Summary statistics of breeding values for trait:  fat
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -37.375 -13.799  -3.095   0.000  14.730  51.130
## Summary statistics of breeding values for trait:  prot
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -24.7325  -4.2736   0.8155   0.0000   6.1330  28.8272
## Summary statistics of breeding values for trait:  scs
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.324984 -0.078694 -0.006315  0.000000  0.081455  0.241742
```