GEOGRAPHY AND REGIONAL DEVELOPMENT 657

Spring 2006

DATA ANALYSIS PROJECTS

1. Porosity Data, all of these files are in simple ASCII format (can be opened in NOTEPAD). They can be *imported* into MINITAB, SPSS or R.

The data set
www.ic.arizona.edu/ic/geog657/data/poros1g.dat

represents measurements of Porosity for 1 x1 mm cells (18 x 18 grid). This is raster data. There are three columns in the data file, the first two columns are coordinates and the third column is the porosity values. The coordinates are for the centers of the 1 x 1 cells

www.ic.arizona.edu/ic/geog657/data/poros2g.dat
was obtained from poros1g.dat by aggregation, i.e. starting from the lower left hand corner the four cell values in the 2 x 2 square were averaged. Similarly then for successive 2 x 2 (non-overlapping) squares. The new coordinates are for the centers of the 2 x 2 cells. (9 x 9 grid)

www.ic.arizona.edu/ic/geog657/data/poros3g.dat
was obtained from poros1g.dat by aggregation, i.e. starting from the lower left hand corner the nine cell values in the 3 x 3 square were averaged. Similarly then for successive 3 x 3 (non-overlapping) squares. The new coordinates are for the centers of the 3 x 3 cells. (6 x 6 grid)

www.ic.arizona.edu/ic/geog657/data/poros6g.dat
was obtained from poros1g.dat by aggregation, i.e. starting from the lower left hand corner the 36 cell values in the 6 x 6 square were averaged. Similarly then for successive 6 x 6 (non-overlapping) squares. The new coordinates are for the centers of the 6 x 6 cells. (3 x 3 grid)

Note that poros6g.dat could have been obtained by aggregation on poros2g.dat or on poros3g.dat

First Computer Assignment Due 7 February
(a) Import the four data files into an appropriate program (e.g., MINITAB, SPSS or R).

(b) For each data set construct and plot a histogram of the porosity values, compute the sample mean and sample variance. Also Incorporate boxplots and stem-and-leaf plots.

Histograms in R
http://www.maths.lth.se/help/R/.R/library/graphics/html/hist.html

Stem and Leaf plots in R using the stem function in the graphics package
http://www.maths.lth.se/help/R/.R/library/graphics/html/stem.html

Index of the graphics package in R
http://www.maths.lth.se/help/R/.R/library/graphics/html/00Index.htm

Boxplots in R using the boxplot function in the graphics package
http://www.maths.lth.se/help/R/.R/library/graphics/html/boxplot.html

(c) For each data set construct a plot of the cell centers and "code" the plotted points according to the associated porosity values. "Coding" can either use colors, different symbols or different sized symbols. Alternatively construct a contour plot for porosity.

See filled.contour function in the *graphics* package in R
http://www.maths.lth.se/help/R/.R/library/graphics/html/filled.contour.html

For the use of *filled.contour*() for these data sets see
http://www.ic.arizona.edu/ic/geog657/data/filled_contour-example.pdf

(d) For parts (a), (b), (c) construct/copy the sequence of steps/instructions used for these computations as well as the plots and results into a word-processor document. *Shft-Print Scrn* should copy a window to the Clipboard and you can then paste it into another program such as MS Word. *This is what you will hand in.* Neither hand-written nor "typed" versions are acceptable.

REFERENCE CARD for R
http://cran.r-project.org/doc/contrib/Short-refcard.pdf

Second Computer Assignment
MULTIVARIATE GRAPHICS
DUE the Thursday following spring break, 24 March

You may use any or all of the following software packages: MINITAB, R, SPSS, ggobi. Prepare a report for your results
A. Be sure you identify the software used on each part
B. Copy the sequence of steps/instructions used for these computations as well as the plots into the report

For one of the following versions of a crime data set
http://www.ic.arizona.edu/ic/geog657/data/2003crimedata.txt
http://www.ic.arizona.edu/ic/geog657/data/crimedata.csv

The first is a tab-delimited file, the second is a comma-delimited file

generate the following plots

a. Scatter Plot Matrix
b. Parallel Coordinate Plot

In addition generate the

c. correlation matrix
then apply Principal Components Analysis (using either the correlation matrix or the covariance matrix). Your results for PCA should include at a minimum; the list of eigenvalues, eigenvectors, percent variance explained for each, biplot for first two Principal Components

Finally how does the information presented the plots and the PCA analysis differ or relate?

Third Computer Assignment
Assignment due 2 May

NOTE: The choices have been modified 12 April 2006

As with prior assignments, be sure to identify the software used, copy screen images or histories to show the steps you followed for the computations.

Do **ONE** of the following four choices

A. Using the data *nztrees* (see the *spatstat* package in R), do the analysis two ways

1. With the full data set
2. With a 5 ft strip trimmed off on the right hand side (See the data description)

(i) Plot the events
(ii) Find the nearest neighbor distance for each event
(iii) Compute and plot the cumulative distribution function for the nearest neighbor dist.
(iv) Compute and plot the kernel smoothed intensity function
(v) Compute and plot the estimated K function
(vi) Generate and plot a uniform random point pattern with 62 points

You obviously can use the *spatstat* package in R but you may also export the data to other software such as Crimestat. You might also use some other package in R

Compare your results when using the full data set with those from the trimmed data set
****
B. Using the data set *south* (found in GeoDa)

Do the analysis twice, once with a spatial lag model and once with a spatial error model

Apply where the response variable is crime. Compare the results

In GeoDa see page 221 of the GeoDa workbook. Reproduce the results shown there

You may also export the data and use R

In R see the functions *lagsarlm* and *errorsarlm*

****

C. Using the data set ***columbus*** (found both in the *maptools* and *spdep* packages in R and in GeoDa)

You may use either R or GeoDa
(I.e., you do not need to use both)

You may use MINITAB or R or GeoDa for the global regression

Apply global regression with crime as the response variable

Apply a spatial lag model with crime as the response variable
See *lagsarlm* in R

Compare the results

****

D. Using the data set ***columbus*** (found both in the *maptools* and *spdep* packages in R)

Apply geographically weighted regression with crime as the response variable

See the **spgwr** function in R

Compare the results with those from global regression.