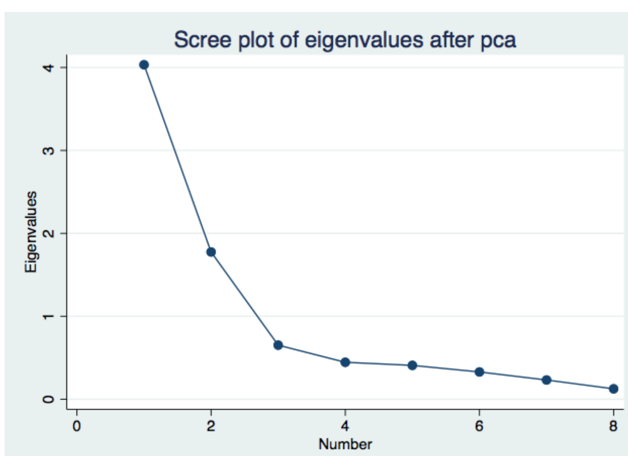
```
* APSTA.GE 2004
* Homework 5
* Jennifer Jackson

use "/Users/jenniferjackson/Desktop/Assignment _5.dta"

* Problem 1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

tabstat stature hip_ht leg_length arm_length sitting_ht knee_girth elbow_girth ///
 wrist_girth, stats(n, mean, sd, skewness) format(%6.0g)

correl stature hip_ht leg_length arm_length sitting_ht knee_girth elbow_girth wrist_girth

* Problem 2 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

pca stature hip_ht leg_length arm_length sitting_ht knee_girth elbow_girth wrist_girth

* Problem 3 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* Because the unexplained values are zero, the total amount of variance per
* variable equals one.

* Problem 4 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* The eigenvalues sum to the number of diagonal elements in the input matrix.
* Because the input matrix here is the correlation matrix, the sum is the same
* as the number of components: seven.

* Problem 5 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* If all the loadings of Composite 1 are positive, that means that's the best
* way to explain the variance among the variables. As for the other composites,
* the signs will have to be adjusted because the sum of the cross-products of
* the factor loadings for each pair of components is zero.

* Problem 6 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

screeplot

* m = 2 factors
```



* Problem 7 ~~~~~~

pca stature hip_ht leg_length arm_length sitting_ht knee_girth elbow_girth wrist_girth, comp(2)

* Problem 8 ~~~~~~

* Because the unexplained variance values are so small, that means that the
* first two components can account for the majority of the variance.

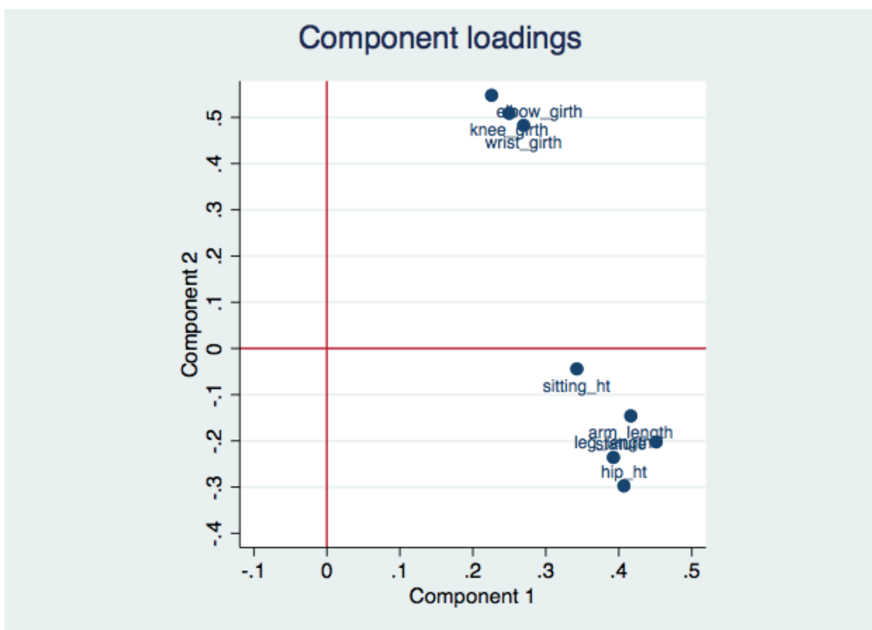
* Problem 9 ~~~~~~

* Only considering the first two components, the sitting height variable still
* has about 52% of its variance unexplained. I.e., sitting height is the least
* well accounted for.

* Problem 10 ~~~~~~

loadingplot, xlab(-.1(.1).5) ylab(-.4(.1).5) aspect(1) yline(0) xline(0)

* The x axis kind of lines up with the variables, but the y axis does not line
* up at all.

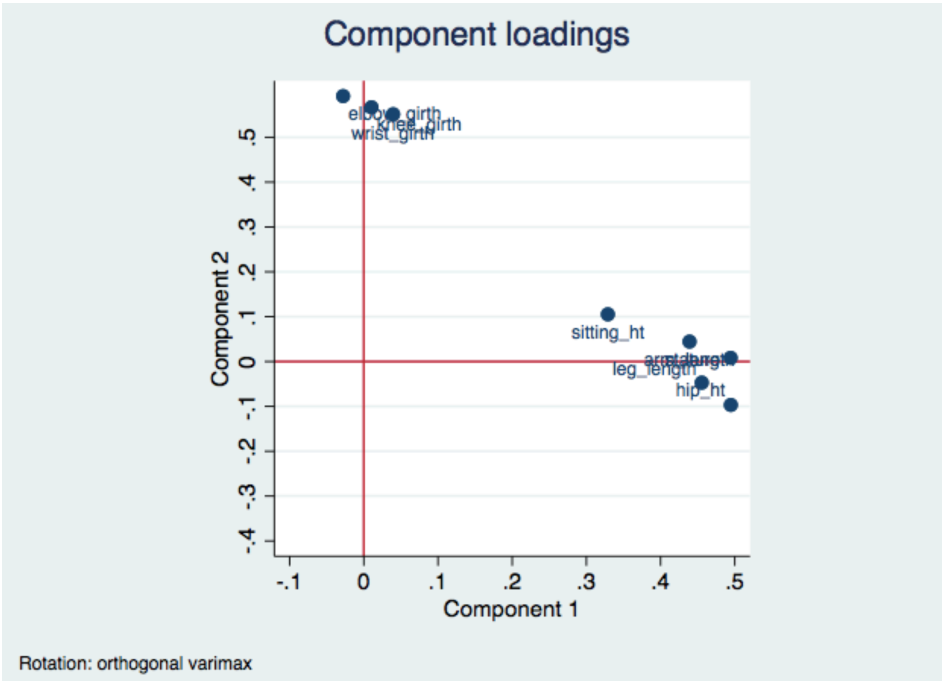


* Problem 11 ~~~~~~

rotate, orthogonal varimax normalize comp(2)

```
* The unexplained amounts of variance remain the same.

* Problem 12
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* Comp1 Rotated
display 0.4944^2 + 0.4947^2 + 0.4555^2 + 0.4390^2 + 0.3289^2 + 0.0104^2 + (-0.0279)^2 +
0.0396^2
* = 1

* Comp2 Rotated
display 0.0081^2 + (-0.0968)^2 + (-0.0471)^2 + 0.0444^2 + 0.1052^2 + 0.5668^2 + 0.5917^2 +
0.5513^2
* = 1

* Comp1 Unrotated
display 0.4512^2 + 0.4070^2 + 0.3926^2 + 0.4164^2 + 0.3425^2 + 0.2497^2 + 0.2257^2 +
0.2696^2
* = 1

* Comp2 Unrotated
display (-0.2023)^2 + (-0.2974)^2 + (-0.2358)^2 + (-0.1460)^2 + (-0.0442)^2 + 0.5089^2 +
0.5477^2 + 0.4825^2
* = 1

* Problem 13
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* Given e(r_T) is the Varimax transformation matrix
matrix B = e(r_T)'
matrix C = e(r_T) * B
matrix list C
* shows the identity matrix (give or take some rounding)

* Problem 14
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

* Rotation redistributes the variance, so the first loading is no longer much
* larger than the rest.

* Problem 15
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

loadingplot, xlab(-.1(.1).5) ylab(-.4(.1).5) aspect(1) yline(0) xline(0)
```

* Yes, the variables are much more closely aligned with the axes after rotation.



* Problem 16
~~~~~~

* I would say Component 1: Joint Girth and Component 2: Length of Limbs and Torso

* To summarise, the majority of variation among physical attributes can be
* accounted for by joint girth and limb/torso length.
~~~~~~