

MSCR 509: High Dimensional Analysis

Homework 10

Anish Shah

April 13, 2020

Question 1

A medical anthropologist is interested in investigating whether there are any differences between male Egyptian skulls of different time periods from the area of Thebes. Various measurements were made on male Egyptian skulls from two different time periods: 4000 BC and 150 AD.

In the EGYPT data set, thirty skulls are available from each time period. For each skull, the following four variables were measured:

- Maximal Breadth of Skull (MB)
- Basibregmatic Height of Skull (BH)
- Basialveolar Length of Skull (BL)
- Nasal Height of Skull (NH)

Part A

Using descriptive statistics, present the data in a table to address the question of Egyptian skulls of different time periods from the area of Thebes.

Table 1: Summary descriptives table by groups of 'Period'

	[ALL] N=60	150 AD N=30	4000 BC N=30	p.overall	p.trend
MB	134 (5.73)	136 (5.35)	131 (5.13)	0.001	0.001
BH	132 (4.97)	130 (4.97)	134 (4.47)	0.010	0.010
BL	96.3 (6.14)	93.5 (5.06)	99.2 (5.88)	<0.001	<0.001
NH	51.0 (3.28)	51.4 (3.72)	50.5 (2.76)	0.329	0.329

Part B

Perform a univariate analysis to determine the association of each variable with the time period (4000BC vs 150 AD) and present in a Table. Interpret results.

Part C

Perform a multivariate analysis to develop a model for predicting the time period in terms of four variables. Present results in a table and interpret your results.

Table 2: OR of Being from 150 AD instead of 4000 BC

	<i>Dependent variable:</i>			
	Period			
	(1)	(2)	(3)	(4)
MB	1.19*** (1.07, 1.36)			
BH		0.86** (0.76, 0.96)		
BL			0.82*** (0.71, 0.91)	
NH				1.08 (0.93, 1.28)
Constant	0.00*** (0.00, 0.0001)	359,622,960.00** (123.10, 10,362,049,049,250,756.00)	268,710,917.00*** (6,151.00, 144,039,127,961,967.00)	0.02 (0.0000, 50.87)
Observations	60	60	60	60
Log Likelihood	-35.75	-38.12	-34.16	-41.09
Akaike Inf. Crit.	75.51	80.24	72.32	86.18

Note:

*p<0.1; **p<0.05; ***p<0.01

Multivariate Prediction of Time Period

Dependent variable:

Period

MB	1.22*** (1.07, 1.44)
BH	0.88 (0.74, 1.01)
BL	0.84** (0.71, 0.95)
NH	1.01 (0.82, 1.24)
Constant	831.50 (0.00, 313,546,479,028,556.00)

Observations 60
Log Likelihood -28.03
Akaike Inf. Crit. 66.05

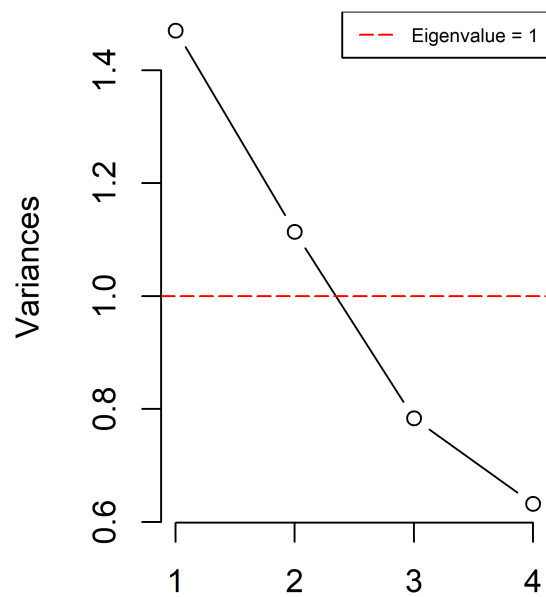
===== Note: $p < 0.1$;
 $p < 0.05$; $p < 0.01$

Based on a multivariate analysis, MB and BL are helpful in predicting the time period of the skull. As MB increases, the OR for being from 150 AD increases, and as BL increases, the OR for being from 150 AD decreases.

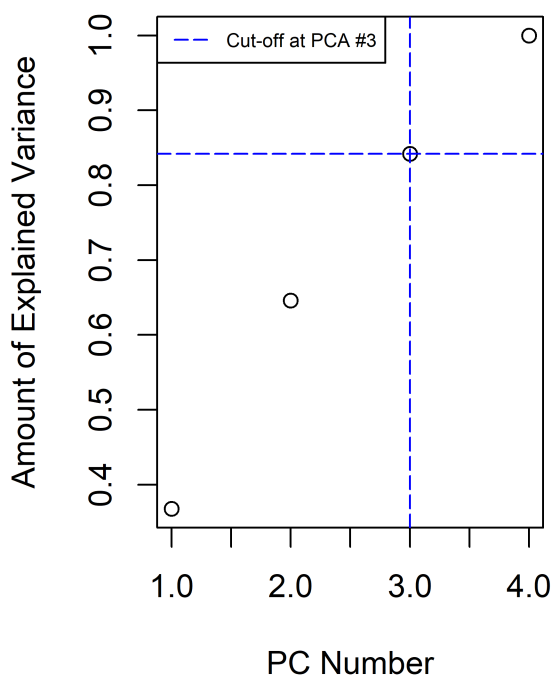
Part D

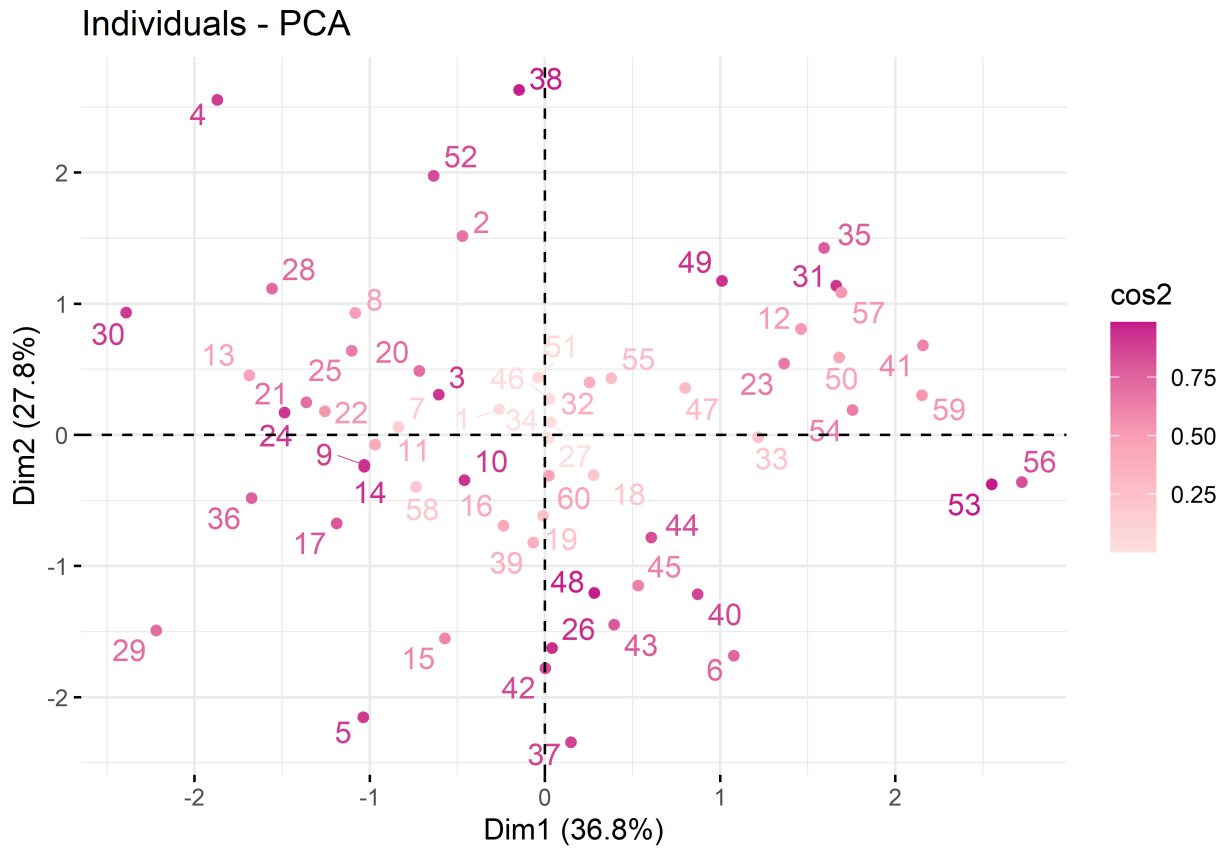
Perform a PCA analysis on the four variables and comment on the variation explained by each principal component. How many PCAs would you select to describe the measurements of skulls?

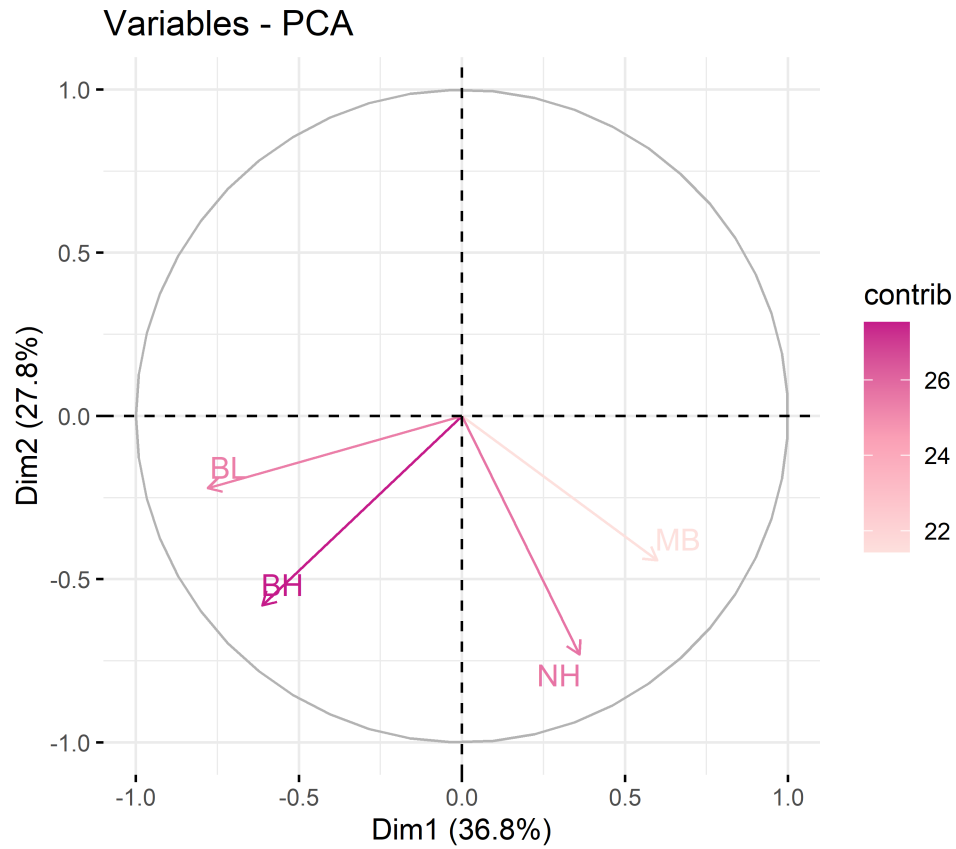
Screeplot of All PCs



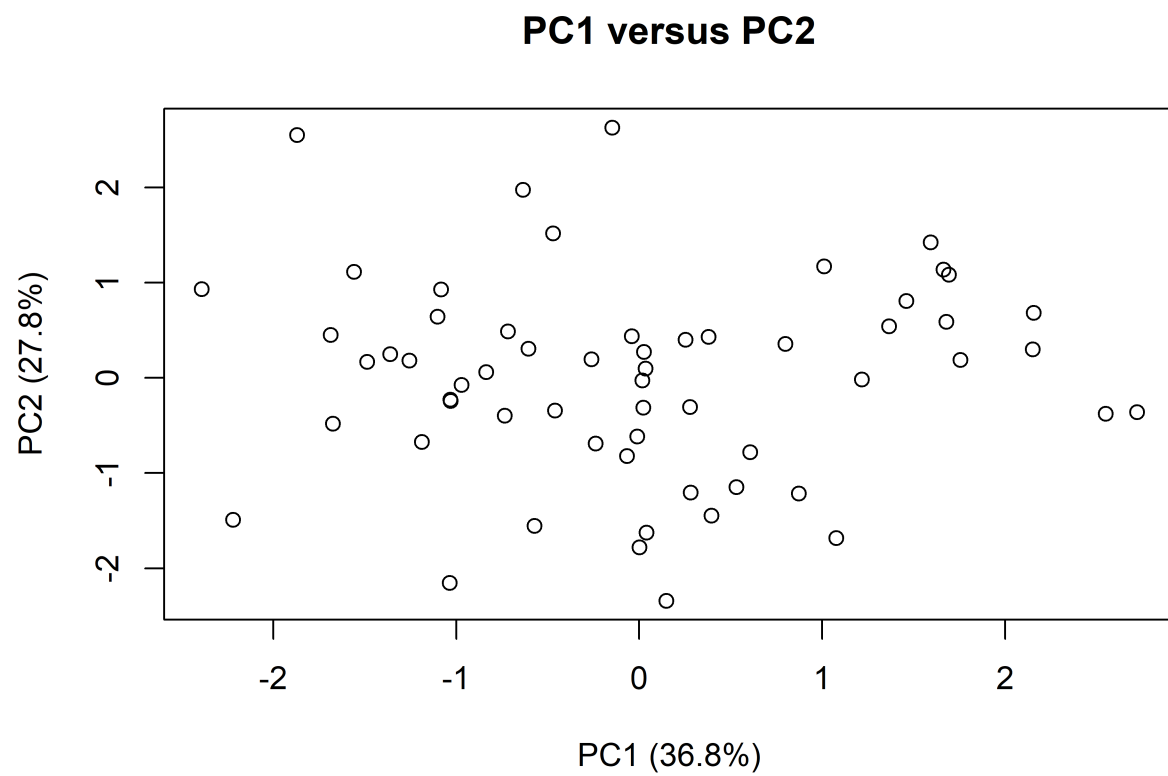
Cumulative Variance Plot



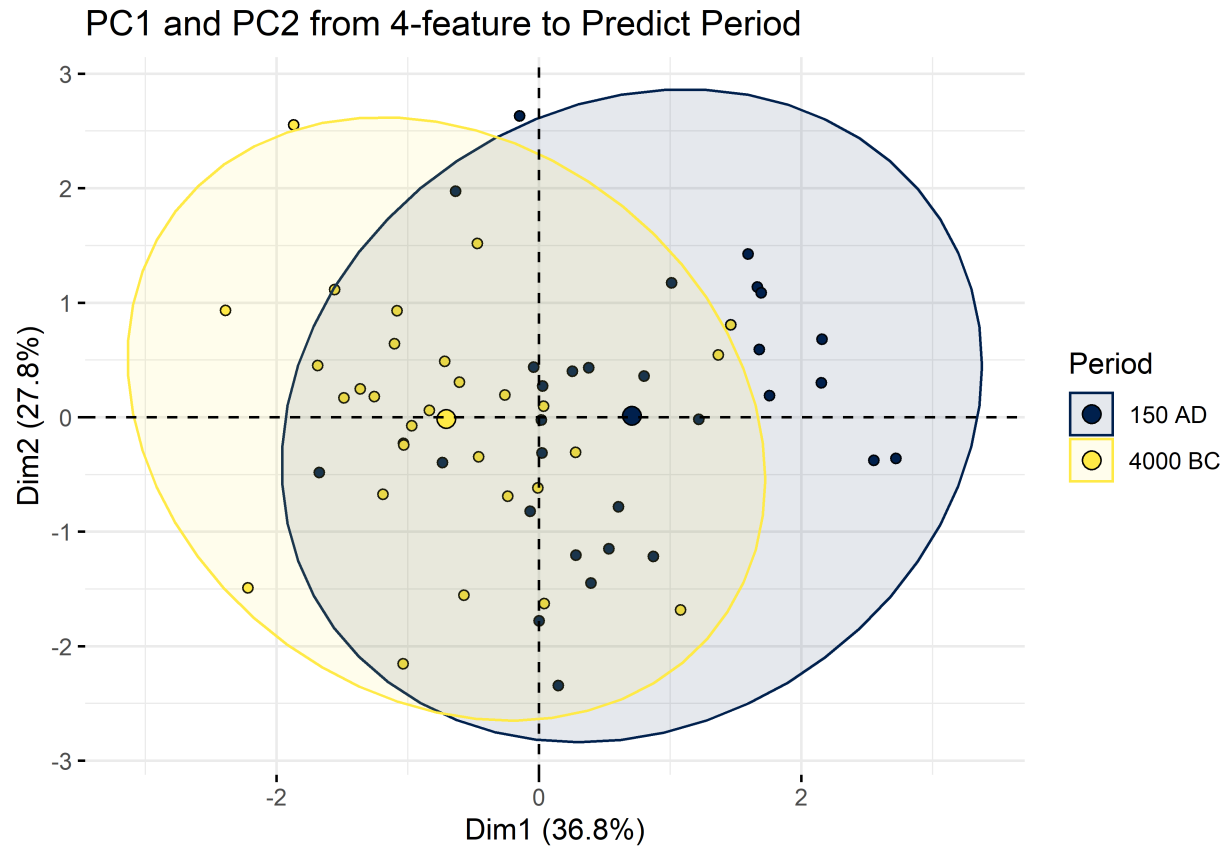




This suggests that most of our variance of findings can explained by the first 2 PC, but the 3rd PC brings us past the eigenvalue of 1. Thus, at minimum, teh 4th PC is not very helpful in explaining hte variance total.



Just the first two components explain about 64.6% of the total variance of the Egyptian skulls.



Based on this information, I would likely just use 2 principal components to help identify the period of the skulls.