Guidance for Hypothesis Testing Framework

Introduction

Hypothesis testing is a cornerstone of scientific research, providing a structured framework for evaluating evidence and making decision based on data. By systematically comparing observed results with expectations under a null hypothesis, researchers can draw meaningful conclusions about populations, interventions, and relationships between variables. This process is essential for ensuring the validity and reliability of findings, particularly in fields that rely on evidence-based practices.

The **WHRI Analytical Framework** serves as a comprehensive guide for researchers and analysts to conduct hypothesis-driven studies. It emphasizes consistency, accuracy, and reproducibility, aligning with WHRI's commitment to high-quality research. From formulating research questions to interpreting and reporting results, this framework covers each step in detail, ensuring that all aspects of hypothesis testing adhere to best practices.

This document provides clear definitions, guidance on selecting appropriate study designs and statistical methods, and instructions for analyzing and interpreting data. By following this framework, researchers can confidently approach hypothesis testing, enhance the rigor of their respective fields.

Relevant documents

WHRI Standard Operating Procedure (SOP): Research Question Development

WHRI guidelines for research question development

WHRI Standard Operating Procedures (SOP): Study protocol development

WHRI guidelines for Study protocol development

Definitions

- ***Hypothesis testing*** is a statistical method used to make decisions or draw conclusions about a population based on sample data, by evaluating evidence against a null hypothesis.
- ***Random sample***: subset of individuals or observations selected from a larger population in such a way that every individual or observation in the population has an equal probability of being chosen.

- ***Inferential statistics:*** refers to the branch of statistics that uses sample data to make predictions, decisions, or generalizations about a larger population, often involving hypothesis testing or estimation.
- ***Null Hypothesis (H_o):*** A statement asserting there is no effect or no difference.
- ***Alternative Hypothesis (H_1):*** A statement asserting there is an effect or difference.
- ***One-tailed Hypothesis:*** tests for an effect in a specific direction (e.g., reduction or increase).
- ***Two-tailed Hypothesis:*** tests for an effect in both directions (e.g., either an increase or decrease).
- ***Significance Level (α) (TYPE I error):*** the probability of rejecting the null hypothesis when it is true (commonly set at 0.05). We can identify the type 1 error as false positive.
- ***Type II error (Beta β):*** occurs in hypothesis testing when the null hypothesis (H_o) is incorrectly not rejected, even though the alternative hypothesis (H_1) is true. We can identify type II error as error of omission or as false negative
- ***Test statistical power*** is the probability that a statistical test correctly rejects the null hypothesis (H_o) when the alternative hypothesis (H_1) is true. Power is calculated as $1-\beta$, where β is a probability of making a **Type II error**.
- ***P-value:*** The probability of observing the results, or more extreme, under the null hypothesis.
- ***Test Statistic:*** A value derived from the sample data used to make decisions about the hypothesis.
- ***Study design*** is a process wherein the trial methodology and statistical analysis are organized to ensure that the null hypothesis is either accepted or rejected and the conclusions arrived at reflect the truth
- ***Study population*** refers to all people who enter a research study, regardless of whether they are exposed, are treated, develop the outcome of interest, or drop out of the study before completion

Review previously developed study protocol, data collection procedures and study design proposed

Reviewing a previously developed study protocol is a critical step to ensure that hypothesis testing aligns with the overall research objectives and methodological framework. The process involves a thorough examination of the protocol to confirm its adequacy and relevance for testing specific hypotheses. Below is a structured approach for conducting this review:

Understand the Research Objectives:

Identify the primary and secondary goals to understand the main objectives outlined in the protocol and ensure they are aligned with the hypothesis testing framework.

Confirm specificity by verifying that the research questions are well-defined, measurable, and actionable for hypothesis testing.

Evaluate the study design:

Ensure the study design (e.g., observational, experimental, or quasi-experimental) is suitable for addressing the hypotheses and confirm that the design supports the statistical methods intended for hypothesis testing.

Assess Variable Definitions:

Ensure that the primary and secondary outcomes are clearly defined and measurable. Review that all independent variables (predictors are explicitly outlined, with clear operational definitions.

Identify covariates and confounders included in the protocol to account for potential biases.

By assessing the variables stated in the study protocol decide if confounding approach or exploratory approach should be used for the modelling.

Review Data Collection Procedures:

Confirm that instruments and tools used to collect data are reliable and valid for the variables of interest. In addition, verify that data collection aligns with the protocol, minimizing variability and ensuring standardization.

Validate Analytical Approach:

Review the null (H_0) and alternative (H_1) hypotheses to ensure they are logical, testable, and aligned with the study's objectives. Confirm that the protocol specifies statistical tests that are appropriate for the data type, distribution, and study design. Check whether the protocol accounts for key assumptions of the chosen tests (e.g., normality, independence, and homoscedasticity).

Check for Ethical Considerations:

Ensure ethical guidelines for participant involvement are adhered to, as this may impact the validity of hypothesis testing and confirm that data privacy and security measures are in place.

Identify Potential Challenges and Gaps:

Assess whether the protocol sufficiently addresses logistical challenges, such as sample size adequacy and data completeness. Identify any elements missing from the protocol that could influence hypothesis testing, such as unexpected confounders or biases.

Ensure alignment with reporting requirements:

Confirm that the protocol includes a plan for reporting hypothesis testing results, including statistical outcomes (p-values, confidence intervals) and interpretations. Review provisions for recording and archiving data, analysis scripts, and decisions to facilitate reproducibility.

By systematically reviewing the protocol in the context of hypothesis testing, researchers can ensure the study is methodologically sound and capable of generating meaningful, reliable results. This step is critical to maintain scientific rigor and to align the protocol with the overarching goals of the research.

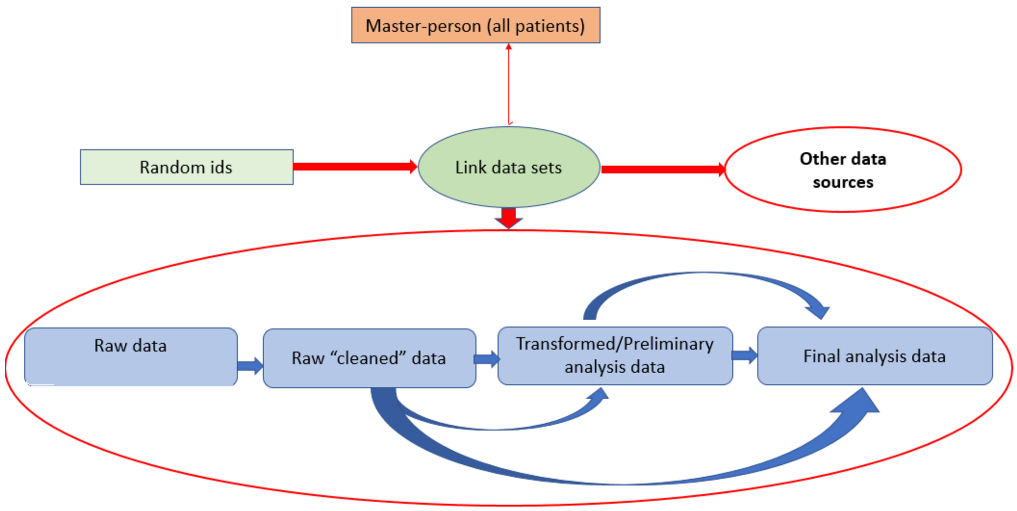
Importance of study design in the context of hypothesis testing

The choice of an appropriate study design is fundamental to the success of hypothesis testing, as it determines the type of data collected, the validity of the findings, and the ability to answer the research question effectively. A well-suited study design ensures that the relationship between variables can be accurately measured while minimizing bias and confounding. For example, observational designs like cohort or case-control studies are ideal for exploring associations, whereas experimental designs such as randomized controlled trials (RCTs) are critical for establishing causal relationships. By aligning the study design with the objectives of hypothesis testing, researchers can ensure that the data collected provides robust and interpretable evidence for or against the proposed hypotheses.

Additionally, an appropriate study design enhances the reliability and generalizability of results. A poorly chosen design may lead to invalid conclusions, either due to systematic biases or insufficient statistical power. For instance, a cross-sectional design may be inadequate for studying changes over time, whereas a prospective cohort study would be more appropriate. Careful consideration of the design also allows for effective control of confounding variables, improving the precision of hypothesis testing. Thus, selecting the right study design is not merely a procedural step; it is a critical decision that directly impacts the credibility, reproducibility, and relevance of the research findings.

Data preparation and validation.

In one of our previous lectures, we described in detail some basic procedures for the data preparation and validation. Below diagram summarize the possible files to be created for the future analysis including hypothesis testing.



Description of the possible data sources to be prepared:

Raw data:

- Should include all the elements as per study protocol/grant proposal
- Each variable in the raw data should be included in the data description table with explanation on how it is related to the study protocol/grant and will be used in the future analysis based on the SAP
- If multiple data tables are included, unique ids should be specified in each table for data linkage

Raw "cleaned data":

- Includes all the elements of the raw data
- All variables transformed to numeric data, except "other specify" type of question (if applicable)
- Data dictionary is created to include variable name, label (description), and variable values applicable

Transformed analysis data:

- Modified raw "cleaned" data sets – change of data structure (from short form to long)
- Any other modifications that alter raw "cleaned" data set structure
- Re-structuring of original data to different analytical tables – patient characteristics, event related data, End of Study points (censoring events), etc.

* Creating and recreating outcomes – separation or combination of variables
* Data type change, from continuous to categorical type, etc.
* Data sets are documented including variables description, appropriate formatting (if applicable)

Final analytical files:

* Data sets created and tailored for specific analysis/sub-project.
* Can be created from raw “cleaned” data sets, “transformed” datasets, or both.
* Multiple analysis datasets can be created for specific project.
* All the data sets are documented including variables description, appropriate formatting (if applicable).
* In case of multiple data sets documentation of how all the data sets are linked is required.

Data validation

When creating different types of data set for future hypothesis testing its important to:

* Check for Missing Values: Identify and address incomplete data to ensure it does not bias the analysis.
* Detect and handle outliers: Identify unusual values and determine whether to include, adjust, or exclude them based on their relevance.
* Validate data consistency: Ensure data entries align with defined formats, ranges, and logical relationships.
* Verify adherence to protocol: Confirm that all data collected matches the study protocol's defined variables and measurements.
* Perform preliminary and descriptive analysis: Summarize data (e.g., means, medians, frequencies) to identify patterns or anomalies before further analysis.

Procedures for hypothesis testing

We already reviewed study protocol, study design, validated the data and prepared analytical files suitable for specific hypothesis testing. Let’s look to appropriate procedures for the hypothesis testing based on research question/s developed.

Formulating Hypotheses:

Once the research question is established, the next step is to translate it into statistical hypotheses. This includes formulating:

Null Hypothesis (H_0): Represents the assumption of no effect or no difference.

Alternative Hypothesis (H_1): Represents the presence of an effect or difference.

Hypotheses must also specify whether they are one-tailed or two-tailed, depending on the directionality of the research question:

One-Tailed Example:

H_0: $\mu_1 \geq \mu_2$ (treatment group mean is greater than or equal to control group mean).

H_1: $\mu_1 < \mu_2$ (treatment group mean is less than control group mean).

Two-Tailed Example:

H_0: $\mu_1 = \mu_2$ (treatment group mean equals the control group mean).

H_1: $\mu_1 \neq \mu_2$ (treatment group mean does not equal the control group mean).

Choosing an appropriate statistical test should be guided by several factors:

- Type of Data: Is the outcome variable continuous, categorical, or count-based
- Number of Groups: Are there one-sample, two-sample, or paired groups
- Data Distribution: Does the data meet parametric assumptions, or is a non-parametric test is needed

For the chosen test to provide valid results, its assumptions must be reviewed and validated. These may include normality, homoscedasticity (equal variances), independence of observations, or linearity in relationships. Failing to meet these assumptions could lead to inaccurate conclusions, so alternative methods or transformations should be considered when assumptions are violated.

When the research involves more complex relationships, analytical models are used:

- Linear Regression: For continuous outcomes.
- Logistic Regression: For binary outcomes.

- Poisson Regression: For count outcomes.
- Multinomial or Ordinal Regression: For categorical outcomes with multiple levels.
- Survival analysis

! Note: of course, more sophisticated models and analysis designs can be used depending on the research question. Please contact our senior statistician at sabina.dobrer@cw.bc.ca if you need to decide on appropriate modelling approach for the analysis.

Each model requires its own set of assumptions, such as the linearity of predictors, lack of multicollinearity, and independence of residuals. Evaluating these ensures the chosen model appropriately captures the relationship between predictors (independent variables) and the outcome (dependent variable). At this point its also important to decide about modelling approach, for example, exploratory or confounding modelling.

- Explanatory modeling is a method that explores why something occurs when limited information is available. It is a "cause and effect" model, investigating patterns and trends in existing data that haven't been previously investigated. For this reason, it is often considered a type of causal research
- Confounding modeling is causal concept that cannot be described in terms of correlations or associations. A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association
- Predictive modeling is a commonly used statistical technique to predict future behavior. It is s a form of data mining that analyzes historical data with the goal of identifying trends or patterns and then using those insights to predict future outcomes

In regression models, it is essential to determine whether specific predictors are meaningfully related to the outcome. This step involves assessing effect sizes, confidence intervals, and statistical significance to ensure that the model provides actionable insights into the relationships being tested.

By following this structured approach, researchers can rigorously test their hypotheses, ensuring the validity and reliability of their findings.

Steps for hypothesis testing and reporting

Hypothesis testing is a systematic process that enables researchers to evaluate the validity of their assumptions about a population based on sample data. By following a series of structured steps, researchers can make informed decisions regarding the null and alternative hypotheses and interpret the findings in the context of their research objectives.

Setting the Significance Level (α)

The significance level, denoted as α, is the threshold for deciding whether to reject the null hypothesis (H_0). It represents the probability of making a Type I error or rejecting H_0 when it is true. The most used value for α is 0.05, but this can vary depending on the study's objectives or field-specific standards. For example, a stricter α (e.g., 0.01) may be used in highly sensitive studies, such as clinical trials.

Computing the Test Statistic

The test statistic is a numerical value calculated from the sample data that quantifies the evidence against H_0. This value is determined using statistical formulas or software and depends on the type of test being conducted (e.g., t-test, chi-square test, regression analysis). The test statistic serves as the basis for determining the p-value, which indicates the strength of evidence against the null hypothesis.

Determining the P-Value

The p-value is a probability measure that shows how likely the observed data (or more extreme data) would occur if H_0 were true. It is derived from the test statistic and the chosen statistical model. A smaller p-value indicates stronger evidence against H_0, suggesting that the alternative hypothesis (H_1) is more likely to be true.

Comparing the P-Value to the Significance Level (α)

The decision to accept or reject H_0 is based on comparing the p-value to the significance level:

If p-value $\leq \alpha$: Reject H_0, providing evidence to support H_1.

If p-value $> \alpha$: Fail to reject H_0, indicating insufficient evidence to support H_1.

For example, if $\alpha = 0.05$ and the p-value is 0.03, H_0 is rejected, as the likelihood of observing the results under H_0 is only 3%. Conversely, if the p-value is 0.08, H_0 is not rejected, as the evidence is insufficient.

Interpretation and Reporting

The final step in hypothesis testing involves interpreting the results in the context of the research question and presenting them clearly. This includes reporting the test statistic value, p-value, and confidence intervals. Of course, other relevant statistics can be reported.

Discussing the practical significance of the findings, as statistical significance does not always translate to meaningful real-world implications.

Acknowledging limitations, such as sample size or assumptions that may affect the reliability of the conclusions.

Clinical relevance and significance should be evaluated. Not everything that is significant is also clinically relevant.

By systematically completing these steps, researchers can ensure that their hypothesis testing is rigorous, transparent, and aligned with the study's goals, ultimately enhancing the credibility and impact of their findings.

Documentation and archiving

Thorough documentation of the hypothesis testing process is crucial for ensuring transparency, reproducibility, and accountability in research. Proper records enable researchers to revisit the analysis, validate results, and share findings with others confidently. This systematic approach safeguards the integrity of the study and facilitates effective communication of the research process.

<u>Recording the Hypothesis Testing Process</u>

The hypothesis testing process should be meticulously recorded to provide a clear and comprehensive audit trail. This includes:

Hypotheses (H_0 and H_1): Clearly state the null and alternative hypotheses, including their directionality (one-tailed or two-tailed).

Chosen Statistical Test: Specify the statistical test selected, explaining why it is appropriate for the data type and research question.

Software and Version Used: Record the statistical software (e.g., R, SAS, SPSS) and its version to ensure consistency and reproducibility of results.

Results and Interpretations: Document the test statistic, p-value, and confidence intervals, along with a concise interpretation of the findings in the context of the research objectives.

Documenting data, scripts, and decisions

- Maintain detailed records of all data manipulations, scripts, and analytical decisions made during the hypothesis testing process. This includes:
- Data Documentation: Ensure all datasets are annotated with clear variable definitions and metadata to facilitate understanding and future use.
- Scripts and Code: Archive scripts or code files used for data preparation and analysis, with comments explaining each step for clarity.
- Decisions: Record all key analytical decisions, such as handling missing values, choosing specific models, or addressing outliers, along with justifications.

Organizing and Securing Outputs

All data, scripts, and analysis outputs should be saved in a secure, well-organized manner to ensure accessibility and protect sensitive information. Use a logical directory structure with clear file names that reflect their contents. Track revisions of scripts and datasets to prevent loss of work or duplication of effort. Store files in secure locations, such as encrypted drives or institution-approved cloud platforms, with appropriate access controls to protect confidentiality.

By adhering to these practices, researchers can ensure that their hypothesis testing process is rigorously documented and easily reproducible, meeting the highest standards of scientific integrity. This meticulous record-keeping also enhances the credibility and impact of their findings in the broader research community.

Summary

The WHRI Analytical framework guidelines provide a structured and detailed approach to designing, conducting, and documenting hypothesis-driven research, emphasizing accuracy, reproducibility, and alignment with study objectives. It covers key elements such as defining clear and measurable research questions, selecting appropriate study designs and statistical methods, and validating data for consistency, reliability, and adherence to protocols. The framework also stresses the importance of formulating well-defined null (H_0) and alternative (H_1) hypotheses and choosing statistical tests and models suited to the data type and research objectives while ensuring all assumptions are met.

Steps for hypothesis testing include setting a significance level, computing the test statistic, determining the p-value, and interpreting results in the context of the research question. The framework outlines procedures for thorough documentation, including recording hypotheses, tests used, results, and decisions, as well as maintaining secure, well-organized data and scripts to ensure transparency and reproducibility. By following these guidelines, researchers can enhance the credibility and impact of their findings while adhering to high standards of scientific rigor.