Research Proposals
how to communicate methodology

W. Cools (ICDS)

February 20, 2019

Research Proposal: Methodology and Statistics

Convince your committee that your study is appropriately designed so that it can be succesful, effective and efficient. Keep in mind that not everyone in your committee is expert in your research domain, some are statisticians.

So you have to make clear

- what is the aim of your study
- how is your study designed to achieve that aim

This draft, based on shared experience in statistical consulting, highlights some issues and remains work in progress. Please share your suggestions for modifications and additions with ICDS (wilfried.cools@vub.be).

Research Aim

Make explicit

- what are the research questions (questions, not stories, not a description of results)
- what are the expected results (eg., treatment outperforms control, an accurate estimate of the influence of age, ...)
- what are the results -at a minimum- to consider the study successful (eg., significant difference)

The aim you specify has consequences for how you need to justify the study, what the implications and requirements are.

Maybe consider the following rough distinction in types of research that may help bring focus in your arguments. The types of design, for example experimental, observational, ... will be discussed in a later section.

The purpose of the research could be

- CONFIRMATORY
  - reason: confirm an expected difference, equality, relation, ... (you know what you are looking for)
  - justification: importance of the scientific question (not simply a curiousity)
  - implication: statistical test or parameter estimation with pre-specified confidence bounds
  - requirement: sample size derivation and relation with cost/availability of observations
- EXPLORATORY
  - reason: to explore data (you are curious what you will observe)
  - justification: the need and importance for the extracted information on substantive grounds

  - implication: description and/or parameter estimation
    - testing/accuracy is typically not the primary aim, but could be secondary
    - less guarantee that you find something interesting
  - requirement: sample size justification in relation to costs (no strict sample size derivation is possible)
    - balance value of new information with costs involved for data collection
    - sample size derivation still required if testing/accuracy is considered primary
- PREPARATORY (how should this be refered to?)
  - reason: to collect data (and monitor procedures) in preparation of future studies
  - justification: the need for such information for implementing succesful -future- studies
  - implication: small scale set-up to show the potential or detect issues
    - phase I and II clinical designs
      - requires decision criteria to proceed or not
    - pilot study: implementation of future study
      - not in itself of interest, not intended for publication

  - requirement: minimal cost to get rough idea (eg., 3 mice per condition)
- TECHNICAL - TECHNOLOGICAL advancement
  - reason: to design, engineer, create, ...
  - justification based on expected contribution versus costs, not statistics
    - proof of concept: feasibility
    - proof of principle: functionality (could it work in principle ?)
    - development application
  - implication: rarely any statistics involved
  - requirement: methodology not related to quantitive research

focus !

Specify -a limited number- of research questions, and be specific. Science is about finding what is looked for, not simply looking everywhere.

Note: Blindly making many statistical tests will lead to detecting differences/relations that are not there, or put in statistical term, will increase type I errors, incorrectly deciding that there is an effect. Adjustments for multiple testing would be required, or better yet, a tentative attitude towards the test results.

p-values and significance

Do not mistake a low p-value (significant result) with an important result.

A p-value is a test result, typically the answer to the question 'could it be due to change' ? Is that what you want or do you want to know how big an effect is ? p-values are not effect sizes, they also depend on sample size. Also consider the parameter estimate and its confidence interval.

definition: stargazing = naively focusing only on what effects are significant

p-values and non-significance, and equality

Absence of proof is not proof of absence.

Note: Failure to show a significant p-value is a failure, not in any way a proof that the effect does not exist. To show equality, maybe consider a non-equivalence test.

interpretation and prediction

Understand the system or simply use it to make good predictions.

Note: Testing (significance) and estimation aim to understand the underlying system, what predictors contribute, and how much. If all you need is making good predictions, shift your focus.

Research Design and Statistics

Garbage in, garbage out.

The potential information depends on - conditions under which the observations are made (eg., independent groups, cross-over, split-plot, ...) - the way research units are assigned/sampled (eg., large random sample, convenience sample, ...)

A poor design will make the study inefficient at best, and fully invalidate the study at worst.

Key example: Statistics in itself can never show a causal relation, an appropriate design is the only design can through an experiment which implies experimental control.

Note: from correlation to causality implies the - exclusion of all confounding variables (requires experiments) - required time-lag for the effects to take place (longitudinal, more or less).

From observation to information and conclusion

Books are written about design and its relation to statistics, too much to discuss here.

Common issues

- how many observations (~ amount of information)
- what are the possible observations (eg., positive continuous, low-medium-high, ...)
- which observations are most likely (eg., very many zero's, mostly low values occasionally high, ...)
- what is the unit of analysis (eg., patient, test score at a given time, ...)
- is the unit of analysis measured repeatedly (between / within)
- how are the units of analysis assigned to various conditions, if any (random, stratified, convenience, ...)
- is the design balanced (same number of observations in each condition, at each time point)
- ...

All questions are relevant to evaluate how to extract the required information and what can concluded based on it.

Number of Observations: Quantity of Information

Observations provide information that is used to test (significance), estimate (confidence interval), predict (cross validation) and understand (qualitative). Observations typically involve a cost (money, time, ethical concerns, ...), so plan your observations accordingly.

Sample sizes are determined depending on the type of research (see above)

- calculate how many observations are required to succesfully conclude the study (significance / high enough accuracy)
- justify the sample size in relation to the cost of observation
- take the minimum that gives a rough idea

Sample sizes are most important for the primary research questions, if multiple exist take the highest value (conservative estimate)

The resulting sample size depends on:

- the statical test in focus (eg., t-test for independent groups at the end of the study)
- the effect size aimed for (eg., .5: difference > 2, assuming pooled population standard error of 4)
- the operational characteristics (type I error α and type II error β)

Note: all these elements should be discussed in a proposal (do not simply state sample size, alpha and power)

The effect size aimed for is based on an effect and an uncertainty:

- effect: should ideally be specified on substantive grounds or at least make reference to common practice, but could be obtained from earlier research
- uncertainty: should ideally be based on earlier data/research

Issue: if you are interested to detect a difference between two groups when you have more than two groups, do not derive sample sizes for the omnibus F-tests.

Issue: sample size are always obtained for future studies, never as a justification for used sample sizes (retrospective power analysis is meaningless).

Conditions of Observation: Quality of Information

Different types of research have different consequences on the quality of the collected information and its processing.

- QUANTITATIVE
  - main question is 'how'
  - focus on population, aim at generalization
  - uses (ideally) representative samples
  - can be descriptive and/or inferential
- QUALITATIVE (holistic/subjective)
  - main question is 'why'
  - focus on understanding of object of research: reasons, opinions, motivations, ...
  - typically explorative / hypothesis generating

note: there is a continuum between the two note: mixed methods formally uses both, can be used consecutively as well

For quantitive research:

The informational value of observations is dependent on the conditions under which they are observed, this is the design of the study.

A general principle is to

- experimentally control the confounding variables (randomisation, blocking, ...) as good as possible
- minimize non-systematic variability (eg., use a reliable tool for measurement)
- maximize systematic variability (eg., use conditions that are as different as possible given the hypothesis)

Control on confounding variables

- randomisation to balance out various disturbances
- enhance comparability (control for unwanted influences)
  - blocking (compare observation with others within a block)
  - repeated measures (compare observation on the same unit, but at different time points)
  - matching (create similar groups, except for the main effect in consideration)
  - cross-over designs (alternate treatments within a unit of observation)
  - ...

Note that an underlying reasoning always is that you want to isolate the effect of interest as much as possible by either avoiding or measuring unwanted influences. Statistically this results in more complex but also better models, typically mixed models.

Selection of Research Units: Generalizability

For generalization, a big enough representative sample is required.

Inference is dependent on the research units under study and how they are sampled.

As an example,
a sample of elderly patients that you know personally does not strictly allow for generalization to

- young people
- non-patients
- people you do not know personally

Example

The aim is to show that the proposed treatment is an improvement over the current standard method.

Participants are randomized into two groups, the treatment and the control group. The control group is given a dummy treatment that mimics as good as possible the actual treatment. A post experiment survey will address whether participants were aware about the use of a dummy treatment. Each participant is measured once, immediately after the (dummy) treatment was administered which results in one score per patient, continuous on a 0 to 10 scale. A t-test for independent means is used to evaluate whether the scores in the treatment group are higher than the scores in the control group.

A sample size was derived for the t-test to detect a minimal difference of 2 in favor of the treatment which was decided upon by our expert panel. The standard method is described in literature to have a population standard deviation of about 4 when used for our type of patients. Because no information is available on the new treatment it is assumed that the same population standard deviation applies. Therefore, a sample size of 51 patients in each of both groups is required for a one-sided test, type I error of .05 and power of .8. Because earlier experiments did show a drop-out of about 10%, $51/9*10 < 57$ patients in each group are included. Note that there is no reason to expect serious deviations from normality. A small but possible risk that is unforeseeable is that there may be much more variance in the treatment group, which would complicate the analyses and increase the required number of participants slightly.

Statistical Analyses Plan

Introduce the statistics that are planned

- in agreement with design
- able to answer the research questions vital for your study

Is it a p-value, a confidence interval, the probability of succesful prediction, ... that is of interest, and how does it answer your questions ?

Reflect on likely challenges and how to deal with them for the analysis of main importance, for example deviations from normality dealt with using log-transformations.

Best practice

- create the data you expect during the design phase -before you have your data-
- verify whether your analysis would be able to extract the required information.

At least

- give the final datafile as expected some serious thought and build the framework by specifying rows and columns before data is collected.

Note: even the most complex and advanced statistics would not make up for a poor design!

Note: do not state that you will be working with SPSS, R -and- Prism without any specific reason to use multiple tools, and think very carefuly about that reason. Using different tools without good reason simply communicates that you do not know how to perform the analyses.