Principles for Statistical Consulting

Fletcher G.W. Christensen

Department of Mathematics & Statistics
University of New Mexico

January 24, 2019

Outline of Presentation

The Motivation

Why Learn About Consulting

- Almost every statistician will, over the course of their career, find themselves consulting with outside investigators on a number of research problems.
- Knowing some of the necessary skills for (and common pitfalls of) consulting can make these interactions far more productive.
- Understanding how to consult also helps us understand how best to approach our own data analyses:
  - What questions do we need to ask?
  - What problems do we need to watch for?

Types of Consulting

Cross-Sectional (Advisory)

- Brief, somewhat superficial conversation between consultant and investigator.
- Statistical advice for data already collected or analyses already carried out.

Longitudinal (Interactive)

- Long-term relationship between statistician and investigator.
- Frequently collaborative and collegial, may lead to the development of new methods or creative application of current methods.

Type-3 Errors

1. Type-1 Error: Deciding that an effect exists when in truth, it does not.
2. Type-2 Error: Deciding that no effect exists when in truth, it does.
3. Type-3 Error: Giving the right answer to the wrong problem.

Type-3 Errors – Examples

Example 1 – Industrial Consulting

1. An engineer studying corrosion wants to estimate the distribution of particle sizes.
2. She believes the particle sizes are normally distributed, but particles of diameter < 0.7 are unobservable using her detection method.
3. She consults with a statistician and explains what she wants: a method for estimating parameters in a truncated Normal. She also shows the statistician a sample of particle sizes from her study:

   25.6, 7.1, 4.2, 3.7, 2.6, 2.0, 1.6, 1.5, 1.3, 1.2, 1.0, 0.9, 0.7
4. The statistician informs her how to estimate parameters for a truncated normal, and both parties leave the consulting session happy.

Type-3 Errors – Examples

Example 2 – Medical Consulting

1. A geneticist is studying the relative biological effects of different kinds of radiation.
2. He counts mutations in samples of organisms when exposed to gamma radiation and neutron radiation. In previous experiments, mutation count has increased linearly with dose.
3. He reports counts to a consulting statistician, along with details of the experiment. His gamma radiation source was "clean", but his neutron radiation source was known to have a gamma-ray contamination of about 7%, so he corrected the mutation counts on the neutron experiment by decreasing them to 93% of their total.
4. The statistician builds a simple linear model for both experiments and compares their slopes.

Type-3 Errors – Source

- These examples come from "Errors of the Third Kind in Statistical Consulting" by A.W. Kimball (JASA, 1957).
- Kimball provides other examples of Type-3 errors in the original paper, along with considerably more discussion of the examples presented, how to avoid them, and how these problems were eventually identified and overcome.
- Kimball's paper (along with the other papers referenced in this presentation) are available on our course website.

A Typical Consulting Session

What Not to Do

- D.I. Bliss ("Communication Between Biologists and Statisticians, A Case Study," TAS, 1969) says his "first rule [for statistical consultants] would be to avoid answering any questions on statistics asked initially by the biologist [because] frequently the statistical function [the biologist] proposes to compute is not the one most pertinent to his problem."
- Investigators very often know what they want from a statistician; far more rarely do they know what they need from one.

A Framework for Consulting

- In "Nonstatistical Aspects of Statistical Consulting" (TAS, 1983), D.A. Zahn and D.J. Isenberg break down the typical consulting session into four phases.
- Each phase requires specific non-statistical skills on the part of the consultant.
- Consulting sessions will often move back and forth between these four phases, though each phase should be addressed for each project the investigator presents.
- The phases are:
  1. Identification of Situation
  2. Definition of Goals
  3. Determination of Actions
  4. Discussion of Expectations

Identification of Situation

"The hardest part of consulting is getting [the investigator] to tell you what they did." – R.R. Christensen

1. What is the current status of the investigator's work?
   - What previous work has the investigator done? What similar work have they (or other researchers) already completed?
   - Have data been collected already? If so, what are the *precise* details of the data collection procedure?
   - What attempts have already been made to solve the problem prompting the investigator to consult with you?
2. What are the practical constraints on the investigator's work?
   - Is the investigator working to a deadline?
   - Are there other collaborators?
   - Does the investigator have unrestricted access to the data?

Identification of Situation

Some good techniques for situation identification:

- *Reflect back* what you're hearing from the investigator. When you think you understand part of the situation, phrase it into your own words and repeat it back to them for confirmation. Often, they will correct misconceptions or offer new details at this point. Sometimes they'll just confirm that you've accurately understood.
- *Ask probing questions* (e.g. "Did you randomize the assignment of subjects to treatments?") Often, investigators aren't thinking about issues that are important for good statistical analysis. Consultants must dig for this information themselves. Be prepared to explain what your questions mean, and why the answers to those questions are important.

Definition of Goals

1. This phase is about finding out what sort of help the investigator wants.
2. Most consulting meetings will either begin in the identification phase or here—though if a meeting begins here, you should leave the goal definitions fairly rough at first and revisit them once you've done more situation identification.
3. At the end of this phase, you want to have a clear picture of *what* the investigator wants to accomplish and *why* they want to accomplish it. Getting answers to both questions is critical.
4. It is also important in this phase to establish *whether* their goals can be addressed by the data that they have, or that they propose to obtain.

Definition of Goals

Some good techniques for goal definition:

- Again, it is good to *reflect back* what the investigator says. This will help clear up any misunderstandings regarding the nature of the help requested and the reasons for requesting it.
- *Take the student role.* Investigators will know the current state of their field, and why their research is important. In the time available, have them teach you as much as they can about the topic.
- *Ask clarifying questions.* "What do you hope to answer with these analyses?" "How will you know if you've answered your research question?"

Determination of Actions

- This is often where investigators want to jump into the consulting session: "Go do this with my data." As discussed before, however, skipping the *Identification* and *Definition* phases often leads to Type-3 errors.
- Properly, this phase is where the consultant helps the investigator to make decisions about what precisely is to be done to address the investigator's goals.
- Actions determined usually fall into three broad categories:
  1. The consultant performs state-of-the-art statistical analyses for the investigator.
  2. The investigator becomes more knowledgable about statistics.
  3. The consultant guides the investigator through performing the necessary analyses.

Determination of Actions

Some potential pitfalls in determining the actions to be taken:

- *The actions to be taken aren't clear:* Usually this results from not having sufficiently addressed the identification and definition phases. Once limitations and goals are well understood, this phase becomes easier.
- *The actions needed are beyond the scope of what the consultant and investigator can do:* Sometimes, because of limitations of resources or ability, it is either impossible or impractical to do what is necessary to address an investigator's goals. Be forthright about your limitations, and ask the investigator to be forthright about theirs. If the necessary actions for a particular goal are not achievable, consider reformulating the goals to yield more achievable actions.

Discussion of Expectations

- Collaborations and consultations often break down when the parties involved have differing expectations about their roles and duties.
- A statistical consultant can be asked to play many different roles, depending on the needs of the investigator:
  - Research assistant
  - Statistical advisor
  - Collaborator
  - Data entry / cleaning specialist
  - Computer programmer
- The consultant may have expectations of the investigator:
  - Committing to a minimum number of consulting sessions
  - Self-educating about areas of statistical methodology
  - Allowing the consultant to use the data set for research

Discussion of Expectations

- Many of the expectations that need discussion can be difficult and sensitive, but they need to be addressed.
- Among these important expectations are:
  - What will be the role and duties of the consultant?
  - What will be the role and duties of the investigator?
  - How much will the consultant be paid for their services?
  - How will the consultant be recognized for their work?
- Consultants and investigators need to come to a point of mutual agreement on these issues.
- If there is continuing disagreement, it may be necessary to revisit the earlier phases to retool the consulting project parameters or acknowledge that the project cannot be performed as outlined.

Investigator Responsibilities – Deming (1965)

- Approval of all statistical methods and procedures.
- Determination of appropriate level of statistical power / precision.
- Approval and implementation of all data collection plans.
- Training and supervision of investigative assistants.
- Determination of the rules for data coding; also, the performance thereof.
- Obligation to report to the consultant at once any departure from agreed procedures.

Consultant Expectations – Deming (1965)

- Formulation of the investigator's problem in statistical terms, subject to investigator approval.
- Explanation of the advantages and disadvantages of various potential data collection plans.
- Explanation of the inferential limitations associated with the data collection plans and statistical methods / procedures approved by the investigator.
- Performance of statistical methods / procedures approved by the investigator.
- Proactive monitoring of the investigation to prevent irresolvable statistical complications from being introduced in data collection and coding.
- Preparation of a report on the results' statistical reliability.

Covariate Considerations

The Consultant's First Step

- A critical part of providing a good consultation is understanding the investigator's problem—all the variables involved and their role in the investigation.
- It may be possible to accomplish this during a first consulting session (c.f. Identification of Situation), but often this will require more effort.
- Strategies that can help you understand the problem:
  - Visit the investigation site and have the investigator walk you through a typical data collection scenario.
  - Schedule a consulting appointment for the sole purpose of discussing variables in depth.
  - Spend time thinking deeply about the investigation.

Dangers of Consultants Lacking Domain Knowledge

- Unwarranted assumptions of process stability during experimentation.
- Undesirable combinations of covariate levels in experimental design.
- Violation / lack of exploitation of known scientific laws.
- Unreasonably large / small designs.
- Inadequate measurement precision of responses or covariates.
- Unacceptable prediction error.
- Undesirable run ordering.

Understanding the Variables

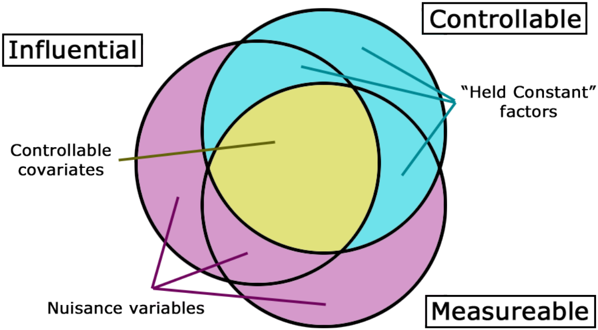

Figure 1. Roles of variables in a designed experiment. Adapted from Coleman & Montgomery, 1993.

Understanding the Variables

Example 3 – An Hypothetical Experiment

Researchers want to conduct a study on how Facebook users engage with Facebook posts.

- Three responses are measured: Does a user "Like" the post? Does a user comment on the post? What form of "Like" action does a user take (if post is liked)?
- What are some controllable covariates in this experiment?
- What are some "held constant" factors in this experiment? Where do they fall in the Venn diagram?
- What are some nuisance variables in this experiment? Where do they fall in the Venn diagram?

1. Experimenter's Name and Organization: **Brief Title of Experiment:**
2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):
3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?:
4. List: (a) each **response variable**, (b) the normal response variable level at which the process runs, the distribution or range of normal operation, (c) the precision or range to which it can be measured (and how):
5. List: (a) each **control variable**, (b) the normal control variable level at which the process is run, and the distribution or range of normal operation, (c) the precision (s) or range to which it can be set (for the experiment, not ordinary plant operations) and the precision to which it can be measured, (d) the proposed control variable settings, and (e) the predicted effect (at least qualitative) that the settings will have on each response variable:
6. List: (a) each factor to be **"held constant"** in the experiment, (b) its desired level and allowable s or range of variation, (c) the precision or range to which it can measured (and how), (d) how it can be controlled, and (e) its expected impact, if any, on each of the responses:
7. List: (a) each **nuisance factor** (perhaps time-varying), (b) measurement precision, (c) strategy (e.g., blocking, randomization, or selection), and (d) anticipated effect:
8. List and label known or suspected **interactions**:
9. List **restrictions** on the experiment, e.g., ease of changing control variables, methods of data acquisition, materials, duration, number of runs, type of experimental unit (need for a split-plot design), "illegal" or irrelevant experimental regions, limits to randomization, run order, cost of changing a control variable setting, etc.:
10. Give current **design preferences**, if any, and reasons for preference, including blocking and randomization:
11. If possible, propose **analysis and presentation techniques**, e.g., plots, ANOVA, regression, plots, t tests, etc.:
12. Who will be responsible for the coordination of the experiment?
13. Should trial runs be conducted? Why / why not?

Figure 2. Predesign master guide sheet from Coleman & Montgomery, 1993. Even in a postdesign consulting scenario, all fields except #10 and #13 remain relevant.

Ethical Considerations

A Starting Point

Extracts from Deming's Code of Professionalism

- A professional seeks recognition and respect for her practice—not for herself alone, but for her colleagues as well.
- She will not follow methods that are indefensible merely to please someone, nor support inferences based on such methods. She will not assent to convenient interpretations unwarranted by statistical theory.
- She will avoid the bias of favoring results of one type over results of another. "As a statistician, I couldn't care less," is the attitude she should bring to all her work.
- She will not deliberately attempt to carry out work that lies beyond her competence—but will, on request, recommend other specialists.

Ethical Conflicts

- In the course of consulting, you may be asked to do something you find ethically dubious.
- These issues can come in many forms. Sometimes you are asked to do something inappropriate: lying, fudging data, deliberately slanting a report.
- Other times, you might be asked to help with research that conflicts with your personal values (e.g. animal testing).
- Sometimes, an investigator may simply refuse to follow your advice on good statistical practice and insist on methods / procedures you know will give faulty results.
- If you find yourself being put in this position, withdraw your participation as gracefully and quickly as you can.

Be Proactive

- On the flip side, take opportunities to model professionalism when they present themselves.
- Don't just respond to investigator requests, reach out to them to check in on their progress and provide timely recommendations.
- Show up: If there are regular project meetings, take the time to attend them and become acquainted with the details of the investigation.
- As a statistician, you have an important role as a methodological gatekeeper. If you see someone doing something methodologically questionable, suggest a better approach to them. If they have a desire to improve, do your best to help them.

In Summary

- Good consulting depends on learning as much about the investigation as possible.
- Make clear plans of action and be sure to establish the expectations for both consultant and investigator.
- Carefully consider all aspects of the data collection scheme and how the data are to be used.
- Cultivate a reputation for professionalism and avoid ethical conflicts.

References

- A.W. Kimball (1957) "Errors of the third kind in statistical consulting," *Journal of the American Statistical Association*, **52**, 133-142.
- W.E. Deming (1965) "Principles of professional statistical practice," *Annals of Mathematical Statistics*, **36**, 1883-1900.
- D.A. Zahn & D.J. Isenberg (1983) "Nonstatistical aspects of statistical consulting," *The American Statistician*, **37**, 297-302.
- D.E. Coleman & D.C. Montgomery (1993) "A systematic approach to planning for a designed industrial experiment," *Technometrics*, **35**, 1-12.