Some students in this class may be interested in obtaining a copy of SPSS (the stats program we will use in this class) for yourself. Here is some info (verified last week, but subject to change):

• SPSS has a "grad-pack" version valid for 4 years that comes with regression modeling and some other add-ons

• This grad-pack version retails for ~$45 to $95 (depending on the version and activation term ordered)

• Here is the link to the IBM page for this software: https://www.ibm.com/us-en/marketplace/spss-statistics-gradpack

• SPSS does not sell the grad pack directly – this is one online retailer that does: untsystem.onthehub.com (go to "statistics" tab)

Also: SPSS has a free demo download good for 30 days - visit https://www.ibm.com/products/spss-statistics

Module 2

Part One: Why Multivariate Statistics?

Why Multivariate Statistics?

- This class is about an advanced set of analytical methods that geographers use all the time
  - Statistics that are multivariate and/or spatial in nature
  - Such statistics have increased in their use in recent years across virtually all fields of physical and social sciences

Why Multivariate Statistics?

- This increase in use has been especially rapid in geography and related disciplines
- A couple of major reasons
  - 1. the difficulty of addressing complex research questions, common to geography, with univariate analysis
  - 2. the availability of numerous good software packages that make multivariate statistics possible

Why Multivariate Statistics?

- Univariate statistics have become inadequate on their own as a base for reading or writing modern-day research
  - **Our challenge for this course:** make sure you feel comfortable selecting and setting up multivariate analyses, and interpreting the results

Why Multivariate Statistics?

- Univariate statistics have become inadequate on their own as a base for reading or writing modern-day research
  - Multivariate stats and calculations are more complicated than univariate stats by at least an order of magnitude (i.e. 10 times)
  - However, the understanding required to use multivariate stats does not require an equivalent leap: familiar concepts are the foundation

Some Key Issues & Concepts

- **Independent Variables (IVs) and Dependent Variables (DVs)**
  - **Q:** What is an independent variable? Dependent?
  - Independent variables are sometimes termed predictor or causal variables because they are often conceptualized as causing change in the dependent variable
  - Univariate stats: one DV, one IV (causality is conceptualized between the two)
  - Bivariate stats: special case – analyze the relationship between two variables, but here neither is conceptualized as causing the other

Some Key Issues & Concepts

- **Independent Variables (IVs) and Dependent Variables (DVs)**
  - Multivariate stats: multiple IVs and (potentially) multiple DVs (in the most complex case)

Some Key Issues & Concepts

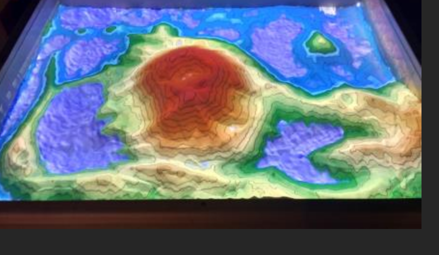
- **Experimental versus Nonexperimental Research**
  - Experimental Research: the researcher controls
    - 1. the levels or conditions of at least one IV
    - 2. all other factors known to influence the DV
  - Nonexperimental Research: the researcher
    - 1. can define what IVs to use and measure
    - 2. does not manipulate or control the levels of the IVs
  - In nonexperimental research, it is difficult to attribute causality to the IV/DV relationship, but researchers often interpret some functional meaning to the relationship (and the statistics work regardless)

Some Key Issues & Concepts

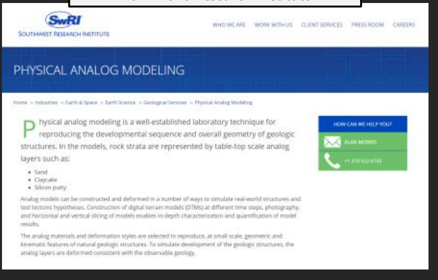
- **Experimental versus Nonexperimental Research**
  - Geographic research: overwhelmingly nonexperimental due to the complexity of what geographers observe
  - **Q:** Is it possible to set up a geographic experiment? What might it look like if it did exist?

Urban Wind Tunnel Modeling at Colorado State University

UNT's 3D Interactive Topographic Drainage Modeling Table



Physical Analog Services Provided by a Non-Profit Research Institute



SwRI
Southwest Research Institute

WHO WE ARE WORK WITH US CLIENT SERVICES PRESS ROOM CAREERS

PHYSICAL ANALOG MODELING

Home > Industries > Earth & Space > Earth Science > Geological Services > Physical Analog Modeling

Physical analog modeling is a well-established laboratory technique for reproducing the developmental sequence and overall geometry of geologic structures. In the models, rock strata are represented by table-top scale analog layers such as:

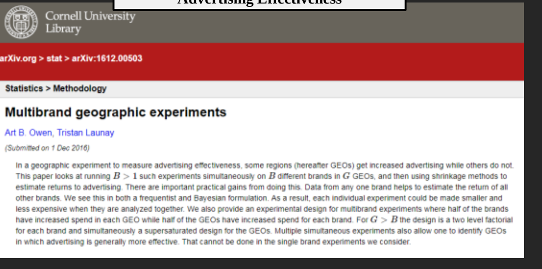
- Sand
- Claycake
- Silicon putty

Analog models can be constructed and deformed in a number of ways to simulate real-world structures and test tectonic hypotheses. Construction of digital terrain models (DTMs) at different time steps, photography, and horizontal and vertical slicing of models enables in-depth characterization and quantification of model results.

The analog materials and deformation styles are selected to reproduce, at small scale, geometric and kinematic features of natural geologic structures. To simulate development of the geologic structures, the analog layers are deformed consistent with the observable geology.

HOW CAN WE HELP YOU?
ALAN MORRIS

Geographic Experiment to Measure Advertising Effectiveness



Cornell University Library

arXiv.org > stat > arXiv:1612.00503

Statistics > Methodology

Multibrand geographic experiments

Art B. Owen, Tristan Launay

(Submitted on 1 Dec 2016)

In a geographic experiment to measure advertising effectiveness, some regions (hereafter GEOs) get increased advertising while others do not. This paper looks at running $B > 1$ such experiments simultaneously on B different brands in G GEOs, and then using shrinkage methods to estimate returns to advertising. There are important practical gains from doing this. Data from any one brand helps to estimate the return of all other brands. We see this in both a frequentist and Bayesian formulation. As a result, each individual experiment could be made smaller and less expensive when they are analyzed together. We also provide an experimental design for multibrand experiments where half of the brands have increased spend in each GEO while half of the GEOs have increased spend for each brand. For $G > B$ the design is a two level factorial for each brand and simultaneously a supersaturated design for the GEOs. Multiple simultaneous experiments also allow one to identify GEOs in which advertising is generally more effective. That cannot be done in the single brand experiments we consider.

UNT

Some Key Issues & Concepts

- **Multivariate Statistics in Nonexperimental Research: How to Approach?**
  - In highly complex situations involving many interrelated variables (i.e. most of geography), one simple option would be to use a series of univariate or bivariate analyses
  - Benefit: help us understand the micro relationships among variable pairs (analyze many pairs)
  - Problem: this would not help us understand overall relationships among groups of variables simultaneously

UNT

Some Key Issues & Concepts

- **Multivariate Statistics in Nonexperimental Research: How to Approach?**
  - One common outcome of using multivariate statistics is the ability to see relationships between one group of variables and another
  - It is this capability that is a primary reason for going to multivariate statistics

UNT

Some Key Issues & Concepts

- **Garbage In, Roses Out?**
  - One issue we need to be much more aware of with multivariate stats is the reliability of our data
  - Especially with multivariate analysis, results aren't necessarily closely tied to easily-observed data characteristics
    - Very hard to pick up errors/problems with a dataset simply by examining the results of a multivariate analysis
    - Very easy to make "garbage data" look great by running it through some multivariate hoops

UNT

Some Key Issues & Concepts

- **Garbage In, Roses Out?**
  - This means we need to be very careful when screening our data for errors before analysis
  - See my handout on data screening: there a few basic guidelines to be aware of

Data Screening Handout

Some Key Issues & Concepts

- **A few more useful definitions**
  - Nominal, ordinal, interval, and ratio data: make sure you are aware of the various data types
  - Samples and populations: as geographers, we need to focus on nonexperimental interpretations, applications, and issues relate to these terms
    - Take precautions to ensure we have a representative sample of the population: define/understand the population, sample appropriately from it
    - Investigate relationships among variables in a predefined population

Some Key Issues & Concepts

- **A few more useful definitions**
  - Descriptive and inferential statistics: know the difference
    - Descriptive statistics: describe a sample or population
    - Inferential statistics: test hypotheses on a sample in order to make general statements about the population it comes from
    - Key point: descriptive statistics are not necessarily an "unpreferred alternative" to inferential statistics; if you can describe a population completely, do it!

Some Key Issues & Concepts

- **A few more useful definitions**
  - Orthogonality: perfect nonassociation between two variables
    - If two variables are orthogonal, knowing the value of one gives no clue as to the value of the other
    - Surprisingly enough to some people, orthogonality is often desirable in multivariate work
    - If all IVs are orthogonal (no overlap in what they explain), then adding each to an analysis will add greatly to the prediction of any DVs

Some Key Issues & Concepts

- **A few more useful definitions**
  - Number and nature of variables to include: in general, the fewer, the better
    - Fewer variables: more general, but less explanation
    - More variables: more explanation, but also more specific (model is hard to generalize and poor as a tool for understanding)

Module 2

Part Two: Introduction to Matrix Algebra

Matrix Algebra

- **What on earth are we doing, and why?**
  - Matrix algebra is the foundation for multivariate statistical analysis

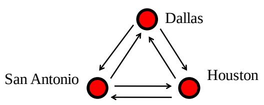
Concept of a "Correlation Matrix":

	Income	Age	Height
Income	1.0	0.9	0.4
Age	0.9	1.0	0.7
Height	0.4	0.7	1.0

Matrix Algebra

- **What on earth are we doing, and why?**
  - Matrices are also powerful in allowing geographers to model a variety of other place-based phenomena, such as encoding the networks that connect places

Transportation geography: an airline network



Matrix Algebra

- **What on earth are we doing, and why?**
  - Matrices are also powerful in allowing geographers to model a variety of other place-based phenomena, such as encoding the networks that connect places

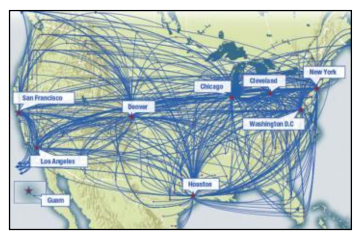
A "connectivity matrix" for the airline network

Flights /day	Dallas	Houston	San Antonio
Dallas	-	44	12
Houston	46	-	16
San Antonio	13	15	-

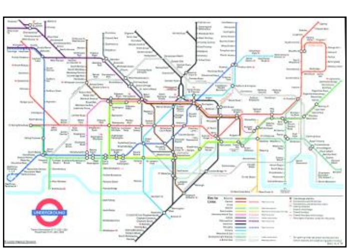
A More Complex Case: The North American Route Map for American Airlines

Current map (following integration of AA and US Airlines systems)

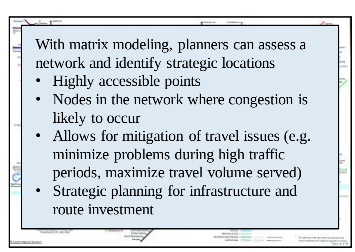
A Related Case: The North American Route Map for United Airlines



Another Case: The London Subway Map



Another Case: The London Subway Map



With matrix modeling, planners can assess a network and identify strategic locations
- Highly accessible points
- Nodes in the network where congestion is likely to occur
- Allows for mitigation of travel issues (e.g. minimize problems during high traffic periods, maximize travel volume served)
- Strategic planning for infrastructure and route investment

Matrix Algebra

- **What on earth are we doing, and why?**
  - Transportation geography is just one context for this use of matrix modeling
  - Any situation where there are links or relationships within a system of places or classifications can use matrix modeling as well (**Q:** can you think of any?)
    - Trade links, social links, material flows, telephone calls, family ties, spread of a technology,…

Matrix Algebra

A "transition matrix" for studying land cover changes in a forest setting

Forest Cover Classes in 2007	Forest Cover Classes in 2017 (hectares)		
	Closed Forest	Open Forest	Long Fallow
Closed Forest	1937.6	447.9	12.4
Open Forest	467.2	283.1	16.7
Long Fallow	18.9	0.3	183.4

Matrix Algebra

- **Important to note that research into the use of matrices as modeling tools in geography is still highly current**
  - Ongoing refinement in the nature of the coefficients used in the matrix as they represent relations/interactions
  - For example, we will deal with a "spatial weight matrix" in our discussion of spatial statistics (in two weeks), but there continue to be attempts to improve on that concept

geographical analysis

Geographical Analysis (2018) 50, 76–96

An Introduction to the Network Weight Matrix

Alireza Ermagun[1], David Levinson[2]

[1]Department of Civil and Environmental Engineering, Technological Institute, Northwestern University, Evanston, Illinois 60208, [2]School of Civil Engineering, University of Sydney, Sydney New South Wales 2006, Australia

This study introduces the network weight matrix as a replacement for the spatial weight matrix to measure the spatial dependence between links of a network. This matrix stems from the concepts of betweenness centrality and vulnerability in network science. The elements of the matrix are a function not simply of proximity, but of network topology, network structure, and demand configuration. The network weight matrix has distinctive characteristics, which are capable of reflecting spatial dependence between traffic links: (1) elements are allowed to have negative and positive values capturing the competitive and complementary nature of links, (2) diagonal elements are not fixed to zero, which takes the self-dependence of a link upon itself into consideration, and (3) elements not only reflect the spatial dependence based on the network structure, but they acknowledge the demand configuration as well. We verify the network weight matrix by modeling traffic flows in a 3 × 3 grid test network with 9 nodes and 24 directed links connecting 72 origin-destination (OD) pairs. Models encompassing the network weight matrix outperform both models without spatial components and models with the spatial weight matrix. The network weight matrix represents a more accurate and defensible spatial dependency between traffic links, and offers the potential to augment traffic flow prediction.

geographical analysis

Geographical Analysis (2018) 50, 76–96

An Introduction to the Network Weight Matrix

Alireza Ermagun[1], David Levinson[2]

[1]Department of Civil and Environmental Engineering, Technological Institute, Northwestern University, Evanston, Illinois 60208, [2]School of Civil Engineering, University of Sydney, Sydney New South Wales 2006, Australia

This study introduces the network weight matrix as a replacement for the spatial weight matrix to measure the spatial dependence between links of a network. This matrix stems from the concepts of betweenness centrality and vulnerability in network science. The elements of the matrix are a function not simply of proximity, but of network topology, network structure, and demand configuration. The network weight matrix has distinctive characteristics, which are capable of reflecting spatial dependence between traffic links: (1) elements are allowed to have negative and positive values capturing the competitive and complementary nature of links, (2) diagonal elements are not fixed to zero, which

origin-destination (OD) pairs. Models encompassing the network weight matrix outperform both models without spatial components and models with the spatial weight matrix. The network weight matrix represents a more accurate and defensible spatial dependency between traffic links, and offers the potential to augment traffic flow prediction.

Focus here on transport network modeling applications, but can be applied to other settings

Matrix Algebra

- **So, this section is a brief overview of some terminology and basic operations we can do with matrices**
  - SPSS and other statistical programs (R, SAS) do these matrix manipulations in the background so we don't have to
  - It's worthwhile to see what's "inside the black box" before we simply trust that we're getting the right results out at the end

Matrix Algebra

A matrix is simply an "array of numbers"

For example,

$$A = \begin{bmatrix} 1 & 2 & 11 \\ 17 & 18 & 15 \end{bmatrix}$$

A_{23} is the element in row 2, column 3

A_{12} is the element in row 1, column 2

The matrix A above has "dimension" 2 x 3 (# of rows x # of columns)

Matrix Algebra

We could have written our previous sample "correlation matrix" as:

$$C = \begin{bmatrix} 1.0 & 0.9 & 0.4 \\ 0.9 & 1.0 & 0.7 \\ 0.4 & 0.7 & 1.0 \end{bmatrix}$$

C is a "square matrix": the number of rows equals the number of columns (very common with statistical work)

Matrix Algebra

More terms: A^T is the "transpose" of matrix A

If $A = \begin{bmatrix} 1 & 2 & 11 \\ 17 & 18 & 15 \end{bmatrix}$ (2 rows x 3 columns)

Then $A^T = \begin{bmatrix} 1 & 17 \\ 2 & 18 \\ 11 & 15 \end{bmatrix}$ (3 rows x 2 columns)

The first row of A becomes the first column of A^T

Matrix Algebra

Another idea: matrix addition/subtraction

Can add or subtract matrices as long as they have the same dimension (rows x columns)

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} 9 & 12 \\ 4 & 14 \\ 15 & 4 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} 10 & 14 \\ 7 & 18 \\ 20 & 10 \end{bmatrix}}_{C}$$

Matrix Algebra

- **Why add two matrices?**
  - **Sample application:** combining two connectivity matrices to give a total
  - Remember the airline example: what if we have one matrix for *American Airlines* and another for *United Airlines*, then combine

Flights /day	Dallas	Houston	San Antonio
Dallas	-	44	12
Houston	46	-	16
San Antonio	13	15	-

Matrix Algebra

Another idea: matrix multiplication (important in multivariate statistical calculations)

Can multiply matrices as long as the number of columns in the first matrix equals the number of rows in the second

$$A \cdot B = C$$

(m rows) x (n columns) · (n rows) x (p columns) = (m rows) x (p columns)

Matrix Algebra

How to do matrix multiplication? Let's just deal with the simplest possible case.

$$A \cdot B = C$$

where all matrices (A, B, and C) are of dimension 2x2 (smallest possible matrix)

Matrix Algebra

$$A \cdot B = C$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

so $$C = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

Matrix Algebra

- **In general for A · B = C**
  - **Each element (n,m) of Matrix C** is the sum of the elements of row n of Matrix A multiplied by the elements of column m of Matrix B
  - Take the case

$$A \cdot B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \cdot \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix}$$

Matrix Algebra

- **Element (1,1) of C (the multiplied matrix) then would be calculated from**

$$\underset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}} \cdot \underset{B}{\begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix}} = \underset{C}{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}}$$

300

a = (1 x 10) + (2 x 40) + (3 x 70) = 300

Matrix Algebra

- **Element (1,2) of C (the multiplied matrix) then would be calculated from**

$$\underset{A}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}} \cdot \underset{B}{\begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix}} = \underset{C}{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}}$$

360

b = (1 x 20) + (2 x 50) + (3 x 80) = 360

Matrix Algebra

In matrix multiplication,

$$A \cdot B \neq B \cdot A$$

even if both multiplications are possible (in terms of the matrix dimensions)

A matrix can be squared (in other words, multiplied with itself, or symbolically, A·A) if and only if it is a square matrix (# rows = # columns)

Matrix Algebra

- **Why square a matrix?**
  - **Sample application:** squaring a connectivity matrix (again think of the airline example) gives a new matrix with the number of <u>one-stop connections</u> between each city
    - The original matrix gives the <u>direct (nonstop) connections</u>

Flights /day	Dallas	Houston	San Antonio
Dallas	-	44	12
Houston	46	-	16
San Antonio	13	15	-

Matrix Algebra

- **Other special terms**
  - **Square matrix:** as noted earlier, # rows = # columns
  - **Diagonal elements of a matrix:** in a square matrix, the diagonal elements are called the "principal diagonal"

	Income	Age	Height
Income	1.0	0.9	0.4
Age	0.9	1.0	0.7
Height	0.4	0.7	1.0

Matrix Algebra

- **Other special terms**
  - **Symmetrical matrix:** symmetric around this diagonal (as in our example correlation matrix)

	Income	Age	Height
Income	1.0	0.9	0.4
Age	0.9	1.0	0.7
Height	0.4	0.7	1.0

Matrix Algebra

- **Other special terms**
  - **Symmetrical matrix:** symmetric around this diagonal (as in our example correlation matrix)
  - Note that square matrices <u>do not have to be symmetrical</u> (think of our airline matrix)

Flights /day	Dallas	Houston	San Antonio
Dallas	-	44	12
Houston	46	-	16
San Antonio	13	15	-

Matrix Algebra

- **Other special terms**
  - **Diagonal matrix:** matrix where the only non-zero elements are on the diagonal

1	0	0	0
0	12	0	0
0	0	4	0
0	0	0	7

Matrix Algebra

- **Other special terms**
  - **Identity matrix:** a diagonal matrix where the diagonal is all <u>ones</u> (the matrix equivalent of the number 1)

A 4x4 identity matrix I:

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

$$\mathbf{A} \cdot \mathbf{I} = \mathbf{I} \cdot \mathbf{A} = \mathbf{A}$$

Matrix Algebra

- **Inverse of a matrix**
  - Definition of the inverse of a matrix

    $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$ (the identity matrix)

    Think of this the same way as in high school algebra

    $x \cdot x^{-1} = 1$ ↖ or (1/x) $2 \cdot 2^{-1} = 1$ ↖ or (1/2)

Matrix Algebra

- **Inverse of a matrix**
  - Used in solving simultaneous linear equations

Example: $3x_1 + 5x_2 = 15$
$2x_1 + 4x_2 = 8$

In matrix notation this becomes:

$$\begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 8 \end{bmatrix}$$

Or $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$

Matrix Algebra

- **Inverse of a matrix**
  - In general,

    $\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}$ is the solution to the system

    $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$

From our previous example (last slide):

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}^{-1} \cdot \begin{bmatrix} 15 \\ 8 \end{bmatrix}$$

Finding A^{-1} gives us the values of x_1 and x_2.
However, finding A^{-1} is not a trivial matter

Matrix Algebra

- **Inverse of a matrix**
  - Not all matrices are invertible
  - Simple test to determine of A^{-1} exists: calculate |A|
    - **|A| is a number:** the "determinant of matrix A"
  - If |A| ≠ 0, then A^{-1} exists
  - Determinant for the case of a 2x2 matrix:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow |\mathbf{A}| = \mathbf{ad - cb}$$

Matrix Algebra

- **Inverse of a matrix**
  - If |A| ≠ 0, then the inverse of the 2x2 matrix is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \cdot \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Matrix Algebra

- **Inverse of a matrix**
  - The determinant and inverse calculations become much more involved in going to 3x3 matrices
  - You don't even want to think about 4x4 and above (let's just say that computers are a wonderful thing!)