3600 Morrissey Hall Access Code: 1553

Course philosophy

* The result of using the scientific method is NOT absolute truth.
* More interested in your position and opinion rather than just the correct answer.
* Quality of work is more important that arbitrary due dates.
* Course synthesizes statistics and demography.

Trends

* A lot of traditional social science assumes that space doesn’t matter.
* Spatial demography is becoming computational.

Spatial demography

* Creating maps is a prerequisite for developing the spatial model.

Statistics concepts

* Goal is to create dependent variables that are interval or ratio whenever possible.
  + Allows use of the most sophisticated spatial statistics (i.e., high statistical power).
* Distributions
  + Methods are based on normal distribution
  + Real world data is rarely normally distributed
  + Income distribution is an example of positively skewed data (right skewness)
  + High school education is an example of negatively skewed data (left skewness)
* Ideal situation
  + Normal distribution
  + Small standard deviation
  + 68-95-99 rule
* Errors
  + Type I error is error of commission
  + Type II error is error of omission
  + Type II error is preferred to Type I error
  + Type I errors lead to retractions of journal articles
  + U.S. legal system
    - H0: person is NOT guilty (not the same as innocent)
      * Evidence is insufficient to reject the null hypothesis
    - HA: person is guilty
* Generally, cannot achieve a BLUE model.

Demography concepts

* The scientific study of human populations
* Elements of demography
  + Mathematical knowledge of populations
  + General movement of populations
  + The physical, civil, intellectual, and moral state of populations

The Importance of Space

* Y indicates dependent variable
* X indicates independent variable
* The goal is integrating space as part of the model
* More interdisciplinary research
* Many disciplines are afraid of integration of social sciences

Software and Infrastructure

* ArcMap will be going away; will be replaced by ArcPro and ArcGIS online
* QGIS is open source

Spatial Perspectives

* Functional distance
  + Time and effort required to move from point A to point B
  + Often must create additional data
* Spatial position
  + Comparative advantage and disadvantage of one space to another space
  + e.g., location of new MLS stadium
    - St. Charles residents will be the primary users
    - St. Louis City residents will absorb the costs
* Spatial order
  + Spatial hierarchy within a region
  + Relevant to class and stratification

Spatial Data

* Geodatabases are easier to use than shapefiles
* Quality of shapefile is important
* Can produce errors in spatial analysis
* Dr. Sandoval recommends obtaining shapefiles from U.S. Census

Data Sources

* U.S. Census ([www.census.gov](http://www.census.gov))
* Social Explorer ([www.socialexplorer.com](http://www.socialexplorer.com))

Homework

* Choose an MSA region other than St. Louis MSA or choose variables for St. Louis MSA that are different from the lab.
* Dependent variable should be interval/ratio or dichotomous.

Lab

* Download data from SocialExplorer.com
  + Treat the category Hispanic or Latino as a race
  + Download for files STATA
    - Tab delimited data
    - STATA .dct file (i.e., dictionary for how to read text files)
    - STATA .do file (i.e., program)
    - Data dictionary .txt file
  + Download for RStudio
    - Comma separated data .csv file
    - Data dictionary .txt file
* Creating ratio variables using STATA
  + Double click .do file (i.e., R12432906) to open it
  + \* used to add comments
  + Add folder file path in front of .txt filename in code
  + Add code before infile line code
  + Add line code capture log close
  + Add line code set more off
  + Add line code clear
  + Add line code log using analysis.log, replace after infile line code
  + Copy code for calculating ratio variables from lab .do file on Blackboard
  + Paste code for calculating ratio variable into .do file from data download
  + Change folder file path in
    - line code for exporting to Excel file
    - line code for creating new STATA database
  + Run program
    - Control-A to highlight the program
    - Click Execute(do) menu button
  + Creates new files in lab01/data folder
    - part01.xls
    - stl\_part01.dta
* Download base shapefiles
  + Download shapefiles from U.S. Census using TIGER/line shapefiles database
    - 2019 ACS 5-year estimate
  + Export ZIP files into lab01/data subfolders named for each shapefile
    - Illinois Census tract 17 (tl\_2019\_17\_tract)
    - Missouri Census tract 29 (tl\_2019\_29\_tract)
    - Entire USA (tl\_2019\_us\_cbsa)
* Create new shapefile in ArcGIS
  + Open new map in ArcMap
  + Add each shapefile as a layer
    - tl\_2019\_17\_tract (i.e., Illinois)
    - tl\_2019\_29\_tract (i.e., Missouri)
    - tl\_2019\_us\_cbsa (i.e., entire USA)
  + Create new file geodatabase in lab01 folder
    - Right Click on folder 🡪 New 🡪 File Geodatabase
    - Rename geodatabase (e.g., lab01)
  + Merge Illinois and Missouri shapefiles
    - Select layers for Illinois and Missouri to activate them
    - Deactivate layer for Entire USA
    - Geoprocessing 🡪 Merge
    - Drag and drop Illinois and Missouri layers
    - Set output file path to newly created file geodatabase
    - Enter name for new shapefile (e.g., states01)
      * Use feature class file type
  + Select Entire USA shapefile layer to activate it
    - Deactivate other layers
  + Export St. Louis MSA shapefile
    - Select tract
    - Right click on layer 🡪 Data 🡪 Export Data
    - Export to file geodatabase
    - Name shapefile (e.g., stl\_msa)
      * Use file and personal geodatabase feature class file type
  + Clip census track shapefile to St. Louis MSA shapefile
    - Select stl\_msa shapefile layer
    - Select states01 shapefile layer with census tracts
    - Geoprocessing 🡪 Clip
    - Input feature is states01
    - Clip feature is stl\_msa
    - Output to file geodatabase
    - Name shapefile (e.g., stl\_msa\_ct\_00)
      * Use feature class file type
      * GEOID is unique to each case
  + Merge ratio variables data created using STATA into shapefile for St. Louis MSA
    - Select St. Louis MSA census tracts shapefile layer (stl\_msa\_ct\_00) to activate it
      * Deactivate all other layers
    - Add ratio variables data as a new layer
      * Click on Add Data button
      * Select part01.xls file
      * Select Sheet1$
    - Right click on St. Louis MSA census tracts shapefile layer (stl\_msa\_ct\_00)
      * Join and Relates 🡪 Join
      * Choose GEOID as the basis for the shapefile layer
      * Choose Sheet1$ as the table
      * Use FIPS as the basis for the table
  + Convert to permanent shapefile
    - Right click on St. Louis MSA census tracts shapefile layer (stl\_msa\_ct\_00)
    - Data 🡪 Export Data
    - Output to file geodatabase
    - Name shapefile (stl\_msa\_ct\_01)
    - Select type “File and Personal Geodatabase feature classes”
    - Click OK

ArcMap

* Fragile software; has integrity issues.
* Restart the computer if the software becomes unstable and starts crashing a lot.

Education Attainment

* The amount of education attainment of a region.
* Flexibility with how trade school education is handled.
* Individuals still seeking education are not counted in the index.

Income Inequality

* Can apply concept to other topics such as education.

Gini Concentration Ration

* Measure area of inequality in Lorenz curve.
* Calculated as part of Social Explorer data.

Theil Index

* More powerful than the Gini Concentration Ratio
  + Better mathematical properties.
  + Additive across different subgroups or regions.
  + Local scores add up to global score.
* Not used very often.
* Difficult to interpret.
* Part of General Entropy class which is a family of measures.

Creating Index

* Rescale individual components (i.e., variables) on 0 to 1 scale.
* Default is that each variable gets the same weight.

Class Project

* Creating index for project is optional.

Lab

* NOTE: Create data dictionary for newly generated shapefiles.
* Creating Excel file (part02.xls) with additional demographic data
  + Download folder lab02
  + Double click the downloaded .do file with the downloaded tract level demographic data (i.e., R124384020) in STATA to open it.
  + Change the file path in front of the .txt filename in the infile line code (line 11).
  + Change the file path in front of the .xls filename in the export function line code (line 152).
  + Change the file path in front of the .dta filename in the save function line code (line 154).
  + Run program
    - Control-A to highlight the program
    - Click the Execute(do) menu button
* Add additional demographic data to shapefile stl\_msa\_ct\_01
  + Open stl\_msa\_ct\_01 shapefile in ArcMap and select it as an active layer.
    - Deactivate all other layers.
  + Click the Add Data button.
  + Add Sheet$1 of part02.xls as a layer and select it as an active layer.
  + Join the Excel data to the shapefile.
    - Highlight and right click on stl\_msa\_ct\_01
    - Select Join and Relates
    - Select Join
    - Select GEOID as the shapefile field on which to base the join.
    - Select Sheet$1 as the data layer for joining.
    - Select FIPS as the data field on which to base the join.
    - Click on Validate Join to check (optional).
    - Click OK
  + Save as new permanent shapefile to the file geodatabase.
    - Highlight and right click on stl\_msa\_ct\_01
    - Select Data
    - Select Export Data
    - Select the file geodatabase (lab01.gdb) as the output location.
    - Name the shapefile stl\_msa\_ct\_02
    - Select type “File and Personal Geodatabase feature classes”
    - Click OK
* Remove census tracts that will skew results.
  + Add stl\_msa\_ct\_02 as a new layer in ArcMap.
    - Deactivate all other layers.
  + Add a new variable named Keep in the attributes table.
    - Right click on the layer.
    - Select Open Attributes
    - Select Add Field from the drop down menu
    - Name the field “Keep”
  + Set default value for the Keep variable
    - Highlight the Keep variable column in the attributes table.
    - Right click and select Field Calculator
    - Enter “1” in the Keep = formula box (where 1 means “yes”)
    - Click OK
  + Identify census tracts to eliminate and remove them
    - Activate the shapefile layer stl\_msa\_ct\_02
    - Select Customize drop down menu
    - Select Toolbars
    - Select Editor
    - Select Start Editing
    - Select the census tracts to eliminate with the selector
    - Right click on the census tract
    - Select Delete Polygon
    - Click Save
    - Select Stop Editing
  + Save the new shapefile to the gis folder.
    - Highlight and right click on modified stl\_msa\_ct\_02 layer
    - Select Data
    - Select Export Data
    - Select the Labs/data/gis folder as the output location.
    - Name the shapefile stl\_ct
    - Select type “Shapefile”
    - Click OK
* Create indexes for statewide data
  + Open gis.do file from the Labs/data/gis folder in STATA
  + Change the file path in line code 3 to the Labs/data/gis folder
  + Run program
    - Control-A to highlight the program
    - Click the Execute(do) menu button
  + Open merge\_final.do file from the Labs/data/lab02 folder in STATA
  + Change the file path in line code 4, 8, 10, and 14
  + Run program lines 1-31
    - Highlight the program code
    - Click the Execute(do) menu button
  + Insert the table from the output after line code 31
  + Replace the minimum and maximum values for the Theil index in the merge\_final.do file
  + Change the file path in the remainder of the program
    - Line code 66
    - Line code 68
      * NOTE: There appears to be a type on the file name
      * “msa\_stl.dta” instead of “mas\_stl.dta”
  + Run the remainder of the program
    - Highlight the program code
    - Click the Execute(do) menu button
* Add indexes to modified shapefile
  + Open stl\_ct shapefile in ArcMap and select it as an active layer.
    - Deactivate all other layers.
  + Click the Add Data button.
  + Add Sheet$1 of final.xls as a layer and select it as an active layer.
  + Join the Excel data to the shapefile.
    - Highlight and right click on stl\_ct
    - Select Join and Relates
    - Select Join
    - Select GEOID as the shapefile field on which to base the join.
    - Select Sheet$1 of final.xls as the data layer for joining.
    - Select FIPS as the data field on which to base the join.
    - Click on Validate Join to check (optional).
    - Click OK
  + Save the new shapefile to the file geodatabase.
    - Highlight and right click on stl\_ct
    - Select Data
    - Select Export Data
    - Select the file geodatabase (lab01.gdb) as the output location.
    - Name the shapefile stl\_msa\_ct\_03
    - Select type “File and Personal Geodatabase feature classes”
    - Click OK
* Add projection to the shapefile data
  + Open stl\_msa\_ct\_03 shapefile in ArcMap and select it as an active layer.
    - Deactivate all other layers.
  + Click on the Toolbox menu icon
  + Select Data Management Tools
  + Select Projections and Transformations
  + Select Define Projections
  + Select stl\_msa\_ct\_03 shapefile as the input
  + Select UTM-15 as the coordinate system.
  + Click OK
  + Go to layout view
  + Right click and select Properties
  + Select Coordinate System tab
  + Change projection coordinate system to NAD 1983 UTM-15
  + Save changes
* Save as a new permanent shapefile to the file geodatabase.
  + Highlight and right click on stl\_msa\_ct\_03
  + Select Data
  + Select Export Data
  + Select the file geodatabase (lab01.gdb) as the output location.
  + Name the shapefile stl\_msa\_ct\_04
  + Select type “File and Personal Geodatabase feature classes”
  + Click OK

Homework

* Select city with at least 250,000 people or MSA with at least 500,000 people.
* Only include census tracks with at least 100 people.
  + All other variables should be okay.
* Additional social data available at PolicyMap
  + Access through SLU Pius XII Memorial Library under databases

Lecture on Geographic Distribution

* Technique of exploratory spatial analysis.
* Spatial variation
  + Non-spatial models may not be valid.
  + Models that apply in one locality are unlikely to be applicable in another locality.
* Why geographic distribution is examined.
  + Identify the center
  + Determine shape and orientation of data
  + Understand the dispersion of the features
* Three Kinds of Center
  + Mean 🡪 average x-y coordinate for all features
  + Median 🡪 x-y coordinate with shortest distance to all features
  + Central Feature 🡪 feature that has the shortest total distance to all other features
* Weighted mean center
  + Not every case should be treated the same
    - land value
    - distinguishing between degrees of poverty
    - different types of crime
  + Sensitive to outliers
* Median center
  + No single equation to calculate median center
  + Approximated by software through iterative calculation
  + Tends to gravitate toward areas with the most features
* Central feature
  + Sum the distance to all other features for each feature.
  + Select the feature with the shortest distance.
* Measuring concentration and dispersion
  + Default is to use one standard deviation when the boundary is known.
  + Use three standard deviation when we don’t know the boundary.
  + Standard distance
  + Standard deviational ellipse (SDE)
* A central city is a heavily populated city at the center of a large metropolitan area [*Central City*. (n.d.). Oxford Dictionary. Retrieved February 6, 2020 from https://www.google.com/search?client=firefox-b-d&q=what+is+a+central+city].
  + Defined and designated by the U.S. Census bureau.
  + There are three central cities (cc) in the St. Louis region.
    - St. Charles, MO
    - City of St. Louis, MO
    - East St. Louis, IL

Lab

* Ensure that spatial analysis extensions in ArcMap are activated
  + Select Customize dropdown menu
  + Select Extensions
  + Select the extensions you want to activate
  + Click Close
* Create file geodatabase subfolder to organize data from exploratory spatial analysis (ESA)
  + Right click on file geodatabase (i.e., lab.gdb)
  + Select New
  + Select Feature Dataset
  + Name feature dataset (e.g., “esa”)
  + Click Next
  + Select the coordinate system (e.g., NAD 1983 UTM-15)
  + Click Next
  + Click Next to skip the z-coordinate settings
  + Modify XY tolerance settings if necessary (generally, should not be necessary)
  + Click Finish
* Access functions for measuring geographic distribution
  + Select Geoprocessing dropdown menu
  + Select ArcToolbox
  + Select Spatial Statistics Tools
  + Select Measuring Geographic Distributions
* Create shapefiles for mean center
  + Add and activate base layer (i.e., stl\_msa\_ct\_04) for new map in ArcMap
  + Select Geoprocessing dropdown menu
  + Select ArcToolbox
  + Select Spatial Statistics Tools
  + Select Measuring Geographic Distributions
  + Select Mean Center
  + Select stl\_msa\_ct\_04 as the input feature class
  + Select the esa subfolder in the file geodatabase (i.e., lab.gdb/esa) as the output feature class
  + Enter name for the new shapefile (e.g., mc\_wht)
  + Set type to Feature classes
  + Click Save
  + Select the variable upon which to base the calculation (e.g., wht) in Weight Field (optional)
  + Click OK
  + Repeat for other groups as necessary
  + Obtain XY coordinates from attribute table
* Create shapefiles for standard distance
  + Add and activate base layer (i.e., stl\_msa\_ct\_04) for new map in ArcMap
  + Select Geoprocessing dropdown menu
  + Select ArcToolbox
  + Select Spatial Statistics Tools
  + Select Measuring Geographic Distributions
  + Select Standard Distance
  + Select stl\_msa\_ct\_04 as the input feature class
  + Select the esa subfolder in the file geodatabase (i.e., lab.gdb/esa) as the output feature class
  + Enter name for the new shapefile (e.g., sd\_wht)
  + Set type to Feature classes
  + Click Save
  + Select desired standard deviation for Circle Size (e.g., 1\_Standard\_Deviation)
  + Select the variable upon which to base the calculation (e.g., wht) in Weight Field (optional)
  + Click OK
  + Repeat for other groups as necessary
  + Obtain area value from attribute table
  + Create new variable to convert shape area to desired units (optional)
    - Open attribute table
    - Click drop down menu
    - Select Add Field
    - Enter name of field (e.g., area-km)
    - Select the kind of value for Type (e.g., Double)
    - Click OK
    - Right click on the newly added field (variable)
    - Select Calculate Geometry…
    - Select the value to calculate in Property
    - Select the units to use in Units
    - Click OK
* Create shapefiles for standard deviational ellipses (i.e., standard distance taking into consideration geographic features such as rivers and lakes)
  + Add and activate base layer (i.e., stl\_msa\_ct\_04) for new map in ArcMap
  + Select Geoprocessing dropdown menu
  + Select ArcToolbox
  + Select Spatial Statistics Tools
  + Select Measuring Geographic Distributions
  + Select Directional Distribution (Standard Deviational Ellipse)
  + Select stl\_msa\_ct\_04 as the input feature class
  + Select the esa subfolder in the file geodatabase (i.e., lab.gdb/esa) as the output feature class
  + Enter name for the new shapefile (e.g., sde\_wht)
  + Set type to Feature classes
  + Click Save
  + Select desired standard deviation for Circle Size (e.g., 1\_Standard\_Deviation)
  + Select the variable upon which to base the calculation (e.g., wht) in Weight Field (optional)
  + Click OK
  + Repeat for other groups as necessary
  + Obtain area value and rotation from attribute table
    - Rotation is measured counterclockwise
  + Create new variable to convert shape area to desired units (optional)
    - Open attribute table
    - Click drop down menu
    - Select Add Field
    - Enter name of field (e.g., area-km)
    - Select the kind of value for Type (e.g., Double)
    - Click OK
    - Right click on the newly added field (variable)
    - Select Calculate Geometry…
    - Select the value to calculate under Property
    - Select the units to use under Units
    - Click OK

Announcements

* No class next Thursday, February 20, 2020
* Can replicate lab for a different city for Homework 01
* Poverty rate can be used as a dependent variable for Homework01 and Class Project

Spatial Interpolation with Grids

* Interpolation is an estimate.
* Administrative boundaries don’t match social boundaries.
* Administrative boundaries may change over time.
* Used to standardize variables.

Modifiable Areal Unit Problem

* Recommendation to no longer use ZIP code data.
  + Will likely result in a Type I error.
* Problems
  + Scale (i.e., aggregation)
  + Grouping (i.e., Zones)
  + Ecological fallacy
* Generally, the larger the spatial units the stronger the relationship among variables.
* Spatial units are arbitrarily defined; different definitions may introduce biases in the analysis.
* Ecological fallacy is that aggregated results are applied to individuals.
* Solutions
  + Normalize boundaries to a specific administration definition of spatial units (e.g., year 2000 census tract boundaries to year 2010 census tract boundaries).
  + Normalize boundaries to standard space (e.g., grid)

Making a Grid

* The smaller the spatial unit the more accurate the data.
* General rule is to use the most accurate data available
  + e.g., census block data are preferred to census tract data they contain the data for the variable of interest.