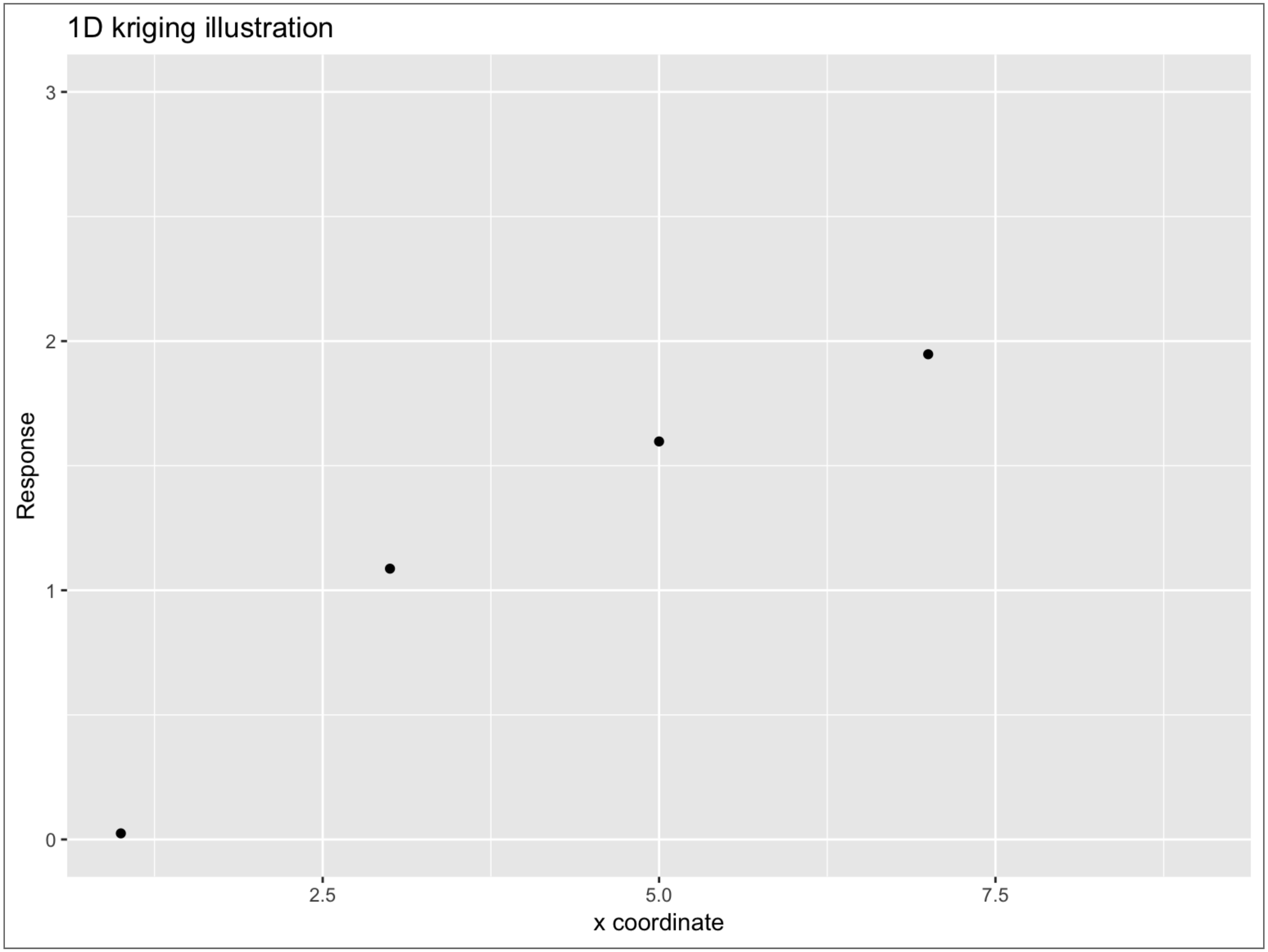
Lecture 9: Point Level Models - Model Fitting, cont..

Class Intro

Intro Questions

- Describe the process for finding the BLUP in this situation:


1D kriging illustration
Response
3
2
1
0
2.5
5.0
7.5
x coordinate

Intro Questions

- For Today:
  - More Model Fitting

Model Fitting

BLUP

- It turns out that the solution for the vector $\boldsymbol{l}$ is
  $$\boldsymbol{l} = \Gamma^{-1}\left(\boldsymbol{\gamma_0} + \frac{(1 - \mathbf{1}^T\Gamma^{-1}\boldsymbol{\gamma_0})}{\mathbf{1}^T\Gamma^{-1}\mathbf{1}}\mathbf{1}\right),$$
  where Γ is an $n \times n$ matrix with entries $\Gamma_{ij} = \gamma_{ij}$ and $\boldsymbol{\gamma_0}$ is the vector of γ_{0i} values.
- Then the Best Linear Unbiased Predictor is $\boldsymbol{l}^T\boldsymbol{Y}$
- This BLUP also requires an estimate of $\gamma(\boldsymbol{h})$

Kriging Solution

```
# Create Gamma Matrix
x <- krige.dat$x
y <- krige.dat$y
D <- dist(x, upper=T, diag=T) %>% as.matrix()
Gamma = 1 - exp(-D/3)

# Create gamma_0 for both s1* and s2*
d1 <- sqrt((4 - x)^2)
d2 <- sqrt((6 - x)^2)
gamma.01 <- 1 - exp(-d1/3)
gamma.02 <- 1 - exp(-d2/3)
```

Kriging Solution

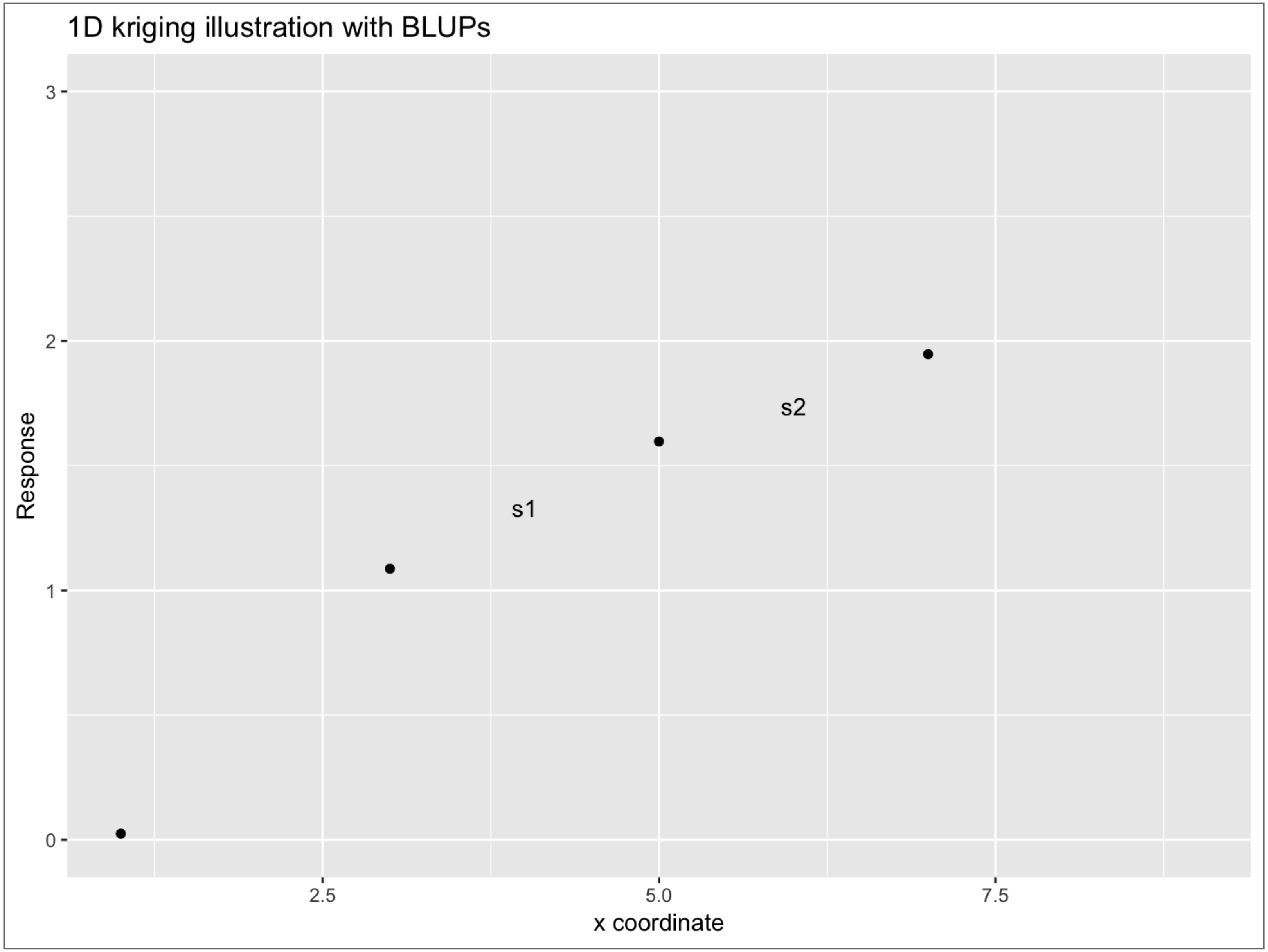
```
# Compute l

l.1 <- solve(Gamma) %*%
  (gamma.01 + c(1 - rep(1,4) %*% solve(Gamma) %*% gamma.01) /
     c(rep(1,4) %*% solve(Gamma) %*% rep(1,4)))
y1.pred <- t(l.1) %*% y

l.2 <- solve(Gamma) %*%
  (gamma.02 + c(1 - rep(1,4) %*% solve(Gamma) %*% gamma.02) /
     c(rep(1,4) %*% solve(Gamma) %*% rep(1,4)))
y2.pred <- t(l.2) %*% y

krige.dat %>% ggplot(aes(x=x, y=y)) + geom_point() +
  xlim(1,9) + ylim(0,3) + ylab('Response') + xlab('x coordinate') +
  ggtitle('1D kriging illustration with BLUPs') +
  annotate('text',x=4,y=y1.pred, label = "s1") +
  annotate('text',x=6,y=y2.pred, label = "s2")
```

Kriging Solution



Kriging with Gaussian Processes

A Gaussian Process

- The BLUP does not contain a distributional assumptions, but rather comes from an optimization framework.
- Now assume that
  $\boldsymbol{Y} = \mu\mathbf{1} + \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma)$
- With no nugget, let $\Sigma = \sigma^2 H(\phi)$, where $(H(\phi))_{ij} = \rho(\phi; d_{ij})$, where d_{ij} is the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$.
- A nugget can be included by modifying Σ to be $\Sigma = \sigma^2 H(\phi) + \tau^2 I$

Minimizing Mean-Square Prediction Error

- **Goal:** find $h(\boldsymbol{y})$ that minimizes $E[(\boldsymbol{Y}(\boldsymbol{s_0}) - h(\boldsymbol{y}))^2|\boldsymbol{y}]$
- $E[(\boldsymbol{Y}(\boldsymbol{s_0}) - h(\boldsymbol{y}))^2|\boldsymbol{y}] = E[(\boldsymbol{Y}(\boldsymbol{s_0}) - h(\boldsymbol{y}) \pm E[(\boldsymbol{Y}(\boldsymbol{s_0}|\boldsymbol{y})])^2|\boldsymbol{y}]$
- $= E\{(\boldsymbol{Y}(\boldsymbol{s_0}) - E[(\boldsymbol{Y}(\boldsymbol{s_0})|\boldsymbol{y}])^2|\boldsymbol{y}\} + \{E[(\boldsymbol{Y}(\boldsymbol{s_0})|y] - h(\boldsymbol{y})\}^2$

Minimizing Mean-Square Prediction Error: Part 2

- As $\{E[(\boldsymbol{Y}(\boldsymbol{s_0})|y] - h(\boldsymbol{y})\}^2 \geq 0$
- we have $E[(\boldsymbol{Y}(\boldsymbol{s_0}) - h(\boldsymbol{y}))^2|\boldsymbol{y}] \geq E\{(\boldsymbol{Y}(\boldsymbol{s_0}) - E[(\boldsymbol{Y}(\boldsymbol{s_0})|\boldsymbol{y}])^2|\boldsymbol{y}\}$
- Hence to minimize $E[(\boldsymbol{Y}(\boldsymbol{s_0}) - h(\boldsymbol{y}))^2|\boldsymbol{y}]$, we set ...
- $h(\boldsymbol{y}) = E[(\boldsymbol{Y}(\boldsymbol{s_0})|\boldsymbol{y}]$
- Hence, $h(\boldsymbol{y})$ that minimizes the error is the conditional expectation of $\boldsymbol{Y}(\boldsymbol{s_0})$
- Note this is also the *posterior mean* of $\boldsymbol{Y}(\boldsymbol{s_0})$

Multivariate Normal Theory

- For consider partioning a multivariate normal distribution into two parts
$$\begin{pmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \end{pmatrix} = N\left(\begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right),$$
where $\Omega_{12} = \Omega_{21}^T$

Conditional Multivariate Normal Theory

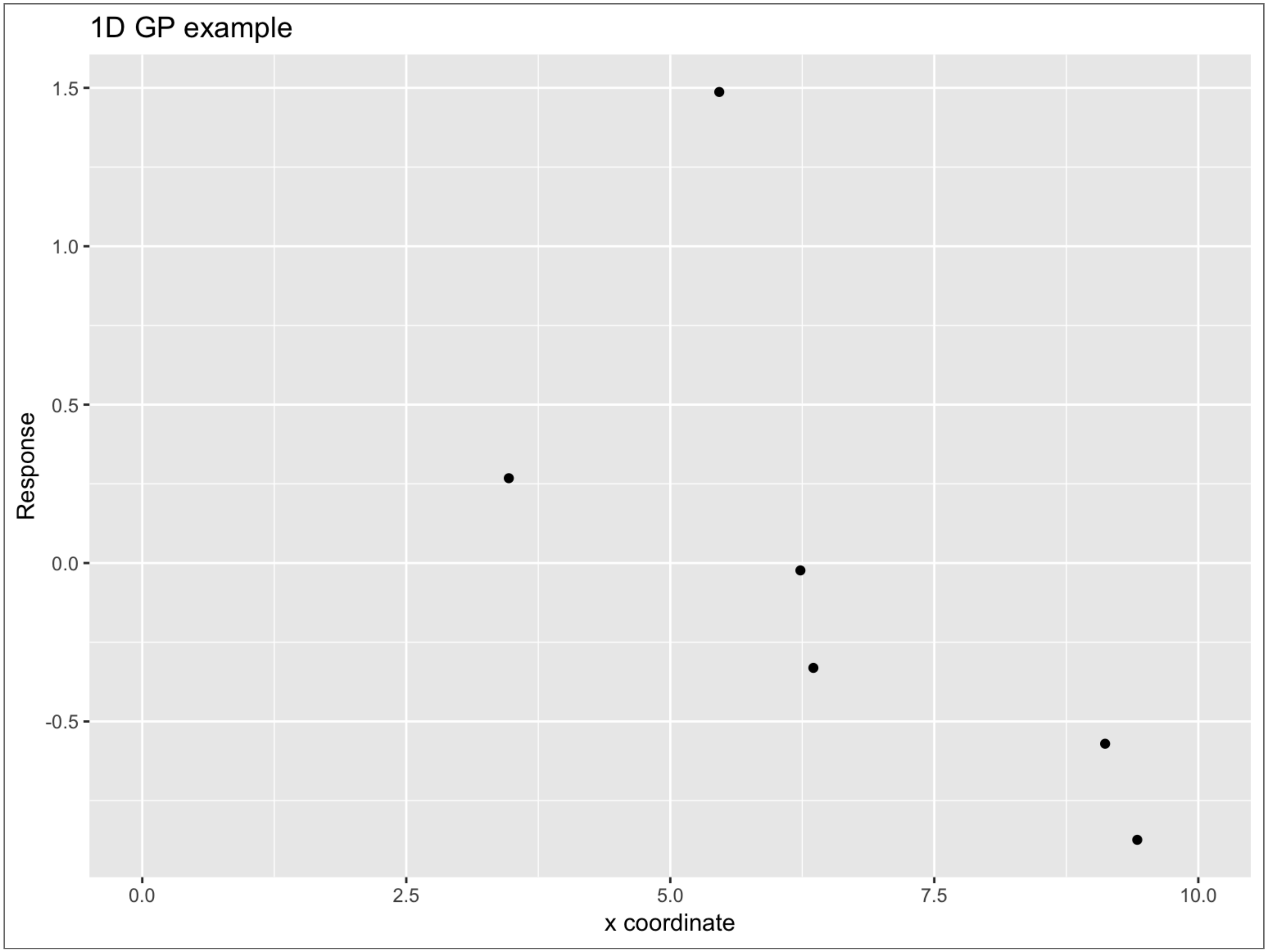
- The conditional distribution, $p(\boldsymbol{Y_1}|\boldsymbol{Y_2})$ is normal with:
- $E[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \boldsymbol{\mu_1} + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{Y_2} - \mu_2)$
- $Var[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$
- Thus with $\boldsymbol{Y_1} = Y(\boldsymbol{s_0})$ and $\boldsymbol{Y_2} = \boldsymbol{y}$
  $\Omega_{11} = \sigma^2 + \tau^2, \quad \Omega_{12} = (\sigma^2 p(\phi; d_{01})), \ldots, p(\phi; d_{0n})))$
  $\Omega_{22} = \sigma^2 H(\phi) + \tau^2 I$

Gaussian Process Exercise:

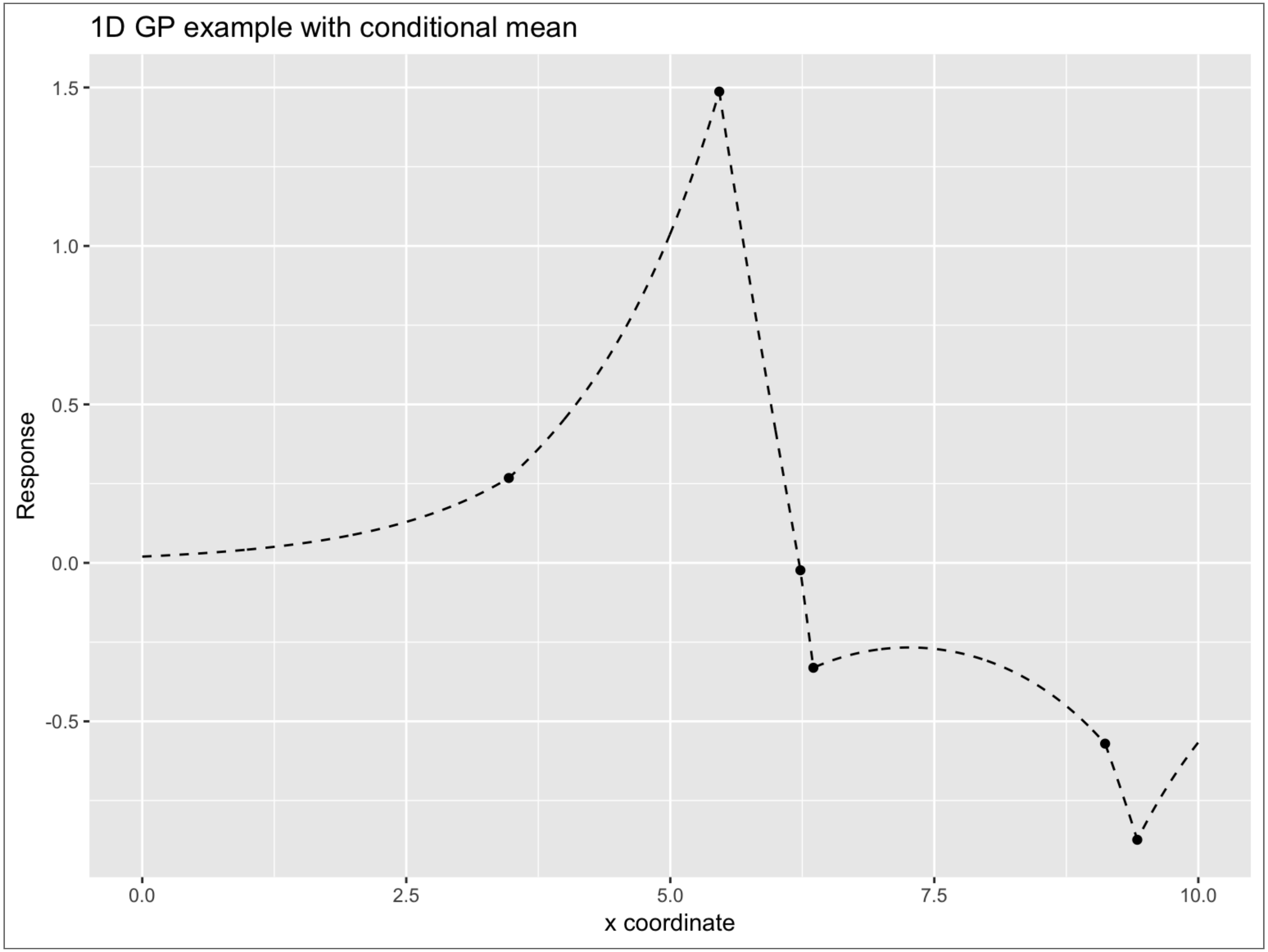
Overview

- Similar to the previous exercise, we will simulate data from a 1D process and make predictions at unobserved locations - which is the interval [0,10].
- In this situation, please plot the mean of the distribution as well as some uncertainty metric.
- You do not need to estimate σ^2, τ^2, and ϕ but can use the known values in the R code.

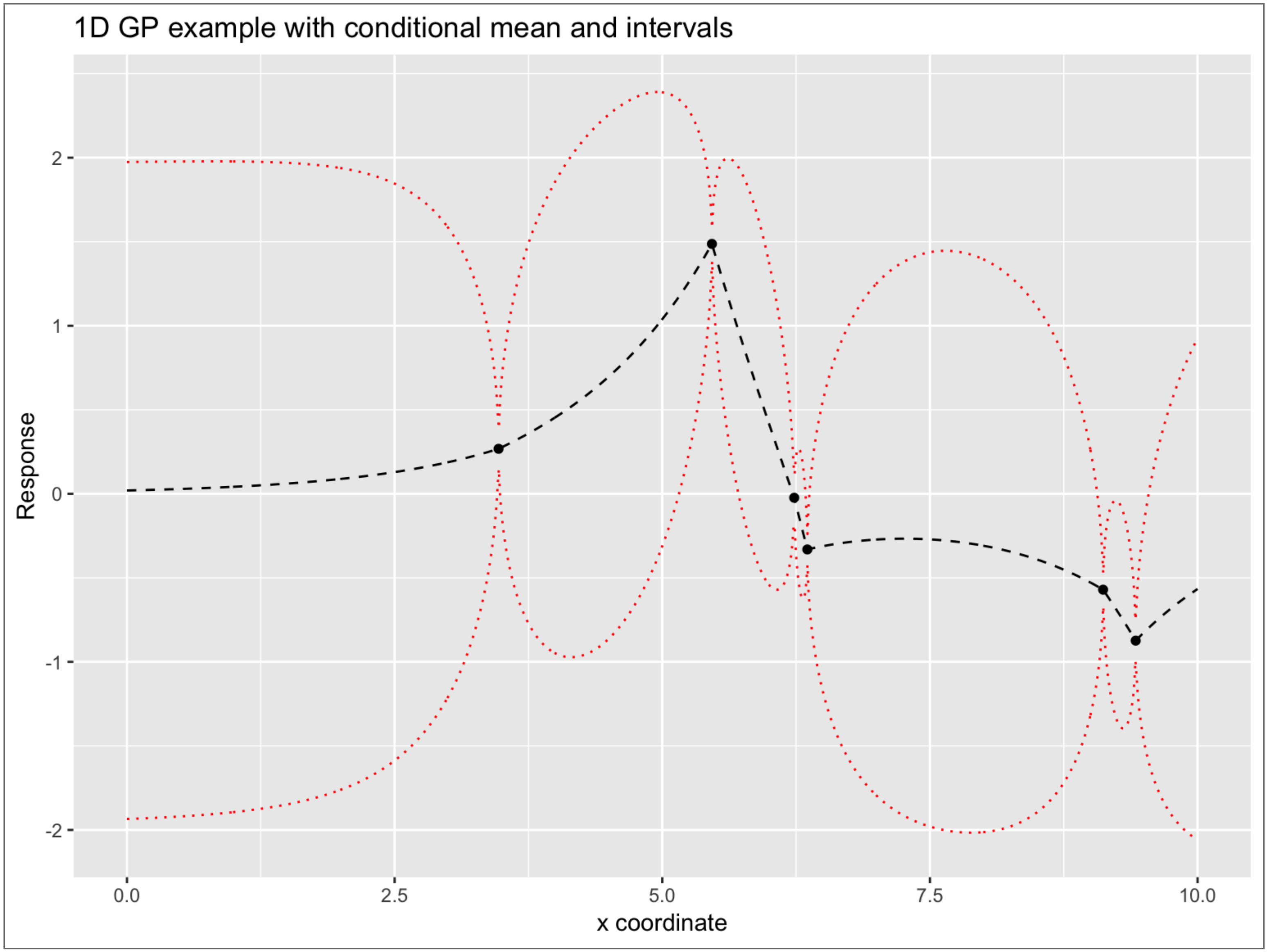
Data Overview



Conditional Expectation



Conditional Expectation and Intervals



Follow up Questions

- Consider the conditional expectation, $E[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \boldsymbol{\mu_1} + \Omega_{12}\Omega_{22}^{-1}(\boldsymbol{Y_2} - \mu_2)$ How is this resultant expectation impacted by $\Omega_{12}\Omega_{22}^{-1}$? How is this resultant expectation impacted by $(\boldsymbol{Y_2} - \mu_2)$?
- Similarly, how does the conditional variance $Var[\boldsymbol{Y_1}|\boldsymbol{Y_2}] = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ change as a function of $\Omega_{12}\Omega_{22}^{-1}\Omega_{21}$?
- If we add a nugget to the previous example, how do the predictions change?
- How does the scenario change if μ_1 and μ_2 are not zero, but rather say $\mu_1 = X_1\beta$ and $\mu_2 = X_2\beta$?

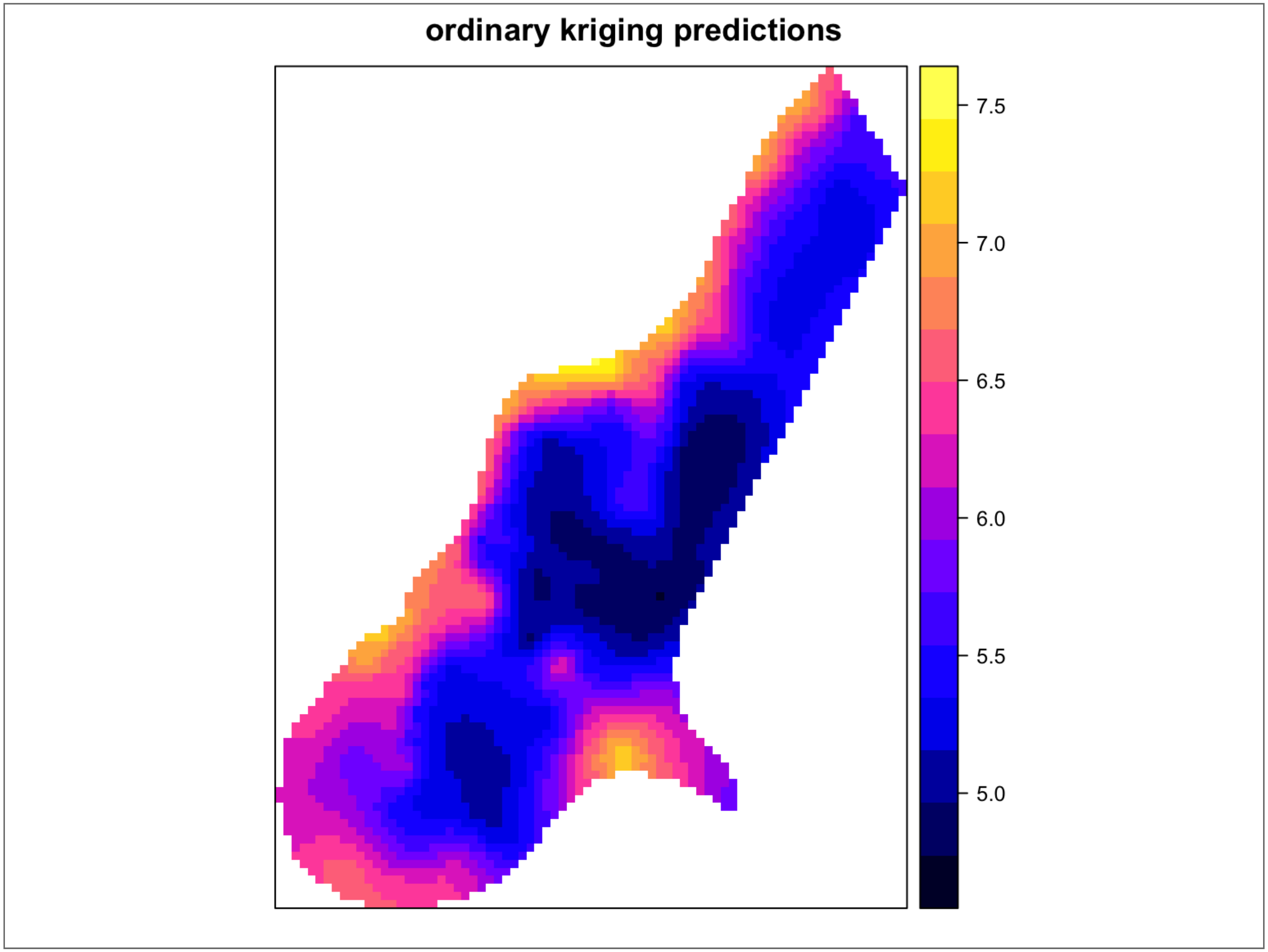
2D Kriging

Discuss how the spatial range could be extended to $\mathcal{R}^2$ rather than $\mathcal{R}^1$. What changes in this situation?

Functions in R for 2D kriging

- The `krige` function in `gstat` contains a function for kriging ; however, this requires a known variogram.

```
## [using ordinary kriging]
```


ordinary kriging predictions
7.5
7.0
6.5
6.0
5.5
5.0

Universal Kriging

- When covariate information is available for inclusion in the analysis, this is often referred to as *universal kriging*
- Now we have
  $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma)$
- The conditional distributions are very similar to what we have derived above, watch for HW question.
- In each case, kriging or universal kriging, it is still necessary to estimate the following parameters: σ^2, τ^2, ϕ, and μ or β.
- This can be done with least-squares methods or in a Bayesian framework.

About those other parameters

We still need

- to choose an appropriate covariance function (or semivariogram)
- and estimate parameters in that function