1.2 Multiple Regression

Examine the linear relationship between 1 dependent (Y) & 2 or more independent variables (X_i)

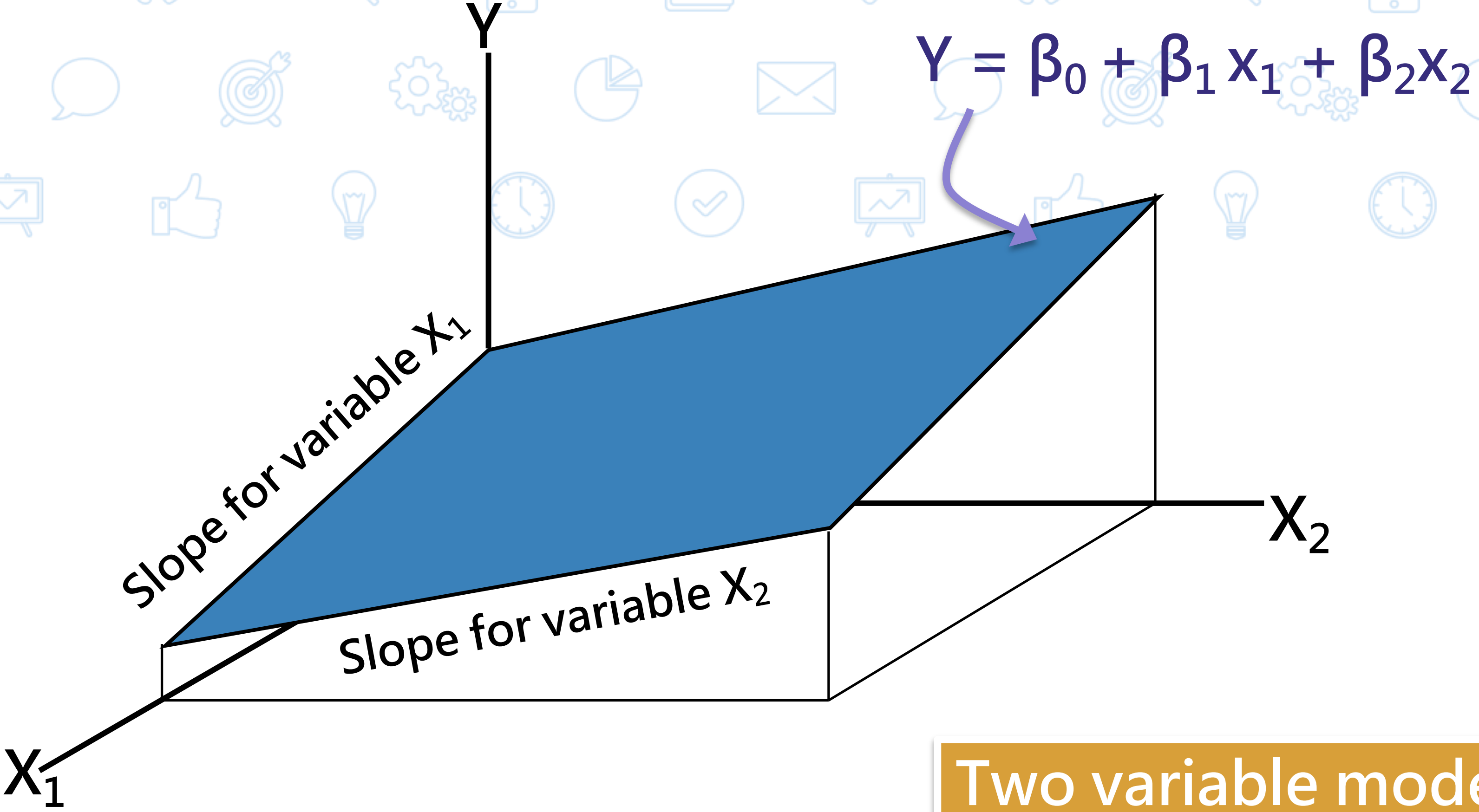
Multiple Regression Model with k Independent Variables

Y-intercept

Population slopes

Random Error

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$



Two variable model

Purposes

Prediction

Explanation

Theory building

Design Requirements

- One y variable (criterion)
- Two or more X variables (predictor variables).
- Sample size: >= 50 (at least 10 times as many cases as X variables)

Simple vs. Multiple Regression

One Y variable predicted from one X variable	One Y predicted from a set of X variables (X_1, X_2 ....X_k)
One regression coefficient	One regression coefficient for each X variable
R^2 proportion of variation in dependent variable Y predictable from X	R^2 proportion of variation in Y variable predictable by set of X variables

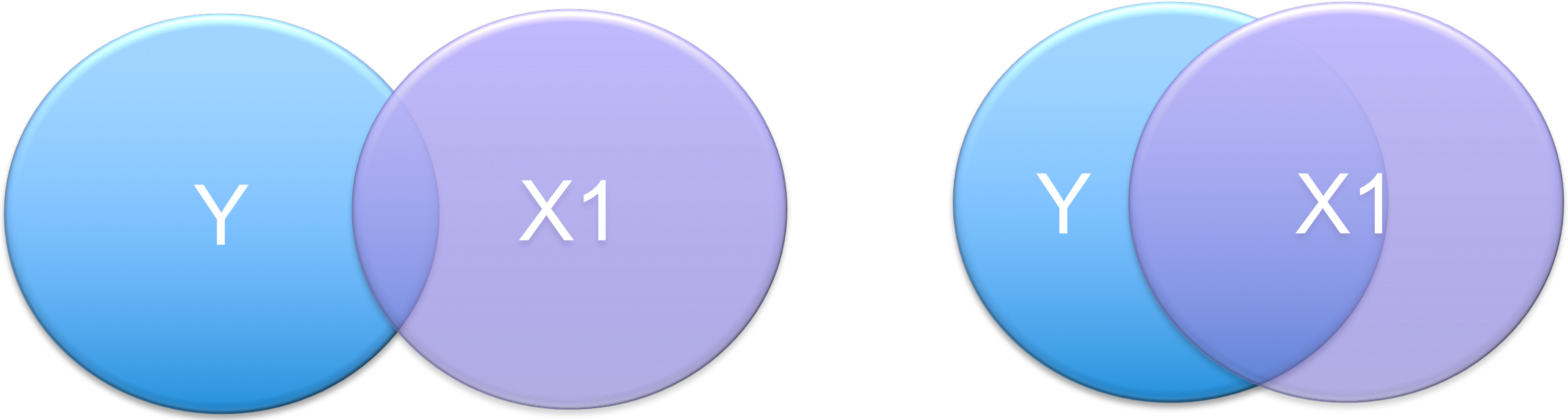
Collinearity and Parsimony

Collinearity

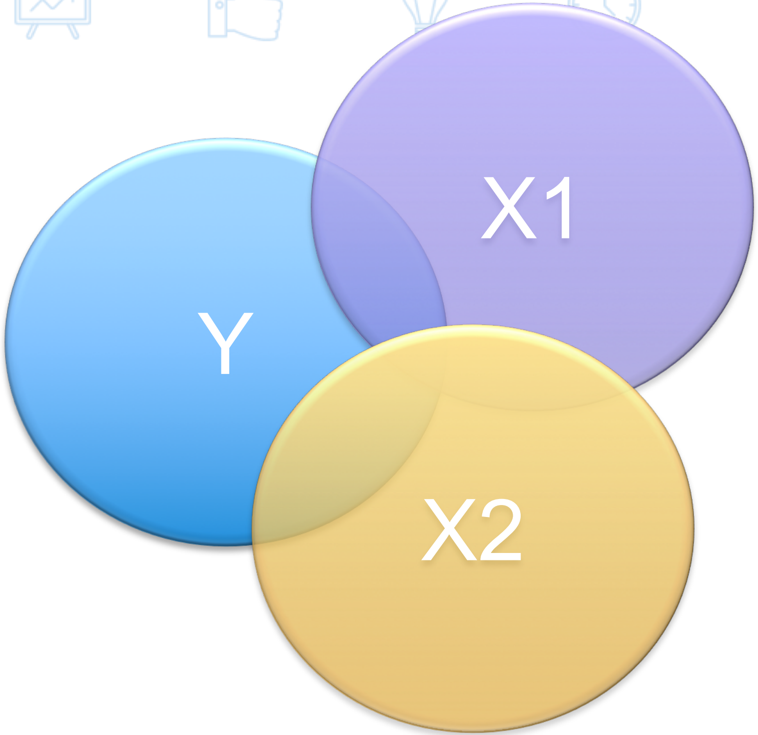
① Two Independent variables are said to be colinear when they are correlated with each other.

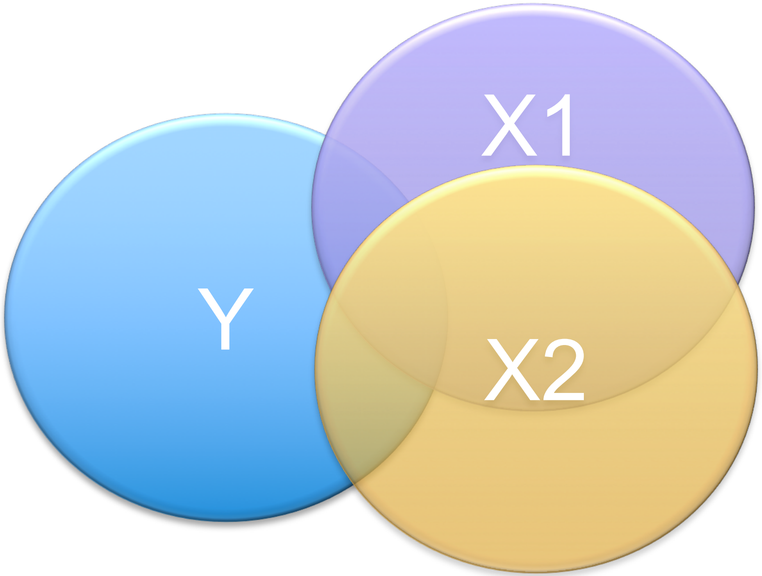
② Remember：Independent variables should be independent of each other.

Collinearity



Collinearity





Can it be avoided ?

YES !

① **Efficiency calculations !**

Ensure you at least have the most efficient design.

② **Variance Inflation Factor(VIF)**

A highly efficient design may still have a dangerously high.

Parsimony

① Avoid adding x associated with each other because often times the addition of such variable brings nothing new to the table

② Prefer the simplest best model, i.e. the **parsimonious model**.

Parsimony

③ Addition of collinear variables can result in biased estimates of the regression parameters.

④ While it' s impossible to avoid collinearity from arising in observational data, experiments are usually designed to control for correlated predictors.

Model Selection

Why not satisfied with the least squares estimates ?

① Prediction accuracy

the least squares estimates often have low bias but large variance. Prediction accuracy can sometimes be improved by shrinking or setting some coefficients to zero. By doing so we sacrifice a little bit of bias to reduce the variance of the predicted values, and hence may improve the overall prediction accuracy.

Why not satisfied with the least squares estimates ?

② Interpretation

With a large number of predictors, we often would like to determine a smaller subset that exhibit the strongest effects. In order to get the "big picture," we are willing to sacrifice some of the small details.

Expert opinion

- Variables can be included in (or eliminated from) the model based on expert opinion.
- If you are studying a certain variable, you might choose to leave it in the model regardless of whether it’ s significant or yield a higher adjusted R^2 .

THANKS!