Test 2 take home part

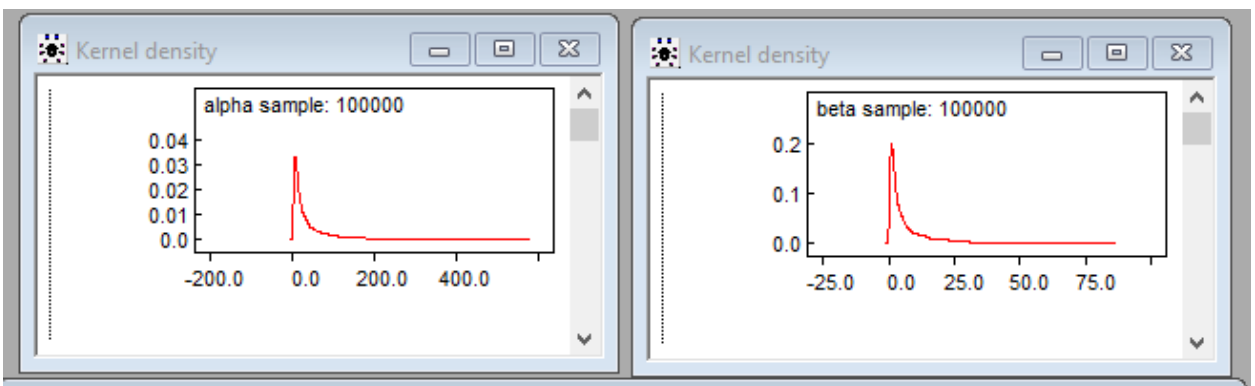
Laha Ale

April 5, 2019

Problem I

```
model
{
   for( i in 1 : N ) {
   lambda[i] ~dgamma(alpha,beta)
   x[i] ~ dpois(lambda[i])
    }

    alpha ~ dgamma(1.0E-3,1.0E-3)
    beta ~ dgamma(1.0E-3,1.0E-3)
}


list(x = c(5, 6, 5, 10, 3, 9, 4, 4, 4, 12), N=10)

list(alpha=2, beta=1)
```

The statistics for alpha and beta as follow:



Node statistics

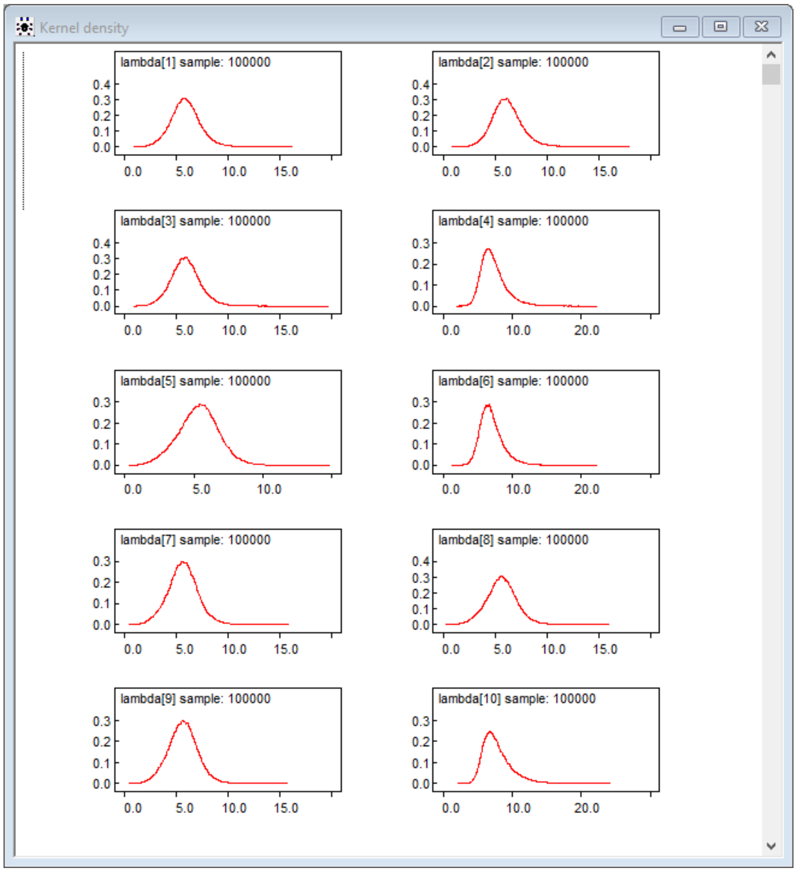
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	49.97	71.91	3.849	2.485	21.14	283.6	10001	100000

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta	8.128	11.67	0.6265	0.3767	3.445	45.51	10001	100000

Both α and β are initial as Gamma distribution. The density curves show relatively smooth. Moth MC errors are less than their standard deviation of 5% or more. CI with 95% credit set are $2.485 < \alpha < 283.6$ and $0.3767 < \beta < 45.51$.

lambda density curve and statistics



Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
lambda[1]	5.882	1.43	0.01483	3.179	5.827	8.916	10001	100000
lambda[2]	6.158	1.465	0.01112	3.521	6.061	9.394	10001	100000
lambda[3]	5.884	1.427	0.01453	3.19	5.834	8.899	10001	100000
lambda[4]	7.235	1.811	0.03727	4.542	6.931	11.67	10001	100000
lambda[5]	5.342	1.455	0.03025	2.415	5.382	8.195	10001	100000
lambda[6]	6.964	1.685	0.02883	4.359	6.716	11.04	10001	100000
lambda[7]	5.615	1.431	0.02203	2.797	5.611	8.532	10001	100000
lambda[8]	5.619	1.428	0.022	2.815	5.616	8.522	10001	100000
lambda[9]	5.616	1.426	0.02197	2.816	5.613	8.492	10001	100000
lambda[10]	7.775	2.087	0.05519	4.856	7.36	12.98	10001	100000

summary statistics show in the above table, for example, λ_1 has $mean = 5.882$ and standard deviation $\sigma_\lambda = 1.43$ MC, error=0.01483 is less than standard deviation of 5%, median=5.827, and 95% credible set $3.179 < \lambda < 8.916$. The distribution of λ are close to normal distribution. Therefore, we can compute $P(\lambda_i|x)$ posterior as $P(\lambda|x) \sim N(\mu_\lambda, \sigma_\lambda)$.

$P(\lambda_1|x) \sim N(5.882, 1.43), P(\lambda_2|x) \sim N(6.158, 1.465)$
$P(\lambda_3|x) \sim N(5.884, 1.427), P(\lambda_4|x) \sim N(7.235, 1.811)$
$P(\lambda_5|x) \sim N(5.342, 1.455), P(\lambda_6|x) \sim N(6.964, 1.685)$
$P(\lambda_7|x) \sim N(5.615, 1.431), P(\lambda_8|x) \sim N(5.619, 1.428)$
$P(\lambda_9|x) \sim N(5.616, 1.426), P(\lambda_{10}|x) \sim N(7.775, 2.087)$

Problem II

```r
library(spdep)
library(maps)
library(maptools)
library(classInt)
library(RColorBrewer)
```

1)

```r
columbus.poly <- readShapePoly(system.file("etc/shapes/columbus.shp", package="spdep")[1])
columbus.coords <- coordinates(columbus.poly)
columbus.knn <- knearneigh(columbus.coords)
columbus.knn2nb <- knn2nb(columbus.knn)
columbus.dist.list <- nbdists(columbus.knn2nb, columbus.coords)
```

```r
columbus.dist.vec <- unlist(columbus.dist.list)
columbus.dist.max <- max(columbus.dist.vec)
columbus.dnn.nb <- dnearneigh(columbus.coords, 0, columbus.dist.max)


#CAR model
columbus.dnn.listw = nb2listw(columbus.dnn.nb, style="B", zero.policy=TRUE)
columbus.dnn.car.out = spautolm(HOVAL~CRIME+NEIG+INC+OPEN+PLUMB+DISCBD,
                                data=columbus.poly, family="CAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.car.out)
```

```
##
## Call: spautolm(formula = HOVAL ~ CRIME + NEIG + INC + OPEN + PLUMB +
##     DISCBD, data = columbus.poly, listw = columbus.dnn.listw,
##     family = "CAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.5129  -8.1015  -4.0272   3.7477  53.7996
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 37.50465   14.63513  2.5626 0.010388
## CRIME       -0.44467    0.19366 -2.2962 0.021665
## NEIG        -0.17553    0.14677 -1.1960 0.231713
## INC          0.39586    0.49127  0.8058 0.420367
## OPEN         0.61076    0.42733  1.4293 0.152931
## PLUMB        1.62740    0.62700  2.5955 0.009444
## DISCBD       3.40167    2.27295  1.4966 0.134500
##
## Lambda: 0.0085685 LR test value: 0.0057607 p-value: 0.9395
## Numerical Hessian standard error of lambda: 0.10361
##
## Log likelihood: -195.9237
## ML residual variance (sigma squared): 173.96, (sigma: 13.189)
## Number of observations: 49
## Number of parameters estimated: 9
## AIC: 409.85
```

```r
#SAR model
columbus.dnn.sar.out = spautolm(HOVAL~CRIME+NEIG+INC+OPEN+PLUMB+DISCBD,
                                data=columbus.poly, family="SAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.sar.out)
```

```
##
## Call: spautolm(formula = HOVAL ~ CRIME + NEIG + INC + OPEN + PLUMB +
##     DISCBD, data = columbus.poly, listw = columbus.dnn.listw,
##     family = "SAR", zero.policy = TRUE)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -19.4905  -8.1299  -3.9008   3.5911  53.8320
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 37.56439   14.65913  2.5625  0.01039
## CRIME       -0.44620    0.19385 -2.3017  0.02135
## NEIG        -0.17537    0.14737 -1.1900  0.23406
## INC          0.40230    0.49168  0.8182  0.41323
## OPEN         0.61093    0.42689  1.4311  0.15239
## PLUMB        1.61595    0.62861  2.5707  0.01015
## DISCBD       3.37332    2.27935  1.4800  0.13889
##
## Lambda: 0.006389 LR test value: 0.0085577 p-value: 0.92629
## Numerical Hessian standard error of lambda: 0.068979
##
## Log likelihood: -195.9223
## ML residual variance (sigma squared): 173.95, (sigma: 13.189)
## Number of observations: 49
## Number of parameters estimated: 9
## AIC: 409.84
```

CAR model with the maximum intercentroid:

$HOVAL = 37.50465 - 0.44467 * CRIME - 0.17553 * NEIG + 0.39586 * INC + 0.61076 * OPEN + 1.62740 * PLUMB + 3.40167 * DISCBD$

SAR model with the maximum intercentroid:

$HOVAL = 37.56439 - 0.44620 * CRIME - 0.17537 * NEIG + 0.40230 * INC + 0.61093 * OPEN + 1.61595 * PLUMB + 3.37332 * DISCBD$

As we can see from the above models the HOVAL has a negative relation with CRIME and NEIG; and positive relation with INC, OPEN, PLUMB, and DISC BD. Based on p-value < 0.05 is significant, parameters CRIME and PLUMB are significant, and the rest of parameters NEIG, INC, OPEN and DISCBD are insignificant.And both models are spatial. According to Log likelihood and AIC, the SAR models is sightly better.

2)

```r
#CAR reduced model by backward elimination
columbus.dnn.listw = nb2listw(columbus.dnn.nb, style="B", zero.policy=TRUE)
columbus.dnn.car.out = spautolm(HOVAL~CRIME+OPEN+PLUMB+PLUMB,
                                data=columbus.poly, family="CAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.car.out)
```

```
##
## Call:
## spautolm(formula = HOVAL ~ CRIME + OPEN + PLUMB + PLUMB, data = columbus.poly,
##     listw = columbus.dnn.listw, family = "CAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.1210  -7.2086  -3.3113   3.3281  59.2100
##
## Coefficients:
```

```
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 60.55667    5.18012 11.6902 < 2.2e-16
## CRIME       -0.74526    0.14233 -5.2363 1.638e-07
## OPEN         0.66525    0.43556  1.5274   0.12667
## PLUMB        1.02051    0.60838  1.6774   0.09346
##
## Lambda: 0.055617 LR test value: 0.48188 p-value: 0.48757
## Numerical Hessian standard error of lambda: 0.070613
##
## Log likelihood: -198.3625
## ML residual variance (sigma squared): 190.41, (sigma: 13.799)
## Number of observations: 49
## Number of parameters estimated: 6
## AIC: 408.73
```

```r
#CAR reduced model by backward elimination
columbus.dnn.sar.out = spautolm(HOVAL~CRIME+OPEN+PLUMB+PLUMB,
                                data=columbus.poly, family="SAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.sar.out)
```

```
##
## Call:
## spautolm(formula = HOVAL ~ CRIME + OPEN + PLUMB + PLUMB, data = columbus.poly,
##     listw = columbus.dnn.listw, family = "SAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.8929  -7.2534  -4.0990   3.5766  59.1877
##
## Coefficients:
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 60.88264    5.24868 11.5996 < 2.2e-16
## CRIME       -0.75071    0.14468 -5.1888 2.117e-07
## OPEN         0.67303    0.43177  1.5587    0.1191
## PLUMB        0.98423    0.61434  1.6021    0.1091
##
## Lambda: 0.0396 LR test value: 0.6483 p-value: 0.42072
## Numerical Hessian standard error of lambda: 0.046713
##
## Log likelihood: -198.2793
## ML residual variance (sigma squared): 189.84, (sigma: 13.778)
## Number of observations: 49
## Number of parameters estimated: 6
## AIC: 408.56
```

```r
#CAR reduced model by backward elimination
columbus.dnn.listw = nb2listw(columbus.dnn.nb, style="B", zero.policy=TRUE)
columbus.dnn.car.out = spautolm(HOVAL~CRIME+PLUMB,
                                data=columbus.poly, family="CAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.car.out)
```

```
##
```

```
## Call: spautolm(formula = HOVAL ~ CRIME + PLUMB, data = columbus.poly,
##     listw = columbus.dnn.listw, family = "CAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.9951  -8.4417  -4.4656   5.5334  59.0249
##
## Coefficients:
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 62.83945    4.98445 12.6071 < 2.2e-16
## CRIME       -0.77689    0.14169 -5.4829 4.183e-08
## PLUMB        1.27498    0.59808  2.1318   0.03302
##
## Lambda: 0.043601 LR test value: 0.27925 p-value: 0.59719
## Numerical Hessian standard error of lambda: 0.075773
##
## Log likelihood: -199.4887
## ML residual variance (sigma squared): 200.14, (sigma: 14.147)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 408.98
```

```
#CAR reduced model by backward elimination
columbus.dnn.sar.out = spautolm(HOVAL~CRIME+PLUMB,
                                data=columbus.poly, family="SAR",
                                listw=columbus.dnn.listw,
                                zero.policy=TRUE)
summary(columbus.dnn.sar.out)
```

```
##
## Call: spautolm(formula = HOVAL ~ CRIME + PLUMB, data = columbus.poly,
##     listw = columbus.dnn.listw, family = "SAR", zero.policy = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.5596  -8.4925  -4.9135   5.9344  59.0233
##
## Coefficients:
##             Estimate Std. Error z value  Pr(>|z|)
## (Intercept) 63.05601    5.04452 12.4999 < 2.2e-16
## CRIME       -0.78067    0.14342 -5.4433 5.231e-08
## PLUMB        1.25746    0.60203  2.0887   0.03674
##
## Lambda: 0.029954 LR test value: 0.36867 p-value: 0.54373
## Numerical Hessian standard error of lambda: 0.047643
##
## Log likelihood: -199.444
## ML residual variance (sigma squared): 199.88, (sigma: 14.138)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 408.89
```

As we can see from above, first we move out NEIG and INC based highest p-values; and find out OPEN and DISCBD are still have p-value > 0.05; therefore, we move out those parameters to get the final model that all of the parameters have the p-value < 0.05.

CAR reduced model:

$HOVAL = 62.83945 - 0.77689 * CRIME + 1.27498 * PLUMB$

SAR reduced model:

$HOVAL = 63.05601 - 0.78067 * CRIME + 1.25746 * PLUMB$

According to the p-value of Lambda, the above two models are not spatial. It makes sense that CRIME has a negative relationship with HOVAL(House value) because people do not want to buy a house located high crime ratio places for safety, and PLUMB positive relationship with HOVAL because of water supply is essential to daily life.

3)

```
library(geoR)
library(spBayes)
X_mean = mean(columbus.poly$X)
print(paste("mean of X=",X_mean))
```

```
## [1] "mean of X= 39.464285755102"
```

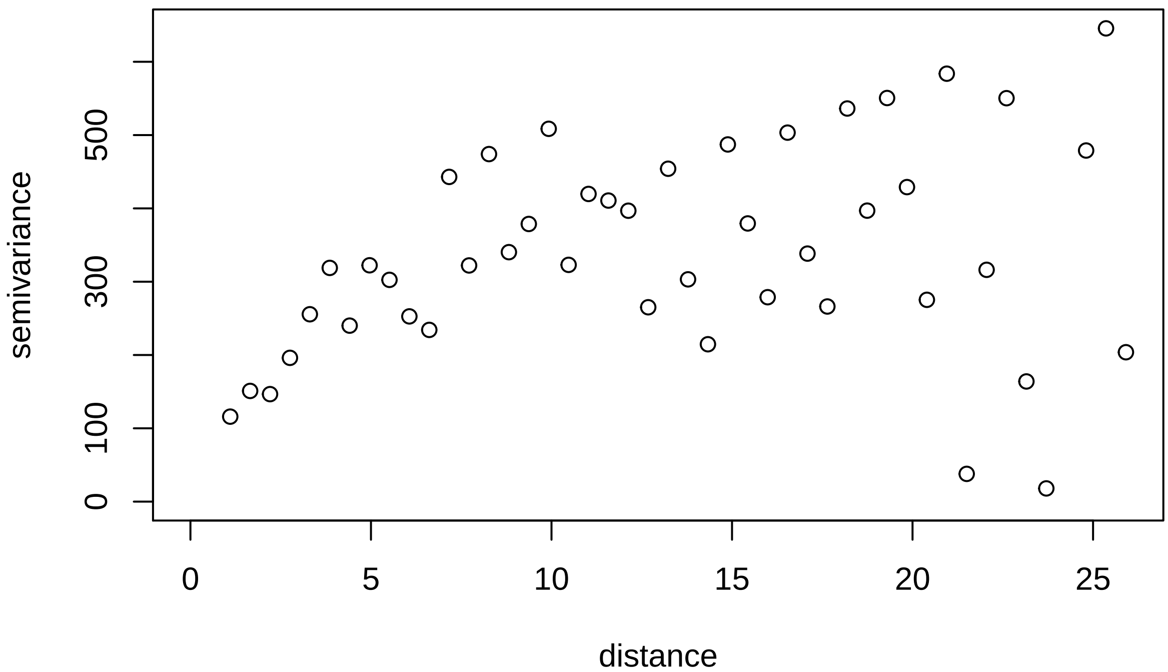
```
Y_mean = mean(columbus.poly$Y)
print(paste("mean of Y=",Y_mean))
```

```
## [1] "mean of Y= 32.3726528571429"
```

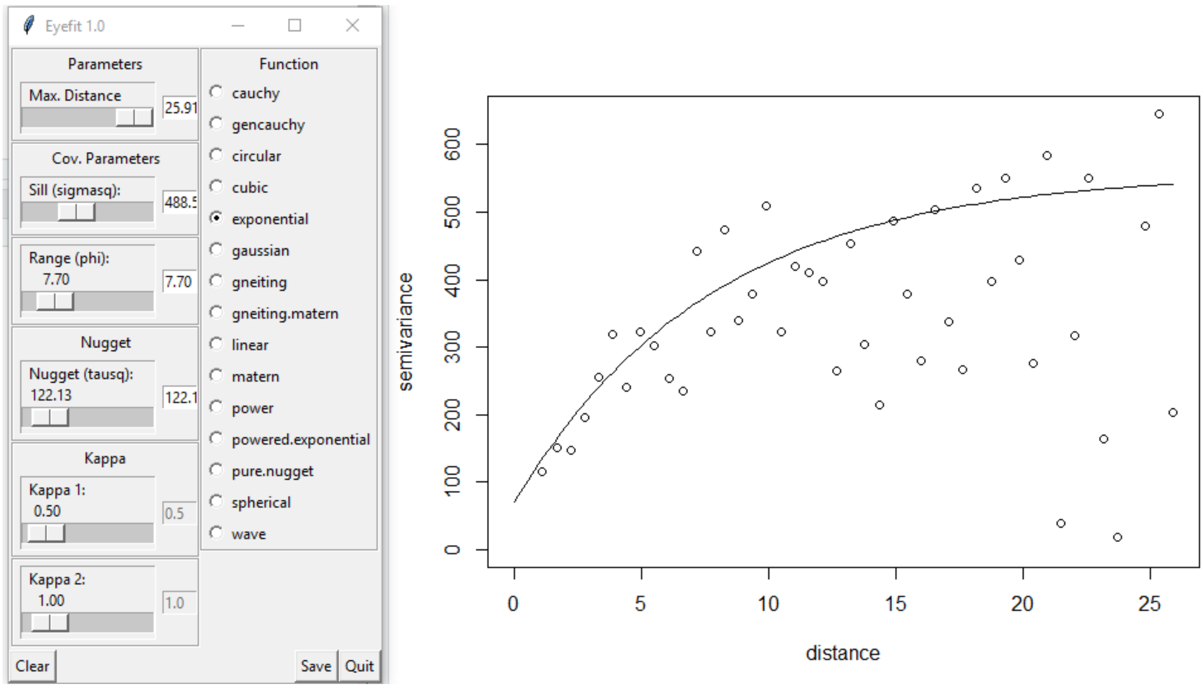
```
coords <- as.matrix(cbind(columbus.poly$X,columbus.poly$Y))
HOVAL <- columbus.poly$HOVAL

bins = 50
max.dist <- max(iDist(coords))
HOVAL.vario <- variog(coords = coords, data = HOVAL,
                       uvec = (seq(0, max.dist, length = bins)))
```

```
## variog: computing omnidirectional variogram
```

```
plot(HOVAL.vario)
```




```r
#adjust paramters with eyefit
eyefit(myscps.vario,silent=TRUE)
```



According step the above results, the following parameters should setting as $\sigma^2 = 488.5, \Phi = 7.7$ and nugget=122.1.

```
library(spdep)

point<-krige.conv(coords = coords,
                  data = HOVAL,loc=c(length(HOVAL),1),
                  krige=krige.control(cov.pars=c(331.5,8.4),
                                      cov.model="exponential",
                                      nugget=139.5))
```

```
## krige.conv: model with constant mean
## krige.conv: Kriging performed using global neighbourhood
```

```
point
```

```
## $predict
##     data
## 47.46833
##
## $krige.var
## [1] 555.0933
##
## $beta.est
##    beta
## 48.0682
##
## $distribution
```

```
## [1] "normal"
##
## $message
## [1] "krige.conv: Kriging performed using global neighbourhood"
##
## $call
## krige.conv(coords = coords, data = HOVAL, locations = c(length(HOVAL),
##     1), krige = krige.control(cov.pars = c(331.5, 8.4), cov.model = "exponential",
##     nugget = 139.5))
##
## attr(,"sp.dim")
## [1] "2d"
## attr(,"prediction.locations")
## c(length(HOVAL), 1)
## attr(,"parent.env")
## <environment: R_GlobalEnv>
## attr(,"data.locations")
## coords
## attr(,"class")
## [1] "kriging"
```

```r
pred_low <-point$predict - 2*sqrt(point$krige.var)
pred_high <-point$predict + 2*sqrt(point$krige.var)
print(paste("The PI is between",pred_low,"and",pred_high))
```

```
## [1] "The PI is between 0.347497727950859 and 94.5891670440649"
```