Least squares and weighted least squares

Rasmus Waagepetersen

October 27, 2023

Least squares versus weighted least squares

Consider general linear mixed model (vector-matrix form):

$$Y = X\beta + ZU + \epsilon$$

Least squares estimate: $\hat{\beta} = (X^\mathsf{T} X)^{-1} X^\mathsf{T} Y$

Linear mixed effects model estimate is *weighted least squares* estimate

$$\hat{\beta} = (X^\mathsf{T} \Sigma^{-1} X)^{-1} X^\mathsf{T} \Sigma^{-1} Y$$

where Σ is matrix of variances and covariances.

If data has nice structure (balanced), least squares and weighted least squares may coincide (e.g. orthodont data) but this is generally not the case.

Weighted least squares is superior to least squares (more accurate) because it takes into account correlations between observations.

Case study: fraction test results for 4th grade students

Computing with fractions is a key obstacle for primary school students

TRACK project is a randomized trial to test a new math teaching system against current practice.

125 schools were randomly allocated to current practice or new TRACK teaching system. 12717 students participate in the trial.

Currently we have data for Autumn 22 (baseline) and Spring 23 (one school year of treatment).

Fixed effects model

Parameters: main effect of teaching method (b), main effect of participating in school year (a), interaction effect (g).

Expected values in two-way table:

	A22 (baseline)	S23
Current	μ	$\mu + a$
New	$\mu + b$	$\mu + b + a + g$

Which parameter is the key parameter ?

Could we say anything about the parameter b ahead of analysing the data ?

Least squares analysis (ordinary two-way ANOVA)

```
> fitlm=lm(Broker~treatment+Test+Test*treatment,data=scores)
> summary(fitlm)
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     15.7001     0.6249  25.122  < 2e-16 *** #mu
treatment       -4.5520     0.8946  -5.089 3.67e-07 *** #b
Test            10.1875     0.3987  25.552  < 2e-16 *** #a
treatment:Test   2.1282     0.5691   3.740 0.000185 *** #g
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.31 on 10115 degrees of freedom

> 14.31^2
[1] 204.7761
```

Residual variance 204.8

Conclusions ? Any potential problems with this analysis ?

Mixed model analysis

```
> fitbrok=lmer(Broker~treatment+Test+Test*treatment
                  +(1|instnr/Klasse)+(1|ID),data=scores)

 Groups        Name        Variance Std.Dev.
 ID            (Intercept) 126.401  11.243
 Klasse:instnr (Intercept)   5.581   2.362
 instnr        (Intercept)  16.806   4.099
 Residual                   57.057   7.554
Number of obs: 10119, groups:  ID, 5621; Klasse:instnr, 272; instnr, 119

                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)      14.3192     0.7148  178.7729  20.033  < 2e-16 *** #mu
treatment        -3.1664     1.0153  180.5600  -3.119  0.00211 **  #b
Test             10.5437     0.2210 4758.4397  47.700  < 2e-16 *** #a
treatment:Test    1.7947     0.3167 4770.6955   5.667 1.54e-08 *** #g

> 126.401+5.581+16.806+57.057#Total variance
[1] 205.845
> (126.401+5.581+16.806)/205.845#Intra student correlation
[1] 0.7228157 #correlation for two observations for same student
```

Same total variance (NB: does not make sense to add standard deviations)

Large correlation 0.72 between two test for same student

Still, results in this case quite similar to least squares analysis.

Analysis based on differences

Similar to paired t-test we may consider the difference between spring and autumn test for each student.

This eliminates parameter b and random effects !

Difference for student with current:

$$\begin{aligned} Y_{i2} - Y_{i1} =& \mu + a + U_i + U_{class(i)} + U_{school(i)} + \epsilon_{i2} \\ & -(\mu + U_i + U_{class(i)} + U_{school(i)} + \epsilon_{i1}) = a + \tilde{\epsilon}_i \end{aligned}$$

Similarly, for student with new:

$$Y_{i2} - Y_{i1} = a + g + \tilde{\epsilon}_i$$

Least squares analysis based on differences

```
> fitdiffbroklm=lm(diffbrok~treatment.x,data=scoresdiff)
> summary(fitdiffbroklm)

Call:
lm(formula = diffbrok ~ treatment.x, data = scoresdiff)

Residuals:
    Min      1Q  Median      3Q     Max
-52.664  -6.664  -0.336   6.664  60.336

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.6640     0.2247  47.464  < 2e-16 *** #a
treatment.x   1.6723     0.3191   5.241 1.67e-07 *** #g
```

Now treatment is effect of TRACK teaching method (similar estimate as in previous analysis.

Note: number of difference observations 4498 is less than half of previous number of observations.

Issue with difference approach: loss of observations in cases where one test is missing for a student.

Not so easy to generalize if more than two tests (if several differences they will be correlated).

If we use Y_{i1} as a covariate (regression variable) for several following tests, then we use Y_{i1} as a surrogate for the ability of the student but Y_{i1} is a noisy measurement of this ability.

Mixed model analysis based on differences

We can still try adding random effect for class and school:

```
> fitdiffbrok=lmer(diffbrok~treatment.x+(1|instnr.x/Klasse.x),data=scoresdiff)
> summary(fitdiffbrok)
Random effects:
 Groups            Name        Variance Std.Dev.
 Klasse.x:instnr.x (Intercept) 12.454   3.529
 instnr.x          (Intercept)  3.262   1.806
 Residual                      99.305   9.965
Number of obs: 4498, groups:  Klasse.x:instnr.x, 249; instnr.x, 115

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  10.6798     0.4612 104.8952  23.155   <2e-16 *** #a
treatment.x   1.7133     0.6535 106.2772   2.622     0.01 *   #g
```

Class and school variances rather minor compared to residual variance

Standard errors higher than for linear regression due to estimated correlation within classes and schools. *p*-value for treatment bigger but still significant at 5% level.

Correlations at different level of hierarchy

Denote: σ^2, $\sigma^2_{student}$, σ^2_{class} and σ^2_{school} variances for noise, student, class and school.

Correlation between two tests for same student:

$$\frac{\sigma^2_{student} + \sigma^2_{class} + \sigma^2_{school}}{\sigma^2_{student} + \sigma^2_{class} + \sigma^2_{school} + \sigma^2}$$

(tests share all random effects except noise)

Correlation between two test for two students within the same class

$$\frac{\sigma^2_{class} + \sigma^2_{school}}{\sigma^2_{student} + \sigma^2_{class} + \sigma^2_{school} + \sigma^2}$$

(tests only share class and school random effects)

Correlation between two test for two students in different classes but same school

$$\frac{\sigma^2_{school}}{\sigma^2_{student} + \sigma^2_{class} + \sigma^2_{school} + \sigma^2}$$

(tests only share school random effect)

For last mixed model analysis based on differences there is no student effect. The two last correlations are

```
> (12.454+3.262)/(12.454+3.262+99.305)
[1] 0.1366359
> 3.262/(12.454+3.262+99.305)
[1] 0.02836004
```

Both correlations quite small - differencing may have worked.

Tests for variances ?

One may be interested in testing whether a variance is zero.

Unfortunately the theory for this is incomplete and it seems that no both generally and easily applicable approach is available.