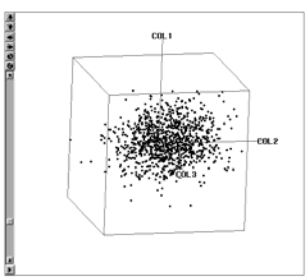
STAT 530/J530
September 15th, 2005

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

3-D Example




```
PROC IML;
sigma = {1 .2 2,
         .2 1 2,
         2 2 10};
mu = {0 0 0};
n = 1000;
seed = 91505;
q=NROW(sigma);
MUMAT=REPEAT(mu,n,1);
SROOT=ROOT(sigma);
Z=NORMAL(REPEAT(seed,n,q));
x=Z*SROOT+MUMAT;
CREATE mvnormdata FROM x;
APPEND FROM x;
QUIT;
```


Goal

Find the coefficients (a's) of the x's so that:

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

has the largest possible variance subject to the condition that the length of the coefficient vector is 1.

Example Cont.

```
library(MASS)
mu<-c(5,0,-1)
sigma<-
  matrix(c(1,0.2,2,0.2,1,2,2,2,10),
ncol=3,byrow=T)
x<-mvrnorm(n=1000,mu,sigma)

coef<-princomp(x,cor=F)$loadings[,1]

pc1<-princomp(x,cor=F)$scores[,1]
```


Example Cont.

We should be able to check that the values are correct using the formula before… so we get:

```
pc1[1]

x[1,1]*coef[1]* +x[1,2]*coef[2]
    +x[1,3]*coef[3]*
```

Are they all off?!?

Some Matrix Background

The dot product of two vectors a and b is $a_1b_1+a_2b_2+\cdots+a_qb_q$

In vector notation this is:

It relates to distance by:

It relates to orthogonality by:

The Next Principal Component

The coefficient vector should be length 1.

The coefficient vector should be orthogonal to the previous one(s).

It should explain the largest possible amount of variance.

Eigenvalues and Eigenvectors

The eigenvectors of the covariance matrix are exactly the coefficients for principal components…

And the eigenvalues are the variances of the new variables!

Eigenvalues and Eigenvectors?!?

Consider the equation:

$$(\Sigma-\lambda I)\mathbf{x}=0$$

λ is an eigenvalue

x is an eigenvector

We typically make the length of x be 1.

We typically order them from largest λ to smallest.

Example Cont.

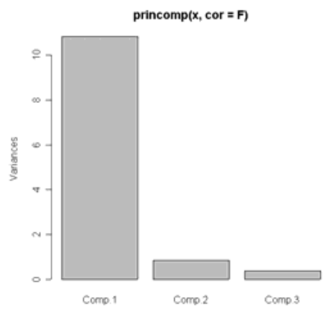
```
eigen(cov(x))$vectors
princomp(x,cor=F)$loadings

eigen(cov(x))$values
round(var(princomp(x,cor=F)$scores),2)
```

STAT 530/J530 B.Habing Univ. of S.C. 10

Scree Plot

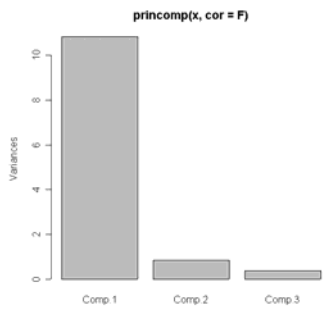
```
plot(princomp(x,cor=F))
```



STAT 530/J530 B.I 11

Oildata

```
plot(princomp(x,cor=F))
```



STAT 530/J530 B.I 12