How can I create a model for a spatially autocorrelated outcome in R?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a model for a spatially autocorrelated outcome in R?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=161747

Creating a model for a spatially autocorrelated outcome in R involves utilizing statistical techniques to account for the inherent spatial relationships present in the data. This can be achieved through the use of specialized packages and functions, such as spatial regression or geostatistical methods, which take into consideration the spatial dependence of the outcome variable. By incorporating spatial autocorrelation into the model, a more accurate and robust analysis can be conducted, leading to more reliable and meaningful results.

How do I model a spatially autocorrelated outcome? | R FAQ

We often examine data with the aim of making predictions. Spatial data analysis is no exception. Given measurements of a variable at a set of points in a region, we might like to extrapolate to points in the region where the variable was not measured or, possibly, to points outside the region that we believe will behave similarly. We can base these predictions on our measured values alone by kriging or we can incorporate covariates and make predictions using a regression model.

In R, the lme linear mixed-effects regression command in the nlme R package allows the user to fit a regression model in

which the outcome and the expected errors are spatially autocorrelated. There are several different forms that the spatial autocorrelation can take and the most appropriate form for a given dataset can be assessed by looking at the shape of the variogram of the data and choosing from the options available.

We will again be using the thick dataset provided in the SAS documentation for proc variogram, which includes the measured thickness of coal seams at different coordinates (we have converted this to a .csv file for easy use in R). To this dataset, we have added a covariate called soil measuring the soil quality. We wish to predict thickness (thick) with soil quality (soil) in a regression model that incorporates the spatial autocorrelation of our data.

The code below installs and loads the nlme package and reads in the data we will use.

```
install.packages("nlme")
library(nlme)
spdata <- read.table("https://stats.idre.ucla.edu/stat/r/faq/thick.csv", header = T, sep = ",")
```

The lme command requires a grouping variable. Since we do not have a grouping variable in our data, we can create a dummy variable that has the same value for all 75 observations.

```
dummy <- rep(1, 75)
spdata <- cbind(spdata, dummy)
soil.model <- lme(fixed = thick ~ soil, data = spdata, random = ~ 1 | dummy, method = "ML")
summary(soil.model)
```

```
Linear mixed-effects model fit by maximum likelihood
Data: spdata
AIC BIC logLik
```

342.3182 351.5881 -167.1591

Random effects:
Formula: ~1 | dummy
(Intercept) Residual
StdDev: 4.826056e-05 2.247569

Fixed effects: thick ~ soil
Value Std.Error DF t-value p-value
(Intercept) 31.94203 3.1569891 73 10.117878 0.0000
soil 2.25521 0.8655887 73 2.605407 0.0111
Correlation:
(Intr)
soil -0.997

Standardized Within-Group Residuals:
Min Q1 Med Q3 Max
-2.68798974 -0.53279498 0.03896491 0.66007203 2.20612991

Number of Observations: 75
Number of Groups: 1

Next, we can run the same model with spatial correlation structures. Let's

assume that, based on following the steps shown in R FAQ:
How do I fit a variogram model to my spatial data in R using regression commands?, we determined that our outcome
thick appears to have a Guassian spatial correlation form. We can
specify such a structure with the correlation and corGaus options
for lme.

```
soil.gau <- update(soil.model, correlation = corGaus(1,
form = ~ east + north), method = "ML")
summary(soil.gau)

Linear mixed-effects model fit by maximum likelihood
Data: spdata
AIC BIC logLik
91.50733 103.0948 -40.75366

Random effects:
Formula: ~1 | dummy
(Intercept) Residual
```

StdDev: 8.810794e-05 2.088383

Correlation Structure: Gaussian spatial correlation
Formula: ~east + north | dummy
Parameter estimate(s):
range
20.43725
Fixed effects: thick ~ soil
Value Std.Error DF t-value p-value
(Intercept) 40.32797 0.5877681 73 68.61204 0.0000
soil 0.00348 0.0160363 73 0.21693 0.8289
Correlation:
(Intr)
soil -0.102

Standardized Within-Group Residuals:
Min Q1 Med Q3 Max
-2.9882532 -0.7133776 -0.1146245 0.6745696 2.0877393

Number of Observations: 75
Number of Groups: 1

In this example, incorporating the Gaussian correlation structure both
improved the model fit and changed the nature of the

regression model.
Without the spatial structure, soil is a statistically significant
predictor of thick. With the spatial structure, this relationship
becomes not significant. This suggests that after controlling for location
and the known correlation structure, soil does not add much new information.

See also

References