

Lecture Notes: Final Project: Strategy, Approach and Methodology

Inasmuch as the final project may require some statistical tools that we have not specifically covered, this is a kind of meta-comment on how an experience investigator might typically approach the matter. Some specialized knowledge may have to be acquired to accomplish your project goals, but this rather general overview is intended to layout a common way to approach statistical problems. At the abstract level, much of statistical analysis is quite similar in its methodology. The actual analysis becomes completely straightforward once the problem is setup and posed as an interesting, tractable statistical question. In short, I tried to be general, although this pattern, in fact, is autobiographically inspired. Seek individual consultation if help is needed on specific questions concerning statistical techniques that have not been covered in class.

1. Study the problem and determine what data can be accessed or generated with regard to project direction. What information resources can be brought to bear on the problem?
2. Formulate some meaningful questions, speculations, conjectures, and the like.
3. Try to convert, translate, transmogrify or express the items in 2) as *hypotheses*, that is, statements that can be settled as statistically true or false. This becomes a kind of statement that is considered to be a “yes or no” question. This is somewhat similar to the language construction of asking a question by appending a questioning phrase to the end a sentence, as in the following: “declarative statement , *isn't that true?*” Sometimes this is simply expressed by an inflected, questioning intonation at the end of a declarative sentence. Other languages use this construction more commonly to formulate and interrogative by using endings like *nes pas*, *nicht wahr*, or *es verdad ?*

Very frequently, the hypothesis is formulated negatively with respect to the straightforward question that is motivating the analysis. That is we make a negative conjecture, and then expect to deny its validity thus affirming the desired result.

4. Analyze the situation to try to determine the nature of the statistical phenomena involved. What are the random variables? Are the random processes independent? How are they distributed? What are the parameters that specify their governing distributions? Are the underlying distributions characterized by common parameters, e.g., expectation and variance? What *levels of significance* are indicated? Is there sufficient data to support the analysis?

5. Develop appropriate measures and statistical tests that can settle the hypothesis conjecture. It is often necessary to approach the desired problem somewhat obliquely. Often one cannot formulate the intended question in a statistically tractable form, so one develops some reasonably suitable approximation to the desired entity, one that is answerable statistically. Typically, one asserts or establishes necessary conditions to allow one kind of statistic to be supplanted by another, more readily evaluated one. The latter may not be exact in answering the question, or perhaps only resolves the issue asymptotically when numbers are larger than available samples, etc. Whatever kind of *statistical engineering* is performed here, one must appeal to some recognized principles of statistics for justification. This is similar to constructing a proof in the sense that one introduces elucidating comments to make each step of the argument clear and justifiable.

Finally, revisit the tentative hypotheses generated in step 3). Consider revising them, making them sharper, more useful, more meaningful, or more amenable to analysis of the data at hand. Remember that research literally means “re-search, to search again.” This becomes a process cycle by which deeper understanding results in stronger, more incisive, sharper and more readily verifiable statistical hypotheses.

6. Finally we become satisfied that we have iterated through the previous steps enough to have a statistical hypothesis that addresses our needs and pursuits, one that meets the requirements of the investigation, on the one hand for which the data supports the analysis. At this point, we are left with a purely perfunctory task of evaluating the tests that have been indicated in previous steps. The creative work has been introduced before this point.
7. Lastly, we are only left with the art and science of recognizing the meaning of the analysis in terms of accepting or rejecting the test hypothesis and understanding the implications thereof. We interpret the implications of the results relative to their *domain meaning*, that is, *in* terms of the actual problem context we are investigating. This often leads to derivative hypotheses that pose themselves once the *null hypothesis* is resolved positively or negatively.
8. Ultimately, having completed the analysis and arrived at appropriate conclusions, we are brought to the task of generating a report and effectively communicating the results to a more general audience. What kinds of collateral material will give clarity to the process and emphasis to the findings? For example, if one decides to use a graph to portray the results, what kind of graph is the instrument of choice for the situation at hand? Generally one must engineer presentation material so as the message is communicated unambiguously and effectively relates the findings.