



Measuring spatial accessibility to healthcare: An applied example using IPUMS PMA data

Webinar Questions and Answers

The following are adaptations of questions received during the live webinar and their answers. If you have additional questions, please contact IPUMS User Support at ipums@umn.edu.

Data Access & Determinations

Which IPUMS data provide point-level locations of medical facilities?

IPUMS data does not provide point-level locations of medical facilities. The facility location data used in the example shown in this webinar were downloaded from Humanitarian Data Exchange (Maina et al, 2019).

How do gridded raster-based population data compare to other approaches for estimating population at fine-grained geographies (i.e., joining population data at the Census tract with tract boundaries)?

It depends on the setting for your study. Vector-based population data sets (i.e. census estimates joined to administrative unit polygons or centroids) can be available for very small areas of homogeneous size in urban areas of high-income countries. Limitations arise with vector-based data in low- and middle-income countries, especially those with study areas that cross the urban-rural continuum, because administrative unit sizes are based on population density. Administrative units tend to be smaller in urban areas due to higher population density and larger in rural areas due to lower population density. A nationwide study that includes urban and rural areas will have heterogeneous administrative units that introduce error in the form of the Modifiable Areal Unit Problem (MAUP). In these cases, gridded raster-based population datasets are ideal because the spatial unit representing population (grid cells) is homogeneous throughout the study area.

What does PPR stand for?

PPR stands for provider-to-population ratio.

In the 2SFCA, how are population points determined? Are they centroids of administrative boundaries, raster centroids, etc.?

It depends on the data used in the analysis. A vector-based approach will often use centroids of administrative boundaries (geographic or population-weighted), or sometimes geographic points representing cities and towns, to represent population. A raster-based approach that uses a gridded population dataset to represent population will use the grid cell centroids from the raster as population points.

What are the factors that influence facility weights? Is it preference, availability of services, etc.?

Facility weights represent the capacity of the facility to provide services. This can be measured as the number of beds, number of providers, or some other measure.

How do I access GPS point data for PMA?

Access to the displaced PMA cluster GPS points is distributed on a case by case basis from the data producers. Email our user support at ipums@umn.edu for guidance and more details.

Applications and Adaptations of the Model

The example from the webinar is in Kenya. What is required to adapt the model to other countries or regions?

The only adaptation needed is to update the spatial reference for your study area. The python script prompts users to update this as needed.

Raster data can be computationally intensive. Does the provided code scale to large areas (like nations or continents)? What are the computing demands when scaling up to large areas?

Yes, the provided code is scalable to large areas and can theoretically be used at the continent level, however it has only been tested at the country level. It uses a combination of batching, parallel processing, and try/except safeguards to prevent it from freezing up.

How should we interpret PPRs? What does it mean when they are greater than one?

Provider-to-population ratio (PPRs) is the number of providers (capacity) per population in demand of services. It is useful as a measure of supply and demand. Accessibility

score, or the sum of all PPRs within the catchment area of a population point, is a relative measure of accessibility. The code provided multiplies accessibility score by 10,000 for readability because it is a unitless measure.

What limitations, if any, should we consider when using the 2SFCA model to conduct weighted regression analyses?

It is important to account for the spatial nature of the data and consider how the accessibility measure was constructed when using the 2SFCA model in a regression. The 2SFCA model produces spatial data, meaning nearby locations may have similar accessibility values. This is known as spatial autocorrelation, and it violates the assumption that observations are independent in conventional regression. Researchers should test for spatial autocorrelation in their regression residuals and consider spatial regression methods when it is present (Monnet et al., 2021). It is also important to recognize that 2SFCA estimates depend on choices such as catchment size, distance-decay, and population dataset, so sensitivity analyses can help determine whether regression results are robust to those modeling choices (Delamater, 2013; Wan et al., 2012).

A lot of the cost-raster parts are also input data in gridded population datasets such as WorldPop. How might this influence catchment areas and populations? Could this under- or overestimate needs and populations serviced by healthcare posts?

This is a valid concern. The cost raster and gridded population datasets can use some of the same underlying geographic information, such as roads, land cover, or built-up areas. That could introduce some correlated error and potentially affect both the catchment areas and the estimated populations they contain. While it may not be endogeneity in the traditional statistical sense, it is a potential source of bias that should be considered.

How might you account for healthcare providers that are informal, traditional, or otherwise not included in the service provider data used in the model. In some contexts, these providers are widely used because of their proximity and low cost.

A model like this is limited by the data available. While it would be ideal to include informal providers, without location and capacity information, we have no way of measuring how their presence influences accessibility with a model.

What are some potential next steps for the validation exercise using IPUMS PMA data? Did you explore whether associations between your predictors and the outcome vary based on geographic location? Did you assess goodness of fit and check assumptions such as multicollinearity, influential observations, and residual spatial autocorrelation?

A sensitivity analysis such as goodness of fit or use of other accuracy assessment measures would be a great next step.