STAT 530/J530
September 22nd, 2005

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Principal Components

$Y=a_1X_1+a_2X_2+\cdots+a_qX_q$

or

$Y_{(nxq)} = X_{(nxq)} A_{(qxq)}$

Reversing It

If we have Y we should be able to get X... say using regression.

In fact in this case we should get it exactly!

$Y_{(nxq)} A^{-1}_{(qxq)} = X_{(nxq)} A_{(qxq)} A^{-1}_{(qxq)}$

$X_{(nxq)} = Y_{(nxq)} A^{-1}_{(qxq)}$

Checking What Was Lost

```
sect3<-as.matrix(oildata[,12:31])

a<-eigen(cov(sect3))$vectors
y<-sect3%*%a

summary(lm(sect3[,1]~y))$r.squared
```


What if we only kept 4 PC's?

```
y4<-y[,1:4]
summary(lm(sect3[,1]~y4))$r.squared

sect3est<-sect3
for (i in 1:20){
  sect3est[,i]<-
    lm(sect3[,i]~y4)$fitted.values}
```


Lots of Lost Information!

```
s3var<-apply(sect3,2,var)

s3varest<-apply(sect3est,2,var)

cbind(s3var,s3varest)

cor(sect3)[1:4,1:4]
cor(sect3est)[1:4,1:4]
```

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 set to have mean 0.)

$$X_1=a_{11}F_1+a_{12}F_2+\cdots+a_{1k}F_k+\varepsilon_1$$
$$X_2=a_{21}F_1+a_{22}F_2+\cdots+a_{2k}F_k+\varepsilon_2$$
$$\vdots$$
$$X_q=a_{q1}F_1+a_{q2}F_2+\cdots+a_{qk}F_k+\varepsilon_q$$

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 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

$X_{qx1}=\Lambda_{qxk}F_{kx1}+u_{qx1}$

Then

$\Sigma=\Lambda_{qxk}\Lambda^T_{kxq}+cov(u)$

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