More Matrix Algebra; Mean Vectors and Covariance Matrices; the Multivariate Normal Distribution

PSYC 943 (930): Fundamentals of Multivariate Modeling

Lecture 11: October 3, 2012

Today's Class

- The conclusion of Friday's lecture on matrix algebra
  - Matrix inverse
  - Zero/ones vector
  - Matrix identity
  - Matrix determinant
  - NOTE: an introduction to principal components analysis will be relocated later in the semester
- Putting matrix algebra to use in multivariate statistics
  - Mean vectors
  - Covariance matrices
- The multivariate normal distribution

DATA EXAMPLE AND SAS

A Guiding Example

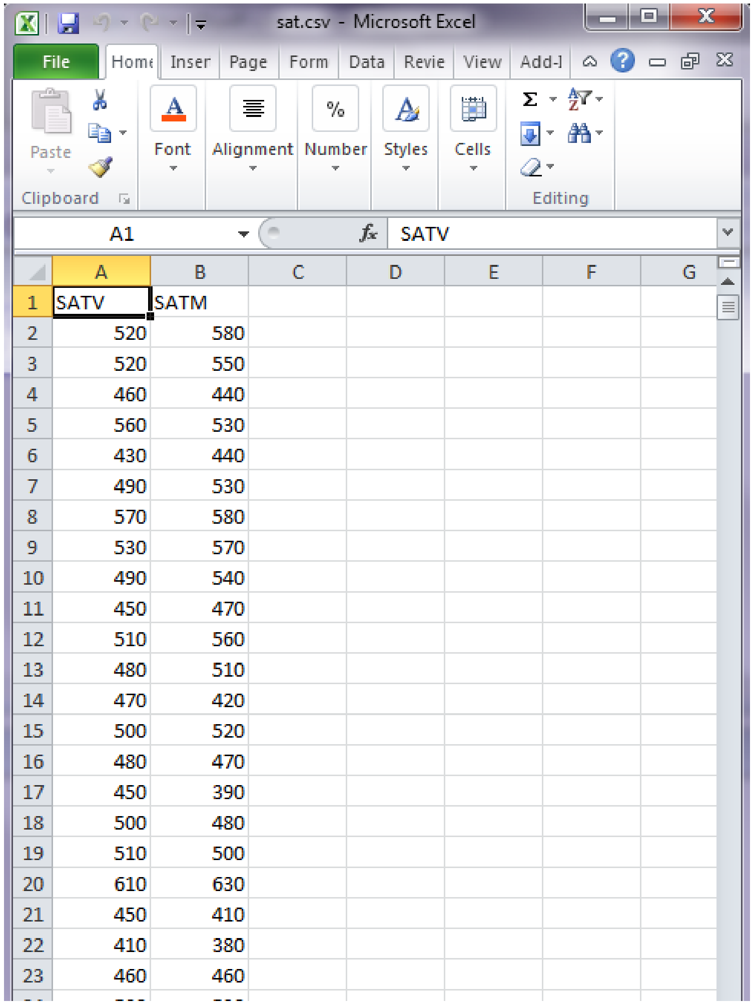
- To demonstrate matrix algebra, we will make use of data
- Imagine that somehow I collected data SAT test scores for both the Math (SATM) and Verbal (SATV) sections of 1,000 students

- The descriptive statistics of this data set are given below:

Statistic	SATV	SATM
Mean	499.3	498.3
SD	49.8	81.2
Correlation		
SATV	1.00	0.78
SATM	0.78	1.00

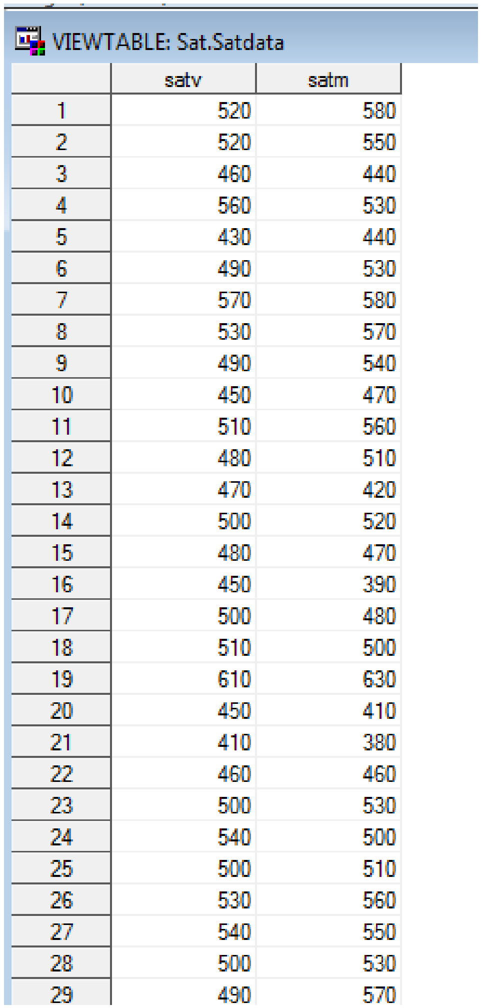
The Data…

In Excel:



	A	B
1	SATV	SATM
2	520	580
3	520	550
4	460	440
5	560	530
6	430	440
7	490	530
8	570	580
9	530	570
10	490	540
11	450	470
12	510	560
13	480	510
14	470	420
15	500	520
16	480	470
17	450	390
18	500	480
19	510	500
20	610	630
21	450	410
22	410	380
23	460	460

In SAS:



	satv	satm
1	520	580
2	520	550
3	460	440
4	560	530
5	430	440
6	490	530
7	570	580
8	530	570
9	490	540
10	450	470
11	510	560
12	480	510
13	470	420
14	500	520
15	480	470
16	450	390
17	500	480
18	510	500
19	610	630
20	450	410
21	410	380
22	460	460
23	500	530
24	540	500
25	500	510
26	530	560
27	540	550
28	500	530
29	490	570

Matrix Computing: PROC IML

- To help demonstrate the topics we will discuss today, I will be showing examples in SAS PROC IML

- The Interactive Matrix Language (IML) is a scientific computing package in SAS that typically used for statistical routines that aren't programed elsewhere in SAS

- Useful documentation for IML:

http://support.sas.com/documentation/cdl/en/imlug/64248/HTML/default/viewer.htm#langref_toc.htm

- A great web reference for IML:

http://www.psych.yorku.ca/lab/sas/iml.htm

MATRIX ALGEBRA

Moving from Vectors to Matrices

- A matrix can be thought of as a collection of vectors
  - Matrix operations are vector operations on steroids
- Matrix algebra defines a set of operations and entities on matrices
  - I will present a version meant to mirror your previous algebra experiences
- Definitions:
  - Identity matrix
  - Zero vector
  - Ones vector
- Basic Operations:
  - Addition
  - Subtraction
  - Multiplication
  - “Division”

Matrix Addition and Subtraction

- Matrix addition and subtraction are much like vector addition/subtraction
- Rules:
  - Matrices must be the same size (rows and columns)
- Method:
  - The new matrix is constructed of element-by-element addition/subtraction of the previous matrices
- Order:
  - The order of the matrices (pre- and post-) does not matter

Matrix Addition/Subtraction

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \\ a_{41} + b_{41} & a_{42} + b_{42} \end{bmatrix} \qquad \mathbf{A} - \mathbf{B} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \\ a_{31} - b_{31} & a_{32} - b_{32} \\ a_{41} - b_{41} & a_{42} - b_{42} \end{bmatrix}$$

Matrix Multiplication

- Matrix multiplication is a bit more complicated
  - The new matrix may be a different size from either of the two multiplying matrices

$$\mathbf{A}_{(r\ x\ c)}\mathbf{B}_{(c\ x\ k)} = \mathbf{C}_{(r\ x\ k)}$$

- Rules:
  - Pre-multiplying matrix must have number of columns equal to the number of rows of the post-multiplying matrix
- Method:
  - The elements of the new matrix consist of the inner (dot) product of the row vectors of the pre-multiplying matrix and the column vectors of the post-multiplying matrix
- Order:
  - The order of the matrices (pre- and post-) matters

Matrix Multiplication

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \\ a_{41}b_{11} + a_{42}b_{21} & a_{41}b_{12} + a_{42}b_{22} & a_{41}b_{13} + a_{42}b_{23} \end{bmatrix}$$

Multiplication in Statistics

- Many statistical formulae with summation can be re-expressed with matrices

- A common matrix multiplication form is: $\mathbf{X}^T\mathbf{X}$
  - Diagonal elements: $\sum_{p=1}^{N} X_p^2$
  - Off-diagonal elements: $\sum_{p=1}^{N} X_{pa}X_{pb}$

- For our SAT example:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \sum_{p=1}^{N} SATV_p^2 & \sum_{p=1}^{N} SATV_p SATM_p \\ \sum_{p=1}^{N} SATV_p SATM_p & \sum_{p=1}^{N} SATM_p^2 \end{bmatrix}$$

$$= \begin{bmatrix} 251{,}797{,}800 & 251{,}928{,}400 \\ 251{,}928{,}400 & 254{,}862{,}700 \end{bmatrix}$$

Identity Matrix

- The identity matrix is a matrix that, when pre- or post- multiplied by another matrix results in the original matrix:

$$\mathbf{AI} = \mathbf{A}$$
$$\mathbf{IA} = \mathbf{A}$$

- The identity matrix is a square matrix that has:
  - Diagonal elements = 1
  - Off-diagonal elements = 0

$$I_{(3\ x\ 3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Zero Vector

- The zero vector is a column vector of zeros

$$\mathbf{0}_{(3\ x\ 1)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- When pre- or post- multiplied the result is the zero vector:

$$\mathbf{A0} = \mathbf{0}$$
$$\mathbf{0A} = \mathbf{0}$$

Ones Vector

- A ones vector is a column vector of 1s:

$$\mathbf{1}_{(3\ x\ 1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- The ones vector is useful for calculating statistical terms, such as the mean vector and the covariance matrix

Matrix “Division”: The Inverse Matrix

- Division from algebra:
  - First: $\frac{a}{b} = \frac{1}{b}a = b^{-1}a$
  - Second: $\frac{a}{a} = 1$
- “Division” in matrices serves a similar role
  - For square and symmetric matrices, an inverse matrix is a matrix that when pre- or post- multiplied with another matrix produces the identity matrix:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$
$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

- Calculation of the matrix inverse is complicated
  - Even computers have a tough time
- Not all matrices can be inverted
  - Non-invertible matrices are called singular matrices
    - In statistics, singular matrices are commonly caused by linear dependencies

The Inverse

- **<u>In data:</u>** the inverse shows up constantly in statistics
  - Models which assume some type of (multivariate) normality need an inverse covariance matrix

- Using our SAT example
  - Our data matrix was size (1000 x 2), which is not invertible
  - However $\mathbf{X}^T\mathbf{X}$ was size (2 x 2) – square, and symmetric

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 251{,}797{,}800 & 251{,}928{,}400 \\ 251{,}928{,}400 & 254{,}862{,}700 \end{bmatrix}$$

  - The inverse is:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 3.61E-7 & -3.57E-7 \\ -3.57E-7 & 3.56E-7 \end{bmatrix}$$

Matrix Determinants

- A square matrix can be characterized by a scalar value called a determinant:

$$\det \mathbf{A} = |\mathbf{A}|$$

- Calculation of the determinant is tedious
  - The determinant for the covariance matrix of our SAT example was 6,514,104.5
- For two-by-two matrices $\left|\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| = ad - bc$
- The determinant is useful in statistics:
  - Shows up in multivariate statistical distributions
  - Is a measure of “generalized” variance of multiple variables
- If the determinant is positive, the matrix is called **positive definite**
  - Is invertible
- If the determinant is not positive, the matrix is called **non-positive definite**
  - Not invertible

Matrix Trace

- For a square matrix $\mathbf{A}$ with V rows/columns, the trace is the sum of the diagonal elements:

$$tr\mathbf{A} = \sum_{v=1}^{V} a_{vv}$$

- For our data, the trace of the **correlation** matrix is 2
  - For all correlation matrices, the trace is equal to the number of variables because all diagonal elements are 1

- The trace is considered the total variance in multivariate statistics
  - Used as a target to recover when applying statistical models

Matrix Algebra Operations (for help in reading stats manuals)

- $(\mathbf{A}+\mathbf{B})+\mathbf{C}=$ $\mathbf{A}+(\mathbf{B}+\mathbf{C})$
- $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
- $c(\mathbf{A}+\mathbf{B})=c\mathbf{A}+c\mathbf{B}$
- $(c+d)\mathbf{A}=c\mathbf{A}+d\mathbf{A}$
- $(\mathbf{A}+\mathbf{B})^T=\mathbf{A}^T+\mathbf{B}^T$
- $(cd)\mathbf{A}=c(d\mathbf{A})$
- $(c\mathbf{A})^T=c\mathbf{A}^T$
- $c(\mathbf{AB})=(c\mathbf{A})\mathbf{B}$
- $\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$
- $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$
- $(\mathbf{AB})^T=\mathbf{B}^T\mathbf{A}^T$
- For x_j such that Ax_j exists:

$$\sum_{j=1}^{N}\mathbf{A}\mathbf{x}_j=\mathbf{A}\sum_{j=1}^{N}\mathbf{x}_j$$

$$\sum_{j=1}^{N}(\mathbf{A}\mathbf{x}_j)(\mathbf{A}\mathbf{x}_j)^T=$$

$$\mathbf{A}\left(\sum_{j=1}^{N}\mathbf{x}_j\mathbf{x}_j^T\right)\mathbf{A}^T$$

MULTIVARIATE STATISTICS AND DISTRIBUTIONS

Multivariate Statistics

- Up to this point in this course, we have focused on the prediction (or modeling) of a single variable
  - Conditional distributions (aka, generalized linear models)

- Multivariate statistics is about exploring **joint distributions**
  - How variables relate to each other simultaneously

- Therefore, we must adapt our conditional distributions to have multiple variables, simultaneously (later, as multiple outcomes)

- We will now look at the joint distributions of two variables $f(x_1, x_2)$ or in matrix form: $f(\mathbf{X})$ (where $\mathbf{X}$ is size N x 2; $f(\mathbf{X})$ gives a scalar/single number)
  - Beginning with two, then moving to anything more than two
  - We will begin by looking at **multivariate descriptive statistics**
    - **Mean vectors and covariance matrices**

- Today, we will only consider the **joint distribution** of sets of variables – but next time we will put this into a GLM-like setup
  - The **joint distribution** will the be conditional on other variables

Multiple Means: The Mean Vector

- We can use a vector to describe the set of means for our data

$$\bar{\mathbf{x}} = \frac{1}{N}\mathbf{X}^T\mathbf{1} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_V \end{bmatrix}$$

  - Here **1** is a N x 1 vector of 1s
  - The resulting mean vector is a *v* x 1 vector of means

- For our data:

$$\bar{\mathbf{x}} = \begin{bmatrix} 499.32 \\ 499.27 \end{bmatrix} = \begin{bmatrix} \bar{x}_{SATV} \\ \bar{x}_{SATM} \end{bmatrix}$$

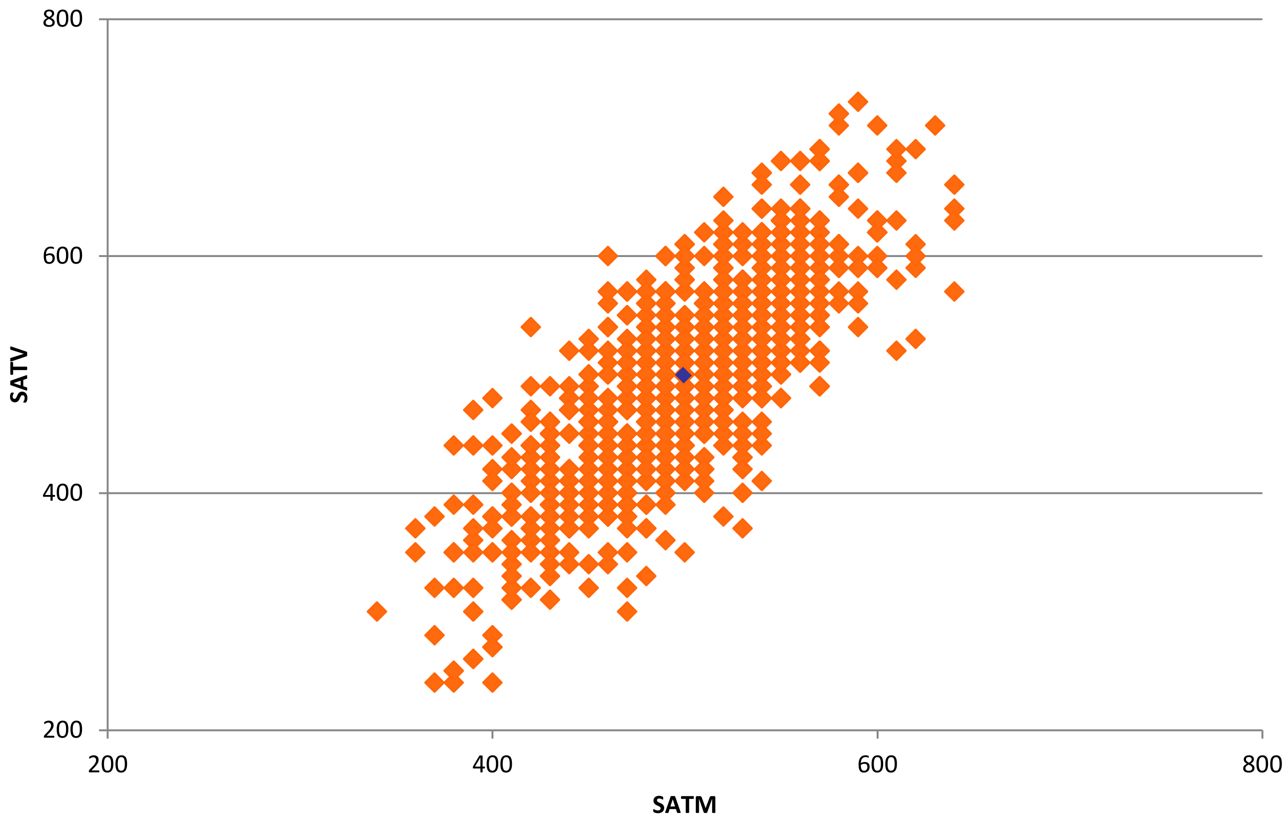
- In SAS PROC IML:

```
*ONES VECTOR WITH SAME LENGTH AS NUMBER OF OBSERVATIONS;
ONES = J(N,1,1); *J function (built in) creates a new matrix with (#rows, #cols, value of element);

*CALCULATION OF THE MEAN VECTOR;
meanvec = (1/N)*t(X)*ONES; *t() function (built in) transposes the matrix in the parentheses;
```

Mean Vector: Graphically

- The mean vector is the center of the distribution of both variables



Covariance of a Pair of Variables

- The covariance is a measure of the relatedness
  - Expressed in the product of the units of the two variables:

$$s_{x_1 x_2} = \frac{1}{N}\sum_{p=1}^{N}\left(x_{p1} - \bar{x}_1\right)\left(x_{p2} - \bar{x}_2\right)$$

  - The covariance between SATV and SATM was 3,132.22 (in SAT Verbal-Maths)
  - The denominator N is the ML version – unbiased is N-1

- Because the units of the covariance are difficult to understand, we more commonly describe association (correlation) between two variables with correlation
  - Covariance divided by the product of each variable’s standard deviation

Correlation of a Pair of Varibles

- Correlation is covariance divided by the product of the standard deviation of each variable:

$$r_{x_1 x_2} = \frac{s_{x_1 x_2}}{\sqrt{s_{x_1}^2}\sqrt{s_{x_2}^2}}$$

  - The correlation between SATM and SATV was 0.78

- Correlation is unitless – it only ranges between -1 and 1
  - If x_1 **and** x_2 both had variances of 1, the covariance between them would be a correlation
    - Covariance of standardized variables = correlation

Covariance and Correlation in Matrices

- The covariance matrix (for any number of variables *v*) is found by:

$$\mathbf{S} = \frac{1}{N}(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T) = \begin{bmatrix} s_{x_1}^2 & \cdots & s_{x_1x_V} \\ \vdots & \ddots & \vdots \\ s_{x_1x_V} & \cdots & s_{x_V}^2 \end{bmatrix}$$

- In SAS PROC IML:

```
*ONES VECTOR WITH SAME LENGTH AS NUMBER OF OBSERVATIONS;
ONES = J(N,1,1); *J function (built in) creates a new matrix with (#rows, #cols, value of element);

*CALCULATION OF THE MEAN VECTOR;
meanvec = (1/N)*t(X)*ONES;

*CALCULATION OF THE COVARIANCE MATRIX;
mean_matrix = ONES*t(meanvec); *for covariance matrix;
cov_matrix = (1/N)*t(X - mean_matrix)*(X - mean_matrix);
```

- $\mathbf{S} = \begin{bmatrix} 2{,}477.34 & 3{,}123.22 \\ 3{,}132.22 & 6{,}589.71 \end{bmatrix}$

```
cov_matrix        2 rows        2 cols     (numeric)

2477.3376 3132.2236
3132.2236 6589.7071
```

From Covariance to Correlation

- If we take the SDs (the square root of the diagonal of the covariance matrix) and put them into a diagonal matrix $\mathbf{D}$, the correlation matrix is found by:

$$\mathbf{R} = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1} = \begin{bmatrix} \frac{s_{x_1}^2}{\sqrt{s_{x_1}^2}\sqrt{s_{x_1}^2}} & \cdots & \frac{s_{x_1 x_p}}{\sqrt{s_{x_1}^2}\sqrt{s_{x_V}^2}} \\ \vdots & \ddots & \vdots \\ \frac{s_{x_1 x_V}}{\sqrt{s_{x_1}^2}\sqrt{s_{x_V}^2}} & \cdots & \frac{s_{x_V}^2}{\sqrt{s_{x_V}^2}\sqrt{s_{x_V}^2}} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & r_{x_1 x_V} \\ \vdots & \ddots & \vdots \\ r_{x_1 x_V} & \cdots & 1 \end{bmatrix}$$

Example Covariance Matrix

- For our data, the covariance matrix was:

$$\mathbf{S} = \begin{bmatrix} 2{,}477.34 & 3{,}123.22 \\ 3{,}132.22 & 6{,}589.71 \end{bmatrix}$$

- The diagonal matrix **D** was:

$$\mathbf{D} = \begin{bmatrix} \sqrt{2{,}477.34} & 0 \\ 0 & \sqrt{6{,}589.71} \end{bmatrix} = \begin{bmatrix} 49.77 & 0 \\ 0 & 81.18 \end{bmatrix}$$

- The correlation matrix **R** was:

$$\mathbf{R} = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1} = \begin{bmatrix} \frac{1}{49.77} & 0 \\ 0 & \frac{1}{81.18} \end{bmatrix} \begin{bmatrix} 2{,}477.34 & 3{,}123.22 \\ 3{,}132.22 & 6{,}589.71 \end{bmatrix} \begin{bmatrix} \frac{1}{49.77} & 0 \\ 0 & \frac{1}{81.18} \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1.00 & .78 \\ .78 & 1.00 \end{bmatrix}$$

In SAS:

```
*DIAGONAL MATRIX OF STANDARD DEVIATIONS FROM COVARIANCE MATRIX:;
*SQRT TAKES STANDARD DEVIATION (COVARIANCE MATRIX HAS VARIANCES);
D_matrix = SQRT(DIAG(cov_matrix));
```

```
D_matrix      2 rows      2 cols     (numeric)

 49.77286         0
        0 81.177011
```

```
*INVERSE OF D_Matrix:;
D_matrix_inv = INV(D_matrix);
```

```
D_matrix_inv      2 rows      2 cols     (numeric)

0.0200913         0
        0 0.0123188
```

```
*CORRELATION MATRIX:;
corr_matrix = D_matrix_inv*cov_matrix*D_matrix_inv;
```

```
corr_matrix      2 rows      2 cols     (numeric)

        1 0.7752238
0.7752238         1
```

Generalized Variance

- The determinant of the covariance matrix is the **generalized variance**

$$\text{Generalized Sample Variance} = |\mathbf{S}|$$

- It is a measure of spread across all variables
  - Reflecting how much overlap (covariance) in variables occurs in the sample
  - Amount of overlap reduces the generalized sample variance
  - Generalized variance from our SAT example: 6,514,104.5
  - Generalized variance if zero covariance/correlation: 16,324,929

```
*GENERALIZED VARIANCE:;
GEN_VAR = DET(cov_matrix);
```

```
GEN_VAR        1 row        1 col      (numeric)

6514104.5
```

- The generalized sample variance is:
  - Largest when variables are uncorrelated
  - Zero when variables form a linear dependency

- <u>**In data:**</u>
  - The generalized variance is seldom used descriptively, but shows up more frequently in maximum likelihood functions

Total Sample Variance

- The total sample variance is the sum of the variances of each variable in the sample
  - The sum of the diagonal elements of the sample covariance matrix
  - The trace of the sample covariance matrix

$$Total\ Sample\ Variance = \sum_{v=1}^{V} s_{x_i}^2 = \text{tr}\ \mathbf{S}$$

- Total sample variance for our SAT example:

```
*TOTAL SAMPLE VARIANCE;
TOT_VAR = TRACE(cov_matrix);
```

```
TOT_VAR      1 row       1 col     (numeric)

9067.0447
```

- The total sample variance does not take into consideration the covariances among the variables
  - Will not equal zero if linearly dependency exists

- <u>**In data:**</u>
  - The total sample variance is commonly used as the denominator (target) when calculating variance accounted for measures

MULTIVARIATE DISTRIBUTIONS (VARIABLES ≥ 2)

Multivariate Normal Distribution

- The multivariate normal distribution is the generalization of the univariate normal distribution to multiple variables
  - The bivariate normal distribution just shown is part of the MVN

- The MVN provides the relative likelihood of observing **all** *V* variables for a subject *p* simultaneously:

$$\mathbf{x}_p = \begin{bmatrix} x_{p1} & x_{p2} & \dots & x_{pV} \end{bmatrix}$$

- The multivariate normal density function is:

$$f\left(\mathbf{x}_p\right) = \frac{1}{(2\pi)^{\frac{V}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}} \exp\left[-\frac{\left(\mathbf{x}_p^T - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_p^T - \boldsymbol{\mu}\right)}{2}\right]$$

The Multivariate Normal Distribution

$$f(\mathbf{x}_p) = \frac{1}{(2\pi)^{\frac{V}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{x}_p^T - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x}_p^T - \boldsymbol{\mu})}{2}\right]$$

- The mean vector is $\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \vdots \\ \mu_{x_V} \end{bmatrix}$

- The covariance matrix is $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 x_V} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2 x_V} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_1 x_V} & \sigma_{x_2 x_V} & \cdots & \sigma_{x_V}^2 \end{bmatrix}$

    - The covariance matrix must be non-singular (invertible)

Comparing Univariate and Multivariate Normal Distributions

- The univariate normal distribution:

$$f(x_p) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

- The univariate normal, rewritten with a little algebra:

$$f(x_p) = \frac{1}{(2\pi)^{\frac{1}{2}}|\sigma^2|^{\frac{1}{2}}} \exp\left[-\frac{(x-\mu)\sigma^{-\frac{1}{2}}(x-\mu)}{2}\right]$$

- The multivariate normal distribution

$$f(\mathbf{x}_p) = \frac{1}{(2\pi)^{\frac{V}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{\left(\mathbf{x}_p^T-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_p^T-\boldsymbol{\mu}\right)}{2}\right]$$

  - When $V = 1$ (one variable), the MVN is a univariate normal distribution

The Exponent Term

- The term in the exponent (without the $-\frac{1}{2}$) is called the **squared Mahalanobis Distance**

$$d^2(\boldsymbol{x}_p) = \left(\mathbf{x}_p^T - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_p^T - \boldsymbol{\mu}\right)$$

  - Sometimes called the statistical distance
  - Describes how far an observation is from its mean vector, in standardized units
  - Like a multivariate Z score (but, if data are MVN, is actually distributed as a χ^2variable with DF = number of variables in X)
  - Can be used to assess if data follow MVN

Multivariate Normal Notation

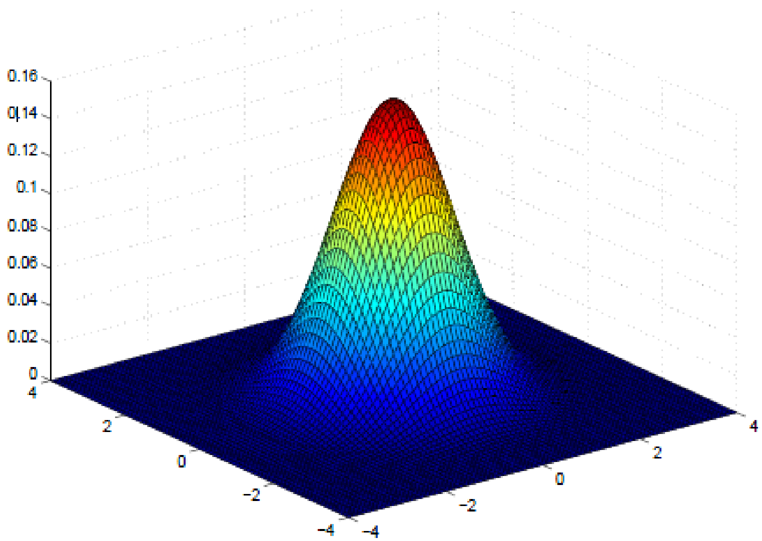
- Standard notation for the multivariate normal distribution of *v* variables is $N_v(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - Our SAT example would use a bivariate normal: $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- **<u>In data:</u>**
  - The multivariate normal distribution serves as the basis for most every statistical technique commonly used in the social and educational sciences
    - General linear models (ANOVA, regression, MANOVA)
    - General linear mixed models (HLM/multilevel models)
    - Factor and structural equation models (EFA, CFA, SEM, path models)
    - Multiple imputation for missing data

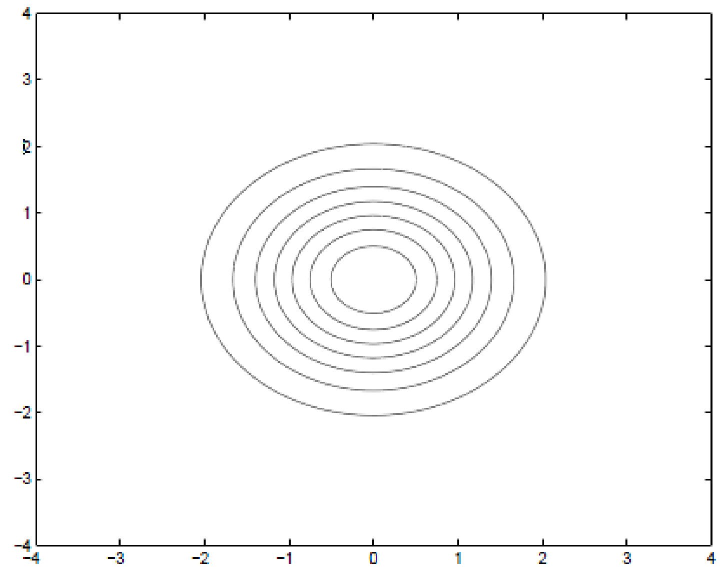
  - Simply put, the world of commonly used statistics revolves around the multivariate normal distribution
    - Understanding it is the key to understanding many statistical methods

Bivariate Normal Plot #1

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma^2_{x_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



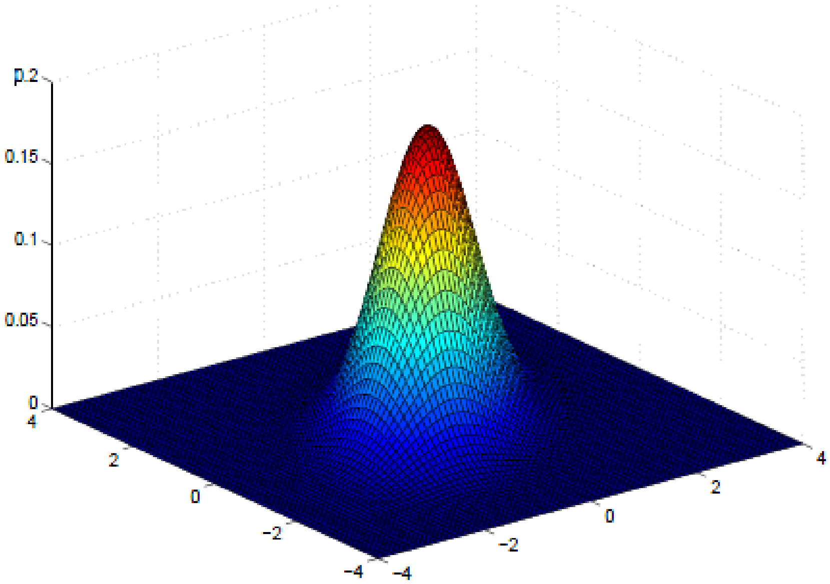
Density Surface (3D)



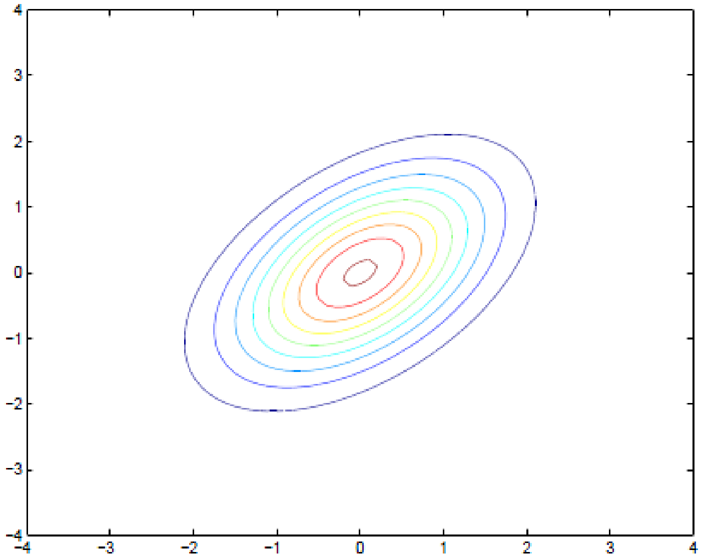
Density Surface (2D):
Contour Plot

Bivariate Normal Plot #2 (Multivariate Normal)

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma^2_{x_2} \end{bmatrix} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$$



Density Surface (3D)



Density Surface (2D): Contour Plot

Multivariate Normal Properties

- The multivariate normal distribution has some useful properties that show up in statistical methods

- If $\mathbf{X}$ is distributed multivariate normally:

1. Linear combinations of $\mathbf{X}$ are normally distributed

2. All subsets of $\mathbf{X}$ are multivariate normally distributed

3. A zero covariance between a pair of variables of $\mathbf{X}$ implies that the variables are independent

4. Conditional distributions of $\mathbf{X}$ are multivariate normal

Multivariate Normal Distribution in PROC IML

- To demonstrate how the MVN works, we will now investigate how the PDF provides the likelihood (height) for a given observation:
  - Here we will use the SAT data and assume the sample mean vector and covariance matrix are known to be the true:

$$\boldsymbol{\mu} = \begin{bmatrix} 499.32 \\ 498.27 \end{bmatrix}; \mathbf{S} = \begin{bmatrix} 2{,}477.34 & 3{,}123.22 \\ 3{,}132.22 & 6{,}589.71 \end{bmatrix}$$

- We will compute the likelihood value for several observations (SEE EXAMPLE SAS SYNTAX FOR HOW THIS WORKS):
  - $\boldsymbol{x}_{631,\cdot} = [590 \quad 730]; f(\boldsymbol{x}) = 0.00000087528$
  - $\boldsymbol{x}_{717,\cdot} = [340 \quad 300]; f(\boldsymbol{x}) = 0.00000037082$
  - $\boldsymbol{x} = \overline{\boldsymbol{x}} = [499.32 \quad 498.27]; f(\boldsymbol{x}) = 0.0000624$

- Note: this is the height for these observations, not the joint likelihood across all the data
  - Next time we will use PROC MIXED to find the parameters in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using maximum likelihood

Likelihoods…From SAS

```
*MULTIVARIATE NORMAL DISTRIBUTION FUNCTION CALCULATIONS;
*CONSTANTS FOR ALL CALCULATIONS:;

PI = CONSTANT('pi'); *the constant pi;
NVAR = NCOL(X); *the number of variables in X;

pi_constant = (2*PI)**(NVAR/2);
sigma_constant = DET(cov_matrix)**(1/2);
sigma_inverse = INV(cov_matrix);


*OBSERVATION #631:;
obs = X[631,];
mean_diff = t(obs)-meanvec;
exponent_term = (-1/2)*t(mean_diff)*sigma_inverse*mean_diff;
likelihood = (1/pi_constant)*(1/sigma_constant)*exp(exponent_term);
```

$$f(\mathbf{x}_p) = \frac{1}{(2\pi)^{\frac{v}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{x}_p^T - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_p^T - \boldsymbol{\mu})}{2}\right]$$

Wrapping Up

- The last two classes set the stage to discuss multivariate statistical methods that use maximum likelihood

- Matrix algebra was necessary so as to concisely talk about our distributions (which will soon be models)

- The multivariate normal distribution will be necessary to understand as it is the most commonly used distribution for estimation of multivariate models

- Friday we will get back into data analysis – but for multivariate observations…using SAS PROC MIXED
  - Each term of the MVN will be mapped onto the PROC MIXED output