5 Preparing and Presenting a Dissertation Proposal

- The purpose for presenting a proposal for a dissertation is that the proposal gives a Ph.D. student the opportunity to present to a group of Ph.D. statisticians

  (a) **what** *the student proposes to study*: What is the research problem in the dissertation?

  (b) **why** *the student proposes the research problem*: What is the importance and relevance of research problem?

  (c) **how** *the student proposes to perform the research*: What statistical methodology does the student propose to use in the dissertation?

  (d) **when** *you intend to do this work*: When do plan to complete the different components of the dissertation?

- The Ph.D. dissertation proposal is a statement of intent. That is, how do you intend to accomplish (a), (b), (c), and (d)?

- Note: **You are not expected to present research results in the proposal.**

- After presenting your proposal, your committee will ask questions and often will recommend changes before you can proceed further with the research.

  – Your committee may add new objectives if they think the proposal does not include enough original research.

  – Your committee can also remove proposed objectives from your proposal if they think it is too long (not manageable) or the objective are not relevant.

- If you successfully defend your proposal, then you have received the approval of your advisor and committee to proceed with the proposed research plus the recommended changes of the committee.

- Consider the proposal to be an 'agreement' or 'contract' between the student and the committee. This 'agreement' or 'contract' is meant to:

  – **Protect the student** from the committee demanding additional research items not in the proposal or changing the objectives of the proposal at a later date. For example, the advisor should not demand that "You also need to include results for 6 experimental factors." when the proposal says results will be presented for ≤ 5 factors.

  – **Protect the committee** from a student who tries to get a Ph.D. that is not of an acceptable quality. For example, the student cannot say "I did know I had to include results for 5 factors." if it is in a written proposal approved by the committee that results will be generated for ≤ 5 factors.

- This is why *clarity* is so important in the proposal (see Section 3.3). If the proposal presentation is clear, then the expectations of the student and the committee are also clear.

- In a successful proposal, the Ph.D. student establishes the following:
  - The student is familiar with statistical literature related to the proposed research problem.
  - The student demonstrates a clear understanding of the proposed methodology to be used.
- In the written proposal and in the defense of your proposal, you want to convince your committee that you are capable of completing all of the necessary research tasks. This includes demonstrating statistical knowledge, technical skills, and the motivation needed to write and defend a dissertation.
- A good dissertation will have a good proposal. Here are two properties of a good proposal:
  - The research objectives are focused. You want to avoid a research problem that is too broad and cannot be clearly defined.
  - The proposed research is manageable with respect to time. I discussed time management in detail in Section 2.4 and Section 3.2. Remember that research usually takes longer than you originally thought.
- When your committee reviews your proposal, they are looking for evidence and reassurance that the dissertation will be a high quality document if all of the stated research objectives are successfully completed.
- The dissertation is a reflection of the quality of Ph.D. program in statistics at that university. Therefore, the committee wants to be sure that the reputation of the Ph.D. program is not hurt by a poor quality dissertation. The proposal is the first step to ensure the reputation of the Ph.D. program is maintained.
- Your committee, in general, will not let you graduate if you have not completed all of the objectives stated in the proposal. Thus, is you pass your proposal defense, your committee considers the completion of the objectives in the proposal as the *minimum requirement* for your graduation.
  - Because the proposal includes objectives and the scope of your research in detail, your committee will examine it to determine if you propose to do too little or do too much.
  - If you are unable to complete all of the stated objectives in the proposal, then your committee will not be happy (and possibly not pass you). Thus, do not include more objectives in the proposal than necessary.

5.1 The contents of the proposal.

- The format for a dissertation proposal varies across universities. Check with statistics faculty at your university to see if there exists a specific structure for the proposal.

- You should also ask for a copies of one or more recent proposals from Ph.D. students who have passed their proposal defense. These can serve as a guide when you have to prepare your proposal.
- Good proposals contain the following:
  (a) Statement of the Research Problem and the Importance of the Research
  (b) Statement of the Research Objectives
  (c) Statement of the Research Scope
  (d) Literature Review
  (e) Statement of the Research Methodology
- Note again, you are not expected to present results in the proposal.
- I recommend that $\frac{1}{2}$ to $\frac{2}{3}$ of your proposal should be devoted to the statements of the research problem objectives, scope, and methodology [(a), (b), (c), and (e)]. The remainder should be the literature review [(d)].
- In general, the proposal will provide details to clarify any issues concerning **what** research and **when** you plan to do the research in terms of objectives, **how** you are going to do the research in terms of methodology, and **why** your research objectives are relevant and important in statistics.
- For more information on relevance and importance, read Section 3.1 and Section 3.3 again.
- Note: Almost everything in your proposal will also be included in your dissertation.
- You should not announce the proposal defense until your proposal is almost complete and your advisor says that it is acceptable.
- You should also schedule a practice presentation with your advisor to determine if it is too long or too short. You want to be sure you address (a), (b), (c), (d), and (e) in the presentation. A practice presentation will help.

My personal experiences

- I gave two practice presentations before I had my scheduled proposal presentation and defense. The first practice presentation was too long, so I revised it, and gave a practice presentation of the revision.
- At Montana State University, I require any Ph.D. student I am advising to give a practice presentation before the scheduled proposal defense.
- The practice presentation usually is one week before the scheduled proposal defense. This gives the student enough time to make changes and give a second practice presentation (like I did when I was a Ph.D. student).

- When you give your presentation, you should not be reading from a set of notes. You should also be able to give the presentation **without a set of notes**. Without notes, you are demonstrating to your committee that you have an indepth understanding of your proposed research.
- As an advisor, I do not want my Ph.D. student to appear unprepared and give an unprofessional presentation.

5.2 Describe how the research relates to published statistical research.

- A good proposal will have a thorough review of the statistical literature that is relevant to the proposed research problem. You want to know how your research fits into the history of previous and related publications.
- In your literature review, you are demonstrating to your committee how your research problem is related to the existing scholarly research. Therefore, you will need to *understand the importance and relevance of published research* from journals, textbooks, conference proceedings, and other professional publications in statistics.
- To 'understand the importance and relevance of published research' does not mean you can reproduce everything your read. For example, you are not expected to prove an important theorem you read, but should should be able to understand why the theorem is important and relevant.
- However, when you are writing the dissertation, you may need to reproduce results cited in your literature review.
- By reading your proposal and seeing your proposal presentation, your committee will assess how well you understand the statistical literature related to the research topic and how your research is connected to the literature you reviewed.
- In the literature review, you need to focus on the publications that are most relevant to your research and support the importance of it. The literature review is the focus of Chapter 6 of the notes.

Recommendations:

- I recommend that you organize the literature into topics and subtopics. For example:
  - If you want to review the topic 'Space-filling experimental designs', you should create subtopics because 'Space-filling experimental designs' is too broad to summarize in a single section.
  - Subtopics include 'Latin hypercube designs (LHDs)', 'orthogonal arrays (OAs)', 'maximin distance designs', 'minimax distance designs', and 'uniform designs'.
  - You should then ask yourself, are these subtopics still too broad. If so, you can divide the subtopic further. For example, Latin hypercube designs (LHDs) can be divided into "orthogonal column LHDs', 'symmetric LHDs', and 'optimal LHDs'.

- I also recommend that you give a simple example for many of the topics you review, and, when possible, use graphical methods in figures. It is much easier to understand a complex concept if an example is included.
- You should read proposals written by recent Ph.D. students who have passed the proposal defense at your university.

5.3 Justify the originality and relevance of the proposed research.

- In your proposal, you need to convince your committee that your research will make a significant contribution to statistical theory or methodology. Thus, you need to justify the originality and relevance of your proposed research.
- When describing the originality and relevance, you should clearly identify how your research problem "fills some gap of knowledge" in statistics. That is, you propose to solve an unsolved problem that will be of interest to other statisticians.
- Your committee members have to be convinced:
  - That your research objectives, research scope, research methodology, and literature review address a part of statistical theory and methodology should be studied.
  - That you have the skills and ability to produce results that use sound statistical methodology and theory.
  - That you can produce results will "withstand the scrutiny" (pass strict review) of other research statisticians when you submit the results in a paper for publication in a professional statistical journal.
  - That your research problem is at a Ph.D. level (and not at an MS degree level).
- A clear statement of objectives and scope will:
  - Explain the goals and research objectives of the study.

    (What do you hope to find?)
  - Show the original contributions of your study by explaining how your research questions or approach are different from previous research.

    (What will you add to the field of knowledge?)
  - Provide a more detailed account of the points summarized in the introduction.
  - Include a rationale for the study. (Why should we study this?)
- In Section 3.1 for examples of originality and relevance.

5.4 Show that you are prepared to complete the research.

- In your proposal and proposal defense, you want to convince your committee members, that you are academically prepared to complete the research. This includes:

  - Acquiring knowledge and comprehension of important material from the courses in your Ph.D. program of study.
  - Having computing skills (such as R and Matlab) needed to perform the research, or convincing your committee that you can develop those skills quickly.
  - Describing the methods you propose for collecting data, analyzing data, running computer simulations, ...
  - Convincing your committee that these methods are appropriate for your proposed research.

- Note: You can have an original and relevant research problem, but you must also have sufficient technical "competence" to do what is required. Technical "incompetence" can prevent or ruin well-planned research problems.

- Here are several questions your committee will consider:

  - Does the student understand the statistical material in the literature review?
  - Does the student understand the proposed methodology?
  - Does the student have the technical skills required for the research?
  - Is the student knowledgable with respect to any data collection techniques (such as experimental data, survey sampling data, computer simulated data) that are required in the research?
  - Can the student show that the proposed data collection techniques propose are appropriate for the research?
  - Does the student know how to interpret and make sense of the type of data that will be collected?

- You need to provide the committee members with a well-written proposal with enough time to thoroughly read it before the proposal defense.

- Be sure the written proposal is clear and avoid the writing style problems and mistakes reviewed in Section 3.4. For example, use consistent notation throughout the proposal.

- During the proposal defense, be prepared to answer questions from your committee without any help from your advisor.

My personal experiences

Here are problems with proposals that I observed as an advisor and committee member:

- The Ph.D. student was unable to clearly state how the proposed research was related to the literature review.

- The literature review has too many 'general' references (such as textbooks) and not enough 'specialized' references (such as journal articles). See Section 2.2 for comments on 'general' and 'specialized' references.

- The proposal presentation was very poor because the student did not practice the proposal presentation enough times before the proposal defense.
- The notation used by the student was inconsistent.
- There were not enough examples in the literature review.
- The student did not truly understand the scope of the computing skills required to perform the research.

5.5 Recommendations

- Before you begin to write the proposal, I strongly recommend that you **prepare an outline** of the dissertation. This will help with preparing the proposal.
- As a Ph.D. advisor, I require this from my students. Here is an example of an outline.

Outline of Dissertation:

Applications of Number-Theoretic Methods in the Generation of Designed Experiments

Literature Review

- Definition and goals of Response Surface Methodology (RSM)?
- Who uses RSM? Why do they use RSM?
- Common approximating models (first-order, interaction, and second-order models)
- Mixture experiments and the Scheffe mixture models.
- A review of response surface designs.
  - Two-level designs to fit the first-order and interaction RS models (2^k factorial designs, 2^{k-p} fractional factorial designs, Plackett-Burman designs).
  - Common RS designs used to fit the second-order RS model (CCDs, BBDs).
  - Discuss the use of computer algorithms and design optimality criteria (D, G, A, and IV optimal design criteria) to generate efficient RS designs of any design size.
  - Provide a review of mixture experiments (simplex lattice, simplex centroid, and extreme vertices designs).
  - Provide a review of space-filling designs (including why researchers are interested in using them). In this section should provide more details because it is directly related to your research.
    * Define minimax and maximin designs. Provide an example of each.
    * Define an orthogonal arrays (OA). Provide a couple of examples.
    * Define Latin hypercube (LH) designs.

- Review number-theoretic (NT) methods for generating designs, and give two examples for each method.
  - Review the lattice-point (LP) method.
  - Powers modulo a prime methods (PMP) when N is prime.
  - Square-root sequence method (SRS).
  - Powers of the (s+1)st root method (PR).
  - Cyclotomic field method (CF).
  - Halton set method (H1) and Hammersly set method (H2).
- Mean square error rep-point method (RP).
- Review design evaluation criteria.
  - F* discrepancy.
  - Root mean-squared distance (RMSD).
  - Average distance (AD).
  - Maximum distance (MD).

Relevance, Originality, and Objectives

Discuss that very little research has been conducted on the use of NT-methods (originality). The primary objectives will be to:

1. Develop algorithms for generating space-filling designs for each of the NT-methods when the experimental design space is
   - The k-dimensional cube
   - The k-dimensional sphere. This includes a discussion of the subset method and the transformation method.
   - The $(q-1)$-dimensional simplex. This includes a discussion of the transformation method of the points from designs in the $(q-1)$-dimensional cube into the $(q-1)$-dimensional (q-component) simplex.
   - A subspace of the q-dimensional simplex which is defined by upper or lower bound constraints on each of the mixture components.
2. Apply the algorithms to generate new designs, and evaluate the RMSD AD, and MD criteria for each design (originality and relevance).
3. Create a catalog of the best designs for each distance criteria across a range of design sizes and number of design factors (relevance).
4. Prove that the distance criteria are invariant to rotations in the sphere (originality and relevance).
5. Determine the conditions (if any) when the distance criteria are invariant to rotations in the cube or the simplex (originality and relevance).
6. Make recommendations such as designs that are robust across all three criteria (relevance).

Scope of the Ph.D. Research

1. Generate space-filling designs for each of the NT-methods with the following scope:

   - Generate designs having $k = 2, 3, 4, 5$ factors for the k-dimensional cube and the k-dimensional sphere.
   - Generate designs having $q = 3, 4, 5, 6$ mixture components for the simplex.
   - For the designs in the cube and the sphere, do the following:
     - For $k = 2$ factors, consider design sizes $N = 3, ..., 20$.
     - For $k = 3$ factors, consider design sizes $N = 4, ..., 30$.
     - For $k = 4$ factors, consider design sizes $N = 5, ..., 40$.
     - For $k = 5$ factors, consider design sizes $N = 6, ..., 50$.
   - For the mixture designs in the simplex, do the following:
     - For $q = 3$ components, consider design sizes $N = 3, ..., 20$.
     - For $q = 4$ components, consider design sizes $N = 4, ..., 30$.
     - For $q = 5$ components, consider design sizes $N = 5, ..., 40$.
     - For $q = 6$ components, consider design sizes $N = 6, ..., 50$.

2. Present examples of generating designs when the design space is a subspace of the simplex and is defined by upper or lower bound constraints on each of the mixture components.

   - Find examples in the statistical literature of mixture experiments in constrained spaces in the literature for $k = 3, 4, 5$, and 6 components. Generate designs using the NT-methods in the constrained subspaces.
   - Consider design size N to be similar to the N in the examples.

Comments:

- After you have completed an outline, I recommend making a list of references that are related to the items in the Literature Review.
- For example, in the outline there was:

  * Define minimax and maximin designs. Provide an example of each.
  * Define an orthogonal arrays (OA). Provide a couple of examples.
  * Define Latin hypercube (LH) designs.

  Now, you should be able to make a list of references for (i) minimax designs, (ii) maximin designs, (iii) orthogonal arrays, and (iv) Latin hypercube designs.

- If you were taking good notes, this will not take long (especially if you were using the 'keywords' method I recommended in 'Personal Experiences' in Section 2.3 (page 15).