PHP2510

REGRESSION WEEKS 12-14

FORMULATING A STATISTICAL ANALYSIS PLAN

Before writing a single line of code, a good researcher outlines their plan. A Statistical Analysis Plan (SAP) is a formal document that connects your research question to the exact statistical methods you will use to answer it.

PICK & EXPLORE A DATASET

Ideally, you'd write your SAP before even collecting any data! But for this hypothetical, choose from any dataset from this repository. Because we are learning regression techniques, pick strategically so that linear regression modeling can be utilized (which is a bit backwards; in real applications, we will choose a statistical tool based on the data, not vice-versa!).

Explore your dataset, determining the key variables available and their type (i.e. categorical vs continuous). Brainstorm some relevant research questions you could pose.

Finalize your research question, making sure your outcome variable of interest is continuous (not binary or categorical) in nature so you can practice working with linear regression.

DEFINE YOUR PLAN

Fill out the following:

1. Chosen Dataset:

2. Research Question (write a single, clear question in non-technical terms):

3. Key Variables (describe their type and empirical distributions):

    Outcome Variable (Y):

    Predictor Variables (X):

4. Testable Hypotheses (translate your question into formal statements about the population)

    Null Hypothesis:

        In words:

        In mathematical notation:

    Alternative Hypothesis:

        In words:

In mathematical notation:

5. Other Testing Parameters

Is this a two-sided test?

What is your alpha-level?

Any other details to specify?

6. Statistical Tools

What is your chosen analysis method? **Linear Regression**

What is your model (for the mean):

In words:

In mathematical notation:

In R coding language:

What diagnostic tests will be performed?

Will you do any outlier detection? Variable transformations?

If diagnostics fail, what will you do?

7. Caveats

What assumptions are you making (with your model and your sampling)?

EXECUTE & REPORT

Complete your analysis plan as described. Comment on any issues encountered (i.e. with model fit) and interpret your final results. Are they plausible? What did you learn from this exercise?