STA 360/602L: Module 4.3

Multivariate normal model III

Dr. Olanrewaju Michael Akande

Reading comprehension example

- Twenty-two children are given a reading comprehension test before and after receiving a particular instruction method.
  - Y_{i1}: pre-instructional score for student i.
  - Y_{i2}: post-instructional score for student i.
- Vector of observations for each student: $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})^T$.
- Clearly, we should expect some correlation between Y_{i1} and Y_{i2}.

Reading comprehension example

- Questions of interest:
  - Do students improve in reading comprehension on average?
  - If so, by how much?
  - Can we predict post-test score from pre-test score? How correlated are they?
  - If we have students with missing pre-test scores, can we predict the scores?
- We will hold off on the last question until we have learned about missing data.

Reading comprehension example

- Since we have bivariate continuous responses for each student, and test scores are often normally distributed, we can use a bivariate normal model.

- Model the data as $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2})^T \sim \mathcal{N}_2(\boldsymbol{\theta}, \Sigma)$, that is,

$$\boldsymbol{Y} = \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim \mathcal{N}_2 \left[\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right].$$

- We can answer the first two questions of interest by looking at the posterior distribution of $\theta_2 - \theta_1$.

- The correlation between Y_1 and Y_2 is

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2},$$

so we can answer the third question by looking at the posterior distribution of ρ, which we have once we have posterior samples of Σ.

Reading example: prior on mean

- Clearly, we first need to set the hyperparameters $\boldsymbol{\mu}_0$ and Λ_0 in $\pi(\boldsymbol{\theta}) = \mathcal{N}_2(\boldsymbol{\mu}_0, \Lambda_0)$, based on prior belief.

- For this example, both tests were actually designed *apriori* to have a mean of 50, so, we can set $\boldsymbol{\mu}_0 = (\mu_{0(1)}, \mu_{0(2)})^T = (50, 50)^T$.

- $\boldsymbol{\mu}_0 = (0, 0)^T$ is also often a common choice when there is no prior guess, especially when there is enough data to "drown out" the prior guess.

- Next, we need to set values for elements of

$$\Lambda_0 = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

- It is quite reasonable to believe *apriori* that the true means will most likely lie in the interval $[25, 75]$ with high probability (perhaps 0.95?), since individual test scores should lie in the interval $[0, 100]$.

- Recall that for any normal distribution, 95% of the density will lie within two standard deviations of the mean.

Reading example: prior on mean

- Therefore, we can set

$$\mu_{0(1)} \pm 2\sqrt{\lambda_{11}} = (25,75) \quad \Rightarrow \quad 50 \pm 2\sqrt{\lambda_{11}} = (25,75)$$
$$\Rightarrow \quad 2\sqrt{\lambda_{11}} = 25 \quad \Rightarrow \quad \lambda_{11} = \left(\frac{25}{2}\right)^2 \approx 156.$$

- Similarly, set $\lambda_{22} \approx 156$.

- Finally, we expect some correlation between $\mu_{0(1)}$ and $\mu_{0(2)}$, but suppose we don't know exactly how strong. We can set the prior correlation to 0.5.

$$\Rightarrow 0.5 = \frac{\lambda_{12}}{\sqrt{\lambda_{11}}\sqrt{\lambda_{22}}} = \frac{\lambda_{12}}{156} \quad \Rightarrow \quad \lambda_{12} = 156 \times 0.5 = 78.$$

- Thus,

$$\pi(\boldsymbol{\theta}) = \mathcal{N}_2\left(\boldsymbol{\mu}_0 = \begin{pmatrix} 50 \\ 50 \end{pmatrix}, \Lambda_0 = \begin{pmatrix} 156 & 78 \\ 78 & 156 \end{pmatrix}\right).$$

Reading example: prior on covariance

- Next we need to set the hyperparameters ν_0 and $\boldsymbol{S}_0$ in $\pi(\Sigma) = \mathcal{IW}_2(\nu_0, \boldsymbol{S}_0)$, based on prior belief.
- First, let's start with a prior guess Σ_0 for Σ.
- Again, since individual test scores should lie in the interval $[0, 100]$, we should set Σ_0 so that values outside $[0, 100]$ are highly unlikely.
- Just as we did with Λ_0, we can use that idea to set the elements of Σ_0

$$\Sigma_0 = \begin{pmatrix} \sigma_{11}^{(0)} & \sigma_{12}^{(0)} \\ \sigma_{21}^{(0)} & \sigma_{22}^{(0)} \end{pmatrix}$$

- The identity matrix is also often a common choice for Σ_0 when there is no prior guess, especially when there is enough data to "drown out" the prior guess.

Reading example: prior on covariance

- Therefore, we can set

$$\mu_{0(1)} \pm 2\sqrt{\sigma_{11}^{(0)}} = (0, 100) \quad \Rightarrow \quad 50 \pm 2\sqrt{\sigma_{11}^{(0)}} = (0, 100)$$
$$\Rightarrow \quad 2\sqrt{\sigma_{11}^{(0)}} = 50 \quad \Rightarrow \quad \sigma_{11}^{(0)} = \left(\frac{50}{2}\right)^2 \approx 625.$$

- Similarly, set $\sigma_{22}^{(0)} \approx 625.$

- Again, we expect some correlation between Y_1 and Y_2, but suppose we don't know exactly how strong. We can set the prior correlation to 0.5.

$$\Rightarrow 0.5 = \frac{\sigma_{12}^{(0)}}{\sqrt{\sigma_{11}^{(0)}}\sqrt{\sigma_{22}^{(0)}}} = \frac{\sigma_{12}^{(0)}}{625} \quad \Rightarrow \quad \sigma_{12}^{(0)} = 625 \times 0.5 = 312.5.$$

- Thus,

$$\Sigma_0 = \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix}$$

Reading example: prior on covariance

- Recall that if we are not at all confident on a prior value for Σ, but we have a prior guess Σ_0, we can set
  - $\nu_0 = p + 2$, so that the $\mathbb{E}[\Sigma] = \frac{1}{\nu_0 - p - 1}\boldsymbol{S}_0$ is finite.
  - $\boldsymbol{S}_0 = \Sigma_0$

  so that Σ is only loosely centered around Σ_0.

- Thus, we can set
  - $\nu_0 = p + 2 = 2 + 2 = 4$
  - $\boldsymbol{S}_0 = \Sigma_0$

  so that we have

$$\pi(\Sigma) = \mathcal{IW}_2\left(\nu_0 = 4, \Sigma_0 = \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix}\right).$$

Reading example: data

Now, to the data (finally!)

```
Y <- as.matrix(dget("http://www2.stat.duke.edu/~pdh10/FCBS/Inline/Y.reading"))
dim(Y)
```

```
## [1] 22  2
```

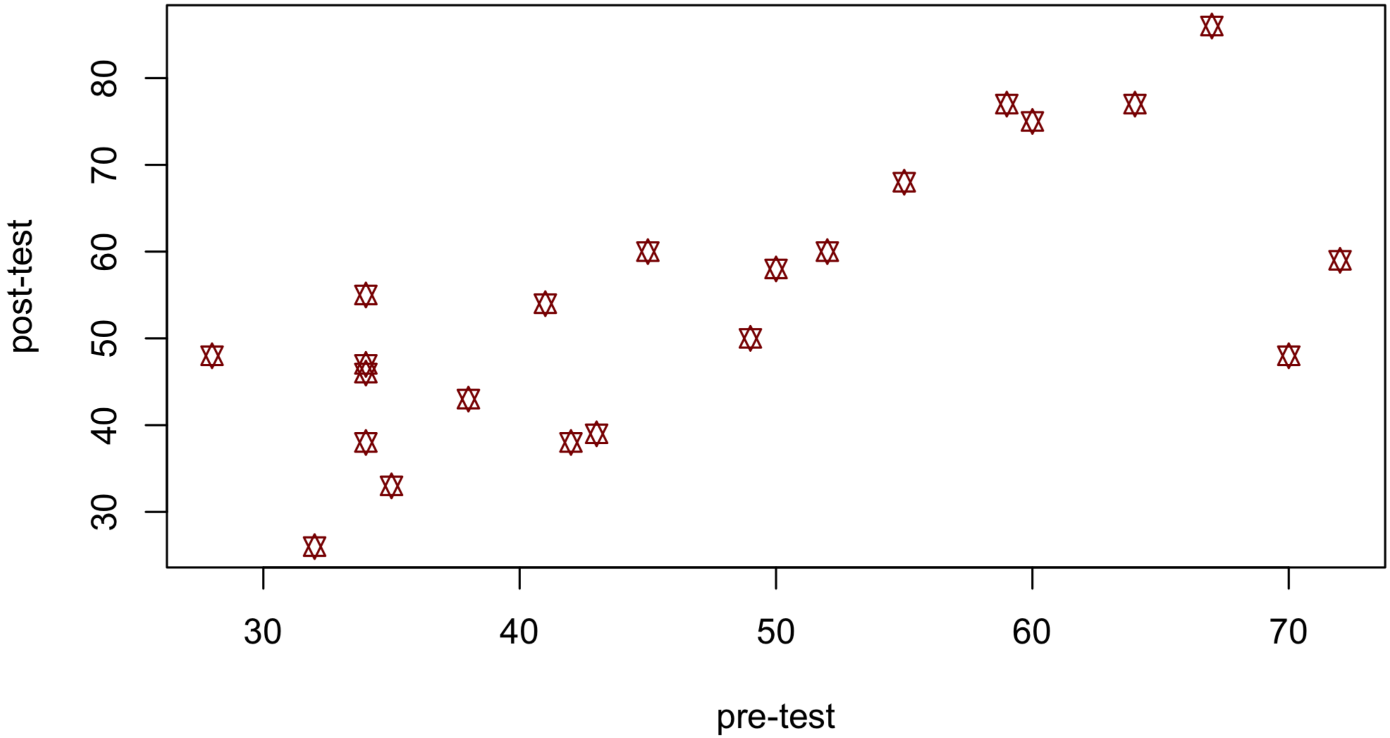
```
head(Y)
```

```
##      pretest posttest
## [1,]      59       77
## [2,]      43       39
## [3,]      34       46
## [4,]      32       26
## [5,]      42       38
## [6,]      38       43
```

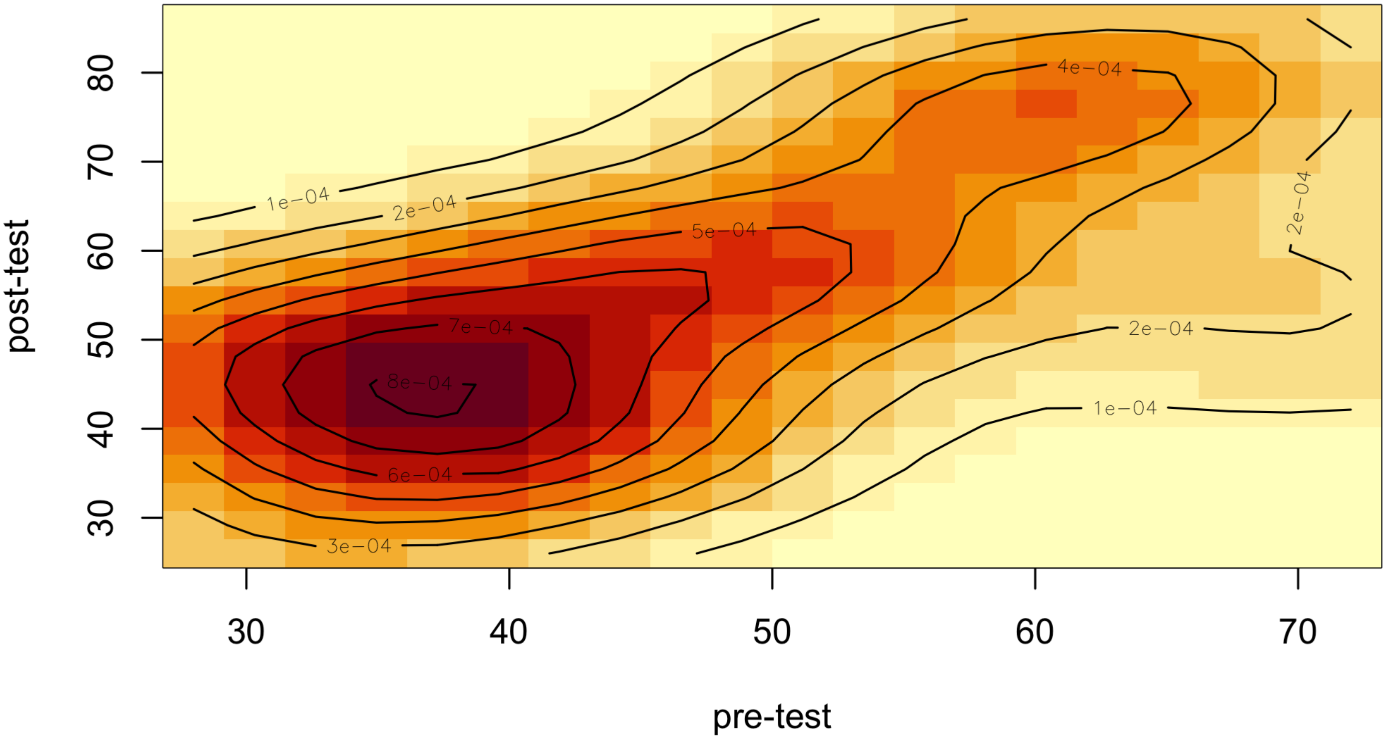
```
summary(Y)
```

```
##     pretest         posttest
##  Min.   :28.00   Min.   :26.00
##  1st Qu.:34.25   1st Qu.:43.75
##  Median :44.00   Median :52.00
##  Mean   :47.18   Mean   :53.86
##  3rd Qu.:58.00   3rd Qu.:60.00
##  Max.   :72.00   Max.   :86.00
```

Reading example: data



Reading example: data



This is just some EDA. We will write the Gibbs sampler and answer the questions of interest in the next module.

What's next?

Move on to the readings for the next module!