An Example Analysis Based on the Aitken Model

© 2010 Dan Nettleton

An Example Experiment

Researchers were interested in comparing the dry weight of maize seedlings from two different genotypes. For each genotype, nine seeds were planted in each of four trays. The eight trays in total were randomly positioned in a growth chamber. Three weeks after the emergence of the first seedling, emerged seedlings were harvested from each tray and weighed together after drying to obtain one weight for each tray. Although nine seeds were planted in each tray, fewer than nine seedlings emerged in many of the trays. Thus, weights were recorded on a per seedling basis, and the number of seedlings that emerged in each tray was also recorded.
```r
# Load the data

d = read.delim("http://www.public.iastate.edu/~dnett/S511/SeedlingDryWeight.txt")

d

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Tray</th>
<th>AverageWeight</th>
<th>NumberOfPerSeedling</th>
<th>Seedlings</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>18</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>14</td>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>19</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>13</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>15</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

y = d[,3]
geno = d[,1]
count = d[,4]

V = diag(1/count)

# Compute V^{-0.5}
A = diag(sqrt(count))

# In general, we could compute V^{-0.5} as follows:
e = eigen(V)
A = e$vectors %*% diag(1/sqrt(e$values)) %*% t(e$vectors)

X = model.matrix(~geno, nrow=8)

X

[,1] [,2]
[1,] 1 0
[2,] 1 0
[3,] 1 0
[4,] 1 0
[5,] 1 1
[6,] 1 1
[7,] 1 1
[8,] 1 1

# Now transform y and X to z and W.
z = A %*% y
W = A %*% X
o = lm(z ~ W - 1)
```
summary(o)

Call: 
  lm(formula = z ~ W - 1)

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
W1   16.103   1.650   9.763  6.64e-05 ***
W2  -4.622   2.376  -1.946   0.0997 .

Residual standard error: 8.883 on 6 degrees of freedom
Multiple R-squared:  0.9590,  Adjusted R-squared:  0.9454
F-statistic: 70.21 on 2 and 6 DF,  p-value: 6.882e-05

#Because V is diagonal in this case, we can alternatively analyze using lm and the weights argument.

o2=lm(y~geno,weights=count)
summary(o2)

Call: 
  lm(formula = y ~ geno, weights = count)

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
  (Intercept) 16.103  1.650   9.763  6.64e-05 ***
  genoB        -4.622  2.376  -1.946   0.0997 .

Residual standard error: 8.883 on 6 degrees of freedom
Multiple R-squared:  0.3868,  Adjusted R-squared:  0.2847
F-statistic: 3.785 on 1 and 6 DF,  p-value: 0.09966

#The unweighted (OLS) analysis is inferior in this case. The OLS estimator of beta is still unbiased, but it's variance is larger than that of the GLS estimator. OLS inferences regarding beta are not, in general, valid.

o3=lm(y~geno)
summary(o3)

Call: 
  lm(formula = y ~ geno)

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
  (Intercept) 15.250  1.750   8.714  0.000126 ***
  genoB        -3.500  2.475  -1.414   0.207031

Residual standard error: 3.50 on 6 degrees of freedom
Multiple R-squared:  0.2500,  Adjusted R-squared:  0.1250
F-statistic:  2.0 on 1 and 6 DF,  p-value: 0.2070