HUDM5124 -- Week 2 assignment: PCA [ALTERNATE VERSION]

In this Canvas module is posted a multivariate data set ("fatality.sav") on auto fatality rates, representing the correlations among six variables measured on a sample of 50 states.

The variables are:

State	name of state
frate	fatalities per 100M vehicle miles driven
drink	min drinking age
gasuse	gasoline use
drivers	total number of drivers (M)
density	population density
temp	mean January daytime low temp

Run a PCA, with varimax rotation (and default parameters otherwise) on these variables. Print out ONLY the derived eigenvalues (and a scree plot) and the rotated components matrix.

Discuss how many components you think should be extracted. Write a brief interpretation of each *rotated* component.

Use SPSS, R, or any other stat package you prefer.

KEY TO ASSN 1: Selected SPSS output:

Correlation Matrix

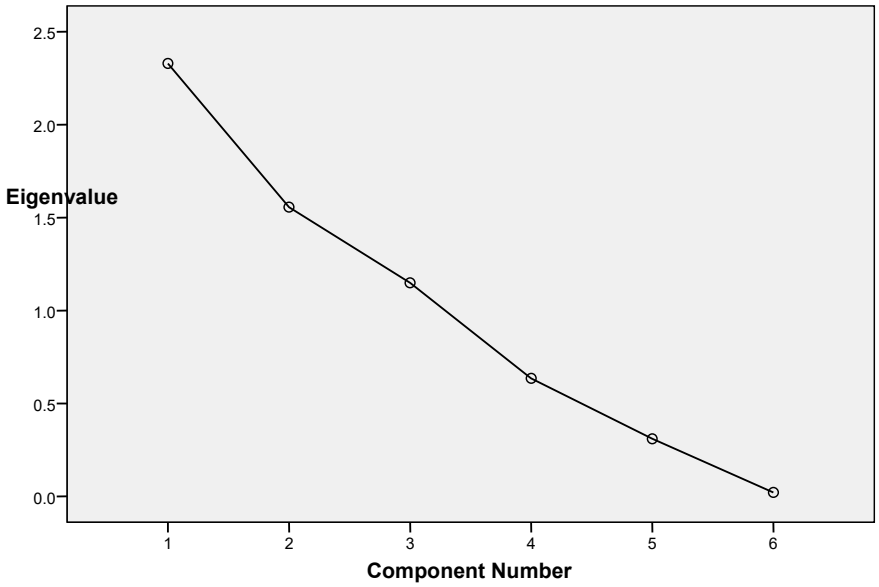
		fatalities per 100M vehicle miles	min drinking age	gasoline use	total number of drivers (M)	population density	mean January daytime low temp
Correlation	fatalities per 100M vehicle miles	1.000	.040	-.213	-.244	-.604	.168
	min drinking age	.040	1.000	.057	.065	-.267	-.135
	gasoline use	-.213	.057	1.000	.977	.133	.352
	total number of drivers (M)	-.244	.065	.977	1.000	.176	.321
	population density	-.604	-.267	.133	.176	1.000	.072
	mean January daytime low temp	.168	-.135	.352	.321	.072	1.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.329	38.822	38.822	2.329	38.822	38.822	2.178	36.301	36.301
2	1.556	25.932	64.755	1.556	25.932	64.755	1.672	27.860	64.161
3	1.149	19.151	83.905	1.149	19.151	83.905	1.185	19.744	83.905
4	.635	10.581	94.487						
5	.310	5.167	99.653						
6	.021	.347	100.000						

Extraction Method: Principal Component Analysis.

Scree Plot



Component Matrix(a)

	Component		
	1	2	3
fatalities per 100M vehicle miles	-.474	.718	.329
min drinking age	-.068	.384	-.792
gasoline use	.915	.318	-.108
total number of drivers (M)	.924	.277	-.135
population density	.466	-.751	.128
mean January daytime low temp	.438	.388	.607

Extraction Method: Principal Component Analysis.
a 3 components extracted.

Rotated Component Matrix(a)

	Component		
	1	2	3
fatalities per 100M vehicle miles	-.132	.902	.133
min drinking age	.153	.200	-.846
gasoline use	.967	-.122	-.027
total number of drivers (M)	.958	-.168	-.045
population density	.084	-.831	.317
mean January daytime low temp	.527	.290	.590

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 6 iterations.

Three components are extracted by the criterion of retaining only those with eigenvalues > 1. Once rotated by the VARIMAX criterion, they can be interpreted as follows:

Component 1: loads high on gas use and number of drivers. This is probably a size-of-state factor

Component 2: loads high on fatality rate and high negative on population density. This seems to suggest that the biggest effect on fatality rate might be population density, with a higher rate in low-density states

Component 1: loads high negative on drinking age, high positive on mean January daytime temperature. It seems to mean that warmer (Sunbelt) states have lower minimum drinking ages (check with a scatterplot?)