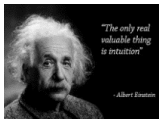
Factor Analysis



Factor Analysis: Intuition

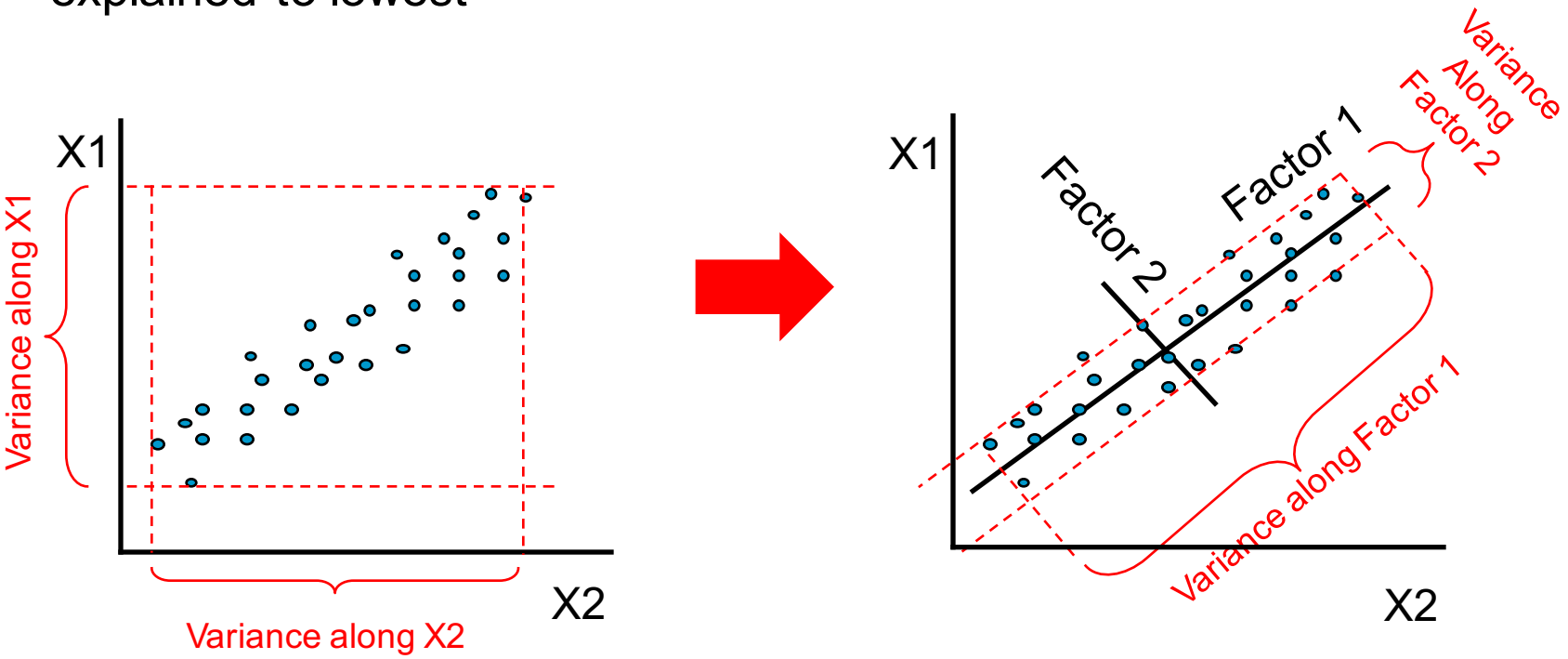
- **Factor Analysis** is a common method used in survey (and other types of) data analysis based on **PCA**
- Factor analysis is **useful** to figure out if you can **group** and **aggregate** (e.g., average) **similar** or highly correlated **variables**, which is a **common** practice with **survey** data
- It is most **useful** in survey research because many questions may elicit similar responses, **highly correlated**. For example:
  - How much do you like this course? How much do you like this professor? **(+)**
  - Do you dislike this product? Would you recommend it to a friend? **(-)**
- **Factor Analysis** looks at the **correlation** and **co-variance** of several variables that have strong **linear associations** and provides groupings of more general variables called **"factors"**

Factor Analysis: Explained

- **Analysis** is about **explaining variance** in the data
- Multivariable variance is **best explained** if we **rotate** the axes in the direction of **higher variance** (**1st Axis**)
- The **second** axis is **perpendicular** to the **1st Axis** and is **rotated** around to the direction of **2nd highest** variance (**2nd Axis**); so on.
- Chances are that **most** of the **variance** in the data is explained along the **first few axes** and the last axes explain very little
- The key **modeling** issue is to find **M factors** out of **P predictors** that explain the **majority** of the data, such that $M << P$
- If so, thhen a substantial **dimension reduction** is accomplished
- Instead of **P variables**, we model **M factors** → 2 methods:
  - **PCA Regression** – use the M **Principal Components** as the new factor variables, but the coefficients don't have much meaning
  - **Aggregation** – finding the variables with high **correlation** within each of the **first M axes** based on their respective **"Factor Loadings"** and **average** them out into new variables

Illustration

- By **rotating** the **axes without** rotating the **data** we are **preserving** the overall **covariance** structure of the data
- A nice property of principal components and factor analysis is that the resulting **factors** are **sequentially ordered** from highest **variance** explained to lowest



KOGOD SCHOOL
of
BUSINESS