hw11.R

joebrew

Sat Nov 15 10:17:29 2014

```r
# Homework 11: Disease Cluster Analysis -SaTscan
# PHC 6937: Spatial Epidemiology
#
# Stl_home is a shapefile which includes St Louis region county
# homicide counts (variable:HC8893) and population (variable:PO8893).
# In this assignment, you are requested to perform a purely spatial
# scan test based on Poisson model to detect any disease clusters
# in the study region.
#
# Steps: Import the data into SatScan and set up the session to run the analysis.
# Present the Output of the cluster information for evaluation
# (including the primary cluster and secondary cluster if necessary)


######
# Attach packages
#######
library(gdata)
```

```
## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
##
## gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.
##
## Attaching package: 'gdata'
##
## The following object is masked from 'package:stats':
##
##     nobs
##
## The following object is masked from 'package:utils':
##
##     object.size
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(maptools)
```

```
## Loading required package: sp
## Checking rgeos availability: TRUE
```

```r
library(rgdal)
```

```
## rgdal: version: 0.9-1, (SVN revision 518)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 1.10.1, released 2013/08/26
## Path to GDAL shared files: /usr/share/gdal/1.10
## Loaded PROJ.4 runtime: Rel. 4.8.0, 6 March 2012, [PJ_VERSION: 480]
## Path to PROJ.4 shared files: (autodetected)
```

```r
library(spatstat)
```

```
##
## spatstat 1.38-1       (nickname: 'Le Hardi')
## For an introduction to spatstat, type 'beginner'
```

```r
library(maps)
library(SpatialEpi)

######
# Set working directory to the haiti directories on local machine
######
if(Sys.info()["sysname"] == "Windows"){
  wd <- 'C:/Users/BrewJR/Documents/uf/phc6194/hw11'
} else {
  wd <- '/home/joebrew/Documents/uf/phc6194/hw11'
}
setwd(wd)

######
# Import the St Louis crime and population shapefile
######
stl <- readOGR("stl_hom", "stl_hom")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "stl_hom", layer: "stl_hom"
## with 78 features and 21 fields
## Feature type: wkbPolygon with 2 dimensions
```

```r
######
# Visually check read-in and projection by mapping
# onto state of Missouri and Illinois
######
map("county", c("missouri", "illinois"),
    fill = TRUE,
    col = "grey",
```

```
    border = "white")
map("state", c("missouri", "illinois"),
    col = "black",
    add = TRUE,
    lwd = 2)
plot(stl,
     add = TRUE,
     col = adjustcolor("red", alpha.f = 0.3),
     border = "darkred")
```
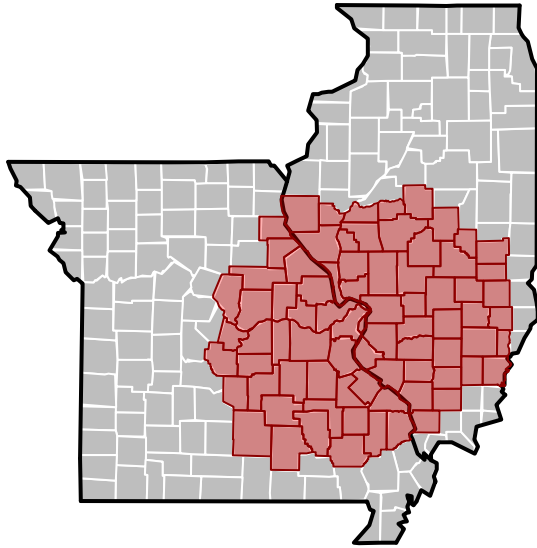


```
######
# Rename some variables for comprehensibility
######
stl$hom <- stl$HC8893
stl$pop <- stl$PO8893

######
# Sanity check: population and homicide (absolute) should correlate
######
plot(log(stl$pop), log(stl$hom),
    xlab = "Population",
    ylab = "Homicides",
    xaxt = "n",
    yaxt = "n",
    main = "Population and homicides (log-log)",
    pch = 16,
    col = adjustcolor("darkorange", alpha.f = 0.8))
axis(side = 1,
    at = log(quantile(stl$pop, probs = c(0, 0.1, 0.4, 0.7, 0.8, 0.9, 1))),
    labels = round(quantile(stl$pop, probs = c(0, 0.1, 0.4, 0.7, 0.8, 0.9, 1)), digits = -3),
    cex.axis = 0.6)
axis(side = 2,
    at = log(quantile(stl$hom, probs = seq(0,1,0.1))),
    labels = round(quantile(stl$hom, probs = seq(0,1,0.1))),
    cex.axis = 0.6)
```
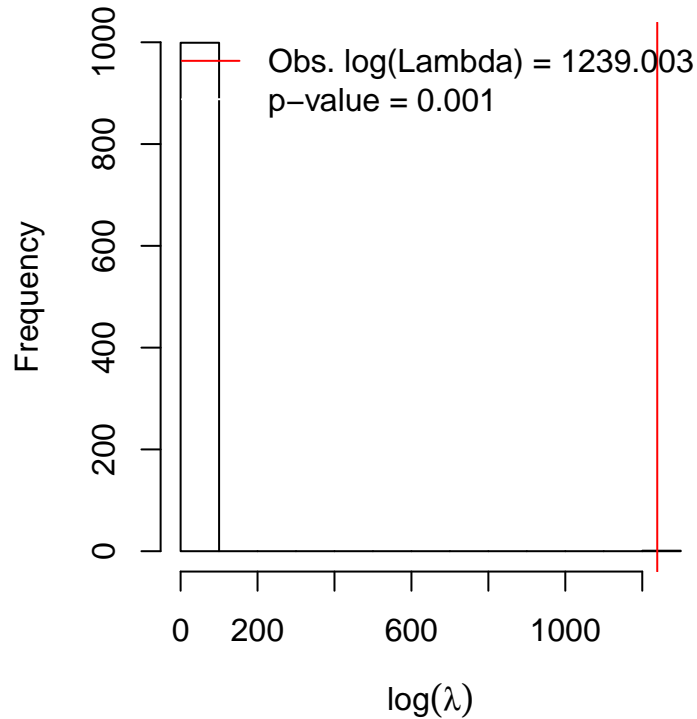
Population and homicides (log–log)
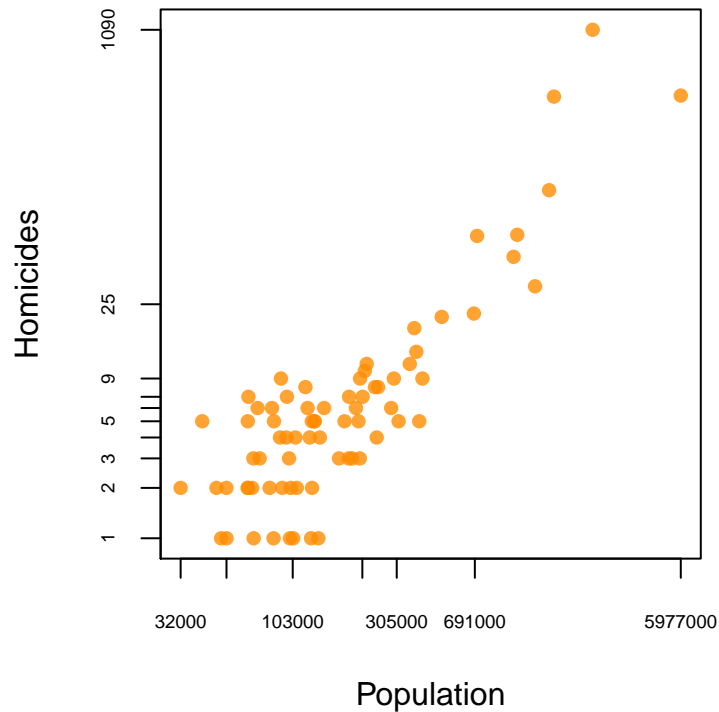

```
######
# Use SpatialEpi's kulldorff function for clustering detection
# (spatial scan)
######
#my_geo <- matrix(coordinates(stl), ncol = 2)
my_geo <- latlong2grid(coordinates(stl))
my_cases <- stl$hom
my_population <- stl$pop
n_strata <- nrow(stl)
expected_cases <- (stl$pop / sum(stl$pop)) * sum(stl$hom) #expected(my_population, my_cases, n_strata)

## Kulldorff using Poisson likelihoods
k_poisson <- kulldorff(geo = my_geo,
                       cases = my_cases,
                       population = my_population,
                       expected.cases = expected_cases,
                       pop.upper.bound = 0.5,
                       n.simulations = 999,
                       alpha.level = 0.05,
                       plot = TRUE)
```

Monte Carlo Distribution of Lambda

Obs. log(Lambda) = 1239.003
p−value = 0.001

Frequency (y-axis): 0, 200, 400, 600, 800, 1000

log(λ) (x-axis): 0, 200, 600, 1000

```r
cluster <- k_poisson$most.likely.cluster$location.IDs.included

## plot
plot(stl, col = "grey", border = "white")
plot(stl[cluster,],add=TRUE,col=adjustcolor("darkred", alpha.f = 0.6))
title(main = "Most Likely Cluster after 999 Monte-Carlo simulations\n(no secondary clusters identified)"
```

Most Likely Cluster after 999 Monte−Carlo simulations
(no secondary clusters identified)
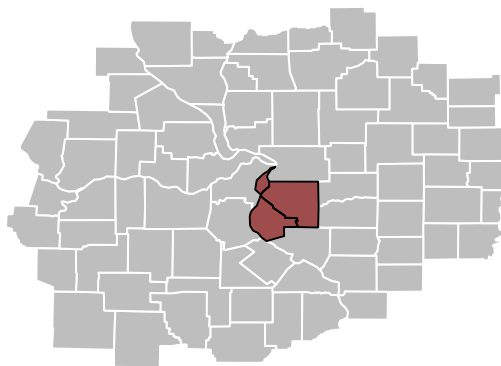
```
######
# PRINT INFORMATION
######
summary(k_poisson)
```

```
##                     Length Class  Mode
## most.likely.cluster    8   -none- list
## secondary.clusters     0   -none- NULL
## type                   1   -none- character
## log.lkhd            2006   -none- numeric
## simulated.log.lkhd   999   -none- numeric
```

```
k_poisson$most.likely.cluster
```

```
## $location.IDs.included
## [1] 46 49 40
##
## $population
## [1] 4093508
##
## $number.of.cases
## [1] 1529
##
## $expected.cases
## [1] 409.2167
##
## $SMR
## [1] 3.736406
##
## $log.likelihood.ratio
## [1] 1239.003
##
## $monte.carlo.rank
## [1] 1
##
## $p.value
## [1] 0.001
```

```
#Cluster:
as.character(stl$NAME[cluster])
```

```
## [1] "St. Clair"      "Monroe"          "St. Louis City"
```