Day9 exercise solutions

Ali Movasati

Nov. 13th, 2024

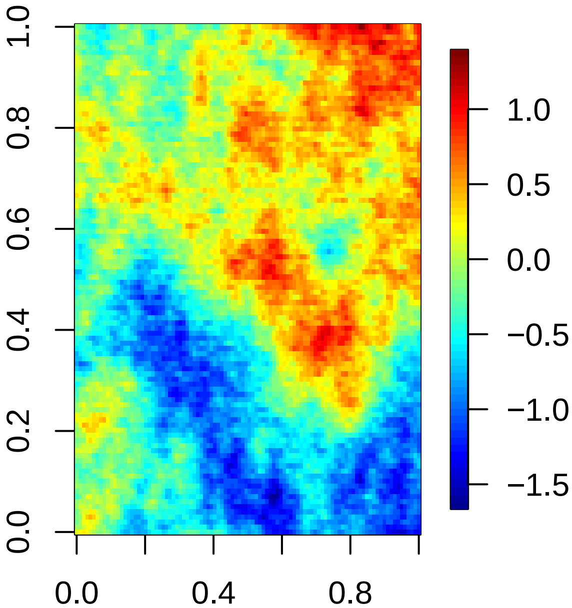
```
# Set global code chunk options
knitr::opts_chunk$set(warning = FALSE)
```

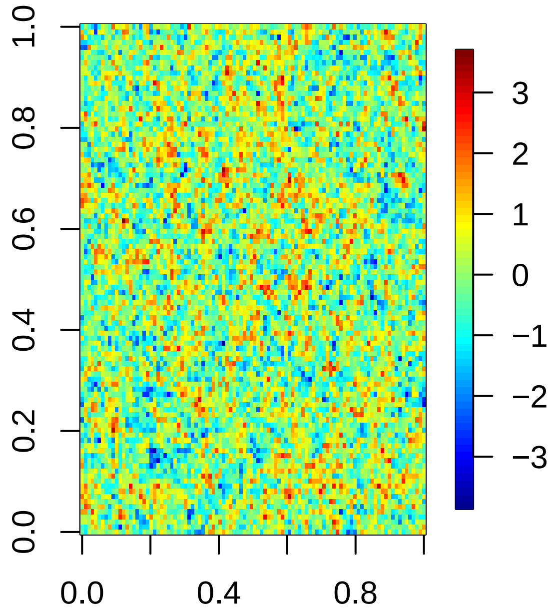
```
# load required libraries
library("skimr")
library("dplyr")
library("magrittr")
library("ggplot2")
library("survival")
library("survminer")
library("fields")

# define functions
`%notin%` <- Negate(`%in%`)
```

Problem 1

1.A)

```
load("/Users/alimos313/Documents/studies/phd/university/courses/stat-modelling/StatModelEx/day11/data/sp

par(mfrow=c(1,2))
image.plot(sim1)
image.plot(sim2)
```






```r
# summary statistics
summary(as.vector(sim1))
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.64094 -0.55859 -0.05631 -0.13214  0.28576  1.37183
```

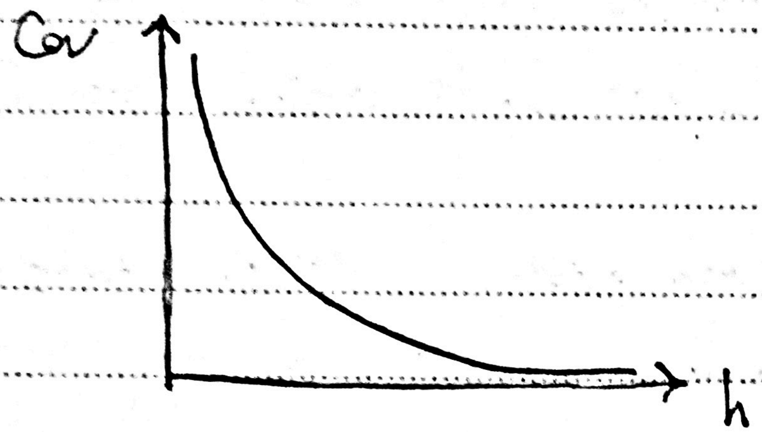
```r
summary(as.vector(sim2))
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.80355 -0.63321  0.01862  0.01869  0.68658  3.64528
```

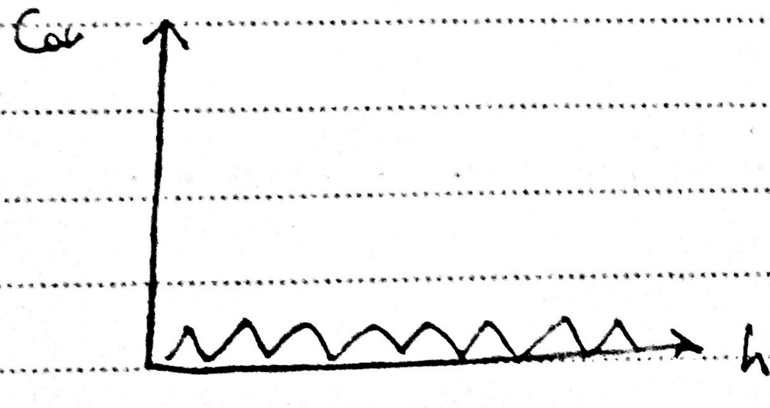
1.B)

subject: date:

Sim1



Sim2



Arman

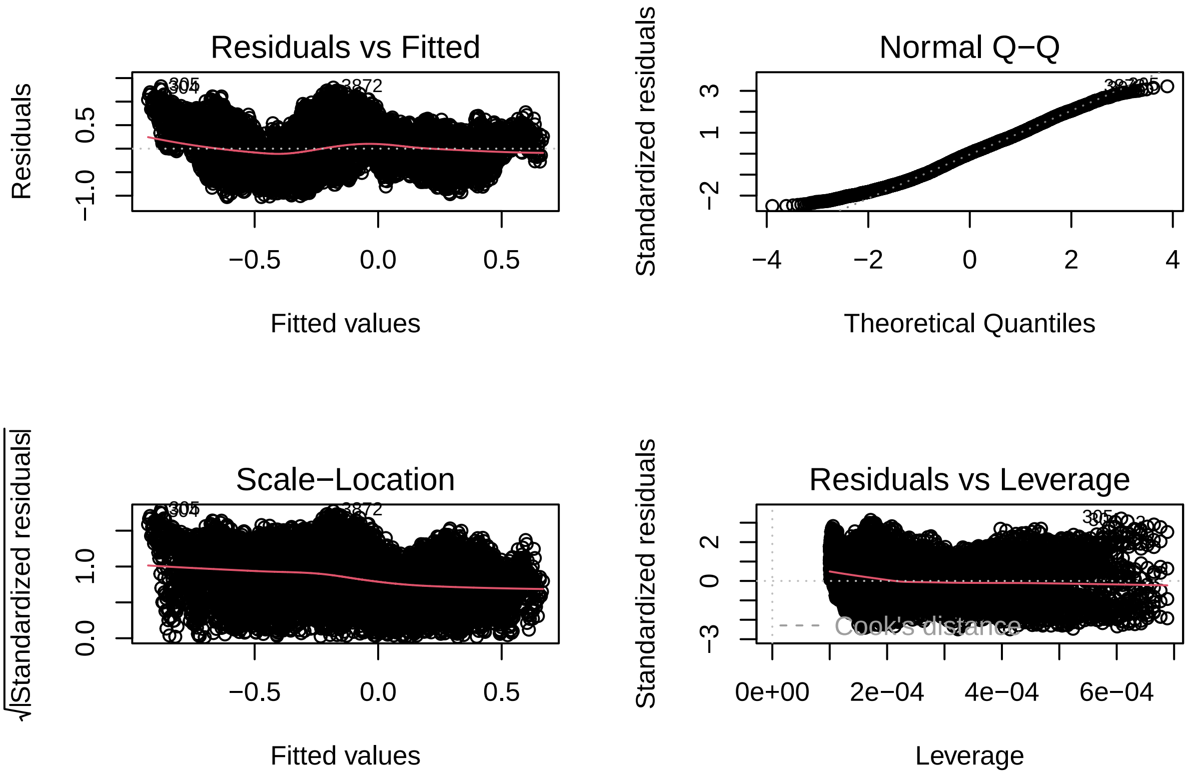
1.C)

```r
coords <- expand.grid(x1 = 1:100, x2 = 1:100)
coords$observed <- as.vector(sim1)

lm_model <- lm(observed ~ x1 + x2, data=coords)
summary(lm_model)
```

```
##
## Call:
## lm(formula = observed ~ x1 + x2, data = coords)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.02896 -0.30692  0.00469  0.27550  1.32699
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.947859   0.011015  -86.06   <2e-16 ***
## x1           0.004133   0.000143   28.91   <2e-16 ***
## x2           0.012020   0.000143   84.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4128 on 9997 degrees of freedom
## Multiple R-squared:  0.4415, Adjusted R-squared:  0.4414
## F-statistic:  3951 on 2 and 9997 DF,  p-value: < 2.2e-16
```

1.D)

```r
par(mfrow=c(2,2))
plot(lm_model)
```



« comments »

1. Assumption of Independent Errors Violation: In spatial data, observations are often spatially correlated; nearby locations tend to have similar values (spatial autocorrelation). This violates the assumption that residuals (errors) are independent of each other.
2. Assumption of Homoscedasticity (Constant Variance of Errors) Violation: Spatial data often exhibit heteroscedasticity, where the variability of the residuals changes across the spatial domain. This can occur if different areas have different levels of variability due to local conditions.
3. Assumption of Linearity Violation: The relationship between predictors and the response variable might not be strictly linear across space. Spatial data often exhibit complex relationships that vary over the spatial domain, introducing non-linear patterns.
4. Assumption of Normality of Errors Violation: The spatial structure or clusters in the data can lead to non-normal error distributions. Clustering or outliers common in spatial data might skew the residuals.
5. Collinearity Among Predictors Violation: Spatial predictors (e.g., longitude, latitude, or environmental variables) are often correlated due to spatial patterns or regional influences. This can introduce multicollinearity.

1.E)

A spatial regression model like kriging or spatial mixed models could be used: Z(s) = + S(s) + Where:

Z(s): Spatial process : Mean trend S(s): Spatial random effect : Independent error term

Problem 2

2.A)

2.B)

2.C)

2.D)

2.E)

2.F)

2.G)