

- Remember: problems are from

SECOND EDITION.

- Also: most problems correspond to an example in the chapter. Understand and imitate.
- Model Selection: This requires judgement and domain-specific knowledge.

What is the purpose of the model?

Example 1: making the case for a public policy decision — simplicity and explainability are priorities.

Example 2: High-Frequency Trading: our code will make decisions $100 \times$ per second. We want to be right 50.1% of the time and we don't care how.

Priorities are completely different.

In example 1 we might eliminate variables aggressively, for lack of statistical or economic significance.

In example 2 we would not: the computer doesn't care if our model is complex.

Note the difference between statistical and economic significance.

Economic / Scientific significance means that the estimated effect size is "large" and/or that there are good reasons to expect the variable to have an effect.

Statistical Significance is what we see in the regression output.

Note that the estimated coefficient, p-value, etc depend on the other variables in the regression. If we change the other variables, everything can change.

Pitfalls and syntax in R.

(1) I (identity operator) "protects"

things like arithmetic operations (e.g. $+$, 2)

from being interpreted by Wilkinson-Rogers notation.

(2) "offset" means "don't estimate this coefficient"

"I am telling you it is - - - ."

See p. 38-39 (among other places).

We looked at the Ch. 3 script

"infer.R".