

Lab 9: Matrices for Models

What does a linear model assume about seedlings that share a plot?

NRES 746, name: _____ group: _____

The scenario

A restoration crew planted pinyon pine seedlings in three burned plots (A, B, and C) in a Great Basin woodland. A year later you measured each seedling's height growth (cm) and the soil moisture (%) next to it: 8 seedlings per plot, 24 in all. The six seedlings below are the first three in plots A and B.

Plot	Seedling	Soil moisture (%)
A	1	8
A	2	10
A	3	12
B	1	9
B	2	11
B	3	13

The crew's model is an ANCOVA: each plot has its own intercept, and all plots share one slope for soil moisture,

$$\text{growth}_i \sim \text{Normal}(\mu_i, \sigma^2), \quad \mu_i = \beta_0 + \beta_B [\text{plot B}]_i + \beta_m \text{moisture}_i$$

where $[\text{plot B}]_i$ is 1 for a plot-B seedling and 0 otherwise. (For the six seedlings there are only two plots.)

How much does soil moisture matter, and what does this model assume about seedlings that grow side by side in the same plot?

Please keep laptops closed until the banner in Part 2. Part 1 and the group check-in are pen and paper.

Part 1: The model as matrices (individual, ~40 min)

1a. Write the model matrix X for the six seedlings, with columns for β_0 , β_B , and β_m in that order.

1b. Take $\beta = (2, 1.5, 0.5)^T$. Compute $X\beta$, and say what each of the three numbers in β means for a seedling.

1c. Compute $X^T X$. What does each entry on its diagonal count or add up?

1d. Suppose the six growth values are independent with $\sigma = 0.5$ cm. Write the covariance matrix V , its inverse, and its determinant. Then suppose plot B is more variable, with $\sigma_B = 1$ cm (plot A stays at 0.5). Write the new V , and write the quadratic form $\mathbf{r}^T V^{-1} \mathbf{r}$ for a residual vector $\mathbf{r} = (r_1, \dots, r_6)^T$ as a sum. How much does a plot-B residual count compared with a plot-A residual of the same size?

1e. Two seedlings grow side by side in the same plot, and their residuals have correlation 0.5 (each with variance 1):

$$V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Find $|V|$ and V^{-1} . Compute $\mathbf{r}^T V^{-1} \mathbf{r}$ for $\mathbf{r} = (1, 1)^T$ and for $\mathbf{r} = (1, -1)^T$, and compare both with independence ($V = I$). Which pattern of residuals is more surprising when the seedlings are correlated, and why does that make sense?

1f. Three seedlings, three proposed correlations:

$$C = \begin{pmatrix} 1 & 0.9 & -0.9 \\ 0.9 & 1 & 0.9 \\ -0.9 & 0.9 & 1 \end{pmatrix}$$

Could this be the correlation matrix of their growth? Explain without computing anything.

Group check-in (~15 min)

Compare your answers with your group, especially 1c, 1e, and 1f, and revise as needed. Agree on what each piece of X , V , and the quadratic form means before you start coding.

Laptops open now.

Part 2: Check it with R (laptop, ~50 min)

Using AI: each question has a recommendation. *Recommended: no AI* marks the core skill of the lab (building and using model and covariance matrices); R's help pages, web search, and classmates are fine. *AI optional* marks work where the results matter but writing the code by hand isn't the point. After each question, circle how you actually used AI.

What to turn in: your filled-in answers. You don't need to submit your code.

The data for all 24 seedlings are on the course website:

```
data_url <-  
  "https://kevintshoemaker.github.io/NRES-746/data/PinyonSeedlings.csv"  
seedlings <- read.csv(data_url)
```

2a. (*Recommended: no AI.*) Build X for all 24 seedlings with `model.matrix(~ plot + moisture, data = seedlings)`. Solve the normal equations, $X^T X \hat{\beta} = X^T \mathbf{y}$, using `%*`, `t()`, and `solve()`, and check your answer against `lm()`. Check: the moisture slope should be 0.44. Fill in the table.

Coefficient	Normal equations	lm()
(Intercept)		
plotB		
plotC		
moisture	0.44	0.44

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2b. (*Recommended: no AI.*) At your $\hat{\beta}$ and the MLE $\hat{\sigma} = \sqrt{\sum r_i^2 / n}$, compute the log-likelihood two ways: with `sum(dnorm(..., log = TRUE))`, and with `mvtnorm::dmvnorm(..., log = TRUE)` using $V = \hat{\sigma}^2 I$.

$\hat{\sigma} =$ _____ cm. `sum(dnorm())`: _____ `dmvnorm()`: _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2c. (*AI optional.*) Check your answers to 1e and 1f with `det()`, `solve()`, and `eigen()`. Then try to draw 1000 random values for three seedlings from a multivariate normal with correlation matrix C from 1f, using `MASS::mvrnorm()`. What happens? What do the eigenvalues of C tell you?

Eigenvalues of C : _____ , _____ , _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2d. Where did you use AI, if anywhere, and what did you have to check or fix? If you didn't use it, what was hardest to get working?

Group discussion and revision (~15 min)

Close laptops. Here is a different covariance matrix someone proposed for the six seedlings in the table (rows and columns in the same order: A1, A2, A3, B1, B2, B3):

$$V = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0 & 0 & 0 \\ 0.6 & 0.6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.6 & 0.6 \\ 0 & 0 & 0 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0 & 0.6 & 0.6 & 1 \end{pmatrix}$$

With your group, say what this V claims about the seedlings, and what could make the 0.6s biologically reasonable. Should the crew treat its 24 seedlings as 24 independent replicates when reporting the effect of plot? Record your answer, any changes you made to earlier answers, new insights, and any “muddiest points” or topics you want to review.

If you finish early. (*Recommended: no AI.*) Plot C’s seedlings look more scattered than the others. Write a negative log-likelihood that gives each plot its own σ (a diagonal V with unequal entries, as in 1d), fit it with `optim()`, and compare its AIC with the one- σ model.

$\hat{\sigma}_A =$ _____ $\hat{\sigma}_B =$ _____ $\hat{\sigma}_C =$ _____

AIC (one σ): _____ AIC (three): _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code