RESEARCH INSTITUTE
NATURE AND FOREST

Herman Teirlinckgebouw
Havenlaan 88 bus 73
1070 BRUSSELS - BELGIUM
T +32 2 430 26 37
www.inbo.be

Statistical Analysis Plan

General information about the Statistical Analysis Plan (StAP) can be found here. If you need any additional information, you can reach out to team BMK through this helpdesk form or by email (bmk@inbo.be).

This document contains 5 sections: metadata, study information, sampling plan, variables and analysis plan. Within each section, the questions of highest importance are indicated with a red asterisk (*). Several examples of a StAP can be viewed here.

1) Metadata

Provide general information about your study and the project it relates to. If your project involves several data collections or studies, write a separate StAP for each of the studies.

Title*

Description*

Describe your project or add a pif-link or SWO-link. Your description should be detailed enough such that the research goals, objectives and experimental setup are clear.

Authors*

Jira number

2) Study information and design

Research question(s) and hypotheses*

Provide your research question(s). When possible translate these questions into specific, concise, and testable hypotheses. List these hypotheses from most important to least important.
Example: Is there a 5% difference in allergy sensitivity between children in green and grey schools? The number of birds (flux) that visits an area reduces after a windmill is built in that area.

Main hypotheses / RQ:

Secondary hypotheses / RQ:

Study type*

Please select one (or more) of the following statements.

- ☐ Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
- ☐ Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes for example surveys, cross-sectional studies, case-control studies
- ☐ Monitoring scheme
- ☐ Meta-Analysis - A systematic review of published studies.
- ☐ Other: ……

Study design*

Describe your study design. The key is to be as detailed as is necessary given the specific parameters of the design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. Typical terminology that can be used here includes parallel group, cross-over, factorial, randomized block, split-plot for experimental studies, retrospective or prospective cohort, case control, cross-sectional studies for observational studies. Also describe how many groups you want to compare, replicates and if there is a follow-up in time (repeated measures). A graphical representation of your design can help to elaborate your experiment.

Example: Full factorial design with two factors which each have two levels (nest protection yes/no and mowing at a later time yes/no). There are 20 replicates for each of the factor level combinations.

Blinding

Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.

- ☐ Blinding is not practically feasible.
- ☐ Blinding is feasible but not needed.
- ☐ Personnel who interact directly with the study workers (field workers, data collectors,...) and the study workers themselves will not be aware of the assigned treatments.
- ☐ Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

3) Random sampling and treatment allocation

In this section we'll ask you to describe how you determine your sample, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

Random sampling*

Please describe the process by which you will collect your data and your inclusion and exclusion criteria. Provide us with the SOP (standard operating procedure).
Describe the statistical population (steekproefkader) from which you will draw your sampling units. Include information about how you will determine sampling units, duration of data gathering efforts, source or location of sampling units.
If you are (partially) using dataset(s) that already exist in this study, include how the original dataset(s) was created.
Example: Oaks (*Quercus robur*) will be selected from a local forest. Only mature oak trees with a minimum trunk diameter of 30 cm will be included. Trees showing signs of disease or physical damage will be excluded.
More information: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study.

Random treatment allocation (Randomization)*

This section is not mandatory when setting up an observational study. If you are planning to allocate a treatment, state how you will randomly allocate this treatment and at what level (define the experimental unit). Typical randomization techniques include: simple, block, stratified, and adaptive randomization. If randomization is required for the study, the method should be specified here.
Example: We will use block randomization, where each sampling unit will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.

Sample size*

Describe the sample size of your study. How many sampling units will be analyzed in the study? This could be the number of individual birds, classrooms, plots, trees or countries included. In a clustered or multilevel design, this could be the number of individuals at each level of the analysis. Indicate if the number(s) you provide is exactly equal to the number of samples or a range, minimum, or maximum. In addition, include a power or precision analysis and list constraints such as time, money, or personnel. Describe how missing data will be taken into account when estimating the sample size (and describe missingness in section 5).

Example: Our target sample size is 280 trees. We will attempt to sample up to 320, assuming that not all trees will be accessible.

4) Variables

In this section you can describe all variables (both manipulated and measured variables) that will later be used in your analysis. This section provides you with the opportunity to describe how each variable will be used.

Outcome and predictor variables *

Precisely define each variable that you will measure. This will include outcome measures, as well as measured predictors or covariates. Mention whether you will transform variables and/or plan to calculate indices. Report the measurement error and detection limit for each variable when needed. When applicable, refer to the published (INBO) protocol.
Example: The primary outcome variable will be foraging efficiency of each wild boar. This will be measured by the number of food items consumed per minute during a 15-minute observation session. Predictor variables will include: age of the wild boar (in months), sex (male/female), body condition score (scaled 1–5, based on fat and muscle reserves) and habitat type (grassland, forest, or mixed).
More information: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise.

DOI published protocol:

Outcome or dependent variables:

Predictor or independent variables:

5) Analysis Plan

In this section, you will describe the analysis for the primary and secondary hypotheses. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis-generating.

Statistical models *

What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, RMANOVA, MANOVA, multiple regression, SEM, etc) and the specification of the model, including the type-I error. This includes each variable that will be included, all interactions, subgroup analyses, pairwise or complex contrasts, and any follow-up tests (for example F-test). State how the FWER (family wise error rate) will be controlled. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Provide enough detail so that another person could run the same analysis with the information provided. Remember that in your final article any test not included here must be noted as exploratory and that you must report the results of all tests.
Example: We will use a repeated measures ANOVA (RMANOVA) with **the** number of food items consumed per minute during a 15-minute observation session as outcome variable. Sex and condition score will be set as predictor variables. This is perhaps the most important and most complicated question within the Statistical Analysis Plan. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you'll test against at those three points.

Outliers and missing data

How will you identify missing data, outliers and influential observations in your data? How will outliers be handled? How will you deal with missing values?
Example: We will verify that each tree will have a measured value at each time point. Outliers will be included in the analysis. Boxplots and histograms will be used to evaluate the distribution of the variables.

Additional information

Note any additional, relevant information about your study which does not fit in any of the previous sections here. This could include preprocessing procedures, initial data analysis, exploratory analysis, reproducibility and documentation.