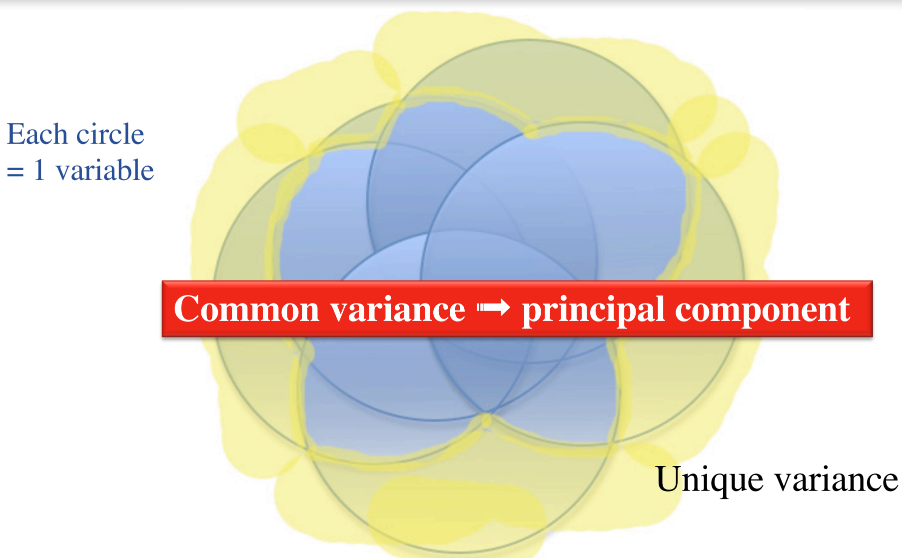
PCA Part II

CLPS 2908
Lecture 11
March 7, 2019

Communing Variables



Communalities

- communality of each variable = h^2
  - = variance in given <u>variable</u> accounted for by selected components (non-uniqueness)

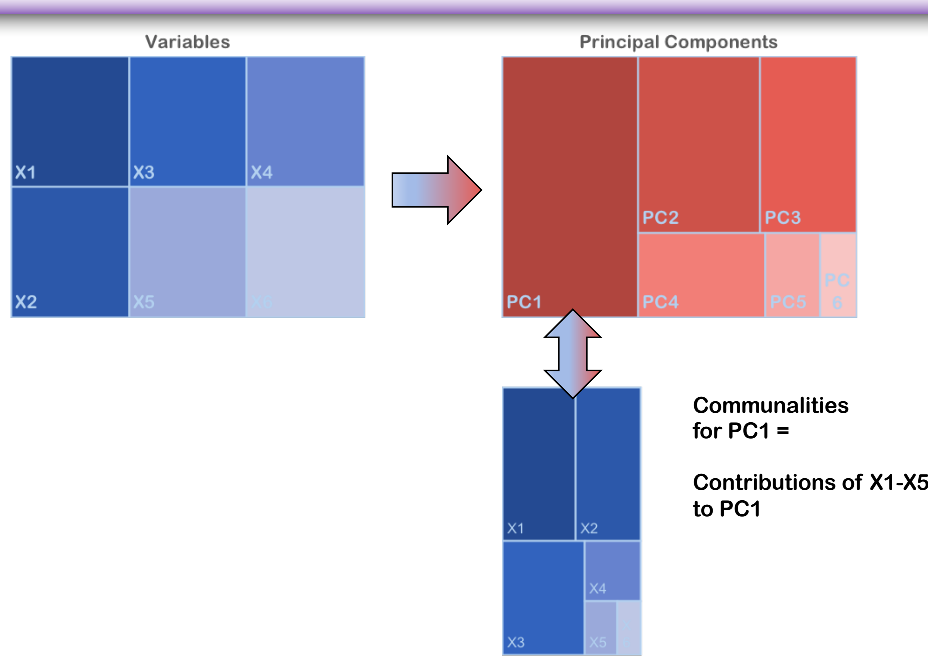
Communalities

	Initial	Extraction
X1	1.000	.625
x2	1.000	.718
x3	1.000	.769
x4	1.000	.792
x5	1.000	.532

(2 components extracted)

  - If $r = p$, what's the sum of all the h_i^2 values?

Variance Rearrangement



Communalities for PC1 =

Contributions of X1-X5 to PC1

Loading Matrix

Component Matrix **(= Factor Loading Matrix)**

Loading matrix A contains correlations between variables and components

	Factor 1	Factor 2
X1	.76362	-.20469
X2	.84590	.04739
X3	-.12137	.86847
X4	.04530	.88857
X5	.72907	.02616

Loading Matrix

Component Matrix **(= Factor Loading Matrix)**

Squared correlations = variance components.
Each X's summed variance components = its communality (what *it* contributes to PC)

	Factor 1	Factor 2	
X1	.76362	-.20469	→ $\hat{h}_1^2 =$ .625
X2	.84590	.04739	
X3	-.12137	.86847	
X4	.04530	.88857	→ $\hat{h}_4^2 =$ .792
X5	.72907	.02616	

Loading Matrix

Component Matrix

PC 1' s summed variance components (squared loadings) = its total variance, λ_1

	Factor 1	Factor 2	
X1	.76362	-.20469	→ $\hat{h}_1^2$
X2	.84590	.04739	
X3	-.12137	.86847	
X4	.04530	.88857	→ $\hat{h}_4^2$
X5	.72907	.02616	

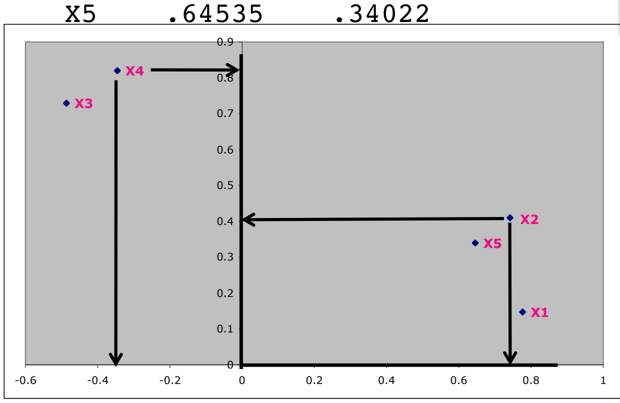
↓ λ_1 = 1.847

↓ λ_2 = 1.589

Loadings = Coordinates

	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
X4	-.34513	.82006
X5	.64535	.34022

Loadings are coordinates in the space of variates



PCA So Far

- Spectral decomposition (original variables → linear combinations)
  $\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$
  - Λ → eigenvalues λ_j (goal = maximal variance for linear combination)
  - **V** → eigenvectors v_j (weighting vector that produces linear combination)
- Extracting r components, whereby normally $r < p$
- Communalities h_i^2 (variable's variance contributing to / shared with / accounted for by components)
- Component ("Factor") **Loadings** a_{ij} $\sum_{j=1}^{r} a_{ij}^2 = h_i^2$ correlation with each component; if squared ➛ shared variance)
- Rotation
- Component/Factor **Scores**

Rotation

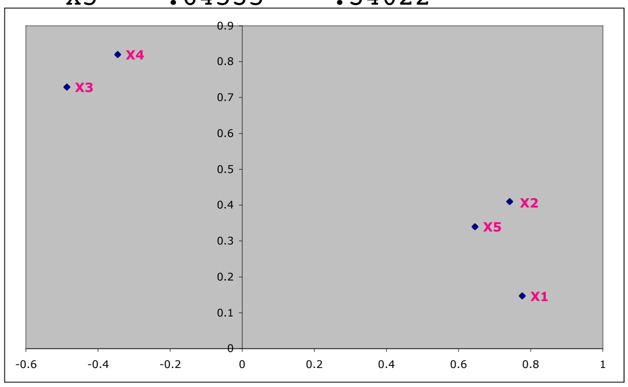
- Changes the reference axes to interpret the meaning of components
- Leaves the relative relationships of variables untouched
- Goal = rearrangement that yields **simple structure**
  - Ideally, each variable loads strongly on only one component.
  - Few components with large variance, each of which is defined by a few variables with large loadings (other variables have near-zero loading)

	C1	C2
X1	1	0
X2	1	0
X3	1	0
X4	0	1
X5	0	1
X6	0	1

Unrotated Loading Matrix

	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
X4	-.34513	.82006
X5	.64535	.34022

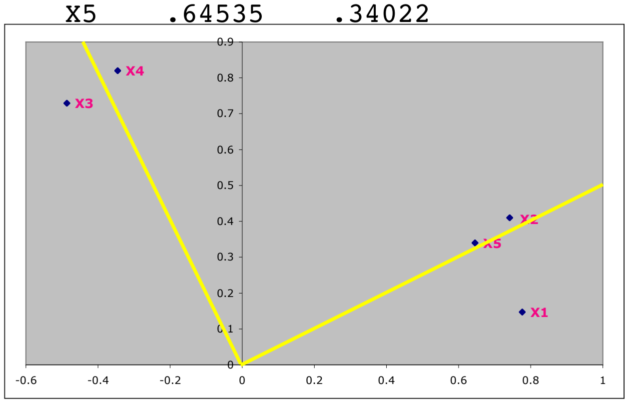
$\left.\right\} \mathbf{A}_{\text{unrotated}}$



Unrotated

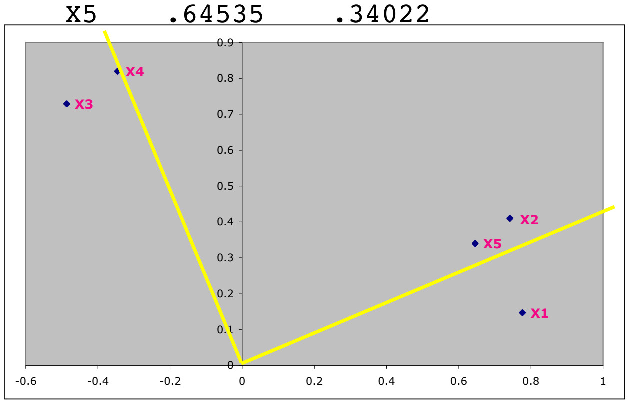
	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
X4	-.34513	.82006
X5	.64535	.34022

Search for better fit of axes and variables



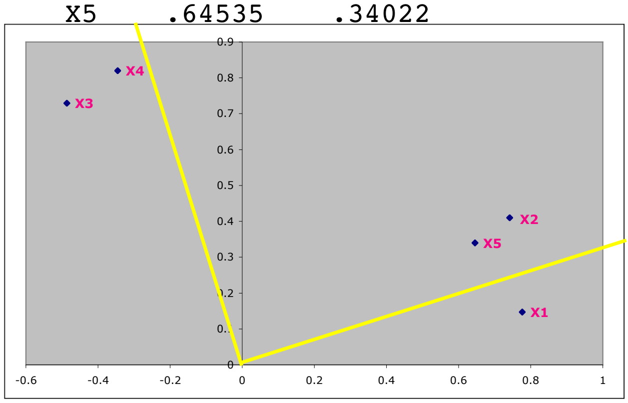
Unrotated

	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
X4	-.34513	.82006
X5	.64535	.34022



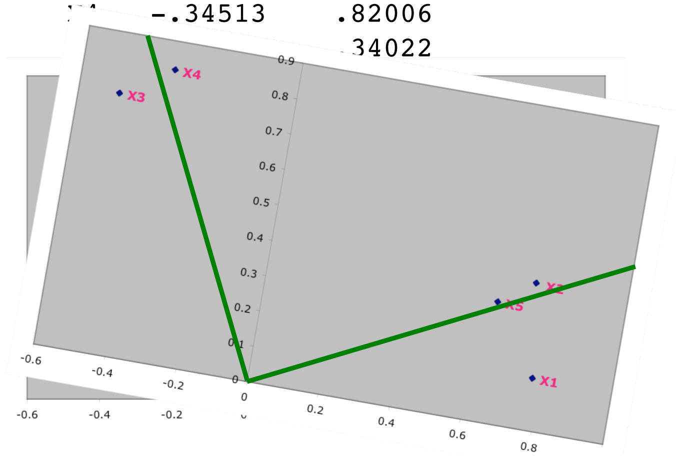
Unrotated

	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
X4	-.34513	.82006
X5	.64535	.34022



Unrotated

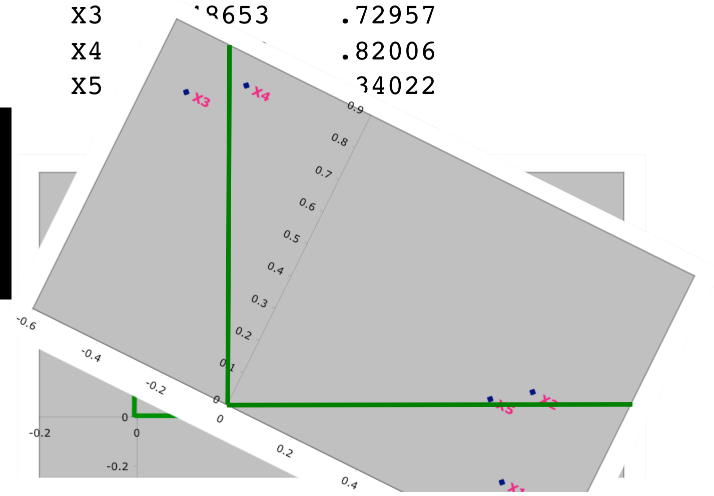
	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	-.48653	.72957
[illegible]	-.34513	.82006
[illegible]	[illegible]	.34022



And Now Rotated…

	Factor 1	Factor 2
X1	.77674	.14728
X2	.74137	.41008
X3	[illegible]8653	.72957
X4	[illegible]	.82006
X5	[illegible]	.34022

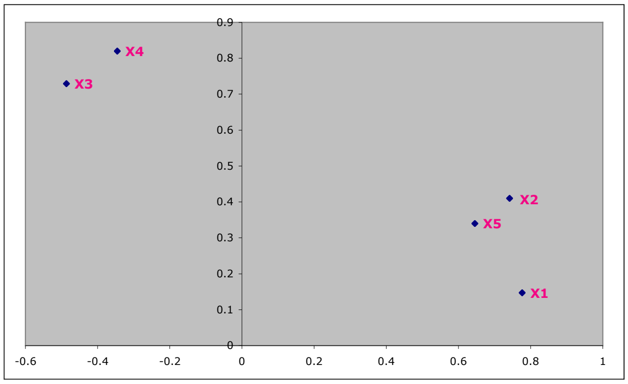
Actual axis rotation: Improved "capture"



Amount of Rotation

Factor Transformation Matrix

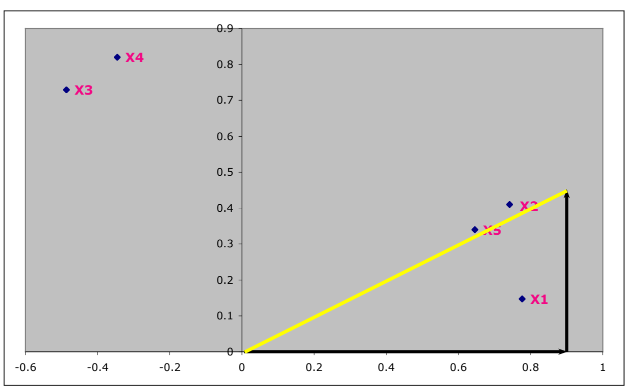
	Factor 1	Factor 2
Factor 1	.90076	-.43433
Factor 2	.43433	.90076



Amount of Rotation

Factor Transformation Matrix

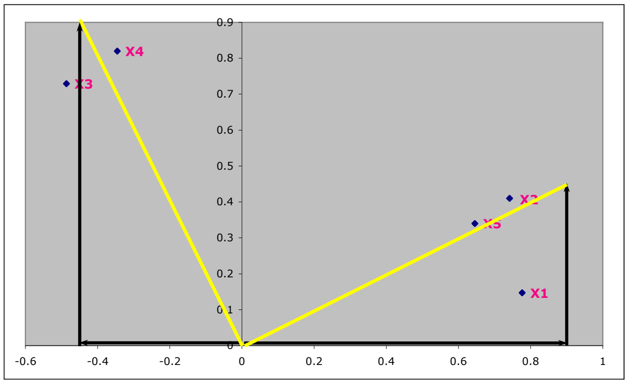
	Factor 1	Factor 2
Factor 1	.90076	-.43433
Factor 2	.43433	.90076



Amount of Rotation

Factor Transformation Matrix

	Factor 1	Factor 2
Factor 1	.90076	-.43433
Factor 2	.43433	.90076



Formal Operation

- Features of Rotation
  - Variables in multidimensional space stay put
  - Only reference axes change
  - Planets-in-space illustration
  - Formal proof of invariance:

$$
\begin{aligned}
R_{xx} &= A\,A' \\
&= (AT)\,(AT)' \\
&= AT\,T'\,A' \qquad T\,T' = I \\
&= A\,A'
\end{aligned}
$$

Component scores

Component scores = "Factor scores" :

- properties of individuals
- reflecting their standing on the r principal components C_j
- Resulting from a transformation of their standing on the original p variables X_i

$$\underset{n\times r}{\mathbf{C}} = \underset{n\times p}{\mathbf{X}} \cdot \underset{p\times r}{\mathbf{W}}$$

Each person's linear combinations of their own variable scores

$$\underset{n\times r}{\mathbf{C}} = \underset{n\times p}{\mathbf{Z}} \cdot \underset{p\times r}{\mathbf{V}}$$

Each person's linear combinations of their **standardized** scores

Output Example