Quantitative Methods
Geography 441

Course Requirements

Exercises

- Must be handed in on time.
- Must be TYPED.
- Must follow the format provided. Some slight variation if fine, as long as it is legible.
- IF you use Microsoft WORD, get used to working with the EQUATION EDITOR.

Quantitative Methods (Geo 441)
Dr. Paul Marr

The formats below are to be used when completing the exercises and tests for this course.

<u>*Format for Hand Calculation:*</u>

Research Question: One hundred and fifty students in Geo 441 were surveyed to determine if the use of the correct answer format was associated with student survival.

Clearly stated research question

H_o : Student survival is independent of using the correct answer format.
H_A : Student survival is contingent upon using the correct answer format.

Null and alternate hypotheses statements

σ = 0.05

Alpha level

<u>Observed</u>

Clearly typed problem

	Dead	Alive	Total
Used Format	17	54	71
Did Not Use Format	67	12	79
Total	84	66	150

<u>Expected</u>

	Dead	Alive
Used Format	39.8	31.2
Did Not Use Format	44.2	34.8

$$\chi^2 = \frac{(17-39.8)^2}{39.8} + \frac{(54-31.2)^2}{31.2} + \frac{(67-44.2)^2}{44.2} + \frac{(12-34.8)^2}{34.8} = 56.4$$

Clearly typed answer

$\chi^2 = 56.4$
Df = 1
$\chi^2_{c\ 0.05,\ 1} = 3.841$

Listing of the test statistic, degrees of freedom, and critical value

Therefore, reject H_o.

Student survival is contingent upon using the correct answer format when completing Geography 441 exercises ($\chi^2 = 56.4$, $p < 0.001$).

Statement of the results, test statistic, and probability

Research Question: Water temperature and dissolved oxygen measurements were sampled for Burd Run during May of 2003 to determine whether temperature influences dissolved oxygen levels.

Clearly stated research question.

H_o : The samples come from a normally distributed population.
H_A : The samples did not come from a normally distributed population..

Null and alternate hypotheses statements

$\sigma = 0.05$

Alpha level

Tests of Normality

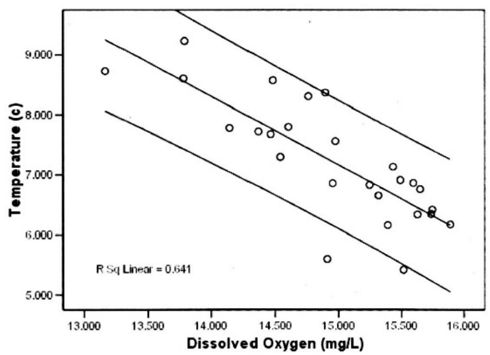
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Temperature (c)	.126	26	.200*	.974	26	.723
Dissolved Oxygen (mg/L)	.128	26	.200*	.935	26	.104

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Computer results in appropriate order

Burd Run DO and Temperature Measurements: May, 2003



Well designed graphic output

Note that the manner in which you place you computer output depends on the tests performed

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.801[a]	.641	.626	.431070

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.961	1	7.961	42.845	.000[a]
	Residual	4.460	24	.186		
	Total	12.421	25			

a. Predictors: (Constant), Temperature (c)

b. Dependent Variable: Dissolved Oxygen (mg/L)

There is a negative linear association between water temperature and dissolved oxygen (R^2=.64, $p < 0.0001$).

Statement of results, statistic, and probability.

The Scientific Method

For the record…

The term *data* is plural.

“Collecting *these data* required considerable personal expenditure.”

The scientific method is used because it is *defensible* AND *repeatable*.

- The scientific method is a defensible technique that helps to minimize external influences.
- Statistical analyses are repeatable.
- The scientific method consists of hypotheses, models, laws, and theories.

Hypothesis – a proposition or statement whose truth or falsity is capable of being tested.

Models – a means of simplifying reality so that relationships between variables can be more clearly studied.

Laws – universal statements of unrestricted range.

Theories – a collection of laws that gives greater insight than produced by individual laws; non-speculative.

Spatial Data Considerations

Spatial Sampling

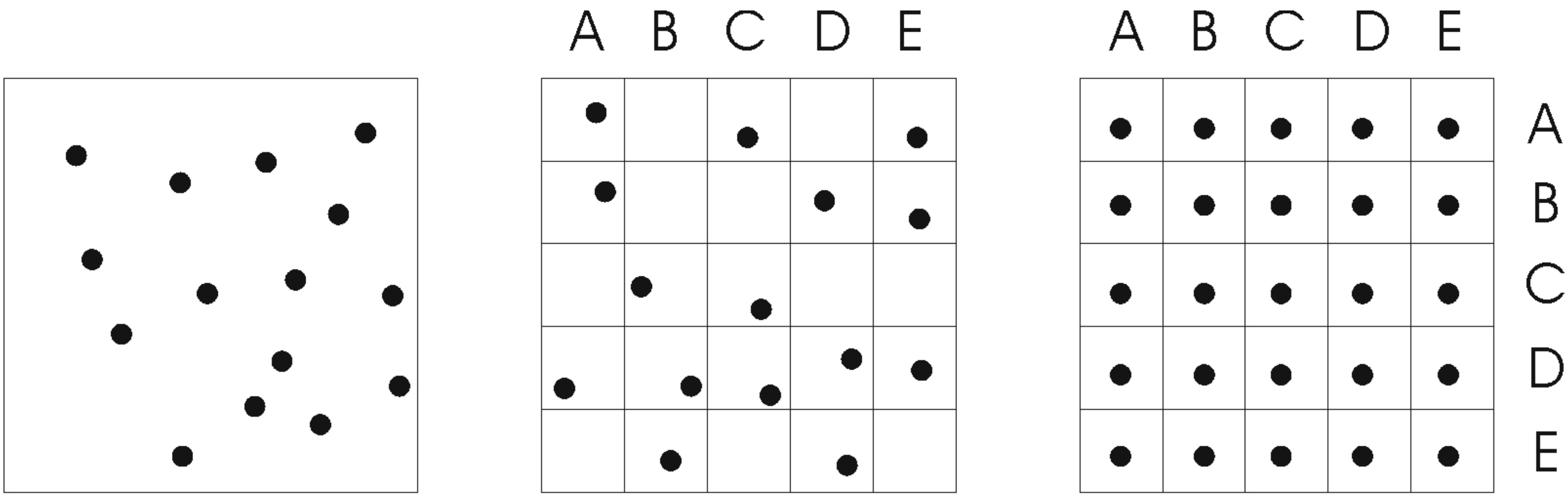
Sample frame: the list of all elements within a population.

e.g. All townships within a county.

Sampling method: the rules which govern how samples of a population are taken.

The three most common are:

1. Random
2. Stratified random
3. Stratified systematic



Random

Stratified Random

Stratified Systematic

Random sampling is often used when it is fairly certain that location does not influence independence.

Stratified random sampling is often used when location does influence independence and the degree of influence is known and can be mapped.

Stratified systematic sampling is often used when location does influence independence but the degree of influence is unknown and cannot be mapped.

Special Considerations with Spatial Data

1. Modifiable areal unit problem

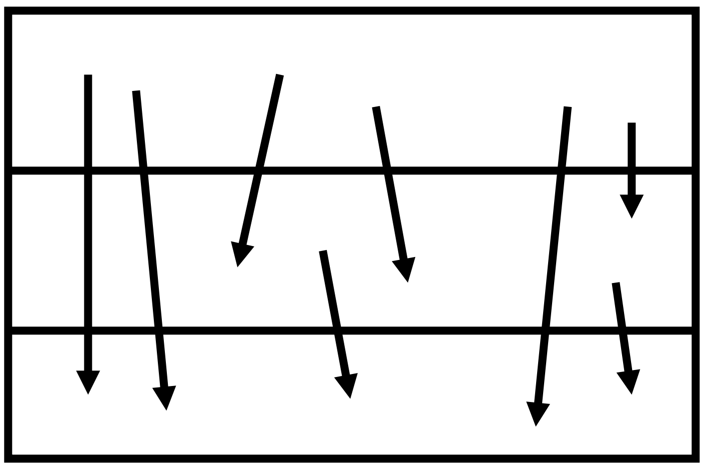
2. Boundary problem

3. Spatial sampling procedures

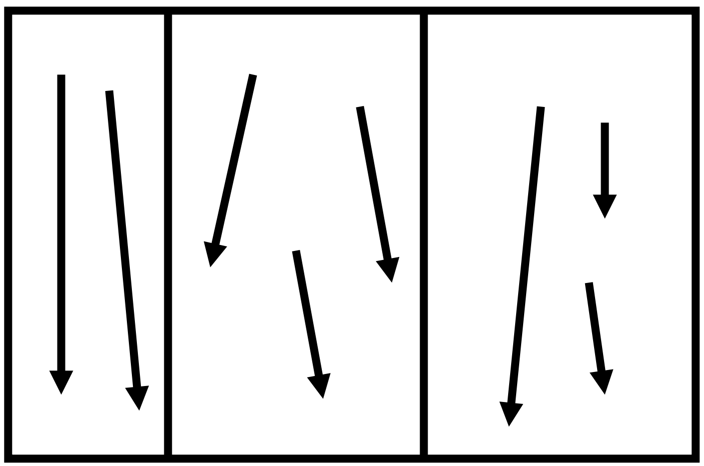
4. Spatial autocorrelation

1. Modifiable areal unit problem:

Changes in areal units influence test results.



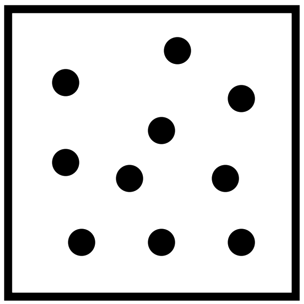
Cross boundary migration



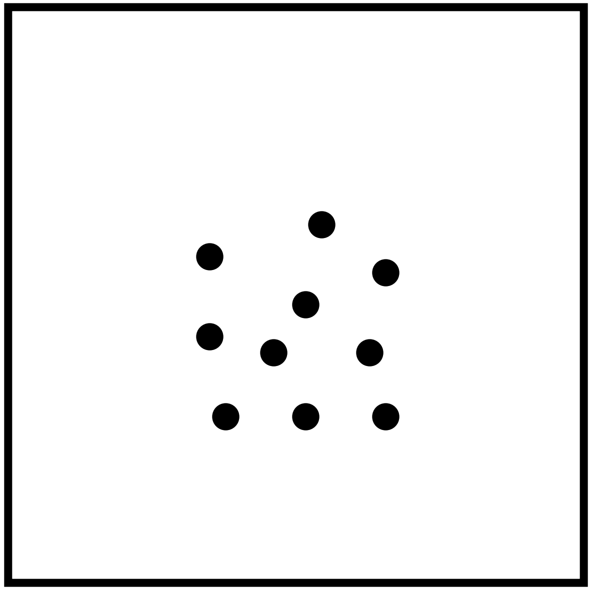
No cross boundary migration

2. Boundary problem

Changes in the study area boundary influence test results.



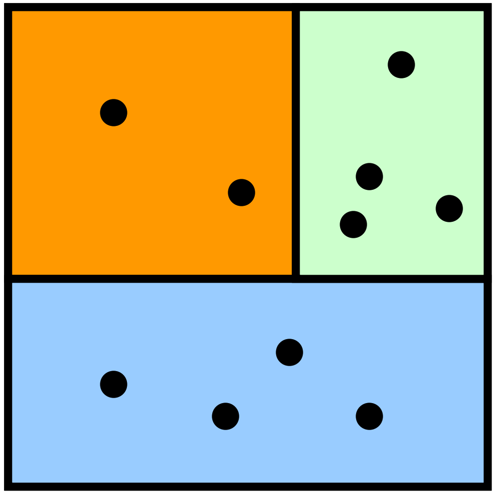
Small study area, even spread of points



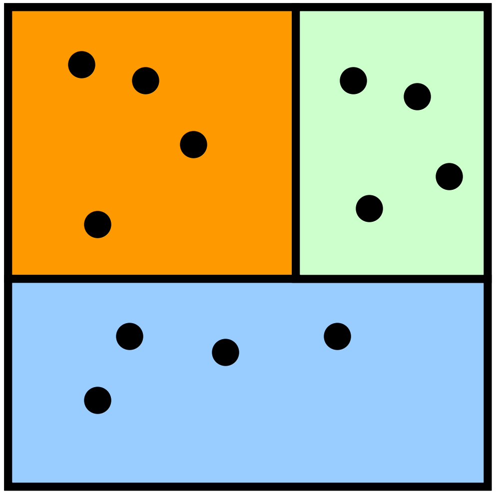
Large study area, clustering of points

3. Spatial sampling procedures

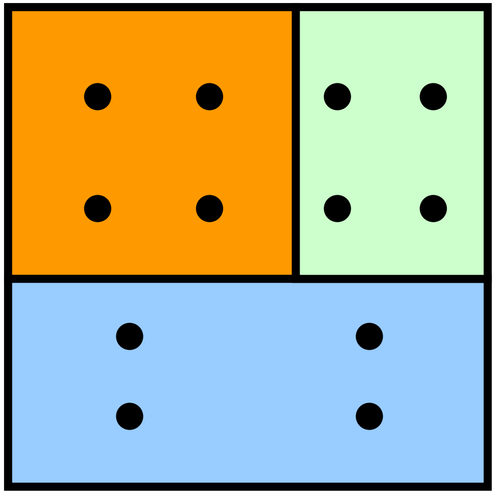
Sampling methods influence test results.



Random



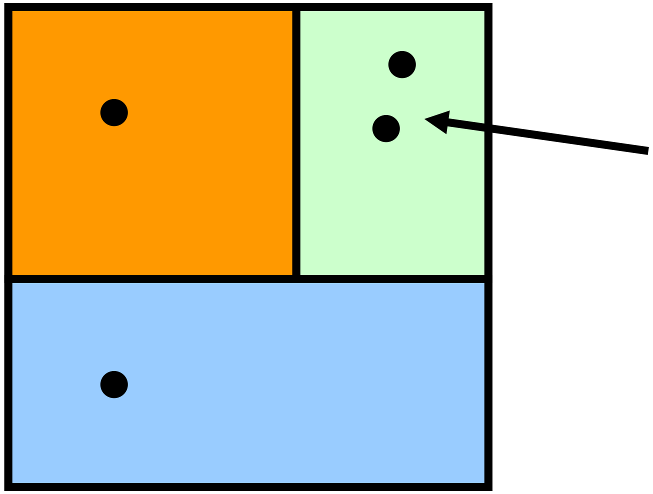
Stratified, Random



Stratified, Regular

4. Spatial autocorrelation

Location near each other often (but not always) influence each other. Non-independence.



These locations have a greater chance of being similar since they are close together.

Spatial Data and Non-independence

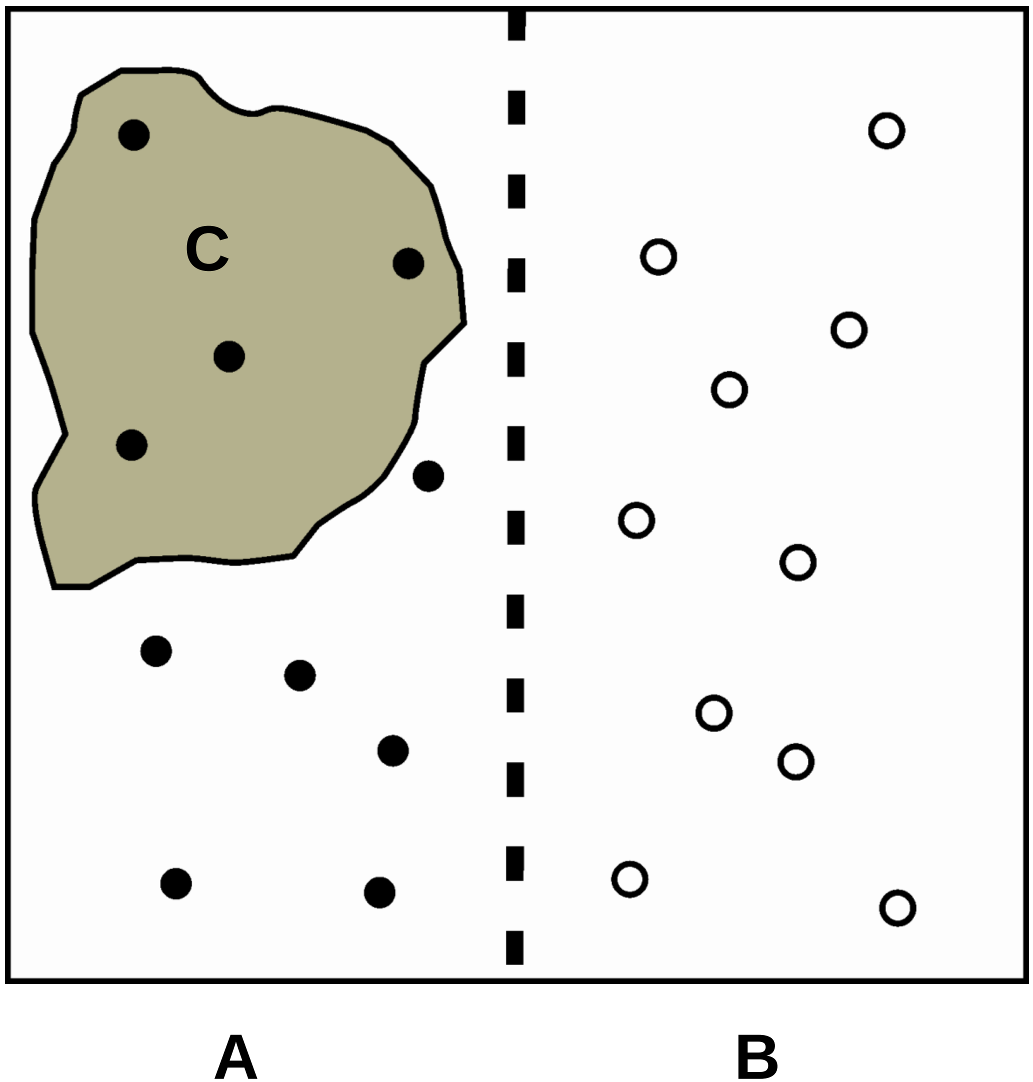
- One and two sample tests assume that the data are independent (i.e. one observation does not influence another).

- Unfortunately, with spatial data one observation often *does* influence another.

For example, two ponds near each other may have have frogs because they both are located in the correct microclimate.

Implications of Non-independence

- Non-independence reduces the effective sample size.
- Inflates the degrees of freedom.
- Reduces the critical value.
- Artificially increases the likelihood of getting significant results.



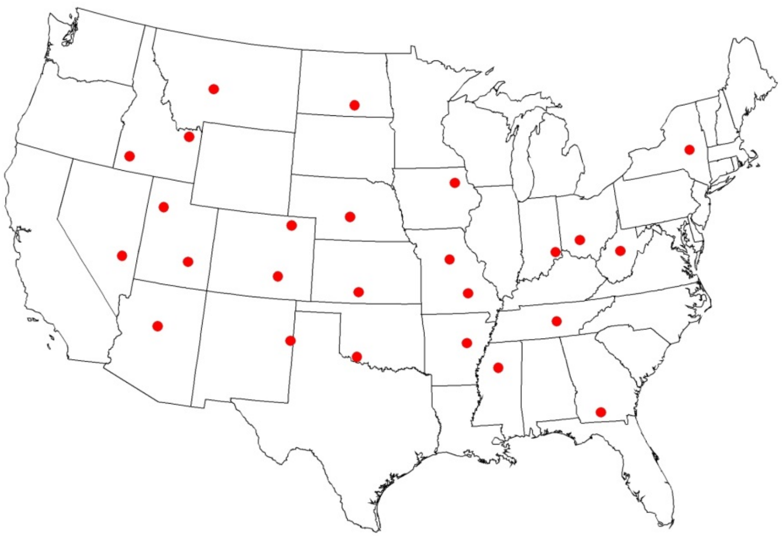
In this example, A and B are study sites. C is a very distinct habitat type.

Habitat C influences the observation values and makes testing between A and B difficult.

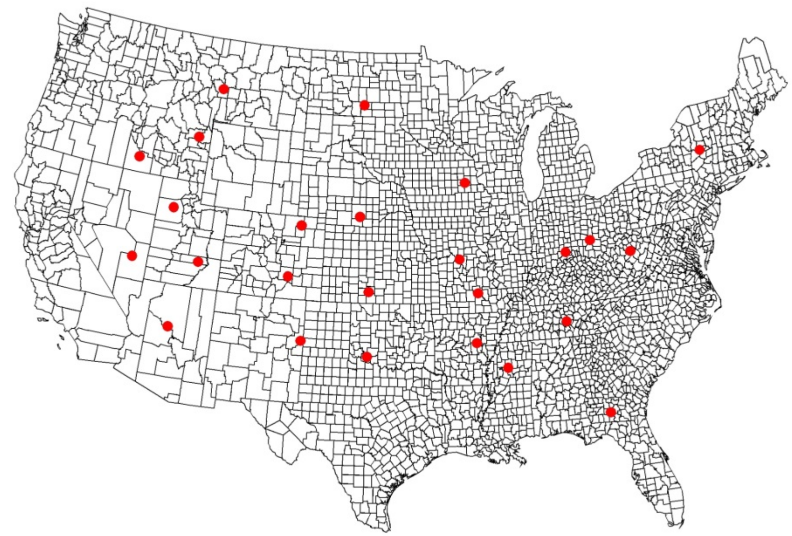
Better to test between A, B, and C.

Spatial Unit Size and Probability Problem

High probability of any particular state having a dot.



Low probability of any particular county having a dot.



- Probabilities tend to be larger with larger spatial units.
  - The probability that an event will occur is higher with larger spatial units.
  - Small areal units tend to have very low frequencies for random events unless the event is extremely common.
  - Under these conditions a normal distribution is almost impossible to attain.

Additional Issues with Spatial Data

1. Ecological Fallacy

Interpretation errors that occur when results derived from aggregate data are applied to disaggregate data.

- It is something akin to trying to determine someone’s grade in a class by looking at the class average.

2. The Problem of Scale

Scale only has meaning within the context of the measuring techniques applied and the questions being asked.

- Does it make sense to measure percent forest cover to 5 decimal places? Are we creating "differences" where none actually exist?

3. Locational Fallacy

Not all processes have a spatial expression. Certain phenomenon are aspatial.

- For example, burglaries have a spatial component and can be mapped. Embezzlement does not.

4. Atomistic Fallacy

An attempt to use individual or disaggregate data to explain aggregate data while ignoring the context of the individuals.

- For example, if we notice that no one is smoking in a bar, then we may conclude that people are smoking less… or it could be a “smoke-free” bar.