Stat 534 Homework 7

Kenny Flagg

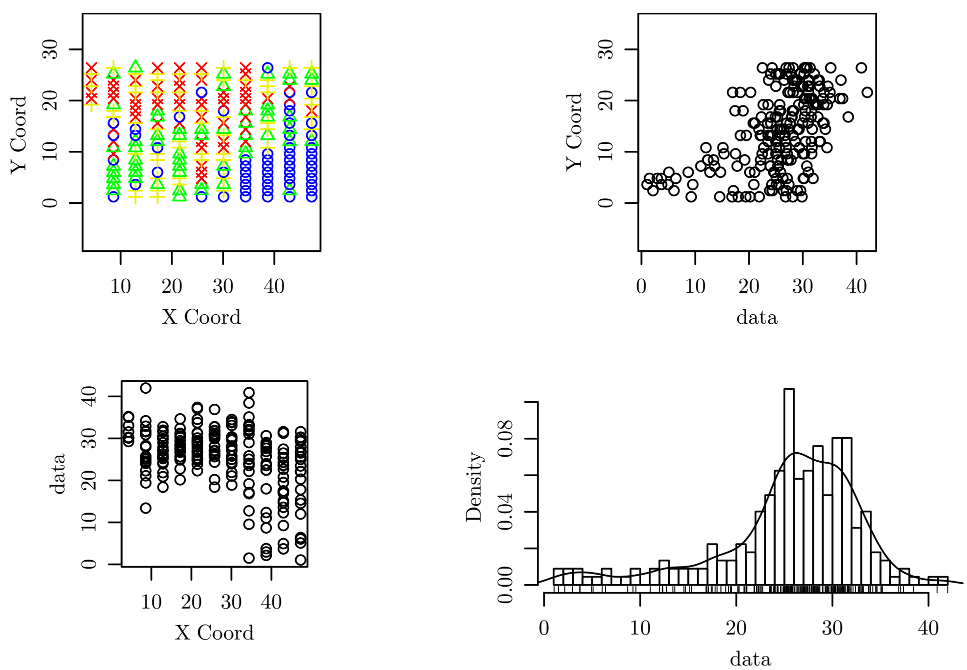
March 6, 2017

1. *A data set (`wheat.txt`) has been sent to you. The data set contains yields of wheat recorded at spatial coordinates. Note that the header is x, y, and z with z being the yields. We will need a couple of different data object types. Pay attention to the R code below. Do not worry about anisotropy.*

```
wheat.dat <- read.table('wheat.txt', header = TRUE)
wheat.geodat <- as.geodata(wheat.dat, coords.col = 1:2, data.col = 3)
wheat.grid <- expand.grid(seq(0, 50, l = 25), seq(0, 30, l = 25))
```

(a) *Plot the data and comment on the results.*

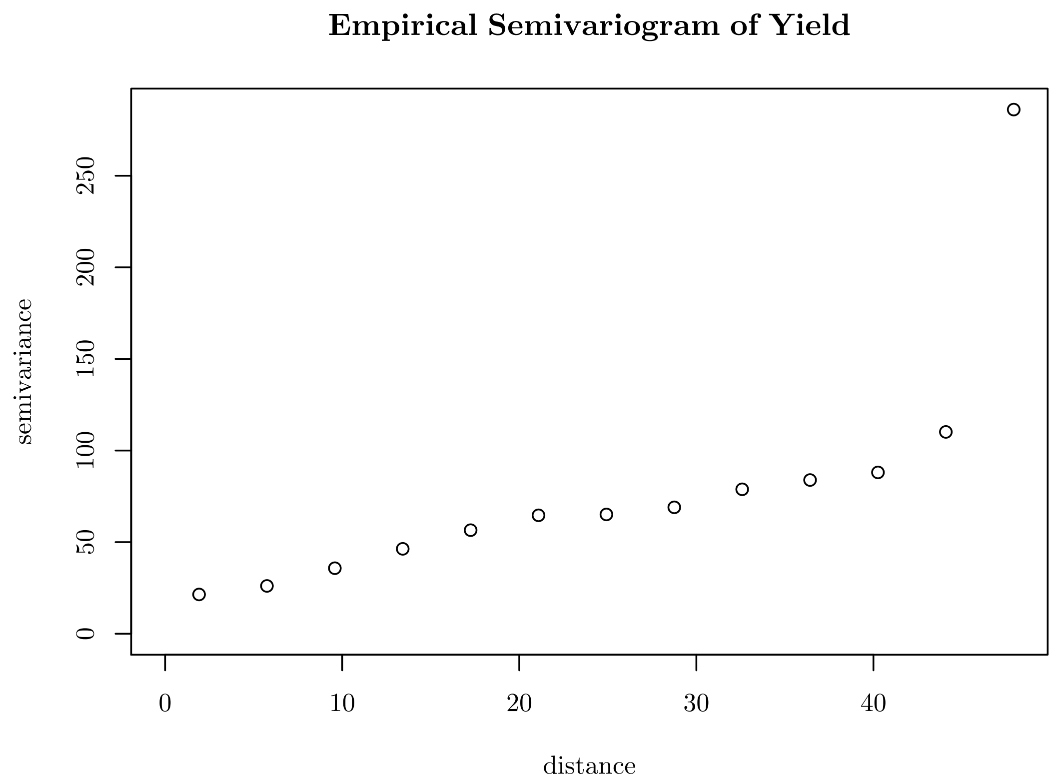
```
plot(wheat.geodat, breaks = 30)
```



There is a trend in wheat yield, with the lowest yields occuring in the bottom left of the region and the highest yields in the top right.

(b) *Produce a plot of the empirical semivariogram of the wheat yields. Can this plot be trusted for estimation of semivariogram parameters to be used in kriging. Why or why not?*

```
yield.semivar <- variog(wheat.geodat)
plot(yield.semivar, main = 'Empirical Semivariogram of Yield')
```



No, this plot cannot be trusted because the process is non-stationary so the empirical semivariogram is biased.

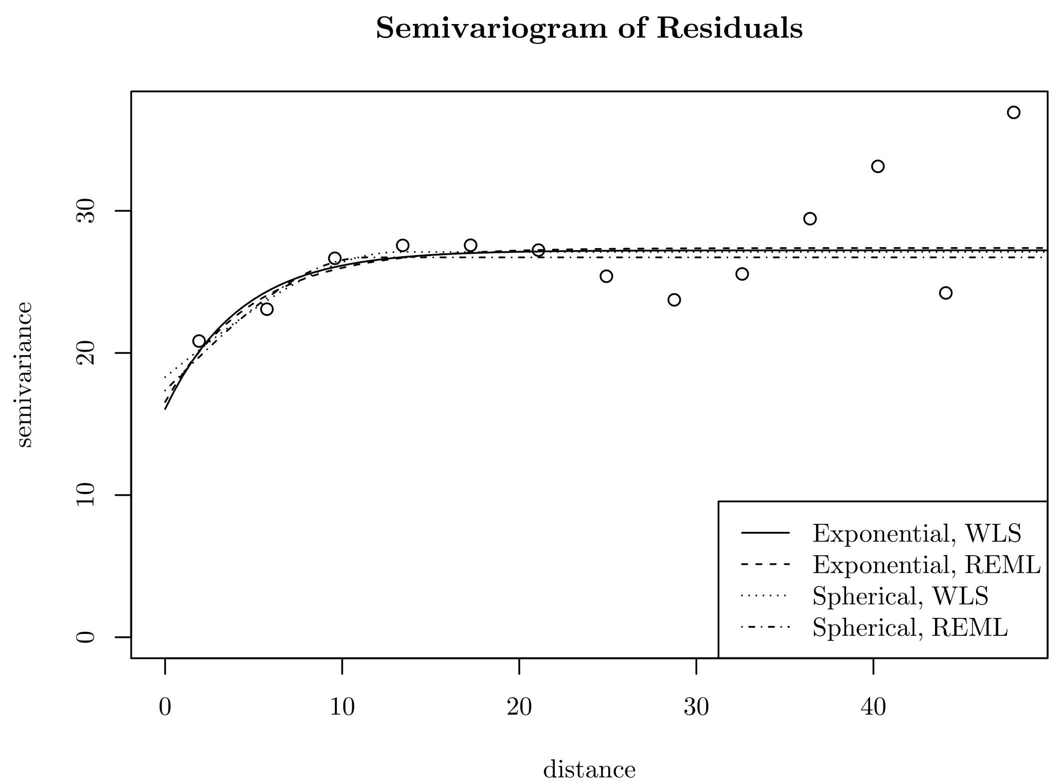
(c) *We will use the* `surf.ls` *function in the spatial library to fit a quadratic trend model to the yields by ordinary least squares and plot the empirical semivariogram of the residuals.*

```
library(spatial, quietly = TRUE)
wheat.ls <- surf.ls(2, wheat.dat) # fits a second order polynomial trend surface
resid.dat <- cbind(wheat.dat$x, wheat.dat$y, residuals(wheat.ls))
resid.geodat <- as.geodata(resid.dat, coords.col = 1:2, data.col = 3)
```

Fit an appropriate semivariogram model to the semivariogram using your method of choice. Justify your final selection.

Based on the empirical semivariogram (plot on next page) I chose initial values of 17 for the nugget, 10 for the (effective) range, and 25 for the total sill (implying a partial sill of 8). I fit exponential and spherical models by Cressie's WLS and by REML.

```
resid.semivar <- variog(resid.geodat)
exp.wls <- variofit(resid.semivar, ini.cov.pars = c(8, 10/3), nugget = 17,
                    fix.nugget = FALSE, cov.model = 'exponential', weights = 'cressie')
exp.reml <- likfit(resid.geodat, ini.cov.pars = c(8, 10/3), nugget = 17,
                   fix.nugget = FALSE, cov.model = 'exponential', lik.method = 'rml')
sph.wls <- variofit(resid.semivar, ini.cov.pars = c(8, 10), nugget = 17,
                    fix.nugget = FALSE, cov.model = 'spherical', weights = 'cressie')
sph.reml <- likfit(resid.geodat, ini.cov.pars = c(8, 10), nugget = 17,
                   fix.nugget = FALSE, cov.model = 'spherical', lik.method = 'rml')
plot(resid.semivar, main = 'Semivariogram of Residuals')
lines(exp.wls, lty = 1)
lines(exp.reml, lty = 2)
lines(sph.wls, lty = 3)
lines(sph.reml, lty = 4)
legend('bottomright', lty = 1:4,
       legend = c('Exponential, WLS', 'Exponential, REML', 'Spherical, WLS', 'Spherical, REML'))
```




```
estimates <- rbind(
  `Exponential, WLS` = c(exp.wls$nugget, exp.wls$cov.pars * c(1, 3)),
  `Exponential, REML` = c(exp.reml$nugget, exp.reml$cov.pars * c(1, 3)),
  `Spherical, WLS` = c(sph.wls$nugget, sph.wls$cov.pars),
  `Spherical, REML` = c(sph.reml$nugget, sph.reml$cov.pars)
)
colnames(estimates) <- c('Nugget', 'Partial Sill', 'Range')
xtable(estimates)
```

	Nugget	Partial Sill	Range
Exponential, WLS	16.06	11.16	12.77
Exponential, REML	16.55	10.84	14.63
Spherical, WLS	18.30	8.80	13.22
Spherical, REML	17.36	9.37	11.35

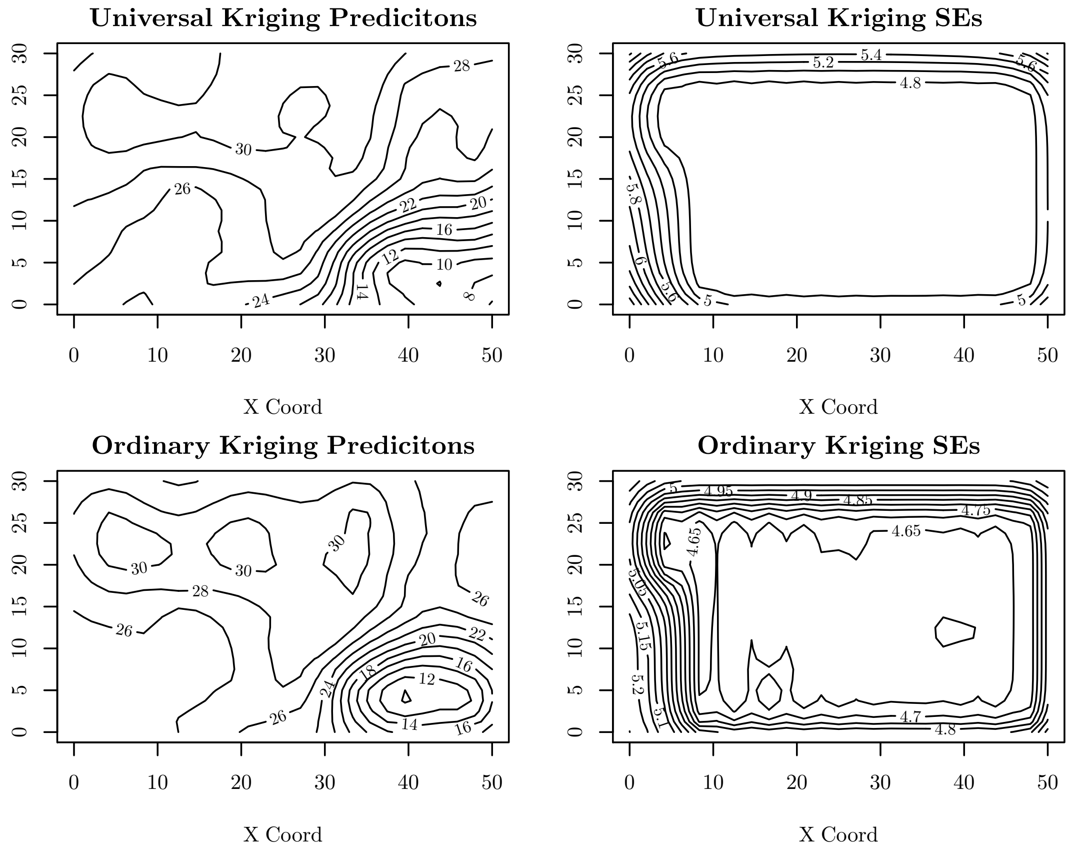
On the plot, all four estimated models look very similar but the table above shows some disagreement about the parameters. (Note that I multiplied the exponential range parameter by 3 so the table shows the effective range for the exponential models and the range for the spherical models.)

I ultimately chose the spherical model estimated by WLS because WLS does not depend on normality and the spherical models fit the points at the lower lags a bit better than the exponential models do.

(d) *Predict yields using universal kriging and ordinary kriging. Use the parameter estimates from the residual semivariogram when you do ordinary kriging. Plot the results along with a plot of the kriging standard errors. Remember to be careful of that range parameter – what you enter depends on which semivariogram model you used.*

```
# I'll just get the parameters straight out of the semivariogram model object.
wheat.uk <- ksline(wheat.geodat, locations = wheat.grid, cov.model = 'spherical',
                   cov.pars = sph.wls$cov.pars, nugget = sph.wls$nugget,
                   trend = 2, m0 = 'kt')
wheat.ok <- ksline(wheat.geodat, locations = wheat.grid, cov.model = 'spherical',
                   cov.pars = sph.wls$cov.pars, nugget = sph.wls$nugget)

par(mfrow = c(2, 2), mar = c(4, 3, 2, 1))
contour(wheat.uk, main = 'Universal Kriging Predicitons')
contour(wheat.uk, val = sqrt(wheat.uk$krige.var), main = 'Universal Kriging SEs')
contour(wheat.ok, main = 'Ordinary Kriging Predicitons')
contour(wheat.ok, val = sqrt(wheat.ok$krige.var), main = 'Ordinary Kriging SEs')
```



Compare the results and comment.

The predicted values from both methods are pretty similar, but the prediction surface from universal kriging smoother than the prediction surface from the ordinary kriging. The standard errors from universal kriging are larger because they incorporate uncertainty from estimating the mean surface.

2. *We looked at this example in class. We have a one-dimensional process with point to point covariance function*

$$C(s_i, s_j) = \exp\left(-\frac{3\,|s_i - s_j|}{5}\right).$$

The nugget effect is 0, the sill is 1 and the practical range is 5. The region B is defined to be the interval $B = (2, 4)$ with $|B| = 2$. The point to block covariance function is

$$\sigma(B, s) = Cov(Z(B), Z(s)) = \frac{1}{2}\int_2^4 \exp\left(-\frac{3\,|u - s|}{5}\right) du.$$

(a) *Find the covariance function.*

There are three cases to consider: $s < 2$, $s > 4$, and $s \in B$.

If $s < 2$ then $u - s > 0\ \forall u \in B$, so

$$\begin{aligned}
\frac{1}{2}\int_2^4 \exp\left(-\frac{3|u-s|}{5}\right) du &= \frac{1}{2}\int_2^4 \exp\left(-\frac{3(u-s)}{5}\right) du \\
&= \frac{1}{2}\left[-\frac{5}{3}\exp\left(-\frac{3(u-s)}{5}\right)\right]_{u=2}^4 \\
&= \frac{5}{6}\left[\exp\left(\frac{3}{5}(s-2)\right) - \exp\left(\frac{3}{5}(s-4)\right)\right].
\end{aligned}$$

If $s > 4$ then $u - s < 0\ \forall u \in B$, so

$$\begin{aligned}
\frac{1}{2}\int_2^4 \exp\left(-\frac{3|u-s|}{5}\right) du &= \frac{1}{2}\int_2^4 \exp\left(\frac{3(u-s)}{5}\right) du \\
&= \frac{1}{2}\left[\frac{5}{3}\exp\left(\frac{3(u-s)}{5}\right)\right]_{u=2}^4 \\
&= \frac{5}{6}\left[\exp\left(\frac{3}{5}(4-s)\right) - \exp\left(\frac{3}{5}(2-s)\right)\right].
\end{aligned}$$

Finally, if $s \in B$ then

$$\begin{aligned}
\frac{1}{2}\int_2^4 \exp\left(-\frac{3|u-s|}{5}\right) du &= \frac{1}{2}\int_2^s \exp\left(-\frac{3|u-s|}{5}\right) du + \frac{1}{2}\int_s^4 \exp\left(-\frac{3|u-s|}{5}\right) du \\
&= \frac{1}{2}\int_2^s \exp\left(\frac{3(u-s)}{5}\right) du + \frac{1}{2}\int_s^4 \exp\left(-\frac{3(u-s)}{5}\right) du \\
&= \frac{1}{2}\left[\frac{5}{3}\exp\left(\frac{3(u-s)}{5}\right)\right]_{u=2}^s + \frac{1}{2}\left[-\frac{5}{3}\exp\left(-\frac{3(u-s)}{5}\right)\right]_{u=s}^4 \\
&= \frac{5}{6}\left[\exp(0) - \exp\left(\frac{3}{5}(2-s)\right) - \exp\left(\frac{3}{5}(s-4)\right) + \exp(0)\right] \\
&= \frac{5}{6}\left[2 - \exp\left(\frac{3}{5}(2-s)\right) - \exp\left(\frac{3}{5}(s-4)\right)\right].
\end{aligned}$$

So the point-to-block covariance function is

$$\sigma(B,s) = \frac{5}{6}\begin{cases} \exp\left(\frac{3}{5}(s-2)\right) - \exp\left(\frac{3}{5}(s-4)\right) & s < 2 \\ 2 - \exp\left(\frac{3}{5}(2-s)\right) - \exp\left(\frac{3}{5}(s-4)\right) & 2 \leq s \leq 4 \\ \exp\left(\frac{3}{5}(4-s)\right) - \exp\left(\frac{3}{5}(2-s)\right) & s > 4 \end{cases}.$$

(b) *Find $\sigma(B,B)$.*

The block-to-block covariance is

$$\begin{aligned}
\sigma(B,B) &= \frac{1}{4}\int_2^4\int_2^4 \sigma(u,s)duds \\
&= \frac{1}{2}\int_2^4 \sigma(B,s)ds \\
&= \frac{1}{2}\int_2^4 \frac{5}{6}\left(2 - \exp\left(\frac{3}{5}(2-s)\right) - \exp\left(\frac{3}{5}(s-4)\right)\right) ds \\
&= \frac{5}{12}\left[2s + \frac{5}{3}\exp\left(\frac{3}{5}(2-s)\right) - \frac{5}{3}\exp\left(\frac{3}{5}(s-4)\right)\right]_{s=2}^{4} \\
&= \frac{5}{12}\left[8 + \frac{5}{3}\exp\left(-\frac{6}{5}\right) - \frac{5}{3}\exp(0) - 4 - \frac{5}{3}\exp(0) + \frac{5}{3}\exp\left(-\frac{6}{5}\right)\right] \\
&= \frac{5}{12}\left[4 - \frac{10}{3} + \frac{10}{3}\exp\left(-\frac{6}{5}\right)\right] \\
&= \frac{5}{12}\left[\frac{2}{3} + \frac{10}{3}\exp\left(-\frac{6}{5}\right)\right] \\
&= \frac{5}{18} + \frac{25}{18}\exp\left(-\frac{6}{5}\right).
\end{aligned}$$