

Assignment 3: Principal Components Analysis & Biplot

The Stata dataset “State Economy Indicators.dta” contains various indicators of the state of the economy in the 50 states plus the District of Columbia. The variables are as follows:

- **state**: State name.
- **approve**: Presidential approval (from Gallup).
- **partisan**: Partisan category (SD=Strong Democrat, LD=Lean Democrat, C=Competitive, SR=Strong Republican)
- **bankruptcyp1k**: Bankruptcy filings per 1000 people.
- **pctgrowth**: Percent economic growth.
- **gsppc**: Gross state product per capita.
- **medfaminc**: Median family income.
- **pctpov**: Percentage living below the poverty line.
- **taxburdenpc**: Tax burden per capita.
- **unemploy**: Unemployment rate.

First, construct a biplot of the economic indicators (**bankruptcyp1k-unemploy**). How many dimensions are necessary to “best” represent the important information contained in the dataset? How would substantively interpret the configuration? Next, conduct a principal components analysis. Interpret the results.

In an effort to control for the economy in estimating the model of approval on partisanship, regress **approve** on **partisan** (treated as categorical) and the economic indicators. Re-estimate the model, replacing the “raw” economic indicators for the principal components. How many components do you need to use? How can you interpret the results?