Factor Extraction

Number of Factors to Extract

- **P variables** have **P** principal components, that is, **P factors** that can be extracted
- But the goal in factor analysis is to **reduce** the model **dimension** to **M << P** factors
- So, **how many** factors are optimal to extract for a model?
  - Use a **business rationale** to group variables into factors and do "Confirmatory Factor Analysis" (**CFA**) to verify or discard the factors
  - Use the **standard threshold** point around **Eigenvalue = 1**
  - Use a given cumulative **variance explained threshold** – a common rule of thumb is to use enough factors to explain **70%** of the variance.

Variance Explained

- It also helps to see how much **variance** is **explained** by each factor
- And to look at the **cumulative variance explained** by the first M factors
- Modeling a **3-Factor** solution, as suggested by the Scree Plot explains **92%** of the **variance**, so we don't need any more factors
- Even a **2-Factor** solution would be fine

	PC1	PC2	PC3
SS loadings	5.47	4.17	1.41
Proportion Var	0.46	0.35	0.12
Cumulative Var	0.46	0.80	0.92
Proportion Explained	0.49	0.38	0.13
Cumulative Proportion	0.49	0.87	1.00

Factor Loadings

- Factor analysis reports **"Factor Loadings"** and **% of Variance** Explained by each **factor**.
- A factor **loading** is a statistic that shows how much a particular **variable contributes** to the **variance** of a factor
- The idea is to group variables with high **Factor Loadings** $F >= 0.5$ rule of thumb – **high** factor **loadings** within a **factor** indicate that variables are highly **correlated** and therefore **group together**
- **Don't** group variables with light loadings $F < 0.5$ → low correlation among the variables → no basis for grouping
- The **next slide** was covered in the **PCA Regression lecture**, but it duplicated here as a **refresher**

Illustration Hitters{ISLR} data

- Factor Loadings for a **3-Factor solution**
- It is **not** as **easy** to **visualize** the variables to group into each factor

	PC1	PC2	PC3
Hitters.AtBat	0.09	0.95	0.17
Hitters.Hits	0.08	0.95	0.14
Hitters.HmRun	0.14	0.48	0.84
Hitters.Runs	0.06	0.94	0.21
Hitters.RBI	0.18	0.74	0.60
Hitters.Walks	0.24	0.78	0.05
Hitters.CAtBat	0.98	0.13	0.02
Hitters.CHits	0.97	0.15	0.01
Hitters.CHmRun	0.84	0.08	0.45
Hitters.CRuns	0.97	0.18	0.03
Hitters.CRBI	0.95	0.11	0.23
Hitters.CWalks	0.94	0.12	0.01

KOGOD SCHOOL
of
BUSINESS