Spatial Statistics Notation

Given that we assume

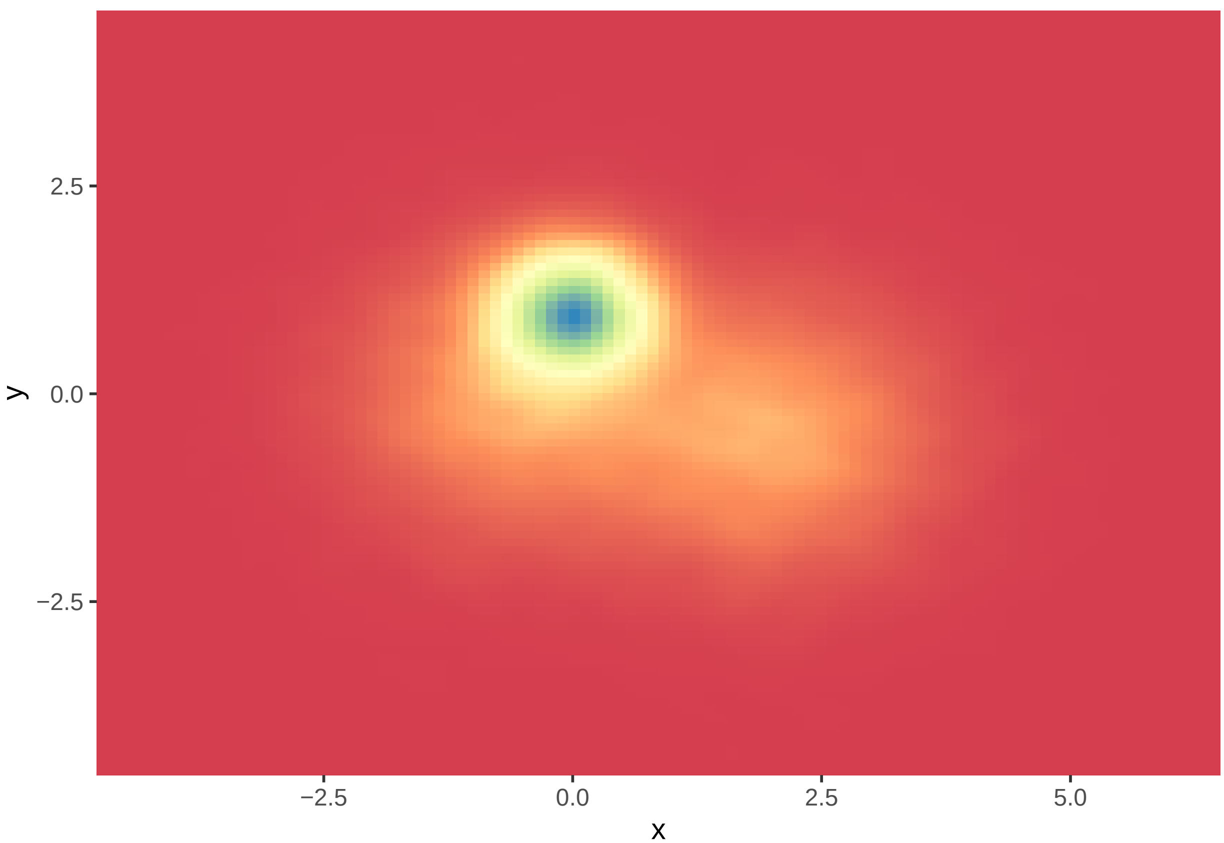
$$\boldsymbol{Y}|\mu, \underline{\theta} \sim MVN_n(\underline{\mu}, \Sigma(\underline{\theta})),$$

Stationarity

Assume the spatial process, $\boldsymbol{Y}(\boldsymbol{s})$ has a mean, $\underline{\mu}(\boldsymbol{s})$, and that the variance of $\boldsymbol{Y}(\boldsymbol{s})$ exists everywhere.

Strict Stationarity

Is this strictly stationary?



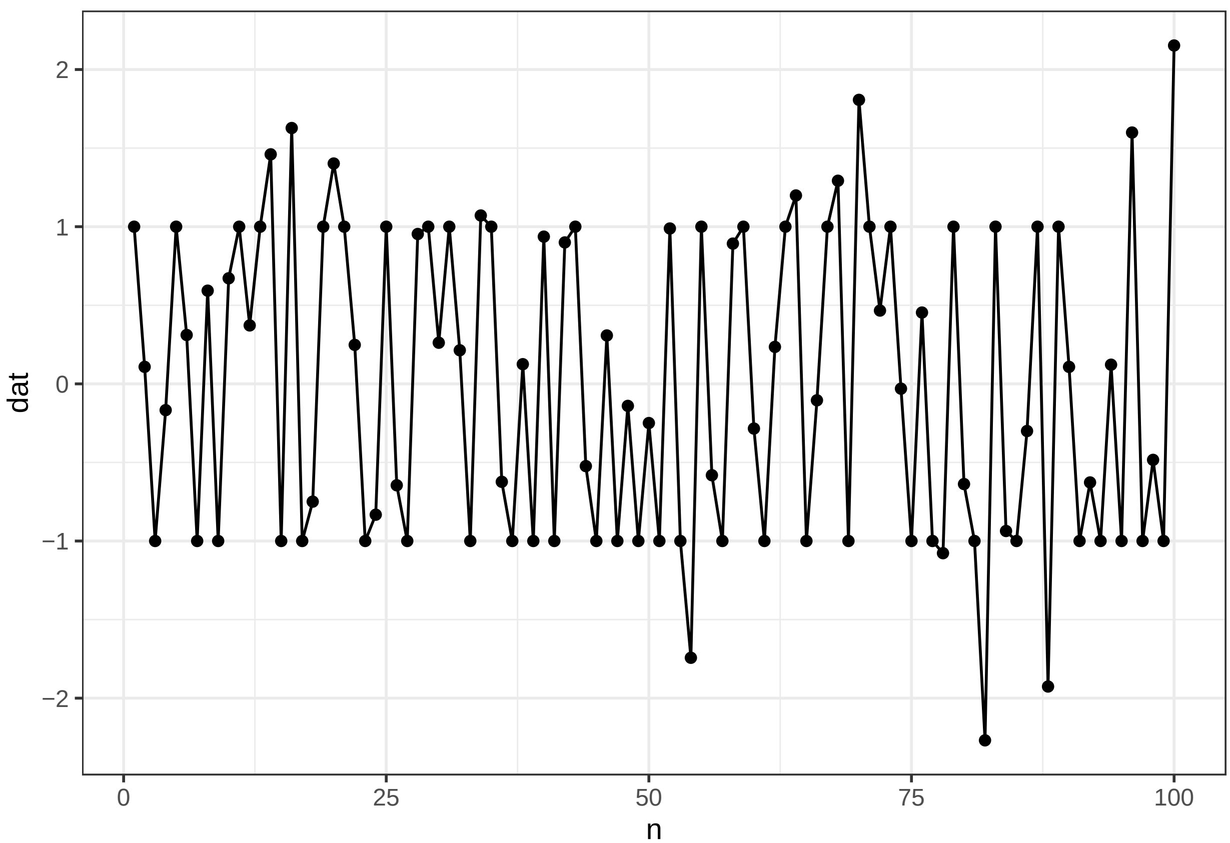
Weak Stationarity

Weak stationarity, or second-order stationarity of a spatial process, requires a constant mean and covariance that is a function of $\boldsymbol{h}$, $Cov(\boldsymbol{Y}(\boldsymbol{s}), \boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h})) = C(\boldsymbol{h})$,

Typically with spatial models, the process is assumed to be mean zero as covariates explain the mean structure.

Weak Stationarity

Is this weakly stationary?



Intrinsic Stationarity

A third kind of stationarity, known as intrinsic stationarity, describes the behavior of differences in the spatial process, rather than the data.

Intrinsic stationarity assumes $E[\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})] = 0$ then define

$$E[\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})]^2 = Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})) = 2\gamma(\boldsymbol{h}),$$

which only works, and satisfies intrinsic stationarity, if the equation only depends on $\boldsymbol{h}$.

Variograms

Intrinsic stationary justifies the use of variograms $2\gamma(\boldsymbol{h})$

Variograms are often used to visualize spatial patterns:

There is a mathematical link between the covariance function $C(\boldsymbol{h})$ and the variogram $2\gamma(\boldsymbol{h})$.

Isotropy

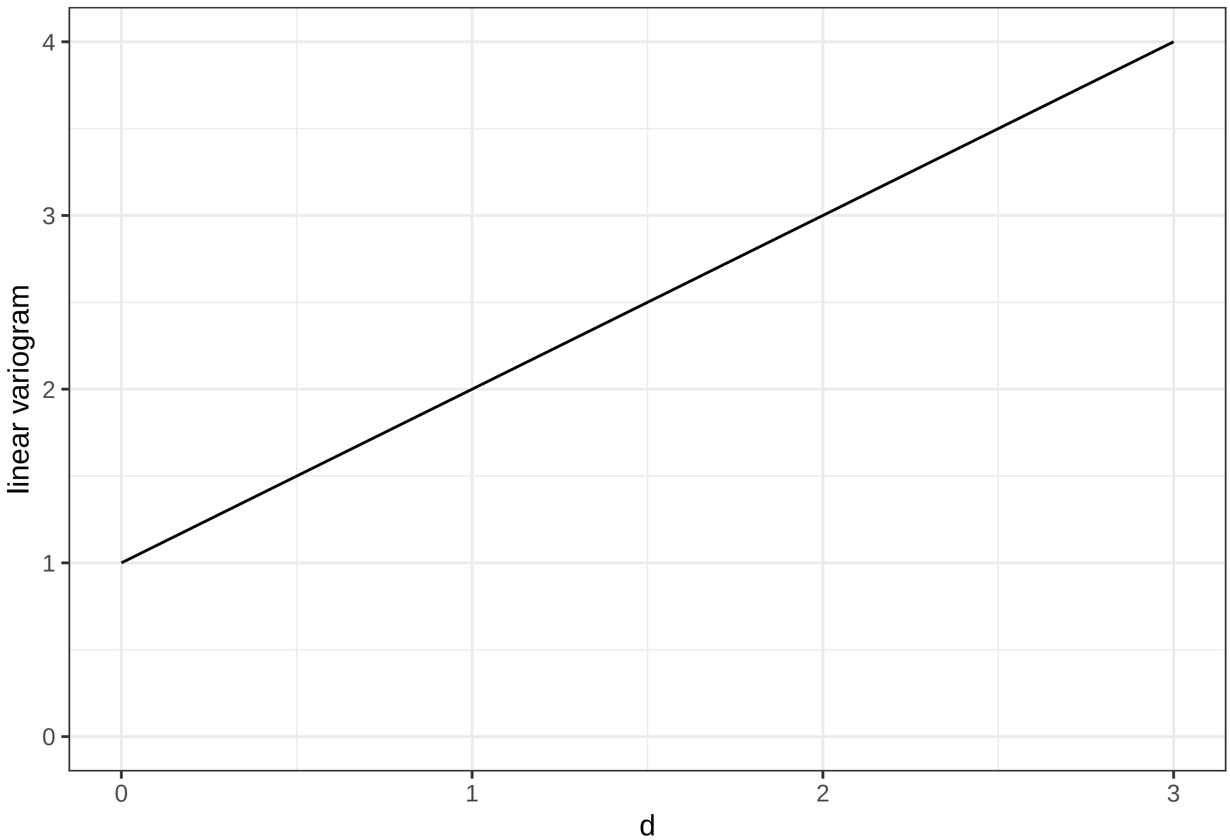
Variogram to the Covariance Function

$$\begin{aligned}
2\gamma(\boldsymbol{h}) &= Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}) - \boldsymbol{Y}(\boldsymbol{s})) \\
&= Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h})) + Var(\boldsymbol{Y}(\boldsymbol{s})) - 2Cov(Var(\boldsymbol{Y}(\boldsymbol{s}+\boldsymbol{h}), \boldsymbol{Y}(\boldsymbol{s}))) \\
&= C(\boldsymbol{0}) + C(\boldsymbol{0}) - 2C(\boldsymbol{h}) \\
&= 2\left[C(\boldsymbol{0}) - C(\boldsymbol{h})\right]
\end{aligned}$$

Given $C()$, the variogram can easily be recovered

Linear semivariogram

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}$$



Nugget, sill, partial-sill, and range

Spherical semivariogram:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \ \text{ if } d \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3\phi d}{2} - \frac{1}{2}(\phi d)^3\right] \\ 0 \ \text{ otherwise} \end{cases}$$

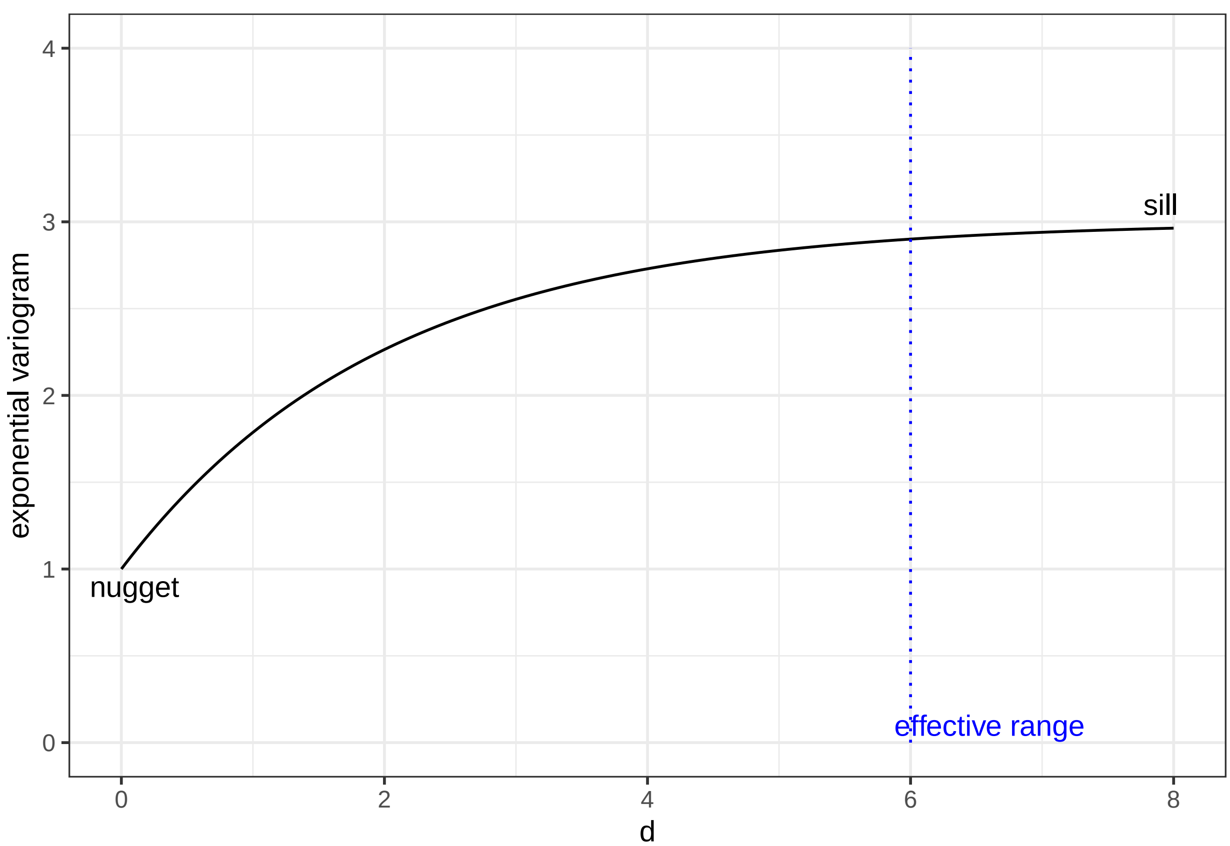
- Sketch, or generate in R, a spherical semivariogram
- On this figure label the nugget, sill, partial sill, and range.

Exponential

We saw the exponential covariance earlier in class, what is the mathematical form of the covariance?

The variogram is

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-d/\phi)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}$$



More Semivariograms: Equations

Gaussian:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 d^2)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}$$

Powered Exponential:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-|\phi d|^p)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}$$

Matérn:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu} d\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu} d\phi)\right] & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} ,$$

where K_ν is a modified Bessel function and $\Gamma()$ is a Gamma function.

Variogram Creation: How?

```
data(meuse)
meuse.small <- meuse %>% select(x, y, copper) %>% as_tibble()
meuse.small %>% head(15) %>% kable()
```

x	y	copper
181072	333611	85
181025	333558	81
181165	333537	68
181298	333484	81
181307	333330	48
181390	333260	61
181165	333370	31
181027	333363	29
181060	333231	37
181232	333168	24
181191	333115	25
181032	333031	25
180874	333339	93
180969	333252	31
181011	333161	27

Variogram Creation: Steps

1. Calculate distances between sampling locations
2. Choose grid for distance calculations
3. Calculate empirical semivariogram

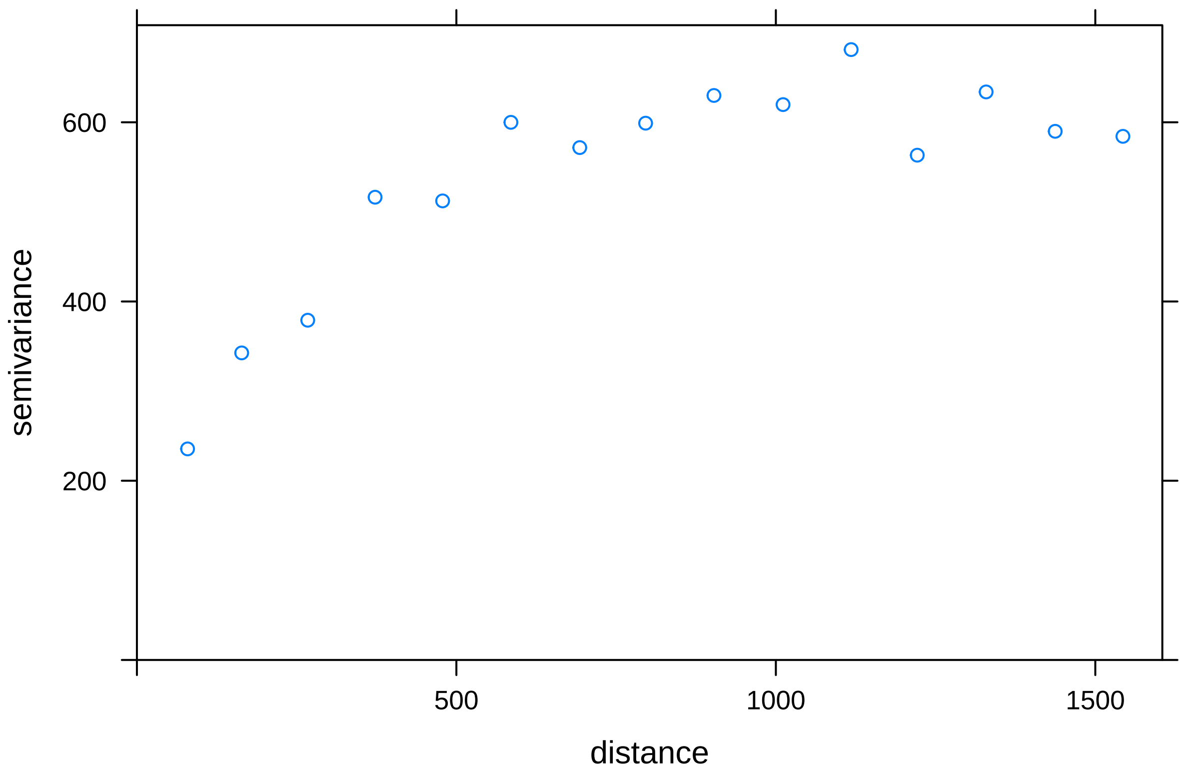
$$\hat{\gamma}(d_k) = \frac{1}{2N(d_k)} \sum_{\boldsymbol{s}_i, \boldsymbol{s}_i, \in N(d_k)} \left[\boldsymbol{Y}(\boldsymbol{s}_i) - \boldsymbol{Y}(\boldsymbol{s}_j)\right]^2,$$

   where

$$N(d_k) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : ||\boldsymbol{s}_i - \boldsymbol{s}_j|| \in I_k\}$$

   and I_k is the k^{th} interval.
4. Plot the semivariogram

Variogram Creation: R function



Variogram Creation: How - Step 1

Calculate Distances between sampling locations

```
#dist(meuse.small)
dist.mat <- dist(meuse.small %>% select(x,y))
```

Variogram Creation: How - Step 2

Choose grid for distance calculations

```
cutoff <- max(dist.mat) / 3 # default maximum distance
num.bins <-  15
bin.width <- cutoff / 15
```

Variogram Creation: How - Step 3

Calculate empirical semivariogram

```
dist.sq <- dist(meuse.small$copper)^2

vario.dat <- data.frame(dist = as.numeric(dist.mat), diff = as.numeric(dist.sq)) %>%
  mutate(bin = floor(dist / bin.width) + 1) %>% filter(bin < 16) %>%
  group_by(bin) %>% summarize(emp.sv = .5 * mean(diff))
vario.dat
```

```
## # A tibble: 15 x 2
##      bin emp.sv
##    <dbl>  <dbl>
##  1     1   236.
##  2     2   347.
##  3     3   348.
##  4     4   488.
##  5     5   499.
##  6     6   577.
##  7     7   553.
##  8     8   623.
##  9     9   600.
## 10    10   665.
## 11    11   603.
## 12    12   673.
## 13    13   557.
## 14    14   643.
## 15    15   574.
```

Variogram Creation: How - Step 4

Plot empirical semivariogram