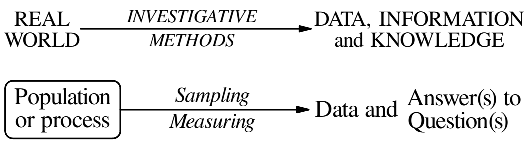
University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: The FDEAC Cycle – Questions and Statistical Questions

Statistical Highlight #87 provides an introductory overview of what statistics is about; its first main point, made in the two schemas shown again at the right, is that statistics is about *answering questions*. As becomes clear in the ensuing discussion in Highlight #87, and its elaboration in Statistical Highlights #88 and #89 for example, this (apparently simple) idea is the starting point for developing the immense sequence of concepts and terminology involved in statistical methods for data-based investigating. However, the emphasis in *this* Highlight #90 is on this foundational idea of answering questions and on clarifying, from among the wide range of question types, those that statistical methods can attempt to answer.



Misplaced emphasis in statistics teaching has resulted in the widespread impression (unfortunately, *not* solely among non-statisticians) that 'statistics' and 'data analysis' are synonymous; this mistake is a source of statistical calamities such as:

- ingoring the question to be answered or at least not formulating it clearly;
- not identifying explicitly the data needed to answer the question;
- wasting resources by overlooking relevant existing data, collecting too little, too much or the wrong data.

Not recognizing, from the inception of data-based investigating, the importance of the question(s) to be answered is epitomized by the travesty of asking a job applicant simply to 'analyze' a data set they are given, sight unseen, to demonstrate their Excel skills.

1. Questions and Statistical Questions

Questions come in wide variety; those that statistical methods can fruitfully address are questions which require *data* to answer them. *We* call these *statistical questions*.

An investigation to try to answer a *non*-statistical question may require answering one or more *statistical* questions – this distinction involves the need for extra-statistical knowledge as against the need for data. For example, a quarterback may ask:

○ How can I increase my proportion of completed passes?

Obviously, *extra*-statistical knowledge is needed to develop a strategy that may produce improvement in passing; THEN:
an investigation using the FDEAC cycle can answer the follow-up question:

○ Has the desired increase in the proportion of completed passes been achieved?

To provide a 'correct' answer, such an investigation would likely need to go beyond the simple-minded comparison of two proportions calculated before and after implementing the improvement strategy. If the desired increase has *not* been achieved, another strategy could be developed, again using extra-statistical knowledge, and its results assessed using the FDEAC cycle.

Another illustration is the question:

○ Is this drug effective in treating disease D?

Again, *extra*-statistical knowledge (medical expertize) is needed to design and run a clinical trial, albeit with *statistical* input about components of the FDEAC cycle that is a statistical basis for such a trial.
Similar considerations apply to answering questions using a poll; in the context of this Highlight #90, in a poll we see clearly:

- the primacy of the questions, AND:
- the need for extra-statistical knowledge to develop the questionnaire, with its added statistical responsibility of being the measuring instrument – for example, see pages HL38.5 and HL38.6 in Statistical Highlight #38.

These illustrations remind us that when answering 'correctly' (*i.e.*, with acceptable limitations in the question context) a substantive question:

∗ there is often collaboration of statisticians with others who provide context and vital extra-statistical expertize;
∗ using the FDEAC cycle directs attention to the essential *initial* requirements for clear formulation and adequate design; both are easily (and routinely) overlooked, especially under the misapprehension that 'statistics' is primarily 'only' data *analysis*.

2. Appendix: Projects, Investigations and Problems

To develop the ideas encompassed by the FDEAC cycle, we distinguish between:

∗ a **project**, which is *broad* and involves *many* questions [one goal of project *formulation* is to *prioritize* these many questions];
∗ an **investigation**, which is *narrower* and involves *one* (*or a few*) question(s) – this question(s) may arise from a project or it may be of interest in its (or their) own right.
  – The question(s) to be answered are the *input* to the Formulation stage of the FDEAC cycle.
  – With*in* the FDEAC cycle, the *output* of each stage is the *input* to the next stage, except that the *knowledge* output from the Conclusion stage *answers* the input *question(s)* to the Formulation stage.

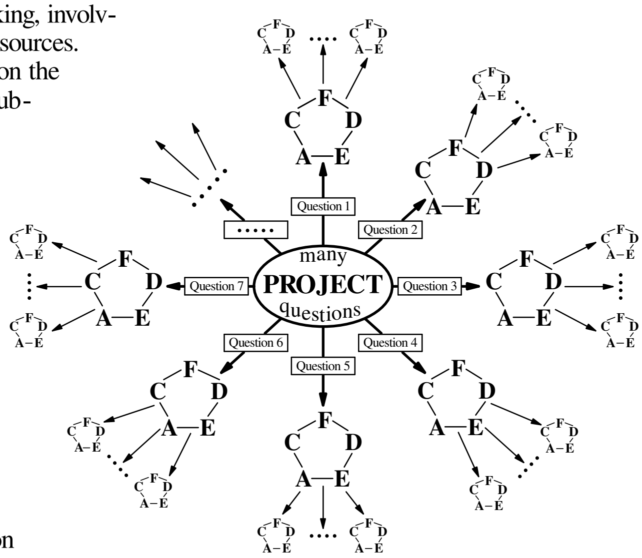
(*continued overleaf*)

> Further discussion of dependencies between stages of the FDEAC cycle, and between four of the stages and the model (which is needed as a basis for *formal* methods of data analysis) is given in Statistical Highlight #88 in Appendix 2 on page HL88.16.

In both projects and investigations, there will inevitably be matters of formulation that require subject-matter expertize – that is, *extra*-statistical knowledge. Examples of projects (phrased as ***How can*** .... questions) are:

- ○ How can the amount of scrap produced by a manufacturing process be reduced?
- ○ How can the quality of drinking water in a province be improved?
- ○ How can the performance on standardized tests of students in a province be improved?
- ○ How can transparent decision-making processes in a large organization be achieved?
- ○ How can a large software system be made less prone to failure or easier to maintain?
- ○ How can satisfaction for the customers of a company providing goods or services be increased?

As indicated in these examples, a project is a *broad* undertaking, involving answers to *many* questions and substantial commitment of resources. The tasks of *formulating* a project and *prioritizing* its questions (on the basis of factors like importance, cost, logical necessity) require subject-matter expertize. The display at the right shows the prioritized questions from a project as inputs to a sequence of FDEAC cycles; each 'main' FDEAC cycle (shown larger) may involve one or more (smaller) sub-cycles to answer questions that arise in answering each 'main' question.



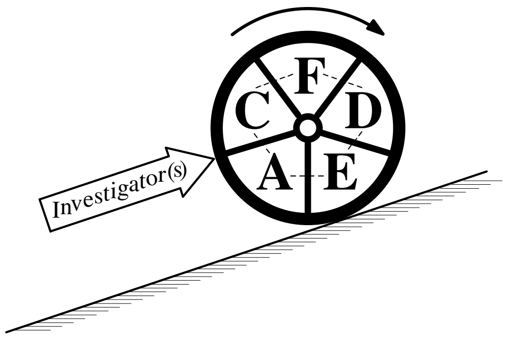
As with answering a statistical question using the FDEAC cycle, a question requiring *extra*-statisatical knowledge is usually also best answered using a *structured process* – for example, the five-stage DefineMeasureAnalyzeImproveControl of Six Sigma. *Statistical* questions arise most obviously in the I stage of DMAIC, but the FDEAC cycle is also relevant to the tasks of project description and question prioritizing in the D stage; statistical considerations of measuring processes, introduced in Section 9 on page HL88.14 in Statistical Highlight #88 and pursued in Statistica Highlight #38, are relevant in the M stage.

- ● To help distinguish contexts which require *distinct* structured processes like the FDEAC cycle and DMAIC, we can say:
  - – the FDEAC cycle deals with *answering (statistical) Questions*, – DMAIC deals with *(extra-statistical) problem solving* [although, in the context of this Highlight #90, *we* might prefer: DMAIC deals with answering *(extra-statistical) Questions*.]

  We *avoid* equivocal phrases like *a task to be done* or *a conclusion to be drawn*.
  Also, these Statistical Highlights use the more descriptive *data-based investigating* in preference to the *empirical problem solving* of the 2004 STAT 231 Course Notes.

To emphasize the *onerous* nature of data-based investigating, the diagram at the right conveys three images relevant to using the FDEAC cycle to obtain Answer(s), with acceptable limitation in the investigation context, to substantive (statistical) Question(s):

- ○ the *effort* of pushing a (heavy) object *up*hill;
- ○ potential *waste* of resources by *premature* cessation of effort;
- ○ the *circular* object is a reminder of the FDEAC *cycle*.



Despite the 'obviousness' of matters like formulating *clear* Question(s) and using measuring processes of acceptable imprecision and inaccuracy in the investigation context, often-poor implementation of these and other components of the FDEAC cycle (*un*necessarily) yields 'wrong' Answer(s).

Experienced investigators realize that *one* misstep in data-based investigating can negate doing correctly *all the other* components of the investigation; an example is malfunction in a measuring process that results in loss of some or all data or yields inaccuracy and/or imprecision that impose unacceptable limitation on Answar(s).