

# Master of Science in Economics

# ECON6027 Spatial Econometrics and Data Analysis

9 November 2023

Final Examination

Instructor: Dr. Shew Fan Liu

Student’s Name:  **\_** (in full)

Student’s ID Number: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# READ ALL INSTRUCTIONS CAREFULLY

1. For each question, please provide separate submissions of the following:
   1. Html outputs
   2. R script file or RMD file
   3. Any other files that you create (eg: .txt, .GAL, etc.)
   4. File name to follow the question number (eg: “Q1.html”, “Q1.Rscript”, etc.)
2. In case you have trouble compiling, copy+paste the outputs into a word document and prepare for submission.
3. Label the answers to the sub-questions clearly.
4. This paper consists of 5 printed pages including this cover page.
5. The examination has a 40% component in the course assessment. Please answer **all** questions. The marks for each part are indicated in brackets.
6. You have **three hours** to complete the paper.

Question 1A: Areal Data (20 marks)

Shared with you are two files:

1. “burglary.Rdata” gives Burglary data for Sheffield: The variables are:
   1. WARD\_NO ward number
   2. WARD\_NM ward name
   3. WARD\_ID census ID
   4. HH number of households
   5. BURG number of burglaries in 1995
   6. TDI Townsend deprivation index (TDI)
   7. TO population turnover (percentage residents having moved within one year of the 1991 census)
2. “wards\_err.GAL” a GAL file to create Sheffield ward neighbours.

A ward is a local authority area, typically used for electoral purposes.

Material deprivation refers to the inability for individuals or households to afford those consumption goods and activities that are typical in a society at a given point in time, irrespective of people's preferences with respect to these items. Material deprivation is measures using the Townsend Deprivation Index whereby a large positive implies deprived and a large negative implies affluent.

<https://en.wikipedia.org/wiki/Townsend_deprivation_index>

Answer the following questions:

1. (5 marks) Load the GAL file and fix the error. Create a weights list object using the (i) row normalised weights style and (ii) binary weights style.
2. (10 marks) Load the burglary dataset and create a new variable (give a summary): **number of burglaries per 1000 households.** 
   1. Conduct a Moran’s test on this variable and comment on the results.
   2. What is the null hypothesis of the above test. Show the calculation to get the null Moran’s statistic.
   3. Generate local Moran’s I statistics for each ward and show the calculation of the global Moran’s I statistic as the sum of local Moran’s I statistics.
   4. How many spatial outliers are there for the variable burglaries per 1000 households?
3. (5 marks) Create a binary variable based on TDI where positive means deprived (False), and negative means not deprived (True). Give a table for the new variable. Conduct a join count analysis on this new variable. State your conclusions clearly.

Question 1B: Spatial Regression (20 marks)

Shared with you are two files:

1. “italy.shapefile”, a shapefile consisting geometries of the Italian regions along with the following attributes,
   1. REGION name of the Italian region
   2. VAR\_UNEMP variation in unemployment
   3. VAR\_RGDP variation in real GDP
2. “it.nb”, a GAL file to generate queen contiguity neighbours of the Italian regions.

Answer the following questions:

1. (5 marks) Load the italy.shapefile and regress **VAR\_UNEMP** on **VAR\_RGDP.** *(The inverse relationship between the variation in unemployment and variation in real GDP is known as Okun’s law in the economics literature.)*
   1. Show the summary of the regression.
   2. Give a scatterplot of the variables along with the fitted line.
   3. Check for homoskedasticity and normality of the regression errors.
2. (5 marks) Load the GAL file and create a neighbourhood connectivity plot on a map of Italy. Conduct Moran’s tests on the OLS residuals and the model variables to check whether spatial spill-over effects need to be considered in the model.
3. (10 marks) Run the usual spatial regressions and conduct appropriate **tests** to identify the most suitable spatial regression model.

Question 2: Geospatial Data (30 marks)

Topo.Rdata provides a data frame with 52 rows and 3 columns, of topographic **height in feet** within a 310 feet square. The objective of this question is to recreate the topology using geospatial interpolation techniques.

1. Give spatial awareness to the dataset. Give a basic plot of your dataset (3 marks)
2. Re-create the Voronoi Tesselation given in Figure 3.1. (5 marks)
3. Create a *square* grid of 10,000 unsampled points for the purpose of interpolation. Give a plot of this grid points. (3 marks)
4. Reconstruct the topology using the inverse distance weighting method where the “idp=1” and “idp=3”. Which idp is more appropriate? (5 marks)
5. Give an ordinary kriging (OK) estimate of the topology by selecting the most suitable parametric model of the semi-variogram. When selecting the most suitable parametric model, consider Stein, Gaussian, Bessel and linear fits and choose the best fit. Your output must include, (10 marks)
   1. Model selected with the fitted semi-variogram plot.
   2. Summary of the OK estimates.
   3. Plot of the reconstructed topology (breaks you use for this plot must be comparable to the Voronoi Tessellation).
6. Give a plot of the three different reconstructed maps: Voronoi tessellation, best IDW estimate, and OK estimate.  (2 mark)
7. For the kind of data given in this question, which method of interpolation is most suitable and why? (2 marks)

A picture containing chart

Description automatically generated

**Figure 3.1**

Question 3: Point Data (30 marks)

Shared with you is the famous “ruspini” dataset.

You are required to prepare a report conducting a comprehensive point pattern analysis of the ruspini clusters.

Label your outputs clearly and remember to give your interim conclusion at each stage and your overall conclusion. Your analysis can include various **calculations**, **tests** and **plots** to support your conclusions.

Remember to conduct suitable edge corrections and alternative competing analysis in each section. Since clustering is clearly present, one tailed tests are preferred over two tailed tests.