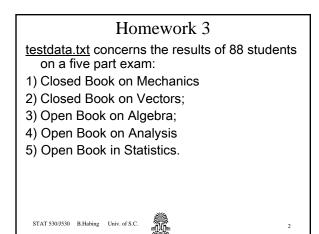


M

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



Homework 3

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

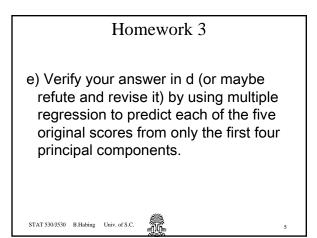
b) Describe how much information would be lost if the data was summarized using only 1, 2, 3, or 4 components.

Homework 3

- c) Find an interpretation for what each of the first four components measure.
- d) Based on your answers to B and C, what information in particular seems to be lost by using only the first four components.

4

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



Factor Analysis

Assuming the X have been standardized:

$$X_{1} = \lambda_{11}F_{1} + \lambda_{12}F_{2} + \dots + \lambda_{1k}F_{k} + u_{1}$$

$$X_{2} = \lambda_{21}F_{1} + \lambda_{22}F_{2} + \dots + \lambda_{2k}F_{k} + u_{2}$$

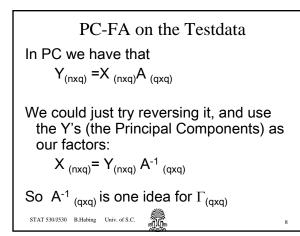
$$\vdots$$

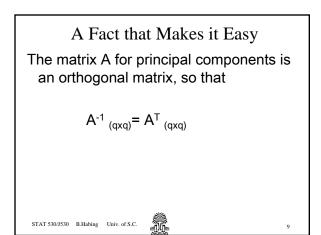
$$X_{q} = \lambda_{q1}F_{1} + \lambda_{q2}F_{2} + \dots + \lambda_{qk}F_{k} + u_{q}$$
STAT 530/550 B.Habing Univ. of S.C.

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

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





And so....

The only problem is that the Y have variances equal to the eigenvalues and we want the F to have variance 1.

So before taking the transpose, multiply each column of A by the standard deviation of that component (the square root of the eigen value).

10

11

12

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

How Well Does it Work?

- What are the communalities? (Sum of the λ_i^2)
- What are the specificities?
 (1-Sum of the λ_i²)
- Does the model give the right correlation?

($\Lambda_{qxk}\Lambda^{\mathsf{T}}_{kxq}\text{+}\text{diag}(\text{specificities})$)

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

Can We Do Better

Principal Factor Factor Analysis says to estimate the specificity and subtract that from the correlation matrix first.

Then do principal components on that "reduced correlation matrix".

For example, could use the largest correlation a variable has with any other.