Suggestions for Researchers

> *The most common error in statistics is to assume that statistical procedures can take the place of sustained effort.*
>
> Good and Hardin (2006) *Common Errors in Statistics*, p. 186

Throughout: Look for, take into account, and report sources of uncertainty.

Specific suggestions for <u>planning</u> research:

- Carefully review the literature *and* any relevant research registries before you embark on new research.
    - Take the file drawer problem into account when writing a literature review.
    - These considerations are especially important when conducting a meta-analysis. The new technique of p-curve (Simonsohn et al, 2013) might prove helpful here.
- Decide what questions you will be studying.
    - Do this before any data collection or analysis!
    - Will the research be exploratory (looking for patterns or effects) or confirmatory (testing or otherwise checking out questions that have been previously formulated, either by you or my someone else), or a combination?
        - Remember that testing a hypothesis using the data that led you to formulate the hypothesis is *not* scientifically or statistically sound; if you plan to do an exploratory study then follow up with a confirmatory study, you will need to use two independent data sets.
    - Clearly define the population that you plan to make inferences about.
    - Decide whether hypothesis testing or estimating parameters (with confidence intervals) – or both -- would be best for addressing the questions of interest.
    - Formulate any hypotheses you wish to test and the alternatives you wish to test them against.
        - See Chapter 2 of Good and Hardin (2012), *Common Errors in Statistics (And How to Avoid Them)* for more detail.
    - Remember that trying to study too many things at once is likely to create problems with multiple testing and power, so it may be wise to limit your study.
- If you will be gathering data, think about how you will gather *and* analyze it *before* you start to gather the data.
    - Read reports on related research, focusing on problems that were encountered and how you might get around them and/or how you might plan your research to fill in gaps in current knowledge in the area.
    - If you are planning an experiment, look for possible sources of variability and design your experiment to take these into account as much as possible.

        - The design will depend on the particular situation.
        - The literature on design of experiments is extensive; consult it.
        - Remember that the design affects what method of analysis is appropriate.
    - If you are gathering observational data, think about possible confounding factors and plan your data gathering to reduce confounding.
    - If you plan to do an exploratory study with a follow-up confirmatory study, think about whether to gather data for both studies together (then randomly split it into data for the exploratory study and data for the confirmatory study) or to wait to gather the data for the confirmatory study after completing the exploratory study.
        - The second option has the advantage that you can learn from mistakes in gathering data for the exploratory study when planning data collection for the second study. It also has the advantage that results of the first study might be useful in deciding sample size for the second study.
        - However, the first option may be necessary for pragmatic reasons if data collection involves difficulties and expenses such as travel to a distant location or an elaborate experimental set-up.
    - Plan to record any time and spatial variables present, or any other variables that might influence outcome, whether or not you initially plan to use them in your analysis.
    - Also think about any factors that might make the sample biased.
        - You may need to limit your study to a smaller population than originally intended.
    - Think carefully about what measures you will use.
        - Be sure you understand what they do and do not measure, and any limitations in their use and interpretation.
        - Ask others to critique your choice of measures.
        - If your data gathering involves asking questions, put careful thought into choosing and phrasing them. Then check them out with a test-run and revise as needed.
    - Think carefully about the details of how you will randomize (for an experiment) or sample (for an observational study).
        - Remember that the appropriate randomization or sampling will depend on the method of analysis planned.
    - Think carefully about whether or not the model assumptions of your intended method of analysis are likely to be reasonable.
        - See Day 4 notes for suggestions on model assumptions and places to be cautious.
        - If model assumptions do not seem reasonable, revise either your plan for data gathering or your plan for analysis, or both.
    - If possible, conduct a pilot study to trouble shoot and obtain variance estimates for a power analysis.
        - Revise plans as needed.

- Decide on appropriate levels of Type I and Type II error, taking into account consequences of each type of error.
- Decide how you will deal with multiple inferences, including "data snooping" questions that might arise later. (See Appendix for Day 4 and http://www.ma.utexas.edu/users/mks/statmistakes/datasnooping.html for suggestions on how to do this.)
- Do a power analysis to estimate what sample size you need to detect meaningful differences.
  - Take into account any relevant considerations such as multiple inference, Intent-to-Treat analysis, or how you plan to handle missing data.
  - Revise plans as needed. For example, you may need to reduce the number of questions you wish to study, or plan for a larger sample size, or choose a "more powerful" method of analysis if appropriate.
  - If your resources do not permit having a large enough sample, consider collaborating with others to carry out a multi-site study.
  - References on power and sample size include:
    - Russ Lenth's Power and Sample Size Page, http://homepage.stat.uiowa.edu/~rlenth/Power/ (advice, java applets, discussion group, FAQ, and link to further references)
    - Currie, Gillian, Sample Size and Statistical Power, http://www.dcn.ed.ac.uk/camarades//files/Critical%20Thinking%20NEBM10032%205.pdf (Focuses on sample size for animal experiments, but gives a good overview for many other types of studies as well.)
    - Dallal, Gerald. What Underlies Sample Size Calculations, http://www.jerrydallal.com/LHSP/sizenotes.htm (Gives a good idea of the whys and wherefores of sample size calculation for many standard types of analyses. However, be aware that for many newer methods, sample size and power calculations are done by simulations. See also his addendum at http://www.jerrydallal.com/LHSP/sizeflaw.htm, and other related pages under Sample Size Calculations at http://www.jerrydallal.com/LHSP/LHSP.htm)
    - Simonson, Uri, The Folly of Powering Replications Based on Observed Effect Size, *Datacolada Blog, Oct. 14, 1913,* http://datacolada.org/2013/10/14/powering_replications/

- If you plan to use existing data, modify the suggestions above, as in the suggestions under Item II(b) at http://www.ma.utexas.edu/users/mks/statmistakes/datasnooping.html
- For additional details and suggestions, see:
  - Chapter 8 of van Belle (2008), *Statistical Rules of Thumb*

    - p. 4 (and references there to later chapters and other sources) of Good and Hardin (2012), *Common Errors in Statistics (And How to Avoid Them)*
- Preregister your detailed research plan. If there is no registry available in your field, consider a site such as The Open Science Project (https://osf.io/getting-started/#registrations) or your own web site.
    - For discussion and examples, see Humphries et al (2013) Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration, Political Analysis (2013) 21:1–20), or Alex Holcombe's blog posting at http://alexholcombe.wordpress.com/2012/08/29/protect-yourself-during-the-replicability-crisis-of-science/)

Specific suggestions for <u>analyzing</u> data:

- Keep careful records of decisions made in data cleaning *and* in data analysis.
- Before doing any formal analysis, ask whether or not the model assumptions of the procedure are plausible in the context of the data.
- Plot the data (or residuals, as appropriate) as possible to get additional checks on whether or not model assumptions hold.
- If model assumptions appear to be violated, consider transformations of the data, or use alternate methods of analysis as appropriate.
    - For discussion of strengths and limitations of some methods, and suggestions on model building, see chapters 4-7 and 11- 15 of Good and Hardin (2012), *Common Errors in Statistics (And How to Avoid Them)*.
    - Keep careful records of what you have tried and your reasoning behind the decisions you have made.
- If more than one statistical inference is used, be sure to take that into account by using appropriate methodology for multiple inference.
    - If your reports of your study results will be spread out over more than one publication, be sure that your methodology for accounting for multiple inference is applies to *all* inferences from one data set (not just those inferences in one publication using that data set).
- If you use hypothesis tests, be sure to calculate corresponding confidence intervals as well.
    - This will help you and readers or users of your research keep uncertainty in mind.
    - But be aware that there may also be other sources of uncertainty not captured by confidence intervals; make notes on them to use in writing up your results.
    - Be sure to take multiple inference into account in calculating confidence intervals – for example, calculate them to give overall confidence level 95% rather than individual confidence level 95%
- When using software, pay attention to choose settings that are appropriate for the details of the method of analysis you have chosen. Do *not* automatically use the default settings!
    - Keep careful records of decisions made in using software.

- For more discussion of record keeping, see:
  - K. Baggerly and D. Berry, Reproducible Research, *AMTAT NEWS Science Policy Column*, January 2011, http://magazine.amstat.org/blog/2011/01/01/scipolicyjan11/
  - A. Gelman (2010) and commentators, *Forensic bioinformatics, or, Don't believe everything you read in the (scientific) papers*, http://www.stat.columbia.edu/~cook/movabletype/archives/2010/10/forensic_bioinf.html, and references therein.
  - If you are using R statistical software, see http://leisch.userweb.mwn.de/Sweave/ , http://phys.org/news/2014-02-scientific-young.html , or http://escholarship.org/uc/item/90b2f5xh .

Specific suggestions for writing up research:

> *Critics may complain that we advocate interpreting reports not merely with a grain of salt but with an entire shaker; so be it. ... Neither society nor we can afford to be led down false pathways.*
>
> Good and Hardin (2006), *Common Errors in Statistics*, p. 119

> *Until a happier future arrives, imperfections in models require further thought, and routine disclosure of imperfections would be helpful.*
> David Freedman (2008, p. 61)

- **Aim for transparency and reproducibility**.
  - Be sure to follow any standards in your field (e.g, the Consolidated Standards of Reporting Trials, http://www.consort-statement.org/; see http://www.equator-network.org/ for a list of other reporting guidelines in the health sciences) or in the journal in which you are publishing.
  - Consider such standards as minimum standards – go beyond them if they do not include all of the following.
  - Include enough detail so the reader can critique both the data gathering and the analysis.
    - Report clearly which measures you used.
      - Give precise definitions of these measures
      - Explain why you chose them.
      - Point out any cautions needed in their use and interpretation.
    - Report how you determined your sample size.
    - Report any data exclusions, and reasons for the exclusion.
    - Look for and report possible sources of bias or other sources of additional uncertainty in results.
      - For example, if your study is underpowered (e.g., because of time or financial limitations, or unforeseen

circumstances), point this out clearly and point out the consequent limitation on the interpretation of your results.
        - For more detailed suggestions on recognizing and reporting bias, see pp. 158 - 160 of Good and Hardin (2012), *Common Errors in Statistics (And How to Avoid Them)*. All of Chapter 8 of that book is a good supplement to the suggestions here.
        - If your pre-planned analysis did not confirm your expectations, but you decide to explore your data by additional analysis, state clearly that the results of these additional analyses are just exploratory and need to be confirmed by additional data.
        - Consider including a "limitations" section, but be sure to reiterate or summarize the limitations in stating conclusions -- including in the abstract.
    - Include enough detail so that another researcher could replicate both the data gathering and the analysis.
        - For example, "SAS Proc Mixed was used" is *not* adequate detail. You also need to explain which factors were fixed, which random, which nested, etc.
        - Refer to the notes you have made when performing the analysis.
    - If space limitations do not permit all the detail needed to be included in the actual paper, provide them in a website referenced in the article.
        - Possibilities include a website maintained by a journal for supplementary information; a website run by your own institution; or an independent website such as figshare.com
        - Be aware that the state of digital curation is evolving, so you may want to post materials on more than one website.
- When citing sources, give explicit page numbers, especially for books.
- Include discussion of *why* the analyses you have used are appropriate
    - i.e., why the model assumptions are well enough satisfied for the robustness criteria for the specific technique, or whether they are iffy.
    - If you encounter page limitations, this might go in a supplementary information website.
- If you do hypothesis testing, be sure to report p-values (rather than just phrases such as "significant at the .05 level") *and also* give confidence intervals.
    - In some situations, other measures such as "number to treat" would be appropriate. See pp. 151 - 153 of van Belle (2008) for more discussion.
- Be sure to report and discuss *all* inferences performed.
    - Negative results (those that are not statistically significant, or not practically significant, or do not meet hopes or expectations) are important to report!
    - Resist any temptation to “spin” your results.
    - If your study results in more than one publication, state so clearly and give references to each publication in the others.

- Be careful to use language (both in the abstract and in the body of the article) that expresses any uncertainty and limitations.
    - Be careful *not* to use language that might falsely suggest certainty. (For example, do *not* say that a result obtained by statistical inference is true or has ben proved.)
    - Words such as "effective" can often be interpreted with too much certainly; avoid or qualify their use.
- If you have built a model, be sure to explain the decisions that went into the selection of that model.
    - See Chapters 11-15 of Good and Hardin (2012), *Common Errors in Statistics (And How to Avoid Them)* for more details on model building.
- Make raw data, related metadata, and any other research materials available publicly.
    - This could be via supplemental materials associated with the journal in which you publish research, a repository in your field, a site such as Figshare (http://figshare.com/faqs), a website maintained by your university or other employer or by a granting agency, or your own website.
- For more suggestions and details, see
    - Chapters 8 and 9 of van Belle (2008)
    - Chapters 8-10 of Good and Hardin (2012)
    - Harris et al (2009)
    - Miller (2004)
    - Robbins (2004)
    - Strasak et al (2007)

And bear in mind the advice of Nobel Laureate in Physics Richard Feynman:

> "The only way to have real success in science ... is to describe the evidence very carefully without regard to the way you feel it should be. If you have a theory, you must try to explain what's good and what's bad about it equally. In science, you learn a kind of standard integrity and honesty.
> *What Do You Care What Other People Think? (1988) p. 217*

> "There is one feature I notice that is generally missing in 'cargo cult science'... It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty — a kind of leaning over backwards... For example, if you're doing an experiment, you should report everything that you think might make it invalid — not only what you think is right about it... Details that could throw doubt on your interpretation must be given, if you know them. ... If you make a theory, for example, and advertise it, or put it out, then you must also put down all the facts that disagree with it, as well as those that agree with it. ... In summary, the idea is to try to give *all* of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another. ... The first principle is that you must not fool yourself -- and you are the easiest person to fool. So you have to be very careful about that.

“Cargo Cult Science”, adapted from a 1974 commencement address at Cal Tech, http://calteches.library.caltech.edu/51/02/CargoCult.pdf

References:
D. Freedman (2008), Editorial: Oasis or Mirage?, *Chance* v. 21 No 1, pp. 59 -61

P. Good and J. Hardin (2012), *Common Errors in Statistics (and How to Avoid Them)*, Wiley

Harris, A. H. S., R. Reeder and J. K. Hyun (2009), Common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals: What editors and reviewers want authors to know, *Journal of Psychiatric Research*, vol 43 no15, 1231 -1234

Miller, Jane (2004), *The Chicago Guide to Writing about Numbers: The Effective Presentation of Quantitative Information*, University of Chicago Press

Robbins, N. (2004), *Creating More Effective Graphs*, Wiley

Strasak, A. M. et al (2007), The Use of Statistics in Medical Research, *The American Statistician*, February 1, 2007, 61(1): 47-55

van Belle, G. (2008) *Statistical Rules of Thumb*, Wiley