The Aitken Model

The Spectral Decomposition Theorem

An $n \times n$ symmetric matrix $\boldsymbol{H}$ may be decomposed as

$$\boldsymbol{H} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}' = \sum_{i=1}^{n} \lambda_i \, \boldsymbol{p}_i \, \boldsymbol{p}_i',$$

where $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]$ is an $n \times n$ orthogonal matrix whose columns $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ are the orthonormal eigenvectors of $\boldsymbol{H}$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_i, \ldots, \lambda_n)$ is a diagonal matrix whose diagonal entries $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are the eigenvalues of $\boldsymbol{H}$ (with λ_i corresponding to $\boldsymbol{p}_i$ for $i = 1, \ldots, n$).

Note that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$ and $\boldsymbol{P}$ $n \times n$ implies $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$.

Homework Problem

Suppose H is a symmetric matrix.

      a) Prove that H is non-negative definite (NND)
          if and only if all its eigenvalues are non-negative.

      b) Prove that H is positive definite (PD)
          if and only if all its eigenvalues are positive.

Symmetric Square Root Matrix

Suppose H is an $n \times n$ symmetric, NND matrix. Then there exists a symmetric, NND matrix B such that $BB = H$.

Proof: Because H is symmetric, it has the spectral decomposition $H = P\Lambda P'$, where the columns of P are the eigenvectors of H and the eigenvalues of H are the diagonal elements of the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

Because H is NND, $\lambda_i \geq 0$ for all $i = 1, \ldots, n$. Thus, $\sqrt{\lambda_i} \in \mathbb{R}^+ \cup \{0\}$ for all $i = 1, \ldots, n$.

Symmetric Square Root Matrix (continued)

Let $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$, and let $\boldsymbol{B} = \boldsymbol{P}\Lambda^{1/2}\boldsymbol{P}'$. Then $\boldsymbol{B}$ is NND because all its eigenvalues $(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ are non-negative. Furthermore,

$$\boldsymbol{B}\boldsymbol{B} = \boldsymbol{P}\Lambda^{1/2}\boldsymbol{P}'\boldsymbol{P}\Lambda^{1/2}\boldsymbol{P}' = \boldsymbol{P}\Lambda^{1/2}\Lambda^{1/2}\boldsymbol{P}' = \boldsymbol{P}\Lambda\boldsymbol{P}' = \boldsymbol{H}. \quad \square$$

The Aitken Model

- $y = X\beta + \epsilon,\ E(\epsilon) = \mathbf{0},\ \text{Var}(\epsilon) = \sigma^2 V$

- Identical to the Gauss-Markov linear model except that

$$\text{Var}(\epsilon) = \sigma^2 V \text{ instead of } \sigma^2 I.$$

- V is assumed to be a known positive definite variance matrix.

- σ^2 is an unknown positive variance parameter.

A Transformation of the Aitken Model

- Let $V^{1/2}$ be the symmetric square root of V. Note V PD implies $V^{1/2}$ is PD and therefore nonsingular.

- Using $V^{-1/2}$ to denote $(V^{1/2})^{-1}$, we have

$$V^{-1/2}y = V^{-1/2}X\beta + V^{-1/2}\epsilon.$$

- With $z = V^{-1/2}y$, $W = V^{-1/2}X$, and $\delta = V^{-1/2}\epsilon$, we have

$$z = W\beta + \delta, \ E(\delta) = \mathbf{0}, \ \text{Var}(\delta) = \sigma^2 I \text{ because}$$

$$
\begin{aligned}
\text{Var}(\delta) &= \text{Var}(V^{-1/2}\epsilon) = V^{-1/2}\sigma^2 V V^{-1/2} \\
&= \sigma^2 V^{-1/2} V^{1/2} V^{1/2} V^{-1/2} = \sigma^2 I.
\end{aligned}
$$

- Thus, after transformation, we are back to the Gauss-Markov model we are familiar with.

- We can apply all the results we have established previously to the Gauss-Markov model

$$z = W\beta + \delta, \ E(\delta) = 0, \ \text{Var}(\delta) = \sigma^2 I.$$

Estimation of $\mathrm{E}(\boldsymbol{y})$ under the Aitken Model

- Note that

$$\mathrm{E}(\boldsymbol{y}) = \mathrm{E}(\boldsymbol{V}^{1/2}\boldsymbol{V}^{-1/2}\boldsymbol{y}) = \boldsymbol{V}^{1/2}\mathrm{E}(\boldsymbol{V}^{-1/2}\boldsymbol{y}) = \boldsymbol{V}^{1/2}\mathrm{E}(\boldsymbol{z}).$$

- Because the Gauss-Markov model holds for $\boldsymbol{z}$, we already know that the best estimator of $\mathrm{E}(\boldsymbol{z})$ is

$$\begin{aligned}
\hat{\boldsymbol{z}} &= \boldsymbol{P_W}\boldsymbol{z} = \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^-\boldsymbol{W}'\boldsymbol{z} \\
&= \boldsymbol{V}^{-1/2}\boldsymbol{X}((\boldsymbol{V}^{-1/2}\boldsymbol{X})'\boldsymbol{V}^{-1/2}\boldsymbol{X})^-(\boldsymbol{V}^{-1/2}\boldsymbol{X})'\boldsymbol{V}^{-1/2}\boldsymbol{y} \\
&= \boldsymbol{V}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2}\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{V}^{-1/2}\boldsymbol{V}^{-1/2}\boldsymbol{y} \\
&= \boldsymbol{V}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.
\end{aligned}$$

- Thus, to estimate $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{V}^{1/2}\mathrm{E}(\boldsymbol{z})$, we should use

$$\boldsymbol{V}^{1/2}\hat{\boldsymbol{z}} = \boldsymbol{V}^{1/2}\boldsymbol{V}^{-1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.$$

- Likewise, if $C\beta$ is estimable, we know the BLUE is the ordinary least squares (OLS) estimator.

$$
\begin{aligned}
C(W'W)^- W'z &= C(X'V^{-1/2}V^{-1/2}X)^- X'V^{-1/2}V^{-1/2}y \\
&= C(X'V^{-1}X)^- X'V^{-1}y.
\end{aligned}
$$

- $C(X'V^{-1}X)^- X'V^{-1}y = C\hat{\beta}_V$ is called a *Generalized Least Squares* (GLS) estimator.

- $\hat{\boldsymbol{\beta}}_V = (X'V^{-1}X)^- X'V^{-1}y$ is a solution to the Aitken Equations:

$$X'V^{-1}Xb = X'V^{-1}y$$

which follow from the Normal Equations

$$\begin{aligned} W'Wb = W'z &\iff X'V^{-1/2}V^{-1/2}Xb = X'V^{-1/2}V^{-1/2}y \\ &\iff X'V^{-1}Xb = X'V^{-1}y. \end{aligned}$$

- Recall that solving the Normal Equations is equivalent to minimizing

$$(z - Wb)'(z - Wb) \text{ over } b \in \mathbb{R}^p.$$

- Note that

$$
\begin{aligned}
(z - Wb)'(z - Wb) &= (V^{-1/2}y - V^{-1/2}Xb)'(V^{-1/2}y - V^{-1/2}Xb) \\
&= ||V^{-1/2}y - V^{-1/2}Xb||^2 \\
&= ||V^{-1/2}(y - Xb)||^2 \\
&= (y - Xb)'V^{-1}(y - Xb).
\end{aligned}
$$

- Thus, $\hat{\boldsymbol{\beta}}_V = (X'V^{-1}X)^- X'V^{-1}y$ is a solution to this generalized least squares problem.

- When V is diagonal, the term "Weighted Least Squares" (WLS) is often used instead of GLS.

- When $V = \text{diag}(v_{11}, \ldots, v_{nn})$, the least squares problem becomes:

  Find $\boldsymbol{b}$ to minimize $(\boldsymbol{y} - X\boldsymbol{b})'V^{-1}(\boldsymbol{y} - X\boldsymbol{b}) = \sum_{i=1}^{n} \frac{1}{v_{ii}}(y_i - \boldsymbol{x}_i'\boldsymbol{b})^2,$

  where $\boldsymbol{x}_i'$ is the ith row of X.

- An unbiased estimator of σ^2 is

$$
\begin{aligned}
\frac{z'(I - P_W)z}{n - \text{rank}(W)} &= \frac{\|(I - P_W)z\|^2}{n - \text{rank}(W)} \\
&= \frac{\|(I - W(W'W)^- W')z\|^2}{n - \text{rank}(W)} \\
&= \frac{\left\|(I - V^{-1/2}X(X'V^{-1}X)^- X'V^{-1/2})V^{-1/2}y\right\|^2}{n - \text{rank}(V^{-1/2}X)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\left\| V^{-1/2}y - V^{-1/2}X(X'V^{-1}X)^- X'V^{-1}y \right\|^2}{n - \text{rank}(X)} \\
&= \frac{\left\| V^{-1/2}(y - X(X'V^{-1}X)^- X'V^{-1}y) \right\|^2}{n - r} \\
&= \frac{\left\| V^{-1/2}(y - X\hat{\beta}_V) \right\|^2}{n - r} \\
&= \frac{(y - X\hat{\beta}_V)'V^{-1}(y - X\hat{\beta}_V)}{n - r} \\
&= \hat{\sigma}_V^2.
\end{aligned}
$$

Inference Under the Aitken Model with Normal Errors

- The Aitken Model with Normal errors:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V}).$$

- Under the Aitken Model with Normal errors, we can back transform to convert known formulas in terms of $\boldsymbol{z}$ and $\boldsymbol{W}$ to formulas in terms of $\boldsymbol{y}$ and $\boldsymbol{X}$ to allow inference about estimable $\boldsymbol{C}\boldsymbol{\beta}$ under the Aitken Model with Normal errors.

An Example

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.
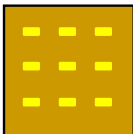
Planted Seeds
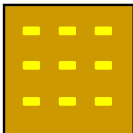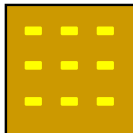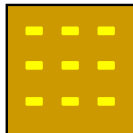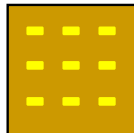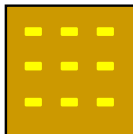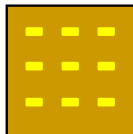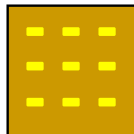
Emerged Seedlings
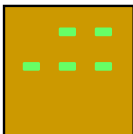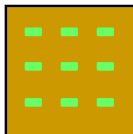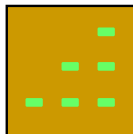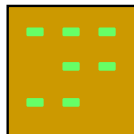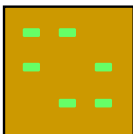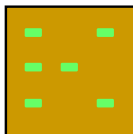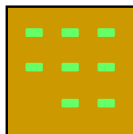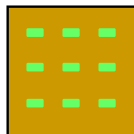
A Model for the Data

Let n_{ij} be the number of seedlings for the jth tray of genotype i ($i = 1, 2; j = 1, 2, 3, 4$).

Let y_{ijk} be the dry weight of the kth seedling in the jth tray of genotype i ($i = 1, 2; j = 1, 2, 3, 4; k = 1, \ldots, n_{ij}$).

Suppose all seedling weights are independent and normally distributed with common variance σ^2 and genotype-specific means, μ_1 for genotype 1 and μ_2 for genotype 2:

$$y_{ijk} \overset{ind}{\sim} N(\mu_i, \sigma^2).$$

A Model for the Data

Now let $y_{ij} = \bar{y}_{ij\cdot} =$ average seedling weight for genotype i and tray j.

It follows that

$$y_{ij} \overset{ind}{\sim} N(\mu_i, \sigma^2/n_{ij})$$

or, equivalently,

$$y_{ij} = \mu_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2/n_{ij}).$$

Model in Matrix and Vector Form

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \end{bmatrix}
$$

$$
\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V})
$$

$$\boldsymbol{V} = \begin{bmatrix} 1/n_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/n_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/n_{13} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/n_{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/n_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/n_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/n_{23} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/n_{24} \end{bmatrix}$$

Alternative Expression of the Same Model

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$$

```
> d=read.delim("http://.../S510/SeedlingDryWeight.txt")
> d
Genotype Tray AverageWeightPerSeedling NumberOfSeedlings
       1    1                        10                 5
       1    2                        18                 9
       1    3                        14                 6
       1    4                        19                 9
       2    5                        13                 6
       2    6                        10                 7
       2    7                        15                 6
       2    8                         9                 8

> y=d[,3]

> geno=factor(d[,1])

> count=d[,4]
```

```
> X=matrix(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
```

```
> V=diag(1/count)
>
> #Compute V^{-.5}
>
> A=diag(sqrt(count))
>
> #In general, we could compute V^{-.5} as follows:
> #
> # e=eigen(V)
> # P=e$vectors
> # lambda=e$values
> # Vn12=P%*%diag(1/sqrt(lambda))%*%t(P)
```

```
> #Now transform y and X to z and W.
>
> z=Vn12%*%y
>
> W=Vn12%*%X
>
> o=lm(z~0+W)
> summary(o)

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 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.883 on 6 degrees of freedom
Multiple R-squared: 0.959,      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)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   16.103      1.650   9.763 6.64e-05 ***
geno2         -4.622      2.376  -1.946   0.0997 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

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)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   15.250      1.750   8.714 0.000126 ***
geno2         -3.500      2.475  -1.414 0.207031
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.5 on 6 degrees of freedom
Multiple R-squared:  0.25,      Adjusted R-squared:  0.125
F-statistic:     2 on 1 and 6 DF,  p-value: 0.207
```

Some Comments on Our Model for the Data

We assumed $y_{ijk} \stackrel{ind}{\sim} N(\mu_i, \sigma^2)$.

This model ignores the trays that were used in this experiment.

It is possible that some trays provided more favorable growth conditions than others due to their position in the growth chamber, soil quality, soil moisture, etc.

The next set of slides describes a natural method for modeling tray effects that could be used if the weights of individual seedlings were available.

Some Comments on Our Model for the Data

We assumed that data on seedling weights were missing completely at random.

This assumption would be violated if, for instance, seeds that did not germinate were likely to result in small plants if they had emerged from the soil.

We also assumed that the pattern of missing data had no impact on the weights of the seedlings that did emerge from the soil.

This assumption would be violated if, for instance, plants in trays with fewer seedlings grow bigger (perhaps due to reduced competition for resources) than plants in trays with many seedlings.

Comments on the Spectral Decomposition Theorem

The Spectral Decomposition Theorem (SDT) introduced in these notes is a powerful and useful theorem.

Proofs are available in several places (including my STAT 611 notes), but it is not necessary to know the proof for STAT 510.

The SDT is a more powerful result than we really need here.

Other matrix decompositions (e.g., Cholesky Decomposition) are computationally more efficient.

We used a symmetric square root matrix (implied by the SDT) in these notes for notational and conceptual convenience.