Dimension Reduction Methods

Instructor: Haoran LEI

Hunan University

Shrinkage Methods

- The methods that we have discussed so far in this chapter have involved fitting linear regression models, via least squares or a shrunken approach, using the original predictors, $X_1, \ldots, X_p$.
- We now explore a class of approaches that *transform* the predictors and then fit a least squares model using the transformed variables.
- We will refer to these techniques as **dimension reduction** methods.

Dimension Reduction Methods: details

- Let $Z_1, \ldots, Z_M$ represent $M < p$ linear combinations of our original p predictors:

$$Z_m = \sum_{j=1}^{p} \phi_{mj} X_j \quad \text{for some constants } \phi_{m1}, \ldots, \phi_{mp}.$$

- We then fit the linear regression model,

$$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \epsilon_i, \quad i = 1, \ldots, N.$$

Advantage of Dimension Reduction

- Dimension reduction serves to constrain the estimated β_j coefficients, since now they must take the form:

$$\beta_j = \sum_m \theta_m \phi_{mj}.$$

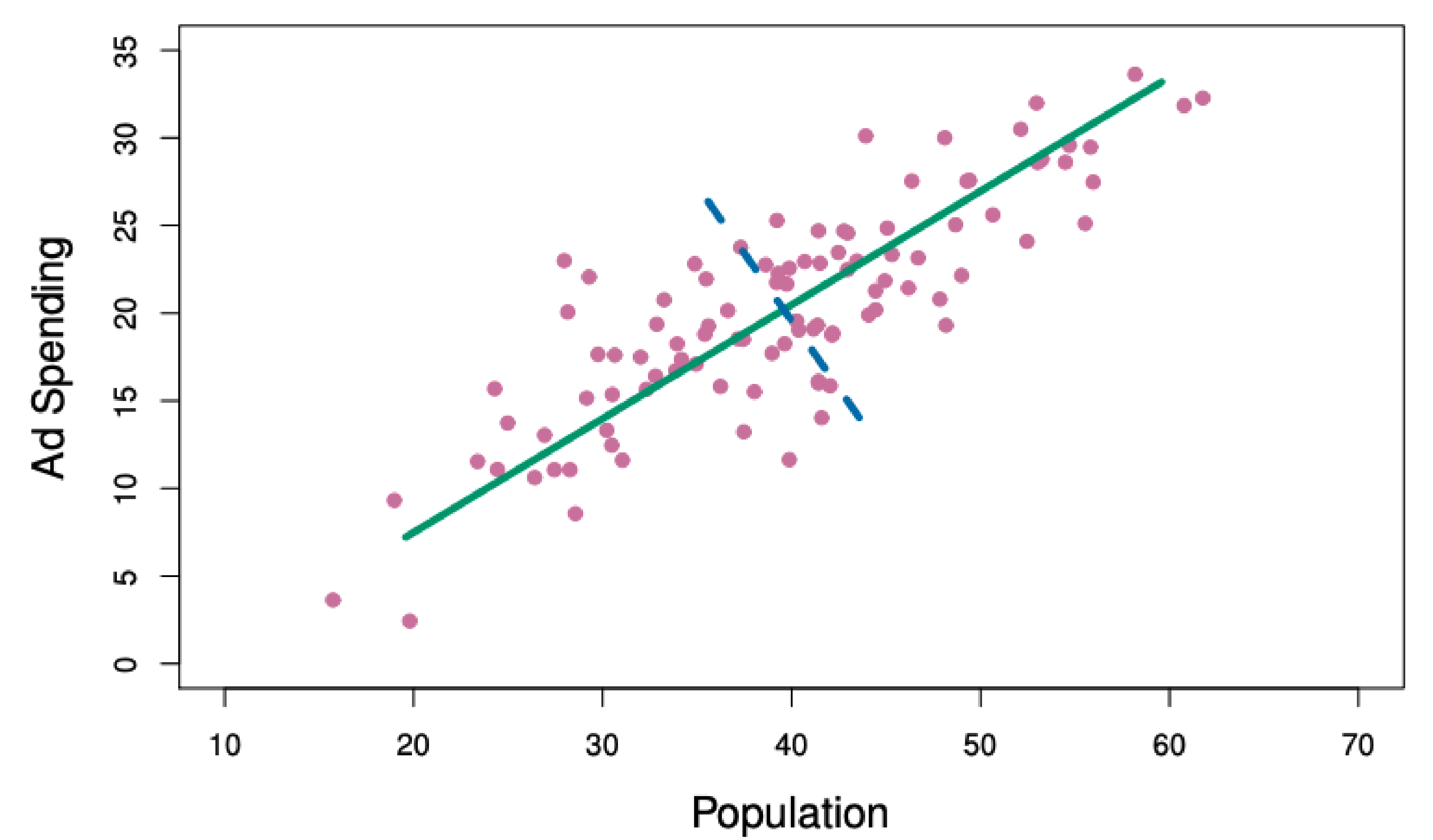
- If the constants ϕ_{m1}, ..., ϕ_{mp} are chosen wisely, then such dimension reduction approaches can often outperform OLS regression.
  - Can win in the bias-variance tradeoff.

Principal Components Regression

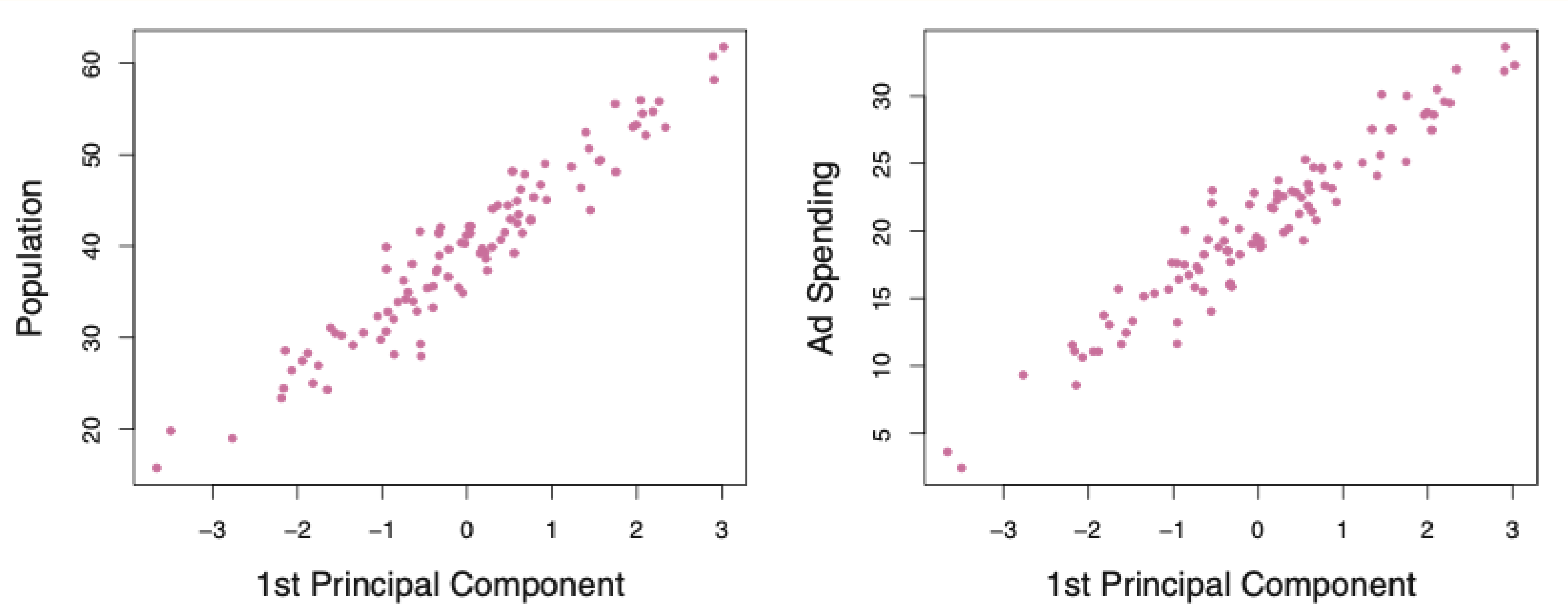
- Here we apply **principal components analysis (PCA)** (discussed in Chapter 10 of the text) to define the linear combinations of the predictors, for use in our regression.
- PCA is indeed a popular unsupervised learning method. Here we use it to "extract the main information" from X's first, denoted by Z's. And then regress y on Z's.

PCA details

- The first principal component is that (normalized) linear combination of the variables with the largest variance.
- The second principal component has largest variance, subject to being uncorrelated with the first.
- And so on ...
- Hence with many correlated original variables, we replace them with a small set of principal components that capture their joint variation.



In the case of $p = 2$, choosing the first main component is equivalent to minimizing the "sum of squared distances."



$$z_1 = \phi_{11} \times (pop_i - \overline{pop}) + \phi_{21} \times (ad_i - \overline{ad})$$

$$\max \mathrm{Var}(z_1) \text{ s.t. } \phi_{11}^2 + \phi_{21}^2 = 1.$$

From PCA to PCR (Principal Components Regression)

- Choosing the number of directions/components M.
- Use PCA to obtain the principal components Z_1,..., Z_M.
- Regress Y on Z_1,..., Z_M.

Use cross-validation to select the optimal M.

Partial Least Squares (PLS)

- Like PCR, PLS is a dimension reduction method, which first identifies a new set of features Z1, . . . , ZM that are linear combinations of the original features, and then fits a linear model via OLS using these M new features.
- But unlike PCR, PLS identifies these new features in a supervised way – that is, it makes use of the response Y in order to identify new features that not only approximate the old features well, but also that are related to the response.
- Roughly speaking, the PLS approach attempts to find directions that help explain both the response and the predictors.

Partial Least Squares (PLS): details

- After standardizing the p predictors, PLS computes the first direction Z_1 by setting each ϕ's equal to the coefficient from the simple linear regression of Y onto X_j. (i.e., $Z_1 = \hat{Y}$).
- Subsequent directions are found by taking residuals and then repeating the above prescription.

Summary of model selection

- Model selection methods are an essential tool for data analysis, especially for big datasets involving many predictors.
- Research into methods that give **sparsity**, such as the **lasso** is an especially hot area.