Homework 4

testdata.txt concerns the results of 88 students on a five part exam:
1) Closed Book on Mechanics
2) Closed Book on Vectors;
3) Open Book on Algebra;
4) Open Book on Analysis
5) Open Book in Statistics.

Homework 4

a) Choose to either use the correlation matrix or the covariance matrix for your principal components analysis and justify your choice.

b) Describe how much information would be lost if the data was summarized using only 1, 2, 3, or 4 components.
Homework 4

c) Find an interpretation for what each of the first four components measure.

d) Which of the five variables are explained very well by the first four components, and which are not?

Factor Analysis

Factor Analysis models are designed to account for this missing data by adding an error term to the model. (Assuming the $X$ have been standardized we can write):

\[
X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1k}F_k + u_1
\]

\[
X_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2k}F_k + u_2
\]

\[
\vdots
\]

\[
X_q = \lambda_{q1}F_1 + \lambda_{q2}F_2 + \cdots + \lambda_{qk}F_k + u_q
\]

Assumptions

• The $u_i$ are independent of each other and of the $F_i$

• The $F_i$ are independent of each other

• Usually set the $F_i$ to have mean 0 and variance 1
The Variances

We have:

\[ X_i = \lambda_{i1}F_1 + \lambda_{i2}F_2 + \cdots + \lambda_{ik}F_k + u_i \]

What can we tell about the variances and covariances?

In Matrix Form

If we write

\[ X_{qx1} = \Lambda_{qdx}F_{kx1} + u_{qx1} \]

Then

\[ \Sigma = \Lambda_{qdx}\Lambda_{kxq}^T + \text{cov}(u) \]

What’s Next

- How does it work?
- Any restrictions on the data?
- How many factors?
- Rotating the data
- Interpreting the results
- Using the results
- Graphically displaying the findings
> beardat<-read.table("http://www.stat.sc.edu/~habing/courses/data/bears.txt",head=T)
> bears<-beardat[,3:7]
> source("http://www.stat.sc.edu/~habing/courses/530/fact.txt")

> fact(bears,method="pf",rotation="none",maxfactors=2)
$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "principal factor - no rotation"

$loadings
          Factor1 Factor2
Head.L    0.937   0.151
Head.W    0.812  -0.367
Neck.G    0.955  -0.084
Length    0.931   0.215
Chest.G   0.948   0.037

$communalities
Head.L  0.901   0.793   0.919   0.914   0.900

$importance
          Factor1 Factor2
variance explained  4.215   0.212
percent explained    0.843   0.042

$residuals
          Min. 1st Qu. Median    Mean 3rd Qu.    Max.
-0.024  -0.009  -0.002  -0.002  0.007   0.017
> fact(bears, method="pf", rotation="none", maxfactors=2)  #CONTINUED
> fact(bears, method="pca", rotation="none", maxfactors=2)
$eigen.values
[1] 4.317 0.378 0.147 0.096 0.063

$method
[1] "principal components - no rotation"

$loadings
   Factor1 Factor2
Head.L   0.945   0.178
Head.W   0.842  -0.528
Neck.G   0.961  -0.034
Length   0.939   0.242
Chest.G  0.954   0.086

$communalities
Head.L  0.924   0.988   0.926   0.940   0.918

$importance
      Factor1 Factor2
variance explained  4.317   0.378
percent explained    0.863   0.076

$residuals
        Min.  1st Qu.   Median   Mean  3rd Qu.    Max.
-0.047   -0.027  -0.020   -0.015   0.000    0.020
> fact(bears, method="pca", rotation="none", maxfactors=2)  # CONTINUED

Variances

<table>
<thead>
<tr>
<th>Variances</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Bar Chart: Variances](image)

Predicted Correlation vs. Original Correlation

![Scatter Plot: Predicted Correlation vs. Original Correlation](image)

**r-squared = 0.484**

Residual Correlation vs. Predicted Correlation

![Scatter Plot: Residual Correlation vs. Predicted Correlation](image)

**Mean = -0.015  s.d. = 0.021**

Frequency of Residual Correlations

![Histogram: Frequency of Residual Correlations](image)