37. Principal Components Analysis (PCA): What is principal components analysis? How does PCA eliminate the problem of multicollinearity? What does it mean for X1 and X2 to be orthogonal? In order to better understand orthogonality, take the building prices data set and perform these steps: (a) Perform a PROC CORR on X1-X9. (b) Create nine orthogonal predictor variables using PCA. Call these variables Z1-Z9. (c) Perform a PROC CORR on Z1-Z9.

• *Statistical Interpretation* - PCA is a transformation of a set of correlated random variables to a set of uncorrelated

(or orthogonal) random variables

• *Linear Algebra Interpretation* - PCA is a rotation of the coordinate system to the canonical coordinate system,

i.e. the natural coordinate system defined by the variation in the data.

• *Numerical Linear Algebra Interpretation* - PCA is a reduced rank matrix approximation that facilitates dimension

reduction.

**Facts and Caveats**

• PCA does not require any statistical assumptions, e.g. the data are not assumed to have a multivariate normal

distribution.

• PCA is a (numerical) linear algebra technique, i.e. it relies on a matrix factorization (the Spectral Decomposition

or Singular Value Decomposition).

• PCA is sensitive to the scale of the data. Most of the time the data should be standardized, i.e. the variables

should have a (0,1) distribution. When the data are standardized our covariance matrix and correlation matrix are

the same matrix.

• If the data are ‘standardized’ to a common scale that is not (0,1), then it should not be standardized to a (0,1)

distribution.

**Why do we use PCA?**

• PCA can be used in its own right to understand the covariance structure in multivariate data with respect to the

measured basis.

• PCA can be used as a method to create a reduced rank approximation to the covariance structure, i.e. PCA can

be used to approximate the variation in *p* predictor variables using *k < p* principal components. This property is

typically referred to as dimension reduction.

• PCA can be used as a means of creating a set of orthogonal predictor variables from a set of raw predictor

variables. Since the principal components created from the original predictor variables are orthogonal, we can use

PCA as a remedy for multicollinearity in regression problems or as a preconditioner to cluster analysis.

**How do we compute the principal components?**

• Consider the *n* × *p* data matrix of predictor variables *X* = [*X*1*, . . . ,Xp*].

• Depending on your software the data may need to be standardized before the principal components are computed.

This is typically true if you use a software to compute eigenvalues and eigenvectors. Statistical software designed

to perform PCA, such as princomp in SAS, will typically internally standardize the data for you.

• Compute the eigenvalue-eigenvector pairs (*\_*1*, e*1)*, . . . ,* (*\_p, ep*) of the square matrix *XTX* where the eigenvalues

are ordered largest to smallest such that *\_i > \_j* for *i > j*.

• Your software will compute the eigenvalue-eigenvector pairs using a matrix factorization called Singular Value

Decomposition or SVD.

• Compute the principal components *Z*1*, . . . ,Zp* using the eigenvalues as the component loadings

• In vector format we can compute each component individual

*Zi* = *X* × *ei*

or we can compute all of the principal components using one matrix computation

[*Z*1*, . . . ,Zp*] = *X* × [*e*1*, . . . , ep*]

**How many principal components should we use?**

• A *p* × *p* matrix will yield *p* principal components if all of the eigenvalues are non-zero.

• One standard approach to selecting the number of principal components to keep is to use the *scree* plot. The

*scree* plot plots the number of components on the x-axis against the proportion of the variance explained on the

y-axis. The suggested number of principal components to keep is the number where the plot forms an ‘elbow’,

i.e. the point where the curve starts to flatten out.

• Another rule for selecting the number of principal components to keep is to use the minimum eigenvalue rule. A

frequently used rule is the *Kaiser Rule*, which recommends that the number of principal components to keep is

equal to the number of eigenvalues grater than one.

• Other rules exist and ad hoc decisions can be made. Keep in mind that in some problems you might keep all of

the principal components.

**–** Example: Keep at least as many principal components needed to explain at least 70% of the total variation

in the data.

**How do I know if I have kept the correct number of principal components?**

• Frequently the scree plot will present some ambiguity in the number of components to keep, e.g. should I keep

four or five principle components?

• The ‘correct’ number of principal components to keep will depend on the application. If you are using PCA as a

preconditioner for regression analysis or cluster analysis, then the effectiveness of these applications under the

alternate choices would determine which number is the best to keep. In this sense the unsupervised learning

problem has been transformed into a supervised learning problem.

• If the PCA is not directly tied to any application, then the choice of the number of components to keep is always

heuristic. Formal inference for the number of components is available under a multivariate normal distribution

assumption.

38.)Principal Components Analysis is described as a method of 'dimension reduction'. How does PCA reduce the dimension of a statistical problem? How do you select the reduced dimension for your problem.

Kaiser rule or the elbow method in the scree diagram.

(39) Are the factor scores always orthogonal? Are they orthogonal after a rotation? (40) If two analysts perform a factor analysis, are they likely to arrive at the same result? If the same two analysts perform a principal components analysis, are they likely to get the same result?

PCA always same results Factor Analysis not necessarily so.

(41) What is the first step in performing a factor analysis? (42) In the context of factor analysis, what is the communality of factors?