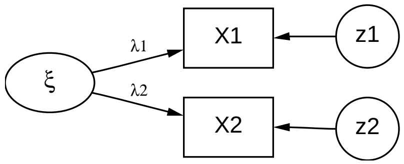
# Exploratory and Confirmatory Factor Analysis

Michael Friendly

Psychology 6140




## Course Outline

1. Principal components analysis
   - FA vs. PCA
   - Least squares fit to a data matrix
   - Biplots
2. Basic Ideas of Factor Analysis
   - Parsimony– common variance $\rightarrow$ small number of factors.
   - Linear regression on common factors
   - Partial linear independence
   - Common vs. unique variance
3. The Common Factor Model
   - Factoring methods: Principal factors, Unweighted Least Squares, Maximum likelihood
   - Factor rotation
4. Confirmatory Factor Analysis
   - Development of CFA models
   - Applications of CFA

## Part 1: Outline





## Why do Factor Analysis?

- Data Reduction: Replace a large number of variables with a smaller number which reflect most of the original data [PCA rather than FA]
  **Example**: In a study of the reactions of cancer patients to radiotherapy, measurements were made on 10 different reaction variables. Because it was difficult to interpret all 10 variables together, PCA was used to find simpler measure(s) of patient response to treatment that contained most of the information in data.
- Test and Scale Construction: Develop tests and scales which are “pure” measures of some construct.
  **Example**: In developing a test of English as a Second Language, investigators calculate correlations among the *item scores*, and use FA to construct subscales. Any items which load on more than one factor or which have low loadings on their main factor are revised or dropped from the test.

# Why do Factor Analysis?

- Operational definition of theoretical constructs:
  - To what extent different observed variables measure the the same thing?
  - Validity: Do they all measure it equally well?

  **Example**: A researcher has developed 2 rating scales for assertiveness, and has several observational measures as well. They should all measure a single common factor, and the best measure is the one with the greatest common variance.
- Theory construction:
  - Several observed measures for each theoretical construct (factors)
  - How are the underlying factors related?

  **Example**: A researcher has several measures of Academic self-concept, and several measures of educational aspirations. What is the correlation between the underlying, latent variables?



# Why do Factor Analysis?

- Factorial invariance: Test equivalence of factor structures across several groups.
  - Same factor loadings?
  - Same factor correlations?
  - Same factor means?

  **Example**: A researcher wishes to determine if normal people and depressive patients have equivalent factor structures on scales of intimacy and attachment she developed.
  The most sensitive inferences about *mean differences* on these scales assume that the relationships between the observed variables (subscales) and the factor are the same for the two groups.



# Two modes of Factor Analysis

- ***Exploratory Factor Analysis***: Examine and explore the interdependence among the observed variables in some set.
  - Still widely used today ($\sim$ 50%)
  - Use to develop a structural theory: how many factors?
  - Use to select “best” measures of a construct.
- ***Confirmatory Factor Analysis***: Test *specific* hypotheses about the factorial structure of observed variables.
  - Does for FA what ANOVA does for studying relations among group means.
  - Requires much more substantive knowledge by the researcher.
  - Provides exactly the methodology required to settle theoretical controversies.
  - Requires moderately large sample sizes for precise tests.



# Principal component analysis vs. Factor analysis

**Principal Components**

- A descriptive *method* for data reduction.
- Accounts for *variance* of the data.
- Scale dependent (***R*** vs. ***S***)
- Components are always uncorrelated
- Components are linear combinations of *observed* variables.
- Scores on components can be computed exactly.

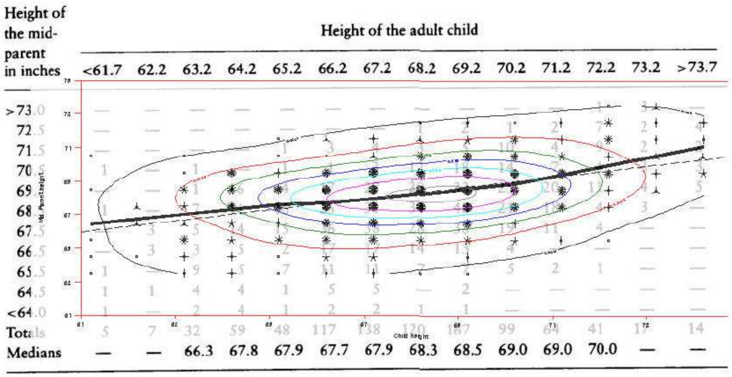
**Factor analysis**

- A statistical *model* which can be tested.
- Accounts for pattern of *correlations*.
- Scale free (ML, GLS)
- Factors may be correlated or uncorrelated
- Factors are linear combinations of *common parts* of variables (unobservable variables)
- Scores on factors must always be estimated (even from population correlations)

# Brief history of Factor Analysis

## Early origins

- Galton (1886): “regression toward the mean” in heritable traits (e.g., height)




- Pearson (1896): mathematical formulation of correlation



# Brief history of Factor Analysis I

## Early origins

- Spearman (1904): “General intelligence,” objectively determined and measured
  - Proposes that performance on any *observable* measure of mental ability is a function of two *unobservable* quantities, or *factors*:
  - General ability factor, **g** — common to all such tests
  - Specific ability factor, **u** — measured only by that particular test
  - “Proof:” tetrad differences = 0 $\rightarrow$ rank(R) = 1
- “Factoring” a matrix
  - Hotelling (1933): Principal components analysis
  - Eckart & Young (1937): Singular value decomposition $\rightarrow$ biplot



# Brief history of Factor Analysis

## Early origins

- Thurstone (1935): *Vectors of the mind*; Thurstone(1947): *Multiple factor analysis*
  - Common factor model— only general, common factors could contribute to correlations among the observed variables.
  - Multiple factor model— two or more common factors $+$ specific factors
  - Primary Mental Abilities— attempt to devise tests to measure multiple facets of general intelligence
- Thurstone (1947): rotation to simple structure
- Kaiser (1953): Idea of analytic rotations (varimax) for factor solutions



# Brief history of Factor Analysis

## Modern development

- Lawley & Maxwell (1973): Factor analysis as statistical model, MLE
  - $\rightarrow$ (large-sample) $\chi^2$ hypothesis test for # of common factors
- Confirmatory factor analysis
  - Jöreskog (1969): confirmatory maximum liklihood factor analysis– by imposing restrictions on the factor loadings
  - Jöreskog (1972): ACOVS model— includes “higher-order” factors
- Structural equation models
  - Jöreskog (1976): LISREL model— separates the measurement model relating observed variables to latent variables from the structural model relating variables to each other.

# Principal components

- ***Purpose***: To summarize the variation of several numeric variables by a smaller number of new variables, called *components*.
- The components are linear combinations— weighted sums— of the original variables.

$$z_1 \equiv \mathrm{PC}_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = \mathbf{a}_1^\mathsf{T}\mathbf{x}$$

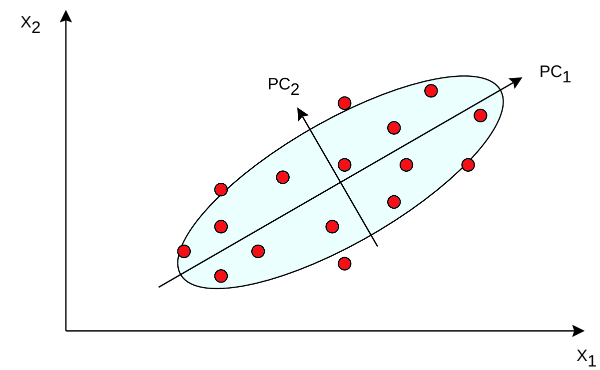
- The first principal component is the linear combination which explains as much variation in the raw data as possible.
- The second principal component is the linear combination which explains as much variation not extracted by the first component

$$z_2 \equiv \mathrm{PC}_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = \mathbf{a}_2^\mathsf{T}\mathbf{x}$$



# Principal components

- The principal component scores are uncorrelated with each other. They represent uncorrelated (orthogonal) directions in the space of the original variables.

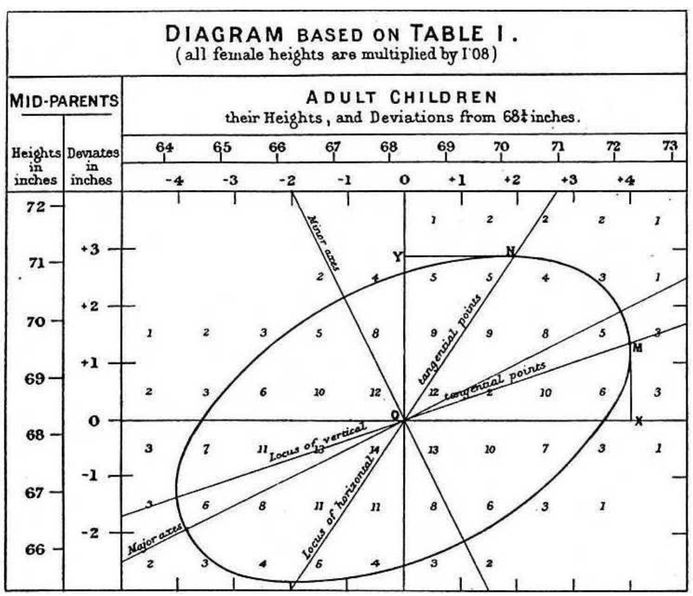



- The first several principal components explain as much variation from the raw data as possible, using that number of linear combinations.



# Principal components

Galton's regresion/correlation/PCA diagram






# Artificial PCA example

- Some artificial data, on two variables, X and Y.
- We also create some linear combinations of X and Y, named A, B and C.

```
A = X + Y
B = 5*X + Y
C = -2*X + Y
```

The data looks like this:

| X | Y | A | B | C |
|---|---|---|---|---|
| 14 | 1 | 15 | 71 | -27 |
| 12 | 2 | 14 | 62 | -22 |
| 11 | 2 | 13 | 57 | -20 |
| 9 | 3 | 12 | 48 | -15 |
| 10 | 3 | 13 | 53 | -17 |
| 11 | 3 | 14 | 58 | -19 |
| ... | ... | ... | ... | ... |
| 1 | 10 | 11 | 15 | 8 |
| 2 | 10 | 12 | 20 | 6 |

- How much of the variance of X and Y do different linear combinations account for?

From simple regression, the *proportion* of variance of X accounted for by any other variable, say A, is just $r^2_{XA}$.
The correlations among these variables are:

| _NAME_ | X | Y | A | B | C |
|---|---|---|---|---|---|
| X | 1.000 | -0.866 | 0.764 | 0.997 | -0.991 |
| Y | -0.866 | 1.000 | -0.339 | -0.824 | 0.924 |
| A | 0.764 | -0.339 | 1.000 | 0.812 | -0.673 |
| B | 0.997 | -0.824 | 0.812 | 1.000 | -0.978 |
| C | -0.991 | 0.924 | -0.673 | -0.978 | 1.000 |

The variances are:

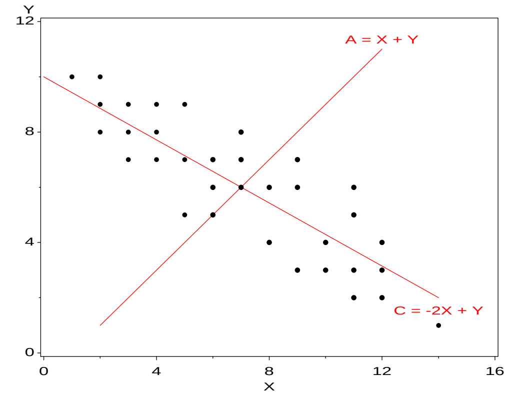
| X | Y | A | B | C |
|---|---|---|---|---|
| 12.757 | 6.000 | 3.605 | 249.160 | 87.330 |

So, the total variance of X and Y is 12.76 + 6.00 = 18.76.
Therefore, the variance of X and Y accounted for by any other variable (say, A) is

$$\begin{aligned} r^2_{XA}\sigma^2_X = (.764)^2(12.76) &= 7.44 \\ r^2_{YA}\sigma^2_Y = (-.339)^2(6.00) &= 0.69 \\ \text{Total} &= 8.13 \rightarrow 8.13/18.76 = 43\% \end{aligned}$$



The plot below shows the data, with the linear combinations, $A = X + Y$, and $C = -2X + Y$.



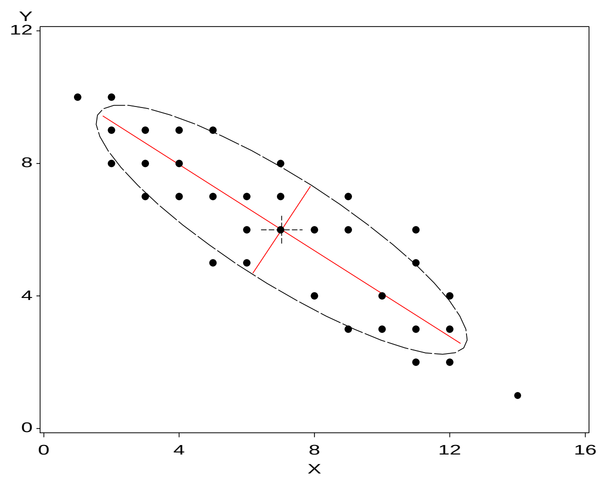

As you may guess, the linear combination $C = -2X + Y$ accounts for more of the variance in X and Y.

$$\begin{aligned} r^2_{XC}\sigma^2_X = (-.991)^2(12.76) &= 12.53 \\ r^2_{YC}\sigma^2_Y = (.924)^2(6.00) &= 5.12 \\ \text{Total} &= 17.65 \rightarrow 17.65/18.75 = 94\% \end{aligned}$$



Principal components finds the directions which account for the most variance.

- Geometrically, these are just the axes of an ellipse (ellipsoid in 3D+) that encloses the data
- Length of each axis $\sim$ eigenvalue $\sim$ variance accounted for
- Direction of each axis $\sim$ eigenvector $\sim$ weights in the linear combination






Using `PROC PRINCOMP` on our example data, we find,

| | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| PRIN1 | 17.6732 | 16.5898 | 0.942237 | 0.94224 |
| PRIN2 | 1.0834 | . | 0.057763 | 1.00000 |

**Eigenvectors**

| | PRIN1 | PRIN2 |
|---|---|---|
| X | 0.838832 | 0.544390 |
| Y | -.544390 | 0.838832 |

- The first principal component, PRIN1 = .8388 X - .5444 Y, accounts for the greatest variance, 17.673 (94.22%).
- The second principal component, PRIN2 = .5444 X + .8388 Y, accounts for the remaining variance, 1.083 (5.78%).
- The two components are uncorrelated, $r$( PRIN1, PRIN1 ) = 0.

# PCA details: Covariances or correlations?

- Principal components can be computed from either the covariance matrix or the correlation matrix.
- Correlation matrix: all variables are weighted equally
- Covariance matrix: each variable is weighted $\sim$ its variance.
- Using the covariance matrix makes sense iff:
  - All variables are measured in comparable units
  - You have adjusted the scales of the variables relative to some external measure of importance
- SAS:

```
PROC PRINCOMP data=mydata options;
  VAR variables;
```

- **options**: COV - analyze the covariance matrix; PLOT=SCREE - produce scree plot



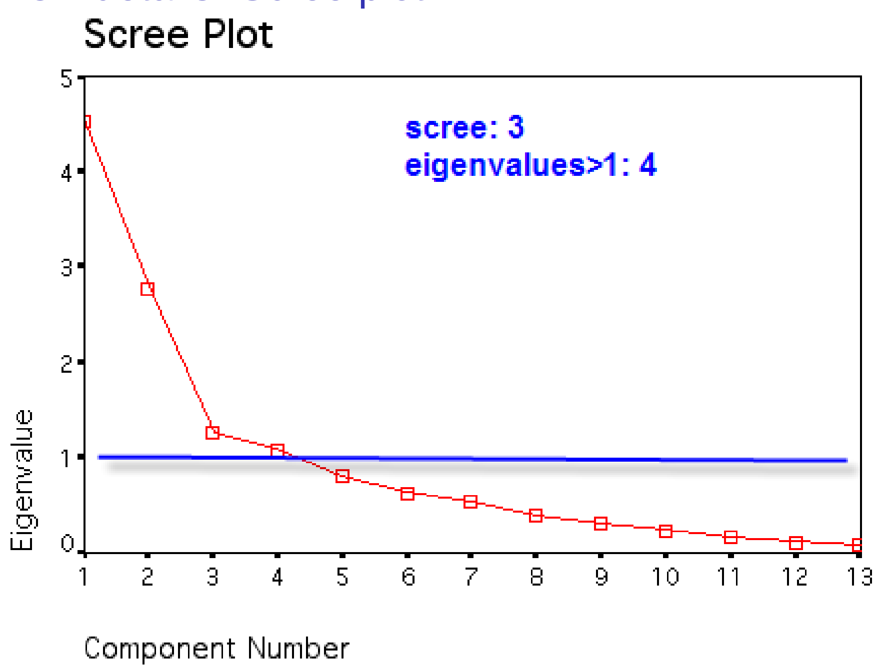
# PCA details: How many components?

- Complete set of principal components contains the *same* information as the original data— just a rotation to new, uncorrelated variables.
- For dimension reduction, you usually choose a smaller number
- Four common criteria for choosing the number of components:
  - Number of eigenvalues $>$ 1 (correlation matrix only)— based on idea that average eigenvalue = 1
  - Number of components to account for a given percentage— typically 80–90% of variance
  - "Scree" plot of eigenvalues– look for an "elbow"
  - How many components are interpretable?
- SAS:

```
PROC PRINCOMP data=mydata
     N=#_components OUT=output_dataset;
  VAR variables;
```



# PCA details: Scree plot



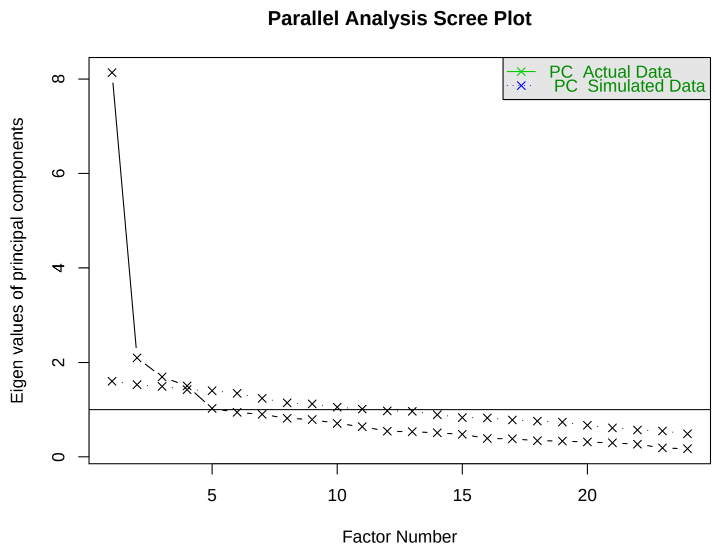



# PCA details: Parallel analysis

- Horn (1965) proposed a more "objective" way to choose the number of components (or factors, in EFA), now called parallel analysis
- The basic idea is to generate correlation matrices of random, *uncorrelated* data, of the same size as your sample.
- Take # of components =the number of eigenvalues from the observed data $>$ eigenvalues of the random data.
- From scree plot, this is where the curves for observed and random data cross.

# PCA details: Parallel analysis

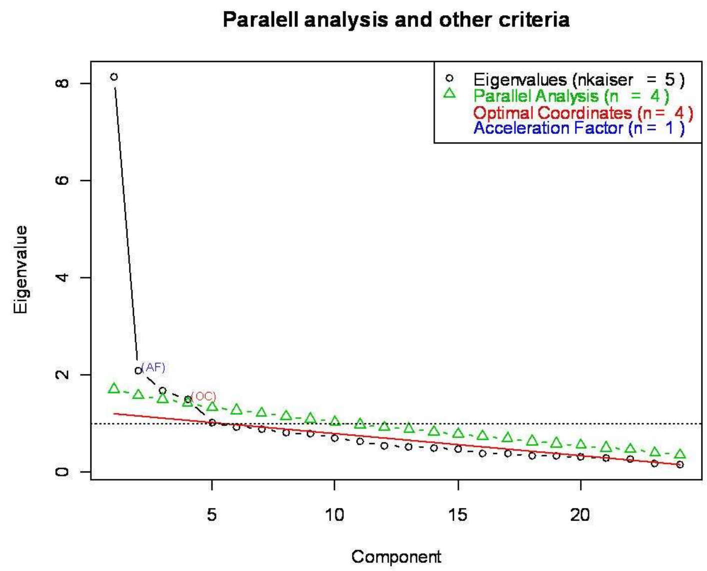
Holzinger-Swineford 24 psychological variables:






# PCA details: Parallel analysis

Holzinger-Swineford 24 psychological variables: Other criteria






# PCA details: Interpreting the components

- Eigenvectors (component weights or “loadings”)
  - Examine the signs & magnitudes of each column of loadings
  - Often, the first component will have all positive signs → “general/overall component”
  - Interpret the variables in each column with absolute loadings $> 0.3 - 0.5$
  - Try to give a name to each
- Component scores
  - Component scores give the position of each observation on the component
  - Scatterplots of: Prin1, Prin2, Prin3 with observation labels
  - What characteristics of the observations vary along each dimension?



# PCA Example: US crime data

```
title 'PCA: Crime rates per 100,000 population by state';
data crime;
   input State $1-15 Murder Rape Robbery Assault Burglary Larceny
         Auto  ST $;
datalines;
Alabama        14.2 25.2  96.8 278.3 1135.5 1881.9 280.7   AL
Alaska         10.8 51.6  96.8 284.0 1331.7 3369.8 753.3   AK
Arizona         9.5 34.2 138.2 312.3 2346.1 4467.4 439.5   AZ
Arkansas        8.8 27.6  83.2 203.4  972.6 1862.1 183.4   AR
California     11.5 49.4 287.0 358.0 2139.4 3499.8 663.5   CA
Colorado        6.3 42.0 170.7 292.9 1935.2 3903.2 477.1   CO
Connecticut     4.2 16.8 129.5 131.8 1346.0 2620.7 593.2   CT
 ...
Wisconsin       2.8 12.9  52.2  63.7  846.9 2614.2 220.7   WI
Wyoming         5.4 21.9  39.7 173.9  811.6 2772.2 282.0   WY
;
proc princomp out=crimcomp;
```

# PCA Example: US crime data

Output:

**Eigenvalues of the Correlation Matrix**

| | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| 1 | 4.11495951 | 2.87623768 | 0.5879 | 0.5879 |
| 2 | 1.23872183 | 0.51290521 | 0.1770 | 0.7648 |
| 3 | 0.72581663 | 0.40938458 | 0.1037 | 0.8685 |
| 4 | 0.31643205 | 0.05845759 | 0.0452 | 0.9137 |
| 5 | 0.25797446 | 0.03593499 | 0.0369 | 0.9506 |
| 6 | 0.22203947 | 0.09798342 | 0.0317 | 0.9823 |
| 7 | 0.12405606 | | 0.0177 | 1.0000 |

- Eigenvalues $> 1$: 2 components
- Differences (numerical version of scree plot): 3 components
- Proportion $> .80$: 2 components
- Interpretability?

PCA: Example

# PCA Example: US crime data

Output:

**Eigenvectors**

| | Prin1 | Prin2 | Prin3 | Prin4 | Prin5 | Prin6 | Prin7 |
|---|---|---|---|---|---|---|---|
| Murder | 0.3002 | -.6291 | 0.1782 | -.2321 | 0.5381 | 0.2591 | 0.2675 |
| Rape | 0.4317 | -.1694 | -.2441 | 0.0622 | 0.1884 | -.7732 | -.2964 |
| Robbery | 0.3968 | 0.0422 | 0.4958 | -.5579 | -.5199 | -.1143 | -.0039 |
| Assault | 0.3966 | -.3435 | -.0695 | 0.6298 | -.5066 | 0.1723 | 0.1917 |
| Burglary | 0.4401 | 0.2033 | -.2098 | -.0575 | 0.1010 | 0.5359 | -.6481 |
| Larceny | 0.3573 | 0.4023 | -.5392 | -.2348 | 0.0300 | 0.0394 | 0.6016 |
| Auto | 0.2951 | 0.5024 | 0.5683 | 0.4192 | 0.3697 | -.0572 | 0.1470 |

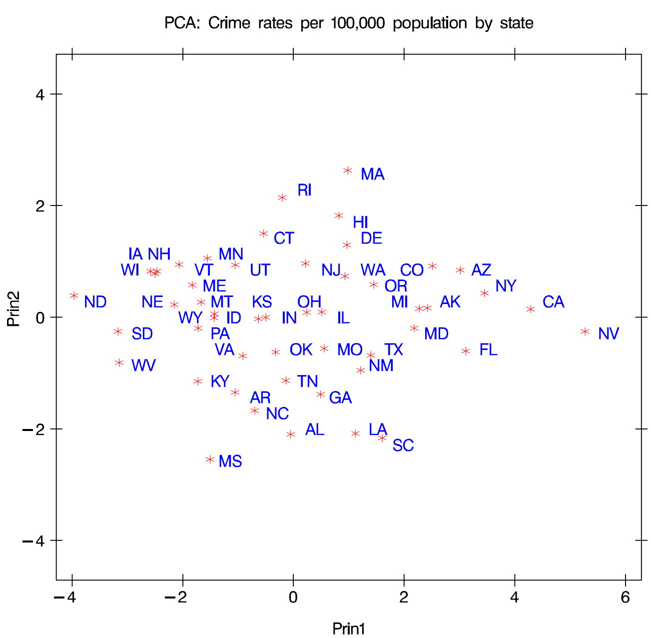
Which variables have large weights on each component?

- Prin1: all positive weights: Overall crime index
- Prin2: Property crimes $(+)$ vs. Violent crimes $(-)$
- Prin3: Robbery, auto vs. Larceny ??

PCA: Example

# PCA Example: Plotting component scores

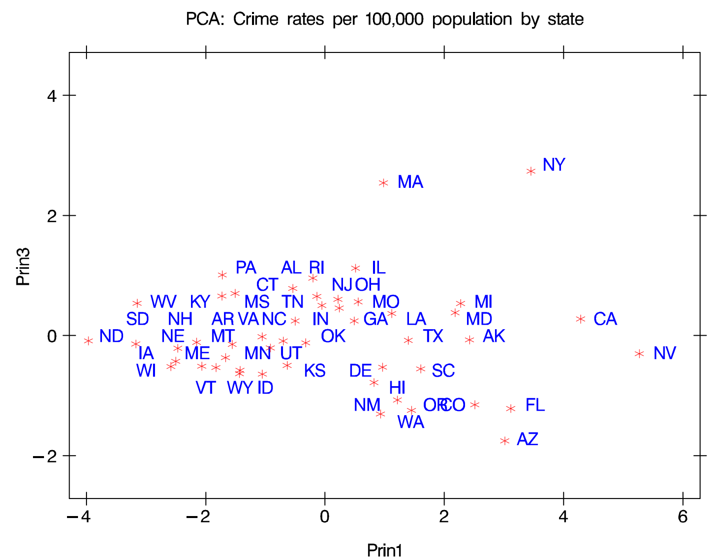
```
%plotit(data=crimcomp, plotvars=prin2 prin1, labelvar=ST);
```




PCA: Example

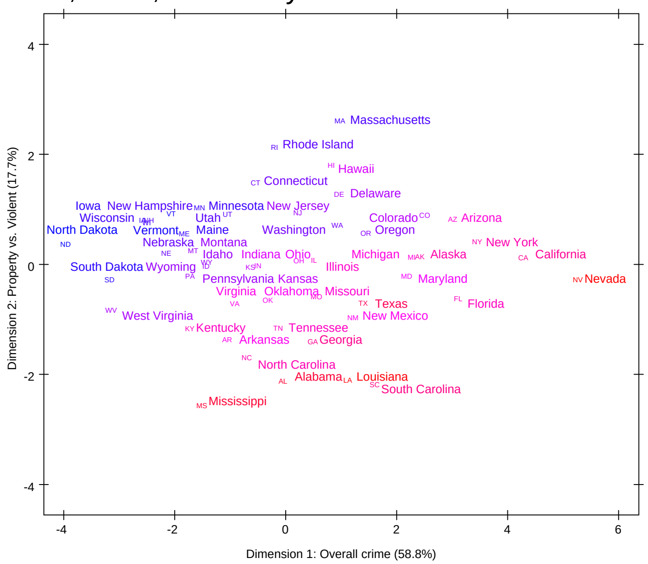
# PCA Example: Plotting component scores

```
%plotit(data=crimcomp, plotvars=prin3 prin1, labelvar=ST);
```



# PCA Example: Plotting component scores, better

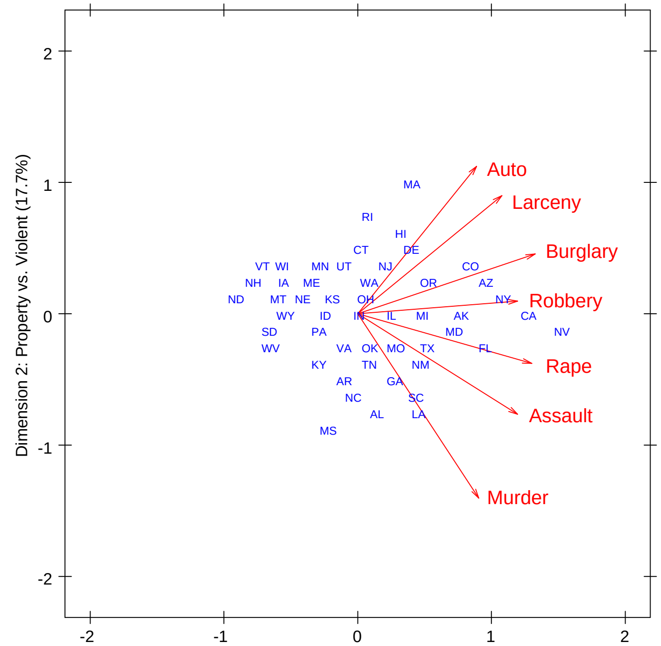
Prin1, Prin2, colored by Murder rate






# PCA Example: Plotting component scores, better still

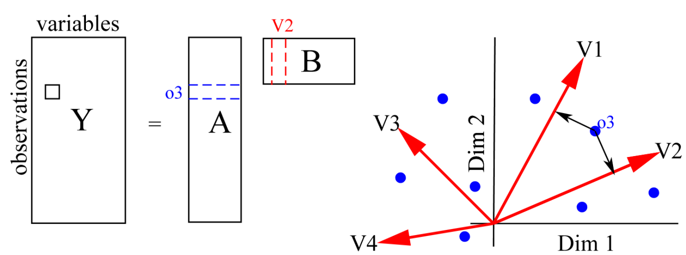
Prin1, Prin2, with variable weights as vectors (Biplot)





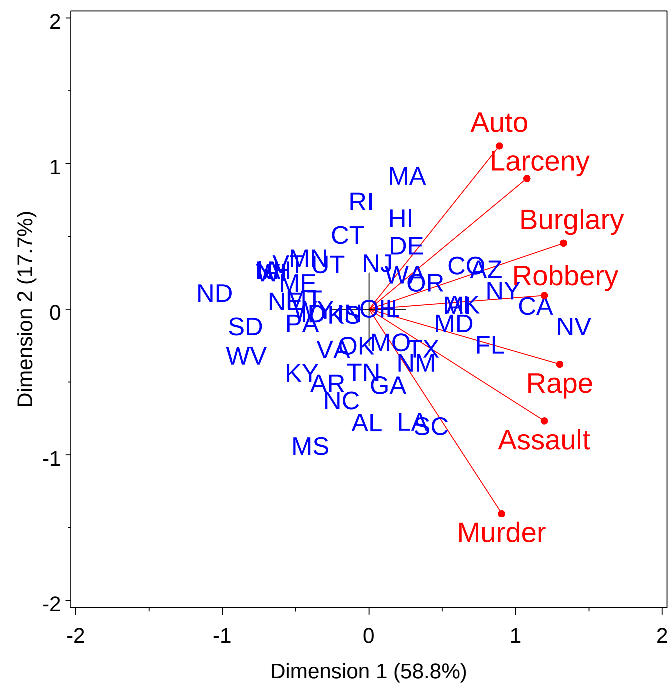

# Biplots: Low-D views of multivariate data

- Display variables *and* observations in a reduced-rank space of $d$ (=2 or 3) dimensions,




- Biplot properties:
  - Plot observations as points, variables as vectors from origin (mean)
  - Angles between vectors show correlations ($r \approx \cos(\theta)$)
  - $y_{ij} \approx \boldsymbol{a}_i^\mathsf{T} \boldsymbol{b}_j$: projection of observation on variable vector
  - Observations are uncorrelated overall (but not necessarily within group)
  - Data ellipses for scores show low-D between and within variation

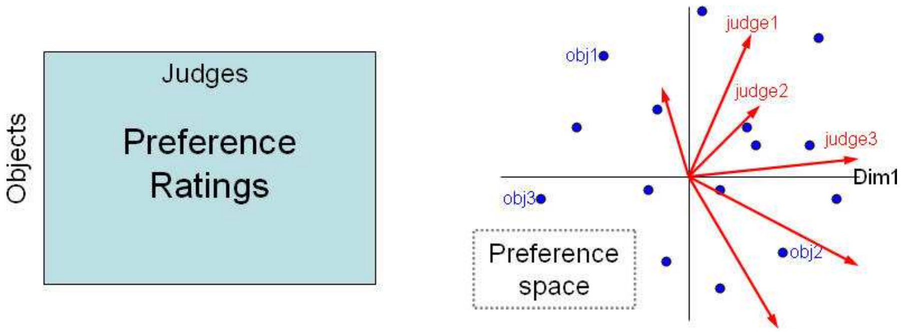





Biplot of US crime data

# Application: Preference mapping I

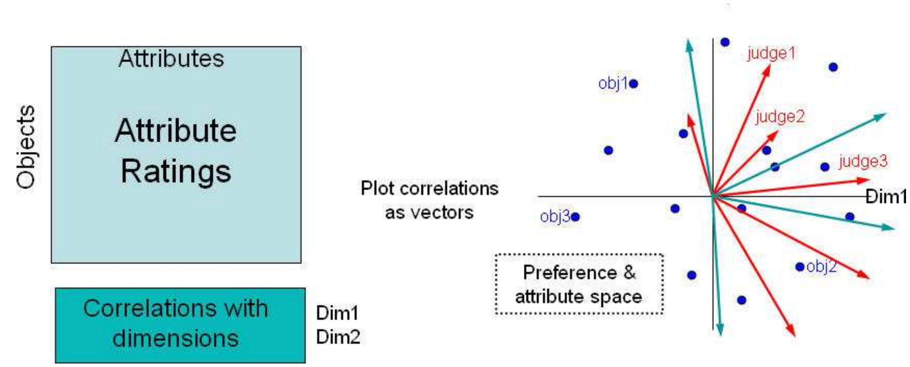
- Judges give “preference” ratings of a set of objects
  - How many dimensions are required to account for preferences?
  - What is the interpretation of the “preference map”?
  - NB: Here, the judges are treated as variables






# Application: Preference mapping II

- Also obtain ratings of a set of attributes to aid interpretation
  - Find correlations of attribute ratings with preference dimensions
  - Project these into preference space






# Example: Car Preference

Preference ratings

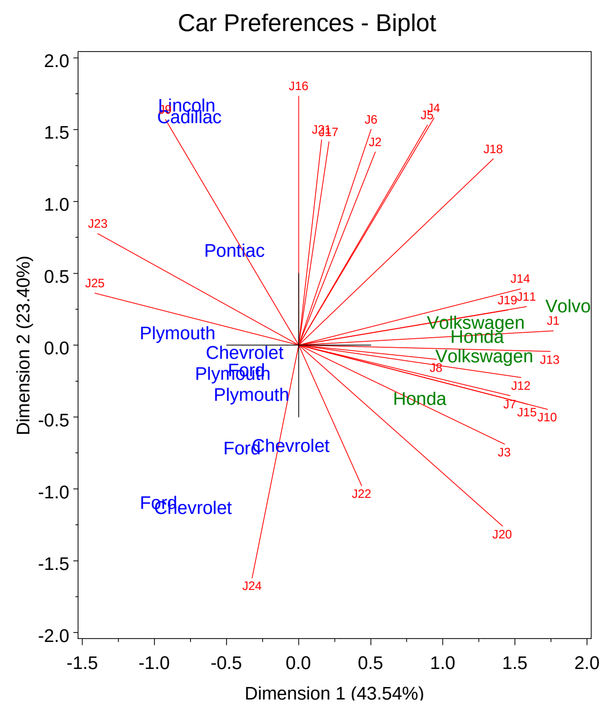
25 judges gave preference ratings for 17 automobile models:

| MAKE | MODEL | J1 | J2 | J3 | J4 | J5 | J6 | J7 | J8 | J9 | J10 | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cadillac | Eldorado | 0 | 8 | 0 | 7 | 9 | 9 | 0 | 4 | 9 | 1 | ... |
| Chevrolet | Chevette | 0 | 0 | 5 | 1 | 2 | 0 | 0 | 4 | 2 | 3 | |
| Chevrolet | Citation | 0 | 4 | 5 | 3 | 3 | 0 | 5 | 8 | 1 | 4 | |
| Chevrolet | Malibu | 0 | 6 | 2 | 7 | 4 | 0 | 0 | 7 | 2 | 3 | |
| Ford | Fairmont | 0 | 2 | 2 | 4 | 0 | 0 | 6 | 7 | 1 | 5 | |
| Ford | Mustang | 0 | 5 | 0 | 7 | 1 | 9 | 7 | 7 | 0 | 5 | |
| Ford | Pinto | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 3 | |
| Honda | Accord | 9 | 5 | 5 | 6 | 8 | 9 | 7 | 6 | 0 | 9 | |
| Honda | Civic | 8 | 4 | 3 | 6 | 7 | 0 | 9 | 5 | 0 | 7 | |
| Lincoln | Continental | 0 | 7 | 0 | 8 | 9 | 9 | 0 | 5 | 9 | 2 | |
| Plymouth | Gran Fury | 0 | 7 | 0 | 6 | 0 | 0 | 0 | 4 | 3 | 4 | |
| Plymouth | Horizon | 0 | 3 | 0 | 5 | 0 | 0 | 5 | 6 | 3 | 5 | |
| Plymouth | Volare | 0 | 4 | 0 | 5 | 0 | 0 | 3 | 6 | 1 | 4 | |
| ... | ... | | | | | | | | | | | |

Analysis & biplot:

```
%biplot(data=Cars, var=J1-J25, id=make);
```






- 2 dimensions: 67% of variance (3: 75%)
- Dim 1 separates Domestic vs. Foreign
- What is Dim 2?
- Clusters of judges vectors suggest market segmentation

# Example: Car Preference

## Attribute ratings

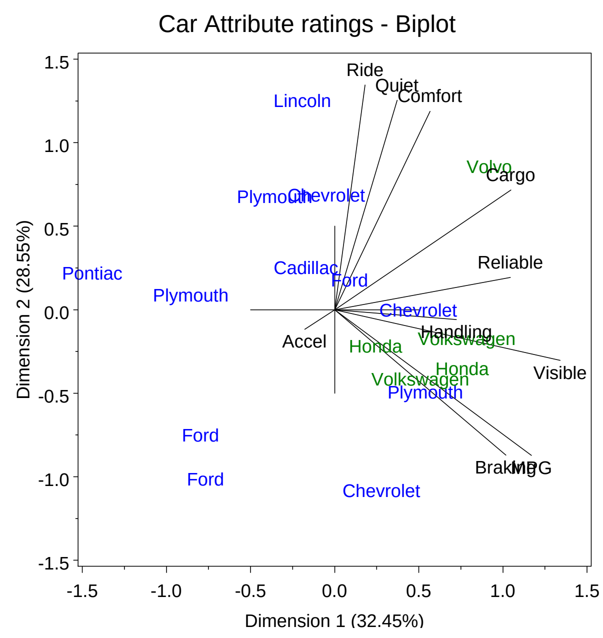
We also have attribute ratings on 10 variables:

| Model | MPG | Rel | Accel | Brake | Hand | Ride | Vis | Comf | Quiet | Cargo |
|---|---|---|---|---|---|---|---|---|---|---|
| Eldorado | 3 | 2 | 3 | 4 | 5 | 4 | 3 | 5 | 3 | 3 |
| Chevette | 5 | 3 | 3 | 5 | 4 | 2 | 5 | 2 | 2 | 3 |
| Citation | 4 | 1 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 5 |
| Malibu | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 4 | 4 |
| Fairmont | 3 | 3 | 2 | 4 | 3 | 4 | 5 | 4 | 3 | 4 |
| Mustang | 3 | 2 | 4 | 4 | 3 | 2 | 3 | 2 | 2 | 2 |
| Pinto | 4 | 1 | 3 | 4 | 3 | 1 | 3 | 2 | 2 | 2 |
| Accord | 5 | 5 | 5 | 4 | 5 | 3 | 3 | 4 | 3 | 3 |
| Civic | 5 | 5 | 4 | 5 | 4 | 3 | 5 | 4 | 3 | 4 |
| Continental | 2 | 4 | 5 | 3 | 3 | 5 | 3 | 5 | 5 | 5 |
| Gran Fury | 2 | 1 | 3 | 4 | 3 | 5 | 3 | 5 | 3 | 5 |
| Horizon | 4 | 3 | 4 | 5 | 5 | 3 | 5 | 2 | 3 | 5 |
| Volare | 2 | 1 | 5 | 3 | 3 | 3 | 5 | 4 | 2 | 4 |
| ... | ... | | | | | | | | | |

Analysis & biplot:

```
%biplot(data=Cars, var=Rel--Cargo, id=make);
```






- Traditional PCA of Attributes × Objects
- Dim 1: Performance
- Dim 2: Comfort
- How do these relate to preference dimensions?



# Example: Car Preference

## Attribute ratings

Calculate correlations of the attribute ratings with the preference dimensions:

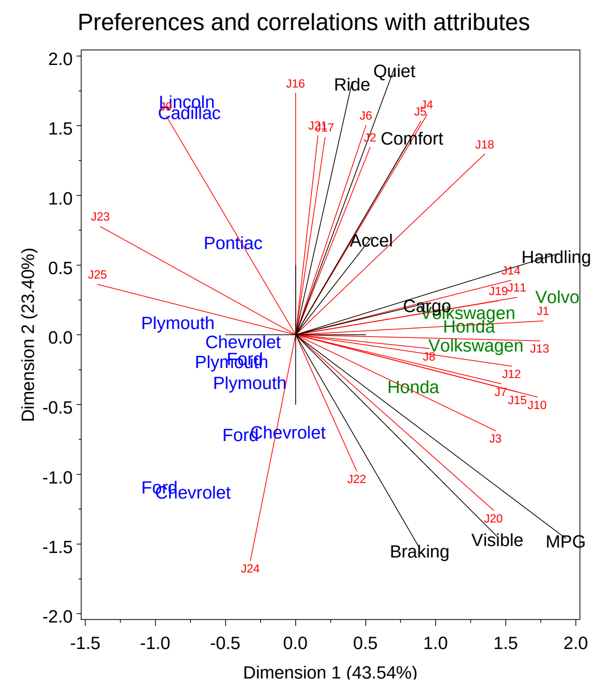
```
data components;
  merge cars biplot(where=(_type_="OBS"));
  run;
proc corr data=components outp=vectors;
  var dim1 dim2;
  with mpg reliable accel braking handling ride visible comfort quiet carg
```

Output:

| | DIM1 | DIM2 |
|---|---|---|
| MPG | 0.60289 | -0.45661 |
| Reliable | 0.69996 | 0.13657 |
| Accel | 0.16958 | 0.21867 |
| Braking | 0.27708 | -0.47862 |
| Handling | 0.58163 | 0.18094 |
| Ride | 0.12654 | 0.56731 |
| Visible | 0.45048 | -0.45278 |
| Comfort | 0.26021 | 0.44702 |
| Quiet | 0.22059 | 0.59791 |
| Cargo | 0.29396 | 0.07101 |

Overlay these as vectors from the origin on the Preference space






Correlations of Attribute ratings with Dimensions overlaid on Preference space

# Summary: Part 1

- **Factor Analysis methods**
  - Exploratory vs. confirmatory
  - PCA (data reduction) vs. FA (statistical model)
- **Principal components analysis**
  - Linear combinations that account for maximum variance
  - Components are uncorrelated
  - All PCs are just a rotation of the *p*-dimensional data
- **PCA details**
  - Analyze correlations, unless variables are commensurate
  - Number of components: Rules of thumb, Scree plot, Parallel analysis
- **Visualizations**
  - Plots of component scores
  - Biplots: scores + variable vectors