ADVANCED BAYESIAN MODELING

Regression Concepts

So far:

$$\begin{aligned} y &= \text{random variables (observed) to be modeled} \\ \theta &= \text{parameters (unobserved)} \end{aligned}$$

(May also need constant values, either from data or pre-chosen)

What if other (observed) variables are needed in the model for y?

For example:

- Research questions could ask how y relates to other variables.
- Other variables could be used to make future predictions ($\tilde{y}$).

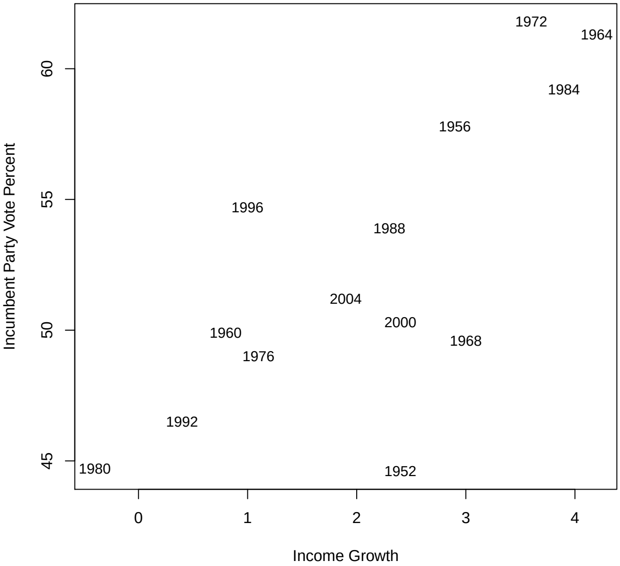
Example: Bread and Peace

In the "Bread and Peace" model of Hibbs (2008) for U.S. presidential elections, the incumbent party's vote share depends, in part, on personal income growth in the previous term:

$$
\begin{aligned}
y_i &= \text{incumbent two-party vote \% in election } i \\
x_i &= \text{weighted-average per capita real income \% growth (previous term)}
\end{aligned}
$$

Presumably, higher income growth is better for the incumbent party.

- How does incumbent party vote % depend on income growth?
- What vote % would the model predict for an upcoming election?



1972
1964
1984
1956
1996
1988
2004
2000
1960
1968
1976
1992
1980
1952
60
55
50
45
0
1
2
3
4
Incumbent Party Vote Percent
Income Growth

Relationship is roughly a straight line:

$$y \approx \beta_1 + \beta_2 x$$

for unknown constants β_1, β_2.

Vertical departures are the **errors**

$$\varepsilon_i = y_i - (\beta_1 + \beta_2 x_i)$$

which are regarded as random.

(One or two points have unusually large absolute error – later.)

Is income growth variable x random?

Arguably, yes: It is uncontrolled and uncertain in advance.

But modeling x is a nuisance: All we care about is how y depends on it.

So regard x as constant by *conditioning* on it ...

Conditional Modeling

Consider n **observations**

$$X_i = (x_{i1}, \ldots, x_{ik}), \quad y_i \qquad\qquad i = 1, \ldots, n$$

X has k **explanatory variables**, or **covariates**, or **predictors**

y represents **response** or **outcome variable**

How does y depend on X?

Idea: Use only the conditional distribution of y given X to make inference about parameter θ, ignoring marginal distribution of X.

Effectively, X is treated as constant.

For (conditionally) independent observations, likelihood is

$$\prod_{i=1}^{n} p(y_i \mid \theta, X_i)$$

(Alternatively, let pairs (X_i, y_i) be *exchangeable*.)

Note: Each y_i depends only on its own X_i.

Linear Regression

Let variable y be continuous and (essentially) unrestricted in range.

The **linear regression** of y on X proposes

$$\mathrm{E}(y_i \mid \theta, X_i) \;=\; \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

where the **coefficients** $\beta_1, \ldots, \beta_k$ are in θ.

Often, the first term is an **intercept**:

$$x_{i1} \;\equiv\; 1 \qquad\qquad i = 1, \ldots, n$$

Commonly assume

- $y_1, \ldots, y_n$ (conditionally) independent
- $\mathrm{var}(y_i \mid \theta, X_i) \;=\; \sigma^2$ (where σ^2 is in θ)

BDA3 calls this *ordinary linear regression*.

Regression equation often written

$$y_i \;=\; \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

$$\mathrm{E}(\varepsilon_i \mid \theta, X_i) \;=\; 0 \qquad \mathrm{var}(\varepsilon_i \mid \theta, X_i) \;=\; \sigma^2$$

Example: Bread and Peace

Assume it is the *mean* incumbent party vote % that depends on income growth:

$$\mathrm{E}(y_i \mid \theta, x_i) = \beta_1 + \beta_2 x_i$$

This is a **simple linear regression**.

We expect $\beta_2 > 0$.