Chapter 19

Analysis for Regional Trials with Unbalanced Data

Jun Zhu

Purpose

To analyze unbalanced data of regional trials for comparing varieties by linear contrast tests.

Definitions

Statistical Model

Analysis of experimental data from regional trials is based on the following linear model, which regards the genotypic effects (G) as fixed and further partitions the random E ($E = Y + L + YL$) and GE interaction effects ($GE = GY + GL + GYL$) ($h = 1,2,...,g$; $i = 1,2,..., n_h$; $j = 1,2,..., n_{hi}$; $k = 1,2,..., r$):

$$y_{hijk} = G_h + Y_i + L_j + YL_{ij} + GY_{hi} + GL_{hj} + GYL_{hij} + B_{k(ij)} + e_{hijk}$$

where, Y = year effect, L = location effect, YL = year × location interaction effect, GY = genotype × year effect, GL = genotype × location effect, GYL = genotype × year × location interaction effect, B = block effect, and e = residual effect.

Analysis

Balanced data of regional trials can be easily analyzed using ANOVA methods. The experimental data of regional trials, however, are quite often unbalanced because of missing genotype records in specific locations and years. Mixed model approaches can then be applied for analyzing unbalanced data of a single trait and multiple traits from regional trials (Zhu, Lai, and Xu, 1993; Zhu, Xu, and Lai, 1993). The phenotypic data of trait y ($f = 1, 2, \ldots, t$) can be expressed in a matrix form of a mixed linear model:

$$\begin{aligned} y_{(f)} &= Xb_{(f)} + U_Y e_{Y(f)} + U_L e_{L(f)} + U_{YL} e_{YL(f)} + U_{GY} e_{GY(f)} + U_{GL} e_{GL(f)} \\ &\quad + U_{GYL} e_{GYL(f)} + U_B e_{B(f)} + e_{\varepsilon(f)} \\ &= Xb_{(f)} + \sum_{u=1}^{8} U_u e_{u(f)} \end{aligned}$$

with variance matrix:

$$\begin{aligned} \text{var}(y_{(f)}) &= \sigma^2_{Y(f)} U_Y U_Y^T + \sigma^2_{L(f)} U_L U_L^T + \sigma^2_{YL(f)} U_{YL} U_{YL}^T + \sigma^2_{GY(f)} U_{GY} U_{GY}^T \\ &\quad + \sigma^2_{GL(f)} U_{GL} U_{GL}^T + \sigma^2_{GYL(f)} U_{GYL} U_{GYL}^T + \sigma^2_{B(f)} U_B U_B^T + \sigma^2_{\varepsilon(f)} I \\ &= \sum_{u=1}^{8} \sigma^2_{u(f)} U_u U_u^T = V_{(f)} \end{aligned}$$

The covariance between traits $y_{(f)}$ and $y_{(f)}$ is

$$C_{(f,f')} = \sum_{u=1}^{8} \sigma_{u(f)/u(f')} U_u U_u^T.$$

Both variance ($V_{(f)}$) and covariance ($C_{(f,f')}$) matrices can be estimated by the MINQUE(1) method (Zhu, 1992; Zhu and Weir, 1996). Comparison of genotypes for trait f can be conducted by testing linear contrast among genotype effects ($\sum_{h=1}^{g} c_h G_{h(f)}$). The linear contrast can be estimated by

$$C_{(f)} = c^T \hat{b} = c^T (X^T \hat{V}_{(f)}^{-1} X)^- X^T \hat{V}_{(f)}^{-1} y_{(f)}$$

with sampling variance $\hat{\sigma}^2(C_{(f)}) = c^T (X^T \hat{V}_{(f)}^{-1} X)^- c$.

If $|C_{(f)} / \hat{\sigma}(C_{(f)})| > z_{(a/2)}$, reject the null hypothesis H_0: $\sum_{h=1}^{g} c_h G_{h(f)} = 0$ and accept the alternative hypothesis H_1: $\sum_{h=1}^{g} c_h G_{h(f)} \neq 0$ at a significance level = α.

To compare the weighted genotypic merits of t traits ($\sum_{f=1}^{t} w_f G_{h(f)}$), the weighted linear contrast can be estimated by

$$C_W = \sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f \hat{G}_{h(f)} = \sum_{f=1}^{t} w_f C_{(f)}$$

with sampling variance

$$\sigma^2(C_W) = \sum_{f=1}^{t} w_f^2 \sigma^2(C_{(f)}) + 2\sum_{f=1}^{t-1} \sum_{f'=f+1}^{t} w_f w_{f'} \sigma(C_{(f)}, C_{(f')})$$

where, $\hat{\sigma}(C_{(f)}, C_{(f')}) = c^T (X^T \hat{C}^{-1}_{(f,f')} X)^- c$ is the covariance between $C_{(f)}$ and $C_{(f')}$.

If $|C_W / \hat{\sigma}(C_W)| > z_{(\alpha/2)}$, reject the null hypothesis H_0: $\sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f G_{h(f)} = 0$ and accept the alternative hypothesis H_1: $\sum_{h=1}^{g} c_h \sum_{f=1}^{t} w_f G_{h(f)} \neq 0$ at a significant level = α.

Originators

Zhu, J. (1992). Mixed model approaches for estimating genetic variances and covariances. *Journal of Biomathematics* 7(1):1-11.

Zhu, J., Lai, M.G., and Xu, F.H. (1993). Analysis methods for unbalanced data from regional trial of crop variety: Analysis for multiple traits (Chinese). *Journal of Zhejiang Agricultural University* 19(3):241-247.

Zhu, J. and Weir, B.S. (1996). Diallel analysis for sex-linked and maternal effects. *Theoretical and Applied Genetics* 92(1):1-9.

Zhu, J., Xu, F.H., and Lai, M.G. (1993). Analysis methods for unbalanced data from regional trials of crop variety: Analysis for single trait (Chinese). *Journal of Zhejiang Agricultural University* 19(1):7-13.

Software Available

Zhu, J. (1997). GENTEST.EXE a computer software for constructing regional test models, GENTESTM.EXE for analyzing single traits of regional tests, and GENTESTW.EXE for analyzing multiple traits of regional tests. *Analysis Methods for Genetic Models* (pp. 285-292), Agricultural Publication House of China, Beijing (program free of charge). Contact Dr. Jun Zhu, Department of Agronomy, Zhejiang University, Hangzhou, China. E-mail: <jzhu@zju.edu.cn>.

EXAMPLE

Unbalanced data (COTTEST.TXT) to be analyzed (Variety = 3, Year = 2, Location = 8, Blk = 1):

Var	Year	Loca	Blk	Yield	Lint%
1	1	1	1	65.7	40.6
1	1	2	1	55.9	40.0
1	1	3	1	83.3	39.5
1	1	4	1	47.0	38.3
1	1	5	1	63.0	40.7
1	1	6	1	26.1	37.6
1	2	1	1	64.7	39.3
1	2	2	1	61.9	40.5
1	2	3	1	58.2	40.0
1	2	4	1	45.3	38.6
1	2	5	1	56.7	38.8
1	2	6	1	44.8	37.3
1	2	7	1	46.7	40.2
1	2	8	1	52.1	39.0
2	1	1	1	64.3	44.0
2	1	2	1	64.2	42.7
2	1	3	1	69.7	43.8
2	1	4	1	34.3	40.3
2	1	5	1	59.4	43.9
2	1	6	1	63.3	42.3
2	1	7	1	59.1	43.3
2	1	8	1	76.2	44.5
2	2	1	1	65.7	42.7
2	2	2	1	78.4	44.5
2	2	3	1	66.6	43.5
2	2	4	1	48.6	41.4
2	2	5	1	70.0	42.3
2	2	6	1	61.0	40.4
2	2	7	1	63.0	45.1
2	2	8	1	73.6	45.0
3	1	1	1	61.4	41.9
3	1	2	1	75.9	38.8
3	1	3	1	75.3	40.0
3	1	4	1	61.3	37.6
3	1	5	1	64.1	40.4
3	1	6	1	57.8	38.2
3	1	7	1	86.8	40.5
3	1	8	1	64.8	40.4
3	2	3	1	72.3	40.0

3	2	4	1	50.7	38.1
3	2	5	1	52.7	38.8
3	2	6	1	63.3	37.8
3	2	7	1	72.0	40.3
3	2	8	1	73.2	40.0

1. Use GENTEST.EXE for generating mating design matrix and data. Before running these programs, create a file for your analysis with four design columns, followed by trait columns. The four design columns are: variety, year, location, and block. There is a limitation (<100 traits) for the number of trait columns.
2. Run GENTESTM.EXE for analyzing each trait. Standard errors of estimates are calculated by jackknifing over locations for stability testing. Always run GENTESTM.EXE before analyzing multiple traits. This program will allow you to choose data transformation based on check variety. You will also need to input coefficients (1, 0, or –1) for conducting linear contrasts for different varieties. The results will be automatically stored in a file named COTTEST.VAR for analysis of single traits.
3. After you finish analysis for each trait, run GENTESTW.EXE for combining analysis of all traits studied. This program will allow you to choose weight coefficients for each trait. The results will be automatically stored in a file named COTTEST.VAR for analysis of multiple traits.

Output 1 for Single Trait Test

```
Traits =, 2
 Variance components =, 6
 File name is cottest.VAR
 Date and Time for Analysis: Thu Jun 22 20:36:15 2000

 Variance Components Estimated by MINQUE(1) with GENTESTW.EXE.
Contrast 1:  +  +  -
Contrast 2:  +  -
Contrast 3:  +  -  -

Analysis of trait Yield
Estimate of Var(Y) =, 0
Estimate of Var(L) =, 43.7906
Estimate of Var(YL) =, 0
Estimate of Var(GY) =, 1.79863
Estimate of Var(GL) =, 31.5667
Estimate of Var(e) =, 55.9106
```

```
Mean of Variety:,     Mean,     S.E.
Mean of Variety 1 =,  55.1,      3.86105
Mean of Variety 2 =,  63.5875,   3.71664
Mean of Variety 3 =,  66.5429,   3.86105

( 3),  V 1 ,  55.1000,       A
( 2),  V 2 ,  63.5875,  a    AB
( 1),  V 3 ,  66.5429,  a     B

Contrast,    C-value,     S.E.,     Standard Normal z-value
(1) This Linear Contrast Test Is for Varieties: (V1, V2) vs. (V3)
Contrast 1,  -14.398222,  7.480348,  1.924806
(2) This Linear Contrast Test Is for Varieties: (V1) vs. (V2)
Contrast 2,  -8.487495,  4.215852,  2.013234
(3) This Linear Contrast Test Is for Varieties: (V1) vs. (V2, V3)
Contrast 3,  -19.930346,  7.480348,  2.664361

Stability Analysis for Variety
Estimates and S.E. are obtained by Jackknifing over environments.

Stability Analysis for Variety 1:
 a = -25.4509, S.E. = 24.3941, 0.95 C.I. is < -73.2633 & 22.3616 >
 b = 1.32148, S.E. = 0.3862, 0.95 C.I. is < 0.564532 & 2.07844 >
 r = 0.83176, S.E. = 0.0856267, 0.95 C.I. is < 0.663932 & 0.999588 >

Stability Analysis for Variety 2:
 a = 8.36712, S.E. = 25.0719, 0.95 C.I. is < -40.7739 & 57.5081 >
 b = 0.879325, S.E. = 0.386812, 0.95 C.I. is < 0.121172 & 1.63748 >
 r = 0.718145, S.E. = 0.157704, 0.95 C.I. is < 0.409044 & 1.02725 >

Stability Analysis for Variety 3:
 a = 16.9362, S.E. = 13.7335, 0.95 C.I. is < -9.98134 & 43.8538 >
 b = 0.804752, S.E. = 0.227599, 0.95 C.I. is < 0.358659 & 1.25085 >
 r = 0.740325, S.E. = 0.104486, 0.95 C.I. is < 0.535531 & 0.945118 >

Stability in Order for Variety
Order by b ( 3),  V 3 ,  a = 16.9362 ,   b = 0.8048 ,  r = 0.7403
Order by b ( 2),  V 2 ,  a = 8.3671 ,    b = 0.8793 ,  r = 0.7181
Order by b ( 1),  V 1 ,  a = -25.4509 ,  b = 1.3215 ,  r = 0.8318

Analysis of trait Lint%
Estimate of Var(Y) =, 0
Estimate of Var(L) =, 0.812176
Estimate of Var(YL) =, 0.495318
Estimate of Var(GY) =, 0
Estimate of Var(GL) =, 0.175569
Estimate of Var(e) =, 0.244097

Mean of Variety:,     Mean,    S.E.
Mean of Variety 1 =, 39.3143,  0.428769
Mean of Variety 2 =, 43.1063,  0.411924
Mean of Variety 3 =, 39.4857,  0.428769

( 3),  V 1 ,  39.3143,  a    A
( 2),  V 3 ,  39.4857,  a    A
( 1),  V 2 ,  43.1063,
```

```
Contrast,  C-value,   S.E., Standard Normal z-value
(1) This Linear Contrast Test Is for Varieties: (V1, V2) vs. (V3)
Contrast 1,  3.449110,  0.552117,  6.247059
(2) This Linear Contrast Test Is for Varieties: (V1) vs. (V2)
Contrast 2,  -3.791962,  0.297600,  12.741823
(3) This Linear Contrast Test Is for Varieties: (V1) vs. (V2, V3)
Contrast 3,  -3.963402,  0.552117,  7.178548

Stability Analysis for Variety
Estimates and S.E. are obtained by Jackknifing over environments.

Stability Analysis for Variety 1:
 a = 7.89727,  S.E. = 4.23156, 0.95 C.I. is < -0.396587 & 16.1911 >
 b = 0.772582, S.E. = 0.103208, 0.95 C.I. is < 0.570295 & 0.974869 >
 r = 0.905533, S.E. = 0.0482835, 0.95 C.I. is < 0.810898 & 1.00017 >

Stability Analysis for Variety 2:
 a = 0.177973, S.E. = 4.68567, 0.95 C.I. is < -9.00594 & 9.36189 >
 b = 1.05051,  S.E. = 0.115322, 0.95 C.I. is < 0.824481 & 1.27654 >
 r = 0.917149, S.E. = 0.0463384, 0.95 C.I. is < 0.826326 & 1.00797 >

Stability Analysis for Variety 3:
 a = 2.71998,  S.E. = 4.52641, 0.95 C.I. is < -6.15178 & 11.5917 >
 b = 0.902994, S.E. = 0.113154, 0.95 C.I. is < 0.681213 & 1.12478 >
 r = 0.937269, S.E. = 0.0332955, 0.95 C.I. is < 0.87201 & 1.00253 >

Stability in Order for Variety
Order by b ( 3),  V 1 ,  a = 7.8973 ,  b = 0.7726 ,  r = 0.9055
Order by b ( 2),  V 3 ,  a = 2.7200 ,  b = 0.9030 ,  r = 0.9373
Order by b ( 1),  V 2 ,  a = 0.1780 ,  b = 1.0505 ,  r = 0.9171

Time Used (Hour) = 0.009722
```

Output 2 for Multiple Trait Test

```
Traits =, 2
 Variance components =, 6
 File name is cottest.COV
 Date and Time for Analysis: Thu Jun 22 20:38:33 2000

 Variance Components Estimated by MINQUE(1) with GENTESTW.EXE.

<W1>: 0.6, <W2>: 0.4,

Analysis for Public Users

Estimated Var for <Yield>
Estimate for Var(Y) =,  -7.04805
Estimate for Var(L) =,  98.8959
Estimate for Var(YL) =, -1.89563
Estimate for Var(GY) =,  4.06199
Estimate for Var(GL) =, 71.2897
Estimate for Var(e) =, 126.267
```

```
Estimated Cov for <Yield> & <Lint%>
Estimate for Cov (Y) =,  -0.0680642
Estimate for Cov (L) =,  29.2247
Estimate for Cov (YL) =, -2.47928
Estimate for Cov (GY) =,  1.12331
Estimate for Cov (GL) =,  1.57541
Estimate for Cov (e) =,   4.52009

Estimated Var for <Lint%>
Estimate for Var(Y) =, -0.280077
Estimate for Var(L) =,  5.20919
Estimate for Var(YL) =, 3.1769
Estimate for Var(GY) =,-0.097269
Estimate for Var(GL) =, 1.12607
Estimate for Var(e) =,  1.56561

Analysis for multiple traits:
Combined Variety Mean:
Mean of Variety:,     Mean,    S.E.
Mean of Variety 1 =,  89.5086, 3.82078
Mean of Variety 2 =, 101.003,  3.68336
Mean of Variety 3 =, 100,      3.82078

 ( 3),  V 1 ,  89.5086,   a   A
 ( 2),  V 3 ,  100.0000,  a   A
 ( 1),  V 2 ,  101.0029,  a   A

Contrast,  C-value,   S.E.  , Standard Normal z-value
(1) This Linear Contrast Test Is for Varieties:
Cont. 1,  -9.488488,  49.575051,  0.191396
(2) This Linear Contrast Test Is for Varieties:
Cont. 2,  -11.494320,  15.667476,  0.733642
(3) This Linear Contrast Test Is for Varieties:
Cont. 3,  -21.985739,  49.575054,  0.443484

Stability Analysis for Variety
Estimates and S.E. are obtained by Jackknifing over environments.

Stability Analysis for Variety 1:
 a = -28.7855, S.E. = 34.6409, 0.95 C.I. is < -96.6817 & 39.1106 >
 b = 1.23013, S.E. = 0.351782, 0.95 C.I. is < 0.540642 & 1.91963 >
 r = 0.827958, S.E. = 0.0706079, 0.95 C.I. is < 0.689566 & 0.966349 >

Stability Analysis for Variety 2:
 a = 11.0869, S.E. = 35.9274, 0.95 C.I. is < -59.3307 & 81.5045 >
 b = 0.917704, S.E. = 0.35883, 0.95 C.I. is < 0.214397 & 1.62101 >
 r = 0.753344, S.E. = 0.1381, 0.95 C.I. is < 0.482668 & 1.02402 >

Stability Analysis for Variety 3:
 a = 21.659, S.E. = 19.4541, 0.95 C.I. is < -16.4711 & 59.7891 >
 b = 0.808956, S.E. = 0.204255, 0.95 C.I. is < 0.408615 & 1.2093 >
 r = 0.781107, S.E. = 0.0871521, 0.95 C.I. is < 0.610289 & 0.951926 >

Stability in Order for Variety
Order of b ( 3),  V 3 ,  a = 21.6590 ,  b = 0.8090 ,  r = 0.7811
Order of b ( 2),  V 2 ,  a = 11.0869 ,  b = 0.9177 ,  r = 0.7533
Order of b ( 1),  V 1 ,  a = -28.7855 ,  b = 1.2301 ,  r = 0.8280

Time Used (Hour) = 0.006667
```