

HW #8 (Last assignment!)

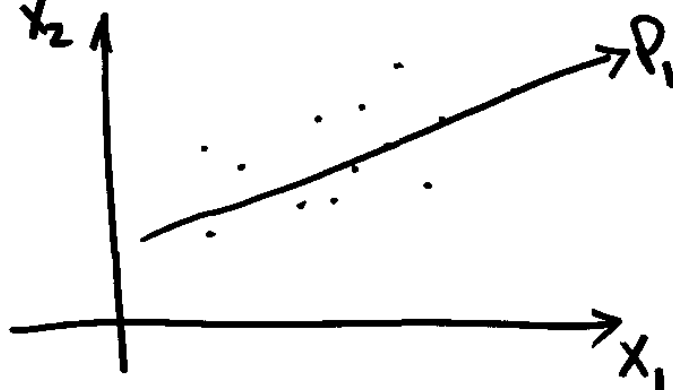
Stat 543

5-19-15

①
Select a statistical procedure, find a data set,
run the analysis, write up a short interpretation
(and be ready to show it to the class)

Principal Components

$P_1 = aX_1 + bX_2$,
so that the dispersion
of points on this
new axis is maximized



P_2, P_3 , etc. must be \perp to all previous
components + again maximize the dispersion

②

Result: You begin with p variables, &
now you have p of new variables.

Benefit: the new variables (the PCs) are
uncorrelated with each other.

The 1st PC contains more "information"
than any single original variable.

How many components to keep?

③

① locate the elbow of the scree plot (2)

② Kaiser's Rule: keep all PCs with eigenvalues > 1 (2)

③ Cumulative proportion — stop when it exceeds a pre-set value (90%) (3)

Factor Analysis

④

- Starts with PC
- Discards the least important components (Kaiser's Rule is the default)
- Geometrically - rotate the remaining axes in all directions (maintaining \perp) to achieve a simpler structure that lends itself to interpretation.

Varimax, Orthomax

⑤

Cluster Analysis (clustering of observations)

	South?	detel?	Party	Congress?	VP?
Reagan	0	1	0	0	0
Carter	1	1	1	0	0
Ford	0	0	0	1	1
Nixon	0	1	0	1	1
Johnson	1	0	1	1	1
Kennedy	0	1	1	1	0

⑥

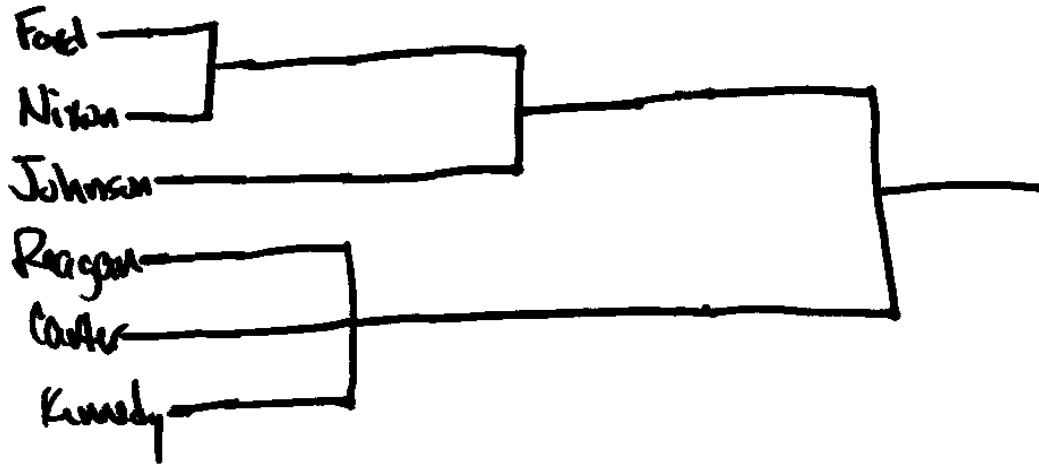
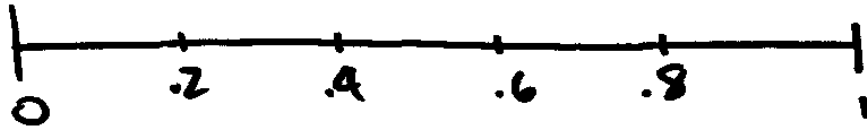
Create similarities or distances

Distances	R	C	F	N	J	K
Reagan	0	.4	.6	.4	1	.4
Carter	.4	0	1	.8	.6	.4
Ford	.6	1	0	.2	.4	.6
Nixon	.4	.8	.2	0	.6	.4
Johnson	1	.6	.4	.6	0	.6
Kennedy	.4	.4	.6	.4	.6	0

Tree diagram (Dendrogram)

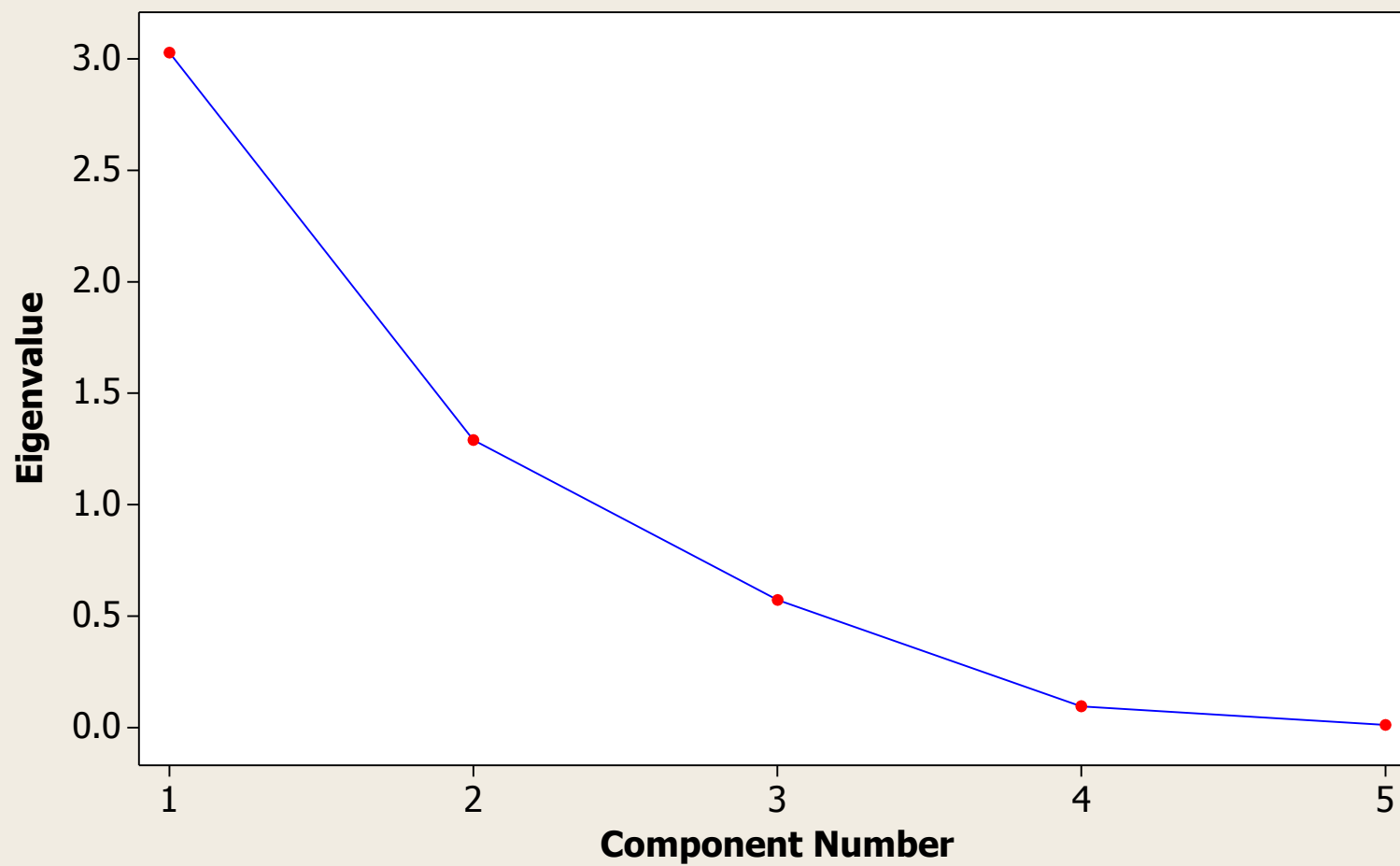
⑦

Farthest neighbor



	C1	C2	C3	C4	C5	C6
	Tract	Population	SchoolYrs	Employment	HealthServEmp	HomeValue
1	1	5.935	14.2	2.265	2.27	2.91
2	2	1.523	13.1	0.597	0.75	2.62
3	3	2.599	12.7	1.237	1.11	1.72
4	4	4.009	15.2	1.649	0.81	3.02
5	5	4.687	14.7	2.312	2.50	2.22
6	6	8.044	15.6	3.641	4.51	2.36
7	7	2.766	13.3	1.244	1.03	1.97
8	8	6.538	17.0	2.618	2.39	1.85
9	9	6.451	12.9	3.147	5.52	2.01
10	10	3.314	12.2	1.606	2.18	1.82
11	11	3.777	13.0	2.119	2.83	1.80
12	12	1.530	13.8	0.798	0.84	4.25
13	13	2.768	13.6	1.336	1.75	2.64
14	14	6.585	14.9	2.763	1.91	3.17

Scree Plot of Population, ..., HomeValue



Correlations: Population, SchoolYrs, Employment, HealthServEmp, HomeValue

	Population	SchoolYrs	Employment	HealthServEmp
SchoolYrs	0.610			
Employment	0.971	0.494		
HealthServEmp	0.740	0.095	0.848	
HomeValue	-0.172	0.186	-0.249	-0.358

Cell Contents: Pearson correlation

Principal Component Analysis: Population, SchoolYrs, Employment, HealthServEm

,

Eigenanalysis of the Correlation Matrix

Eigenvalue	3.0289	1.2911	0.5725	0.0954	0.0121
Proportion	0.606	0.258	0.114	0.019	0.002
Cumulative	0.606	0.864	0.978	0.998	1.000

Variable	PC1	PC2	PC3	PC4	PC5
Population	0.558	0.131	-0.008	-0.551	0.606
SchoolYrs	0.313	0.629	0.549	0.453	-0.007
Employment	0.568	0.004	-0.117	-0.268	-0.769
HealthServEmp	0.487	-0.310	-0.455	0.648	0.201
HomeValue	-0.174	0.701	-0.691	-0.015	-0.014

Scree Plot of Population, ..., HomeValue

Factor Analysis: Population, SchoolYrs, Employment, HealthServEmp, HomeValue

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
Population	0.972	0.149	0.967
SchoolYrs	0.545	0.715	0.808
Employment	0.989	0.005	0.978
HealthServEmp	0.847	-0.352	0.841
HomeValue	-0.303	0.797	0.726
Variance	3.0289	1.2911	4.3200
% Var	0.606	0.258	0.864

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
Population	0.979	0.085	0.967
SchoolYrs	0.591	0.677	0.808
Employment	0.987	-0.060	0.978
HealthServEmp	0.822	-0.406	0.841
HomeValue	-0.251	0.815	0.726
Variance	3.0215	1.2986	4.3200
% Var	0.604	0.260	0.864

Factor Score Coefficients

Variable	Factor1	Factor2
Population	0.328	0.094
SchoolYrs	0.216	0.541
Employment	0.326	-0.018
HealthServEmp	0.261	-0.290
HomeValue	-0.060	0.622