Techniques for studying and explaining correlation and covariance structure:

- Principal Components Analysis (PCA)
- Factor Analysis

Principal Component Analysis

Given:

Random variables $X_1, \ldots, X_p$ with covariance matrix $\mathbf{\Sigma}$

Define: $\boldsymbol{X'} = [X_1, \ldots, X_p]$

Goal:

Find linear combinations

$$Y_1 = \boldsymbol{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

$$Y_2 = \boldsymbol{a}_2'\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p$$

…

$$Y_p = \boldsymbol{a}_p'\boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p$$

such that

$\mathrm{Var}(Y_i) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_i$ is as large as possible, and

$\mathrm{cov}(Y_i, Y_j) = \boldsymbol{a}_i'\boldsymbol{\Sigma}\boldsymbol{a}_j = 0$

Such $Y_1, \ldots, Y_p$ are called ***principal components***

Note:

There are many vectors $\boldsymbol{a}_i$ that are solutions to this maximization problem. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length.

That is, we will assume that:

$$\boldsymbol{a}_i'\boldsymbol{a}_i = a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2 = 1$$

The solution can be found with the Lagrange multiplier technique

Let

$$g(\boldsymbol{a}, \lambda) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} + \lambda(1 - \boldsymbol{a}'\boldsymbol{a})$$

be the objective function.

We need the partial derivatives to be equal to 0.

Now

$$\frac{\partial g\left(\vec{a},\lambda\right)}{\partial\lambda}=\left(1-\vec{a}'\vec{a}\right)=0 \text{ if } \vec{a}'\vec{a}=1$$

and

$$\frac{\partial g\left(\vec{a},\lambda\right)}{\partial\vec{a}}=2\Sigma\vec{a}-2\lambda\vec{a}=\vec{0} \text{ if } \Sigma\vec{a}=\lambda\vec{a}$$

Thus $\vec{a}$ is an eigenvector of Σ and λ is the eigenvalue associated with $\vec{a}$.

Also $Var\left(\vec{a}'\vec{x}\right) = \vec{a}'\Sigma\vec{a} = \vec{a}'\left(\lambda\vec{a}\right) = \lambda\vec{a}'\vec{a} = \lambda$

Hence $Var\left(\vec{a}'\vec{x}\right) = \lambda$ is maximized if λ is the largest eigenvalue of Σ.

Summary:

$Y_1 = \boldsymbol{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$

is the ***first principal component*** if

$$\vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}$$

is the eigenvector (length 1) of $\boldsymbol{\Sigma}$ associated with the largest eigenvalue λ_1 of Σ.

Recall any positive matrix, Σ

$$\Sigma = \left[\vec{a}_1, \cdots, \vec{a}_p\right] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \vec{a}_1' \\ \vdots \\ \vec{a}_p' \end{bmatrix} = PDP'$$

where $\vec{a}_1, \cdots, \vec{a}_p$ are eigenvectors of Σ of length 1 and

$$\lambda_i \geq \ldots \geq \lambda_p \geq 0$$

are eigenvalues of Σ.

$P = \left[\vec{a}_1, \cdots, \vec{a}_p\right]$ is an orthogonal matrix.

$(P'P = PP' = I)$

Theorem 1:

Let the eigenvalues of $\boldsymbol{\Sigma}$ be $\lambda_1 \geq \lambda_2 \geq \ \dots \geq \lambda_p \geq 0$ with corresponding eigenvectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_p$

Then the i^{th} principal component is given by:

$Y_i = \boldsymbol{e}_i'\boldsymbol{X} = \ e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \ \ i = 1, \dots, p$

Furthermore,

$$\mathrm{Var}(Y_i) = \boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i, \qquad i = 1, \dots, p$$

and

$$\mathrm{cov}\left(Y_i, Y_j\right) = \boldsymbol{e}_i'\boldsymbol{\Sigma}\boldsymbol{e}_j = 0$$

Achievements:

1. $Var(Y_1)$ is maximized.
2. $Var(Y_i)$ is maximized subject to Y_i being independent of $Y_1, \ldots, Y_{i-1}$ (the previous i -1 principle components)

Theorem 2:

The total variance is:

$$\mathrm{tr}(\boldsymbol{\Sigma}) = \sigma_{11} + \ldots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i)$$

$$= \lambda_1 + \ldots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

The vector of Principal components

$$\vec{C} = \begin{bmatrix} C_1 \\ \vdots \\ C_p \end{bmatrix} = \begin{bmatrix} \vec{a}_1'\vec{x} \\ \vdots \\ \vec{a}_p'\vec{x} \end{bmatrix} = \begin{bmatrix} \vec{a}_1' \\ \vdots \\ \vec{a}_p' \end{bmatrix}\vec{x} = P'\vec{x}$$

has covariance matrix

$$\Sigma_{\vec{c}} = P'\Sigma P = P'\left(PDP'\right)P = \left(P'P\right)D\left(P'P\right)$$

$$= D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix}$$

An orthogonal matrix rotates vectors, thus

$$\vec{C} = P'\vec{x}$$

rotates the vector $\vec{x}$

into the vector of Principal components $\vec{C}$

Also

$$tr(D) = tr\left(\Sigma_{\vec{c}}\right) = tr\left(P'\Sigma P\right) = tr\left(\Sigma PP'\right) = tr\left(\Sigma\right)$$

$$\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \sigma_{ii}$$

$$\sum_{i=1}^{p} \mathrm{var}\left(C_i\right) = \sum_{i=1}^{p} \mathrm{var}\left(x_i\right) = \text{Total Variance of } \vec{x}$$

The ratio

$$\frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} = \frac{\lambda_i}{\sum_{j=1}^{p} \sigma_{jj}} = \frac{\operatorname{var}(C_i)}{\text{Total Variance of } \vec{x}}$$

denotes the proportion of variance explained by the i^{th} principal component C_i.

Further goals:

- Data reduction: if some of the eigenvalues are "small" we can ignore them and represent the data with fewer variables/principle components.
- Explained variance: the proportion of variance explained by the first *k* principal components is:

$$\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_p}$$

If the ***X*** variables are highly correlated, then the total variance can be explained with a relatively small *k*, like $k = 1, 2$, or 3.

In general, we aim for about 80-90% explained variance.

Also

$$\operatorname{cov}\left(C_i, x_j\right) = \operatorname{cov}\left(\vec{a}_i'\vec{x}, \vec{e}_j'\vec{x}\right) = \vec{a}_i'\Sigma\vec{e}_j$$

$$= \vec{a}_i'\left(\lambda_1\vec{a}_1\vec{a}_1' + \cdots + \lambda_p\vec{a}_p\vec{a}_p'\right)\vec{e}_j$$

$$= \lambda_i\vec{a}_i'\vec{e}_j = \lambda_i a_{ij}$$

where $\vec{e}_j' = \left[0, 0, \cdots, 0, \underset{j}{1}, 0, \cdots, 0\right]$

$$\operatorname{corr}\left(C_i, x_j\right) = \frac{\operatorname{cov}\left(C_i, x_j\right)}{\sqrt{Var\left(C_i\right)}\sqrt{Var\left(x_j\right)}}$$

$$= \frac{\lambda_i a_{ij}}{\sqrt{\lambda_i}\sqrt{\sigma_{jj}}} = \sqrt{\frac{\lambda_i}{\sigma_{jj}}}a_{ij}$$

Remarks:

- The principal components are not scale invariant.
- If one of the variables has much bigger variance than the rest, then the first principle component will largely be represented by that variable. This will be reflected in the eigenvector.

Theorem 3: The correlation between the i^{th} principle component and the j^{th} variable is:

$$\operatorname{cor}(Y_i, X_j) = \frac{e_{ij}\sqrt{\lambda_i}}{\sqrt{\sigma_{jj}}}$$

Example (8.1 on p. 434)

In this example p = 3 and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

It can be shown that:

$\lambda_1 = 5.83, \qquad \boldsymbol{e}_1' = [0.383, -0.924, 0]$

$\lambda_2 = 2, \qquad \boldsymbol{e}_2' = [0, 0, 1]$

$\lambda_3 = 0.17, \qquad \boldsymbol{e}_3' = [0.924, 0.383, 0]$

Therefore, the principle components are:

$Y_1 = 0.383X_1 - 0.924X_2$,

Var(Y_1) = λ_1 = 5.83 and explains $\frac{5.83}{5.83+2+0.17} = 72.9\%$ of the total variance

$Y_2 = X_3$

$Y_3 = 0.924X_1 + 0.383X_2$

$\mathrm{cov}(Y_1, Y_2) = \mathrm{cov}(0.383X_1 - 0.924X_2, X_3) = 0 + 0 = 0$

Example

In this example wildlife (moose) population density was measured over time (once a year) in three areas.

Year	Area 1	Area 2	Area 3	Year	Area 1	Area 2	Area 3
1	11.3	14.1	6.9	13	6.1	9.9	6.8
2	10.4	14	11.2	14	9.7	13.2	6.6
3	9.9	13	8.7	15	8.1	9.4	4
4	8.2	11.4	3.3	16	11.3	11.8	4.9
5	10.1	11.9	8.7	17	8.8	11.5	8.8
6	10.7	13.8	12.5	18	9.4	11.6	5.7
7	11	14.9	8.9	19	7.5	11.4	4.9
8	7.1	8.5	3.7	20	8.8	10.7	7.2
9	14.7	14.5	12.1	21	7.5	11.1	7
10	5.4	9	4.1	22	9.1	13.2	8.9
11	7.3	7.6	5.6	23	6.8	9.8	7.6
12	10.2	10.9	7.3				

picture

The Sample Statistics

$$\vec{\bar{x}} = \begin{bmatrix} 9.10 \\ 11.62 \\ 7.19 \end{bmatrix}$$

The mean vector

$$S = \begin{bmatrix} 4.297 & 3.307 & 3.295 \\ & 3.527 & 3.527 \\ & & 6.566 \end{bmatrix}$$

The covariance matrix

$$R = \begin{bmatrix} 1 & .796 & .620 \\ & 1 & .687 \\ & & 1 \end{bmatrix}$$

The correlation matrix

Principal component Analysis

The eigenvalues of S

$$\lambda_1 = 11.85974, \quad \lambda_2 = 2.204232, \quad \lambda_3 = 0.814249$$

The eigenvectors of S

$$\vec{a}_1 = \begin{bmatrix} .522 \\ .523 \\ .674 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} .582 \\ .359 \\ -.730 \end{bmatrix}, \quad \vec{a}_3 = \begin{bmatrix} .624 \\ -.733 \\ .117 \end{bmatrix}$$

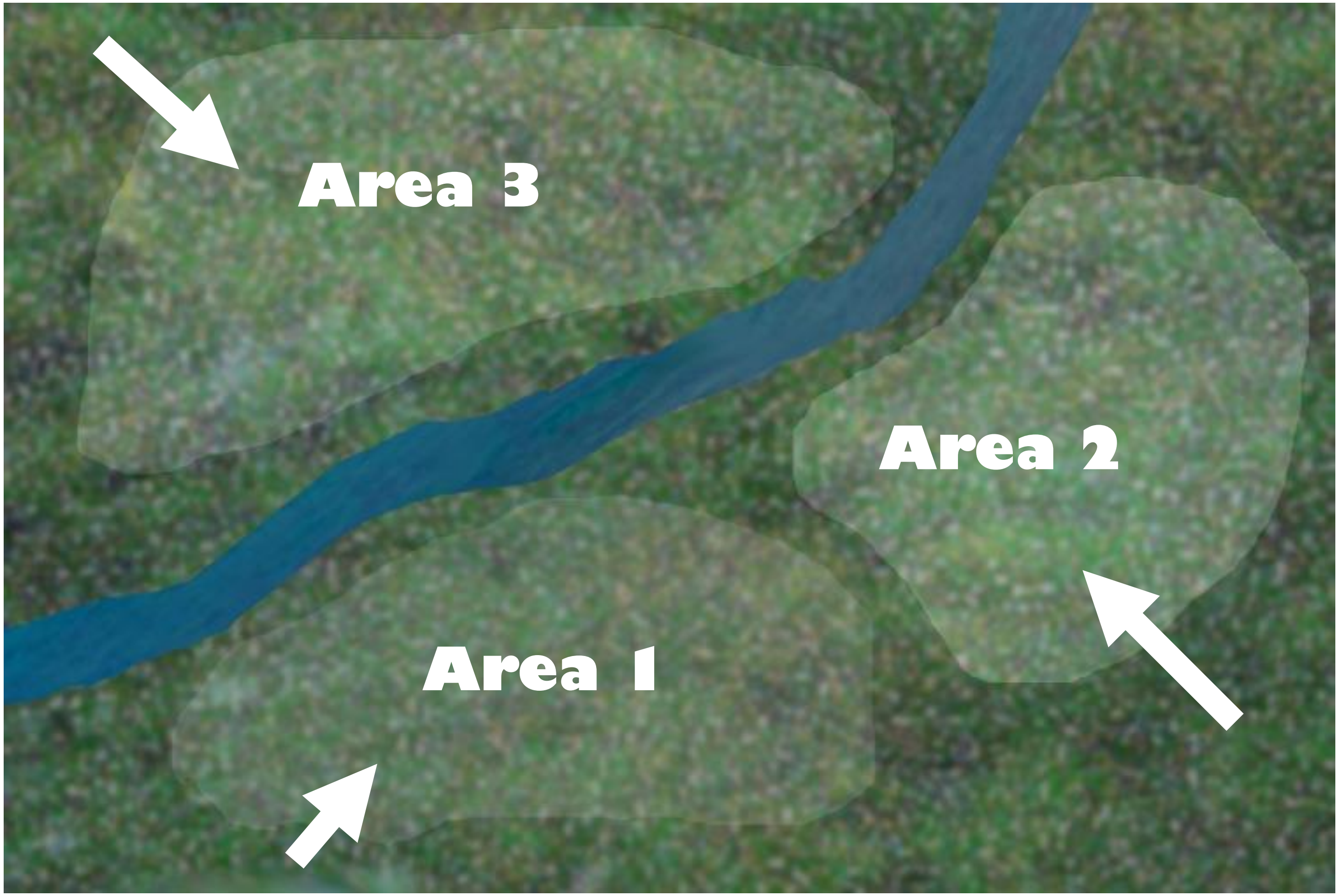
The principal components

$$C_1 = .522x_1 + .523x_2 + .674x_3$$

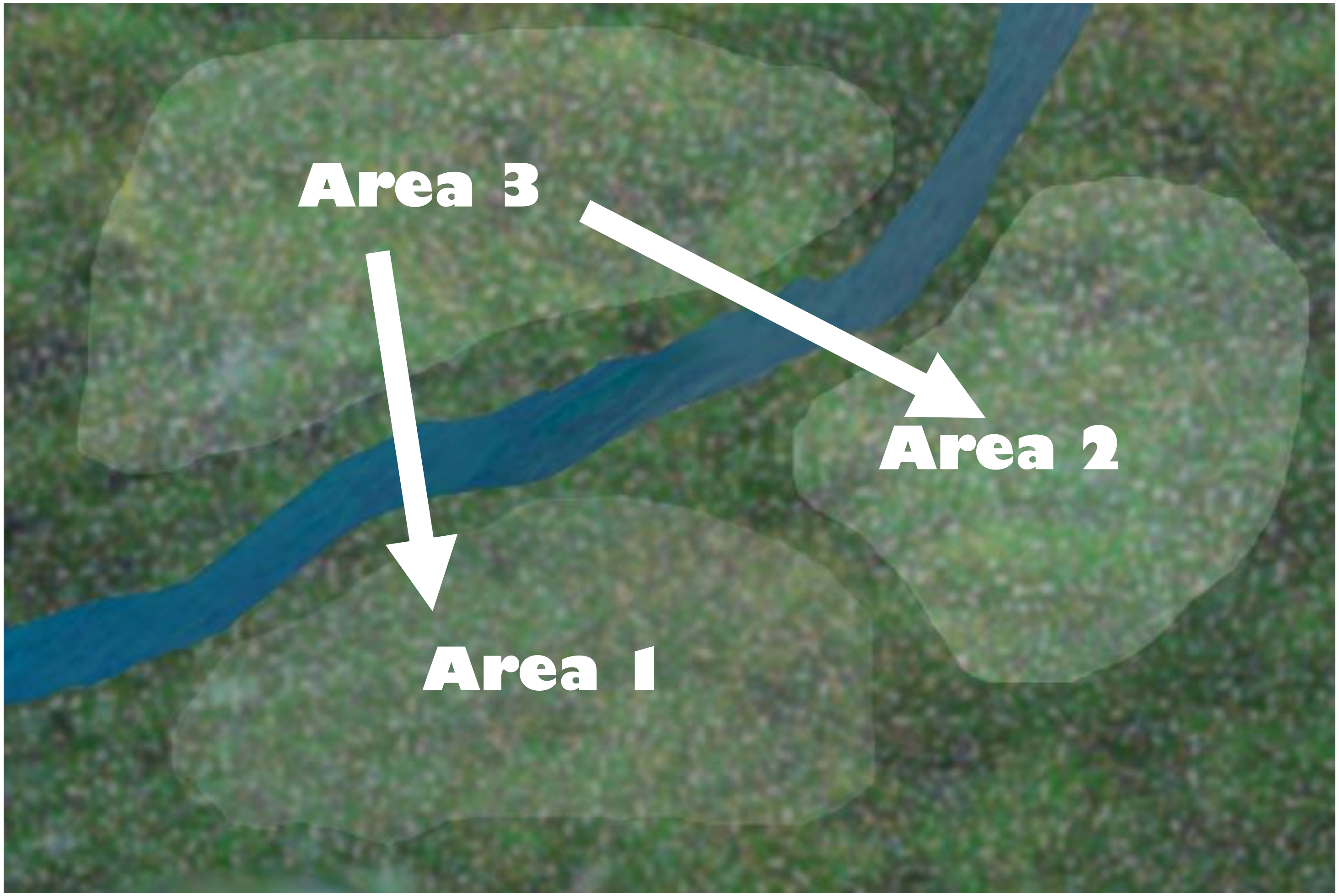
$$C_2 = .582x_1 + .359x_2 - .730x_3$$

$$C_3 = .624x_1 - .733x_2 + .117x_3$$

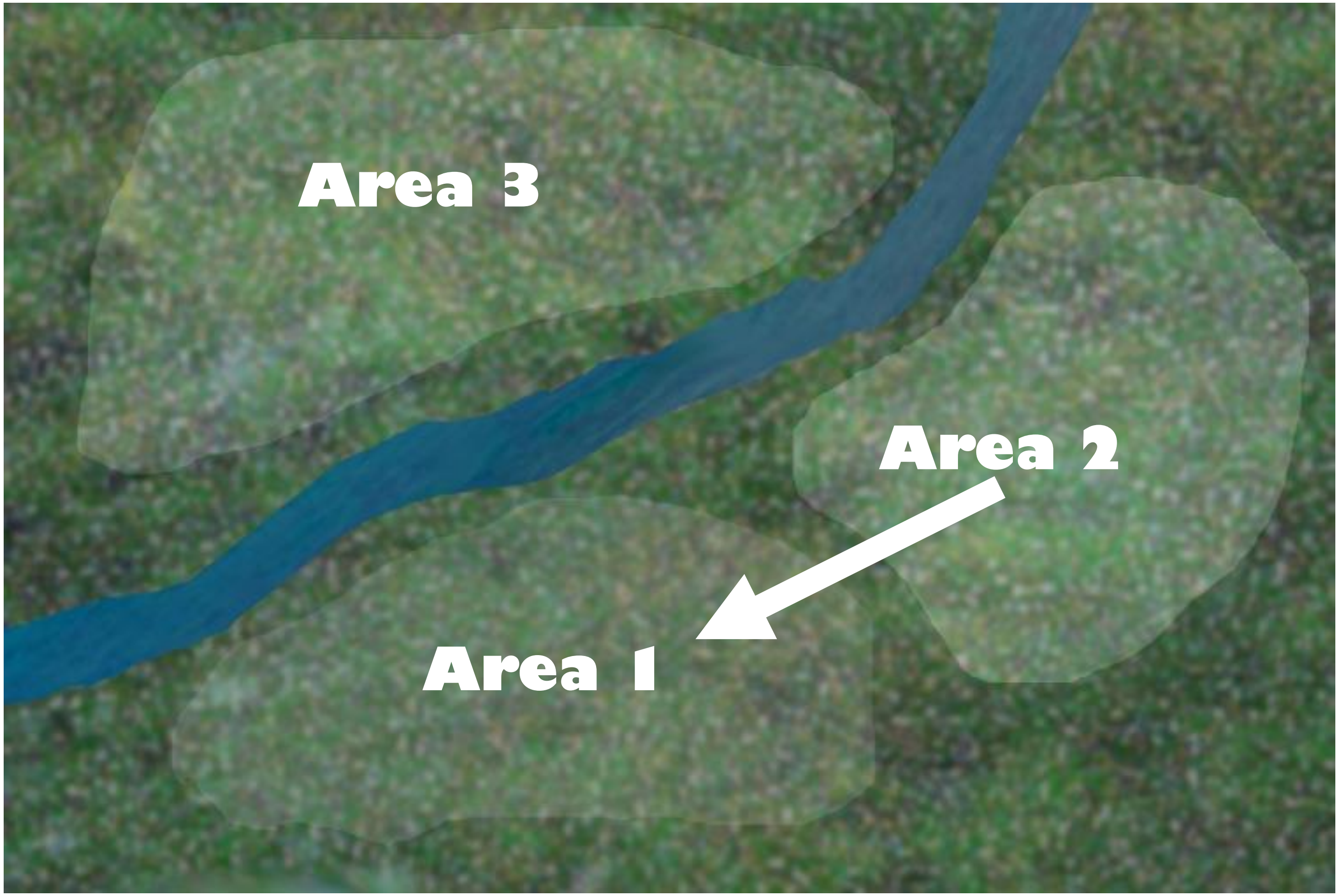
$$C_1 = .522x_1 + .523x_2 + .674x_3$$



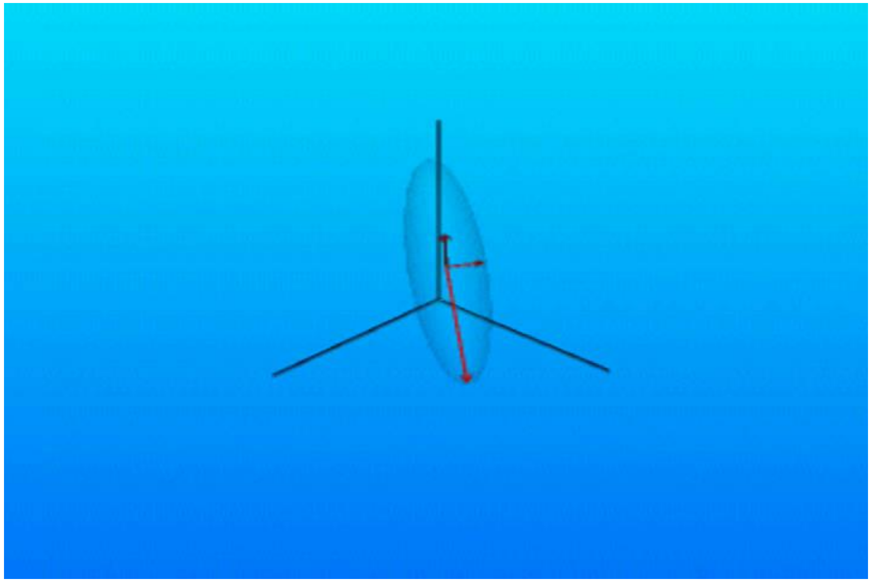
$$C_2 = .582x_1 + .359x_2 - .730x_3$$



$$C_3 = .624x_1 - .733x_2 + .117x_3$$



Graphical Picture of Principal Components



Multivariate Normal data falls in an ellipsoidal pattern.

The shape and orientation of the ellipsoid is determined by the covariance matrix Σ.

The eignevectors of Σ are vectors giving the directions of the axes of the ellopsoid. The eigenvalues give the length of these axes.

The Example

i	λ_i	% variance
1	11.8597	79.71%
2	2.20423	14.82%
3	0.81425	5.47%
Total	14.8782	100%

Comment:

If instead of the covariance matrix, Σ, The correlation matrix ***R***, is used to extract the Principal components then the Principal components are defined in terms of the standard scores of the observations:

$$z_i = \frac{x_i - \mu_i}{\sqrt{\sigma_{ii}}}$$

and $$\operatorname{corr}\left(C_i^*\right) = \sqrt{\lambda_i}\, a_{ij}$$

The correlation matrix is the covariance matrix of the standard scores of the observations.

More Examples

Example: *Bone Lengths of White Leghorn Fowl:*

The correlation matrix of the complete set of six fowl bone measurements had the following form:

Skull Length	1.000	0.584	0.615	0.601	0.570	0.600
Skull Breadth		1.000	0.576	0.530	0.526	0.555
Humerus			1.000	0.940	0.875	0.878
Ulna				1.000	0.877	0.886
Femur					1.000	0.924
Tibia						1.000

Table: Principal Components

Dimension	Component 1	2	3	4	5	6
Skull.Length	-0.35	0.54	0.77	0.05	0.03	0.00
Skull.Breadth	-0.33	0.70	-0.64	0.00	0.01	0.06
Wing.Humerus	-0.44	-0.19	-0.04	-0.52	0.17	-0.68
Wing.Ulna	-0.44	-0.25	0.01	-0.49	-0.15	0.69
Leg.Femur	-0.43	-0.28	-0.06	0.51	0.67	0.13
Leg.Tibia	-0.44	-0.23	-0.05	0.47	-0.71	-0.18
Eigenvalue	4.568	0.714	0.412	0.173	0.076	0.057
% of Total Variance	76.1	11.9	6.9	2.9	1.3	0.9

Identification of Components:

component	*Description*
1	General average of all bone dimensions (Size)
2	Comparison of skull sizewith Wing and Leg lengths
3	Comparison of skull length and breadth (Skull Shape)
4	Comparison of Wing and Leg lengths
5	Comparison of femur and tibia
6	Comparison of humerus and ulna

Example 3: *Weschler Adult Intelligence Scale Subtest Scores*

Table: Principal Components

	Component 1	2	3	4
WAIS subtest:				
Information	0.83	0.33	-0.04	-0.01
Comprehension	0.75	0.31	0.07	-0.17
Arihtmetic	0.72	0.25	-0.08	0.35
Similarities	0.78	0.14	0.00	-0.21
Digit Span	0.62	0.00	-0.38	0.58
Vocabulary	0.83	0.38	-0.03	-0.16
Digit Symbol	0.72	-0.36	-0.26	-0.01
Picture Completion	0.78	-0.10	-0.25	-0.01
Block Design	0.72	-0.26	0.36	0.18
Picture Arrangement	0.72	-0.23	0.04	-0.05
Object assembly	0.65	-0.30	0.47	0.13
Age:	-0.34	0.80	0.26	0.18
Years of Education:	0.75	0.01	-0.30	-0.23
Eigenvalue	6.69	1.42	0.80	0.71
% of Total Variance	51.47	10.90	6.15	5.48
Cum % of Variance	51.47	62.37	68.52	74.01

Identification of Components:

component	*Description*
1	General intellectual Performance
2	Experiential or age factor - bipolar dimension comparing verbal or informational skills known to increase with advancing age to subtests measuring spatial-perceptual qualities and other cognitive abilities known to decrease with age
3	Spatial imagery or perception dimension
4	Numerical facility