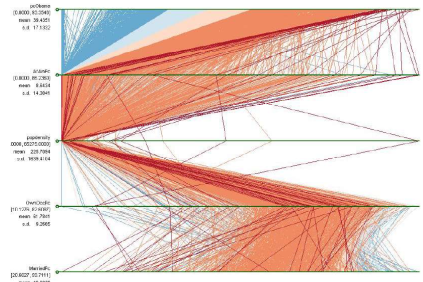
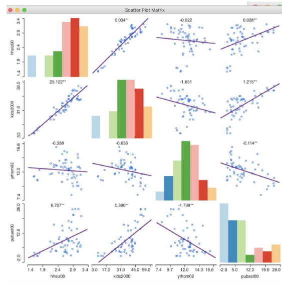
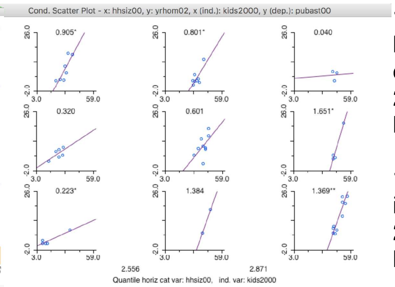
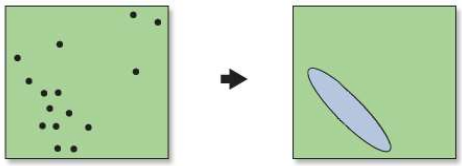
SPATIAL STATISTICS STUDY SHEET

MULTIVARIATE ESDA

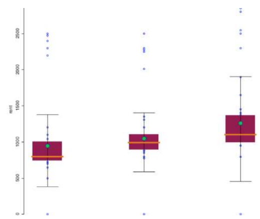




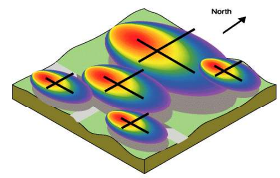
1st Order Effects: direct relationship between events and underlying distribution
2nd Order Effects: relationship between events (attraction/repulsion)

1st Order Stationarity: no variation in intensity/density over space
2nd Order Stationarity: no interaction between events



SD Ellipses: Assess spatial orientation of attribute distribution





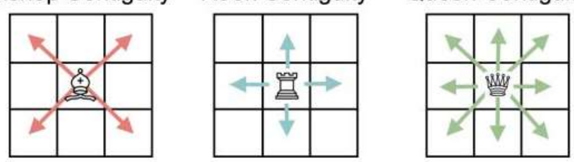
Anisotropy: directional influences on observed values. Similar to spatial trend, but not deterministic. Displays higher autocorrelation in one direction vs. others

AUTOCORRELATION

Positive Autocorrelation:
neighboring areas are alike
Negative Autocorrelation:
neighboring areas are different
No Autocorrelation:
random pattern b/w neighbors

Detected autocorrelation depends on definition of NEIGHBORS (# & direction)

Bishop Contiguity | Rook Contiguity | Queen Contiguity



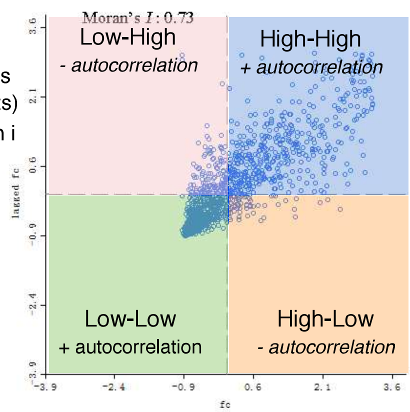
MORAN's I

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

N = total # of observations
W = sum of all w_{ij} (weights)
x_i = value of x at location i
$\bar{x}$ = mean of x
i & j are locations

H_0: random distribution
If p>0.05, cannot reject H_0
If p < 0.05 & +Z: Evidence for **clustering**
If p < 0.05 & - Z: Evidence for **repulsion**

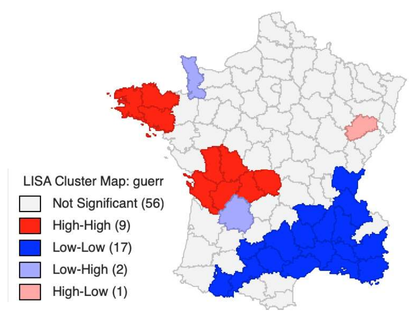
Note: in Geoda, z & p obtained via Randomization → 999 permutations



LOCAL INDICATORS OF SPATIAL AUTOCORRELATION (LISA)

All LISA measures identify locations surrounded by similar (high-high, low-low) or dissimilar (high-low, low-high) features.

Local Moran's I: feature of interest NOT included in neighborhood calculations
Geary's C & Getis-Ord G compare feature value to avg value of feature of interest + its neighbors
Note: interpretation of C is backwards: value <1 = +AC, high clustering



SPATIAL REGRESSION

Adjusted R^2: % of variation accounted for by IVs
p<0.05 indicates significance for predictor
Variance Inflation Factor (VIF) > 7.5:
Evidence of redundancy/collinearity
Multicollinearity Condition > 30:
Remove one or more correlated variables
Jarque Bera: if statistically significant, evidence of non-normality of errors & potentially missing predictor

Heteroskedasticity tests: significance indicates non-stationary DV, heterogeneity in predictive power
Breusch-Pagan: tests non-constant error variance
Koenker-Bassett: robust Breusch-Pagan
White: squares & cross-products of predictors

Better model fit evidenced by lower AIC, higher log-likelihood, and lower Schwarz

If DV is log-transformed:
"A 1 unit increase in x will lead to an $(e^{\beta}-1)*100$ % change in y"
If IV is log-transformed:
"A 1% change in x is associated with a $\beta/100$ unit change in y"
If both IV and DV are log-transformed:
"A 1% change in x is associated with a β% change in y"

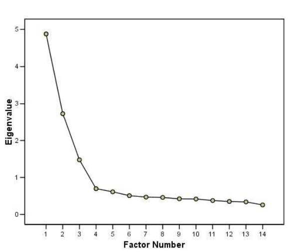
Spatial Error: Error terms assumed correlated due to unobserved neighborhood effects
Spatial Lag: Observations dependent on neighbors' outcome
Spillover/diffusion effects expected
For either specification, don't interpret coefficients or sig. for spatial weights

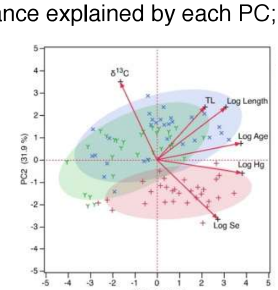
Lag vs. Error Decision Rule:
If neither LM-Error & LM-Lag are sig, keep OLS
If one of LM-Error or LM-Lag is sig, run sig model
If BOTH LM-Error & LM-Lag sig:
Compare Robust LM-Error & Robust LM-Lag
Run whichever model is sig
Can also run model with the largest LM value

Logistic Regression: binary outcome
Output is $ln(\frac{P_a}{1-P_a})$
Interpret coefficients as LOG-LIKELIHOOD
"Ceteris paribus, a 1 unit increase in x is associated with a 20% increase in LL of y"
Interpret odds ratios as LIKELIHOOD
"Ceteris paribus, a 1 unit increase in x is associated with a 3.5 times increase in y"
Autologistic Regression: includes function to describe surrounding values

SPATIAL PCA: reduces data redundancy by grouping data into fewest # of components

Each PC is linear combination of original attributes, contribution indicated by variable LOADINGS
Scree plot indicates % variance explained by each PC; Biplots can help assign meaning to PCs





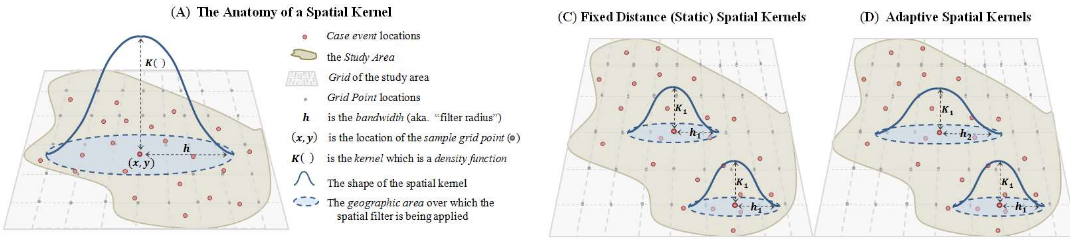
Vectors with small angles = + correlation
Vectors with 90deg angles = not correlated
Vectors with large angles = - correlation
High distance from PC origin =
High influence of variable on a PC

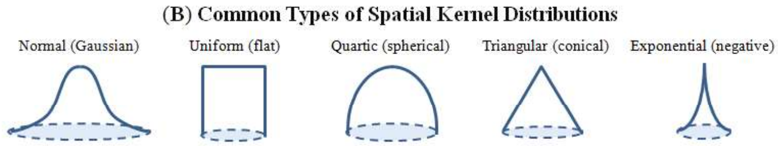
PCA requires external knowledge to group variables into meaningful indices/categories

Geographically Weighted Regression:
allows variation of coefficients over space using spatial kernel centered on observations
Choice of weighting function doesn't matter
Select bandwidth by CV or AIC

GWR is only worth it if it supplies lower residual sum of squares (RSS) than OLS!

KERNEL DENSITY: calculates feature density in neighborhood around feature

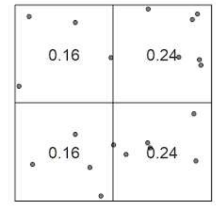




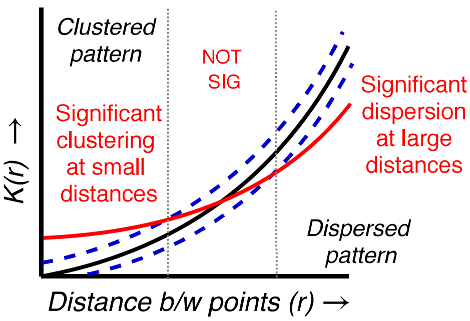
Choice of KD function not critical
Choice of BW and smoothing degree:
- Small BW: spiky results
- Large BW: loss of detail, high smoothing

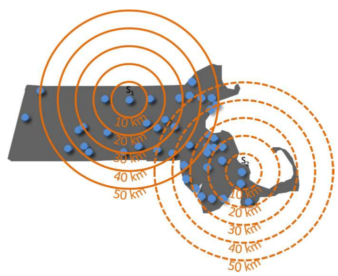
Can be selected using CV methods
High sensitivity to: outliers, edge effects (boundaries of study areas), quadrat size if imposing grid/fishnet

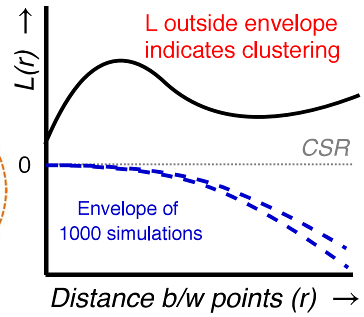
Quadrat Counts can help summarize counts
Variance to Mean Ratio (VMR)
If VMR>1, evidence for clustering since variability is larger than under CSR
Use Pearson's chi^2 test for sig. of differences

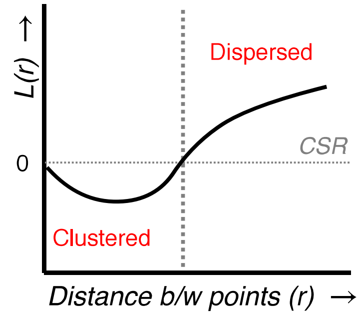


POINT PATTERN ANALYSIS: summarizes spatial dependence over a range of distances



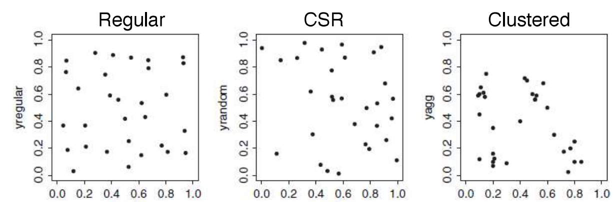




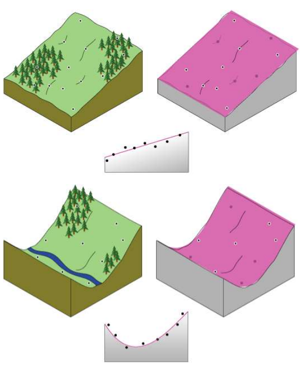


K Function
- Compares observed with expected (defined by Poisson process).
  If # of points within radius > # expected from random process, conclude there is clustering
- Use Monte Carlo simulations for CIs/envelopes
- Can be univariate or bivariate to identify attribute clustering
- Suspect at large distances due to edge effects

L Function: transformed K function where expected values are horizontal



INTERPOLATION

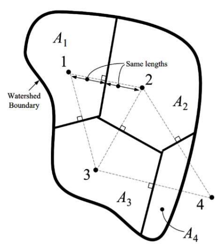


Trend Surfaces

Simple Regression
- Minimize SS residual distance
- Assumes:
  - independent x & y
  - z is normally distributed
  - Error independent of location

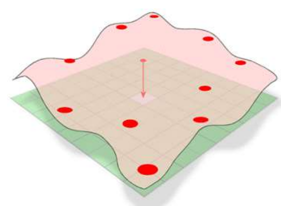
Polynomial Regression
- Goal: minimize deviations b/w sample points & surface
- Assumes:
  - Spatially dependent continuous z

Thiessen/Delauney



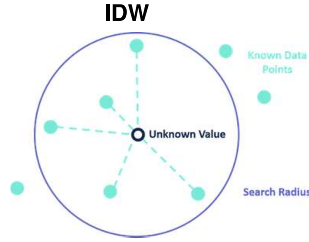
Created by connecting measurement points & drawing perpendicular bisectors; assigns value of nearest sample points

Splines



Bends surface to known values by repeatedly applying smoothing eqn (piecewise polynomial) to minimize 'energy' or 'curvature'

IDW

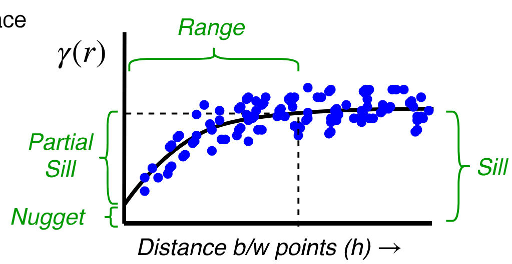


Measured values closest to unknown value have most influence on prediction
Weights based on power function, diminishes rapidly

Kriging: interpolates based on changes in variances over space patterns deduced via semivariogram, which shows avg differences between pairs of point values

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

$Z(s_i)$ = measured value
i = measured location
λ = weight at i
s_0 = predicted location
N = # of measured values



Nugget: non-0 γ when h=0 due to measurement error
Sill: variogram levels out for large h. After sill, there is no correlation b/w points
Sill = variance of dataset
Range: value of h where sill occurs

$\geq$30 point pairs per bin
small nugget
sill at range represents sample variance

Weights matrix calculated as C^{-1} (matrix of differences b/w point pairs) multiplied by D (matrix of distances b/w point pairs)

SIMPLE KRIGING: known mean **ORDINARY KRIGING:** unknown constant mean **UNIVERSAL KRIGING:** unknown variable mean

References

Crawley, M. J. (2012). *The R book*. John Wiley & Sons.
Anselin, L. (2019). GeoDa Workbook.
The Landscape Ecology Toolbox website
ESRI. (2019). ArcMap 10.7 Documentation.
Gimmond, M. (2019). Intro to GIS & Spatial Analysis.
Anselin, L. (2005). Exploring Spatial Data with GeoDa: A Workbook. CSISS.
De Smith, M. J., Goodchild, M. F., & Longley, P. (2018). *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*.
Center for Spatial Data Analysis at University of Chicago. Spatial Analysis Tutorials.

Lecture Notes created by Sumeeta Srinivasan for UEP 294: Spatial Statistics at Tufts University
Study guide developed by Aishwarya Venkat, Doctoral Student at Tufts University