

Ifgi – Muenster – Fall School 2006

Spatial data analysis course

Exam 1

Data Description

The data used in this exam is the infant mortality rate in Rio de Janeiro in 1994, contained in the file “rio_inf_mort.zip”. This file contains the following attributes:

SPRAREA: polygon area

SPRPERIMET: polygon perimeter

TOTBIRTH: total birth numbers

DEATHNEO: total deaths in the first 24 hours

DEATPOST: total deaths in one year of birth

TOT_RATE: total mortality rate

NEO_RATE: mortality rate of newborn babies

POST_RATE: mortality rate within one year of birth

APGAR: quality index of baby

LOWWEIGHT: percentage of low weight at birth

CESAREA: proportion of cesarean births

ADOLESC: proportion of adolescent mothers

ILLITER: proportion of illiterate mothers

SEC SCHOOL: proportion of mothers with more than 8 years of schooling

PROP_SLUMS: proportion of slums in the district

OBJET_ID_1: object id

Data preparation

Open the data in GeoDA, using the OBJET_ID_1 attribute as the index. Create weights using a queen neighborhood. Check the column names and explore the data.

Exam Questions

1. Plot the mortality rates of newborn babies (NEO_RATE) against the total birth numbers (TOTBIRTH) using the “ScatterPlot” function. Also show the mortality rate of newborn babies (NEO_RATE) as a map. What type of graphic do you see? Can you explain what is happening? Which procedure can you use to correct this effect? If needed, apply the correction procedure and analyze the resulting maps and scatter plots.

2. Consider the following variables:

NEO_RATE: mortality rates of newborn babies

APGAR: quality index of baby

LOWWEIGHT: percentage of low weight at birth

CESAREA: proportion of cesarean births

ADOLESC: proportion of adolescent mothers

SECSCHOOL: proportion of mothers with more than 8 years of schooling

Which of these variables are spatially autocorrelated? Are your measures of autocorrelation statistically significant? Are the clusters of high and low values of any of these variables? Can you plot these clusters?

3. Consider the relation between the mortality rates of newborn babies (NEO_RATE) compared to the variation of variables such as: quality index of baby (APGAR), percentage of low weight at birth (LOWWEIGHT), and proportion of adolescent mothers (ADOLESC).

Can you propose a nonspatial model for this relation? Can you propose a spatial regression model? Which model has a better result considering the R^2 value? Can you explain your results?

Bonus question: can you propose a better model?