Lag.correlation.covariance.semivariance.R

thomasadler

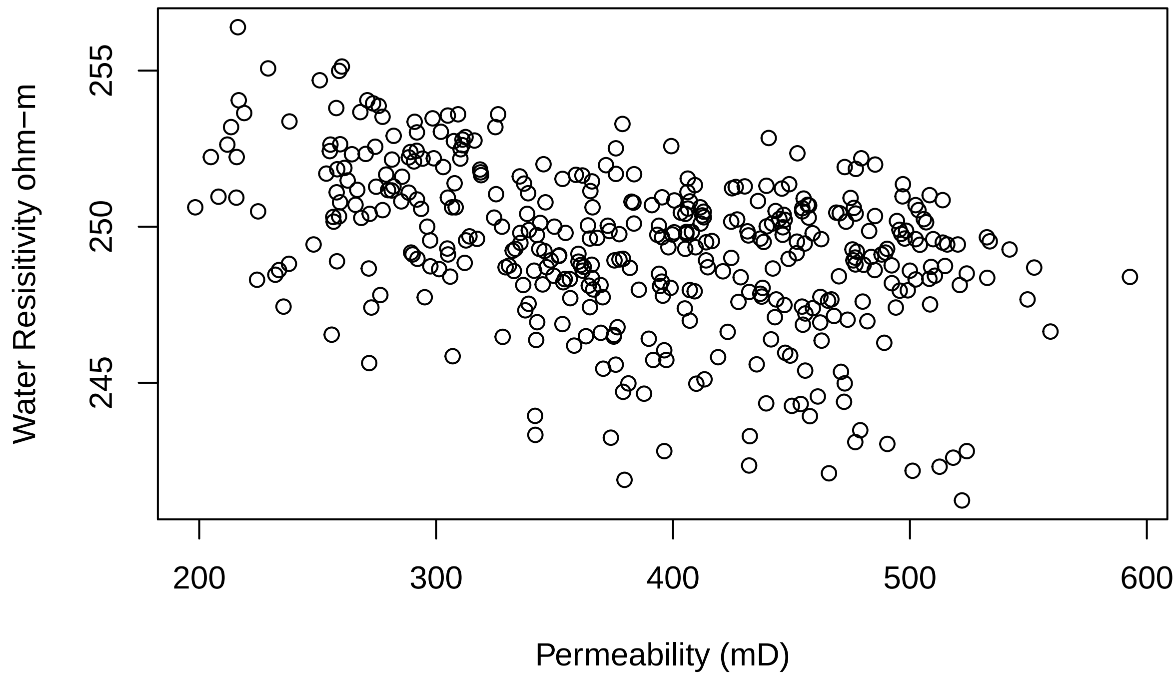
2019-09-12

```r
# Lags, correlation, covariance, and semi-variance

# Created by: Thomas Adler
# Last Updated: September 11th, 2019

# The following script utilizes the "Berea Sandstone" data set to caluclate statistics for
# a given   lag distance.   The Berea   sandstone   is  a   section of  rock    where   resistivity
# and   permeability    were    measured    on   extremely    small    intervals    (~3 mm  increments).
```

```r
library(readxl)

# define file location
myFile <- "~/Desktop/R/CE269 Applied Geostatistics/Homework 2/Burea_subset.xls"

# read data file
myData <- read_excel(myFile)
# view dataset structure
str(myData)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    403 obs. of  4 variables:
##  $ x (mm)                 : num  15 15 15 18 18 18 18 21 21 24 ...
##  $ y (mm)                 : num  33 60 75 27 84 117 123 63 105 21 ...
##  $ Water Resisitivity ohm-m: num  249 249 244 249 247 ...
##  $ Permeability (mD)       : num  552 490 458 360 474 ...
```

```r
# define variables of interest
v2 <- myData$`Water Resisitivity ohm-m`
v1 <- myData$`Permeability (mD)`

# scatter   plot of resistivity and permeability for    all the sample  points
plot(v1,v2,xlab = "Permeability (mD)",ylab = "Water Resisitivity ohm-m")
```




```
# Pearson's Correlation coefficient (r)
r <- pearson(v1,v2)

# Covariance (C) of sample population (n-1)
cov <- covariance(v1,v2)
```

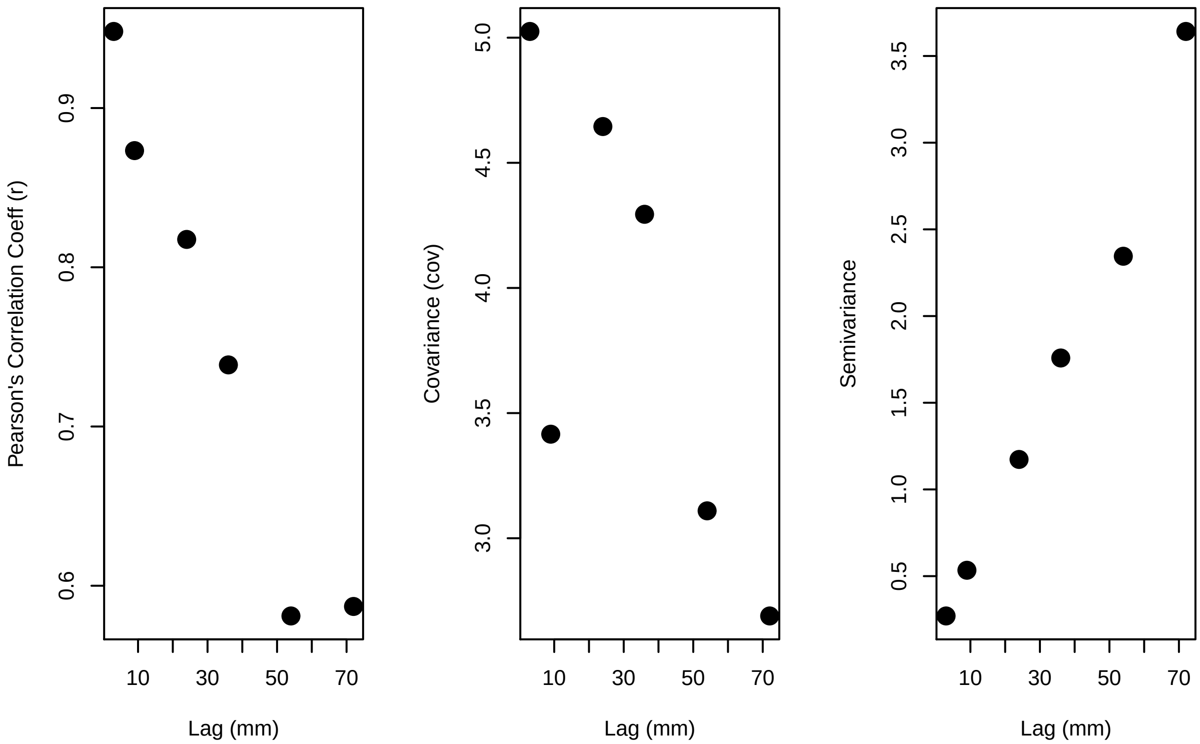
```
# Define grid system
x <- myData$`x (mm)`
y <- myData$`y (mm)`
# Define variable of interest
v <- myData$`Water Resisitivity ohm-m`
# Define direction of interest
di <- x
# Define direction to keep constant
dc <- y
# Define lag
lag = c(3, 9,   24, 36, 54, 72)
```

```
#Create Blank Data Frame
ldf <- data.frame(matrix(ncol = 5, nrow = 0))
n <- c("dlow", "dhigh","vlow", "vhigh", "dist")
colnames(ldf) <- n
#Parameter definitions
# dlow <- coordinate position 1
# dhigh <- coordinate position 2
# vlow <- variable of interest at position 1
# vhigh <- variable of interest at position 2
# dist <- distance (or lag) between positions 1 and 2

#Fill Blank Data Frame
dcu <- unique(dc)
j=1
for (k in 1:length(lag)){
```

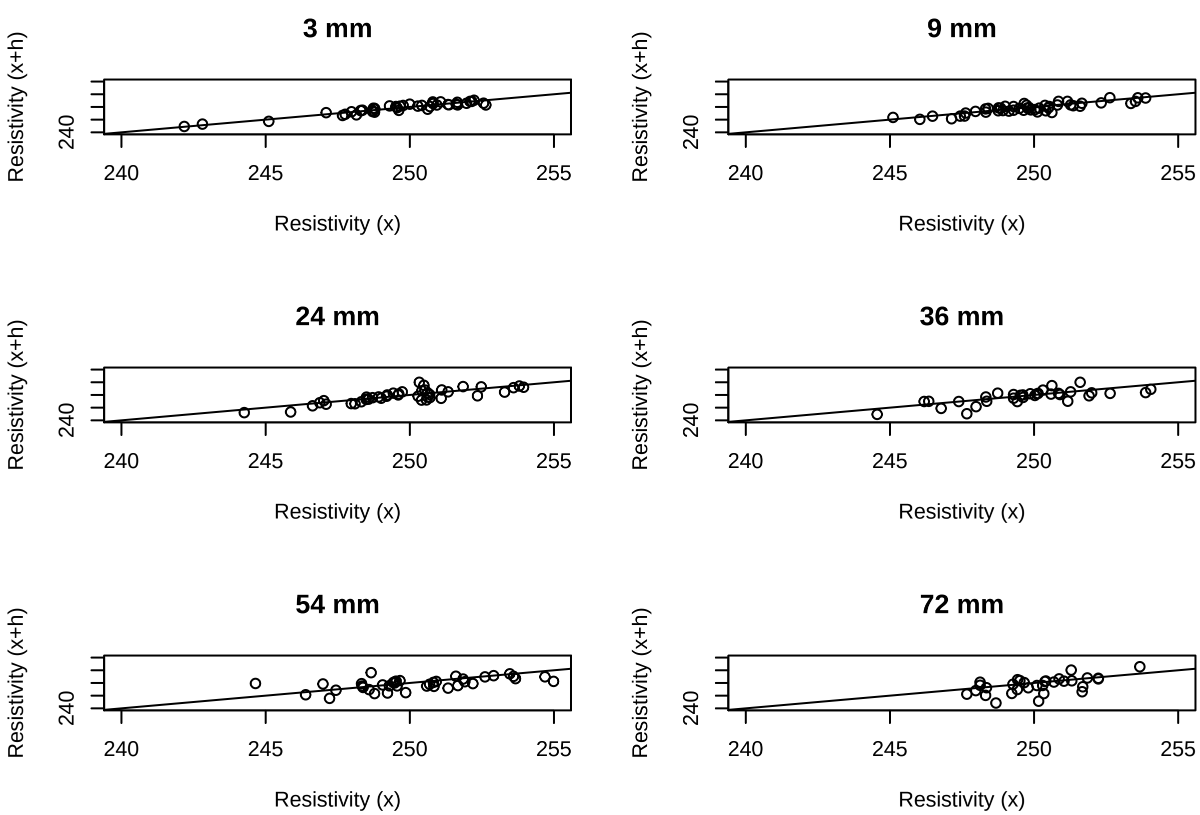
```r
#Create dataset of
for (i in 1:length(dcu)){
# take sample of data set
sub <- data.frame("dir interest" = di, "dir constant" = dc, "v" = v)
# subset the data by keeping one dimension constant
sub <- subset(sub, dc == dcu[i])
if (length(sub$v) > 1){
# Every combination of locations and variables
df1 <- t(combn(sub$dir.interest,2))
df2 <- t(combn(sub$v,2))
df <- cbind2(df1,df2)
df <- data.frame("d low" = df[,1], "d high" = df[,2], "v low" = df[,3], "v high" = df[,4])
# distance between every combination of locations
df$dist <- abs(df$d.low-df$d.high)
# subset the data by distances that equal the desired lag
sdf <-subset(df, dist == lag[k])
if (!dim(sdf)[1] == 0) {
# add data to the lag data frame (ldf)
   l<-dim(sdf)[1]
   ldf[j:(j+l-1),] <- sdf
  j = j+l
}
}
}
}
```

```r
library(dplyr)
#ldf %>% group_by(dist) %>% summarize(pearson(vlow,vhigh))

# Split data by lag
sp <- split(ldf, ldf$dist)

# Calculate Pearson Correlation (p)
p <- lapply(names(sp), function(x) pearson(sp[[x]][["vlow"]],sp[[x]][["vhigh"]]))
p <- unlist(p)

# Calculate Covariance (cov)
cov <- lapply(names(sp), function(x) covariance(sp[[x]][["vlow"]],sp[[x]][["vhigh"]]))
cov <- unlist(cov)

# Calculate semivariance (s)
s <- lapply(names(sp), function(x) semivariance(sp[[x]][["vlow"]],sp[[x]][["vhigh"]]))
s <- unlist(s)

#Combining result
results <- data.frame("Lag"= lag,"p"= p,"cov"= cov,"s"= s)
write.table(results, "~/Desktop/R/CE269 Applied Geostatistics/Homework 2/results.xlsx", sep="\t")
```

```r
# Plotting statistical parameters against lag distance
par(mfrow=c(1,3))
plot(results$Lag,results$p,ylab="Pearson's Correlation Coeff (r)",xlab="Lag (mm)",col = "black",pch = 1
plot(results$Lag,results$cov,ylab="Covariance (cov)",xlab="Lag (mm)",col = "black",pch = 16,cex=2)
plot(results$Lag,results$s,ylab="Semivariance",xlab="Lag (mm)",col = "black",pch = 16,cex=2)
```




```
# Plotting resistivity at different lag steps
names(sp) <- c("3 mm", "9 mm","24 mm","36 mm","54 mm","72 mm")
par(mfrow=c(3,2))
lapply(names(sp), function(x) plot(sp[[x]][["vlow"]],sp[[x]][["vhigh"]],xlab="Resistivity (x)",ylab="Re
```



```
## [[1]]
## NULL
##
## [[2]]
## NULL
##
## [[3]]
## NULL
##
## [[4]]
## NULL
##
## [[5]]
## NULL
##
## [[6]]
## NULL
```