

7 Writing a Dissertation

- Chapter 7 provides guidance for taking all of your research and writing a professional multi-chapter dissertation in statistics. You will need to determine (i) what to include and exclude from all of your notes and results, (ii) how to organize all of your research materials, and (iii) how to present it to your committee.
- When you start to write the first draft of your dissertation, you should already have a lot of well-organized notes. If you do not have a lot of notes or if they are not organized, then you have wasted a lot of time. As I have said many times, time-management and good note-taking skills are necessary for a good dissertation.
- The focus of writing is preparing a professional presentation. You will “communicating” and not finding new results. Remember: **Writing is not exciting.**

7.1 General Comments About Writing a Dissertation.

- As you write the different Chapters of the dissertation, always think about your research objectives and research scope. You want your writing to stay focused on the objectives and scope.
- You should create an Outline with the chapters and the sections and subsections that divide each chapter in your dissertation into its important parts (See Section 5.5 for an example). Typical dissertation chapters are:
 - **Introduction:** The Introduction contains a general background that motivates the relevance of the study and a description of the research problem (including the research objectives and the research scope).
 - **Literature Review:** The Literature Review chapter in a dissertation is an expanded version of the literature review in the proposal (See Chapter 6). That is, it will include almost all of the proposal literature review plus any additional references discovered since the proposal defense.
 - **Methods or Methodology:** In the Methods chapter, you must describe the statistical methodology you used to generate your research results. Remember to include how your methods will generate the results to satisfy your research objectives and research scope.
 - **Results:** In the Results chapter you describe and summarize the research results you found. Remember to include how your results satisfy your research objectives and research scope.
 - **Discussion:** In the Discussion chapter, you explain what your results mean, draw conclusions within the research scope, how they filled a ‘gap’ in statistical knowledge, discuss the practical application of the results, and list any items for future research.
 - * Remember that you are doing more than just summarizing the results. You need to make conclusions and state implications of the results.

- Before you begin writing, it helps to have an idea about what your finished dissertation should look like. Therefore, if possible, look at one or more dissertations before you begin writing and while you write. Examine the structure and organization of the chapters, sections, and subsections. You may be able to use these dissertations to form a model for your dissertation. For example:
 - Can you estimate how many pages your dissertation will be?
 - * Consider the number of pages your Literature Review and Methods summary was for your proposal. It will probably be a several pages longer in the dissertation.
 - * How many tables and figures will there be in the Results section?
 - * How many pages do you expect will be in each item in your Outline of the dissertation?
 - * If you have one Appendix or at least two Appendices, approximately how many pages will be required?
 - Should I change the order of the items in the Outline ?
 - * Is the order of the sections and subsections in each chapter logical and coherent? Or, are you putting unrelated items together?
 - Do your sentences and paragraphs make sense?
 - Have you allowed enough time to make several revisions?
- You cannot leave all or most of the writing to the end of the process. Many Ph.D. students fail to realize that any work that is submitted is considered a draft. You must allow sufficient time for at least two revisions before your advisor allows you to give copies to your committee to read. Then, expect to make at least one revision based on the comments of the committee.
- You must start writing as soon as you can, and continue to write as often as you can. The more time you allow for revisions, the more professional the submitted dissertation will be.
- You can write the different items in the Outline of the dissertation in any order that makes an efficient use of your time. I suggest to begin writing with sections and subsection that you know and understand the best.
- You should also add to the Bibliography every time you cite a reference in any of the chapters.
- A common mistake that Ph.D. students make when writing the dissertation is to assume the committee knows the definition of many of the statistical terms you use. You must have *clarity* in your writing. For example:
 - You refer to using ‘bootstrapping’ in the Methods chapter. You must define ‘bootstrapping’ and describe the specific type of bootstrapping approach (such as ‘accelerated bias-corrected bootstrap methods’).

- You state the ‘the design points are uniformly scattered in the design region’. What do you mean by ‘uniformly scattered’? Provide a definition.

Every statistical term must be defined by using either (i) a reference to a previously published definition or (ii) providing a precise definition for any new terminology you created. Each definition must appear **before the term is used for the first time** in the dissertation.

- Because you will be making revisions, you must be sure that any term used in the revision has been defined earlier in the dissertation. For example, if you include a reference to fractional-factorial designs on the top of page 38 of a revision, you must be sure that a fractional-factorial design is defined *before page 38*.
- Each term should be used in only one way throughout the dissertation.
 - For example, use ‘experimental design space’ or ‘experimental region’, but not both.
- Consistent notation should be used in only one way throughout the dissertation.
 - For example, for a design size choose between N and n . Suppose you choose n . You will then use n throughout the dissertation. Also, do not use n to represent anything else.
 - This will require you to change the notation used in references. For example, if a journal article uses N , there is no problem changing it to n for the dissertation.
- For early drafts, use simple plots (possibly hand-made drawings) for graphics. Do not spend time on making a graph look professional early because there is a high probability that you will have to modify it later. I recommend making high-quality graphics at the end of the writing process.
 - I have seen students spend a lot of time making a complicated flowchart of a computer algorithm for a computer simulation. They are excited to show the advisor the flowchart and how nice it looks. Later the algorithm is modified and a new flowchart has to be made. The student wasted a lot of time making the original flowchart which can no longer be used.
 - I recommend drawing a flowchart or pseudo-code using paper and pencil. Only when it is certain the flowchart or pseudo-code is correct should you prepare a professional-looking version.
 - I have also seen students spend a lot of time entering numbers in tables and then find out there was a minor mistake in their computer program. The student wasted a lot of time entering numbers that could not be used.
- If you are presenting information in a table or figure, then you **must reference and discuss the table or figure in the chapter**. Also, the table or figure must appear after the reference.
- If there is nothing to discuss for a table or figure, then there is no need for the table or figure in the dissertation. It should be removed.

7.2 Writing the Different Chapters

7.2.1 Writing the Introduction

- The Introduction to your dissertation should explain to your committee the research problem in statistics that you are going to investigate. It should include the Research Objectives and the Research Scope.
- Once again, remember the dissertation is your project. Your advisor is there provide guidance, and not to tell you exactly what to do.
- The introduction describes the process you used to investigate the research problem and emphasizes the originality and relevance of your work.
- You should avoid presenting any conclusions that you have made during the research process. You will save conclusions for the Discussion chapter.
- The end of the Introduction usually ends with a ‘map’ that briefly outlines what will be contained in the other chapters or sections of the dissertation. For example:
 - Chapter 2 contains a brief review of response surface methodology with emphasis on response surface designs followed by a literature review. The literature review is focused on space-filling designs including Latin hypercube designs, orthogonal arrays, and an introduction to number-theoretic designs (which are basis of the original research in this dissertation). Chapter 3 ...
- Once you have completed generating all of your research results, you will have a much clearer idea of whether or not you successfully satisfied the research objectives.
- As you write the other chapters of the dissertation, you may find that your research objectives and research scope stated in the proposal have changed a little. Therefore, you will need to revise the objectives and scope for the Introduction to the dissertation with the approval of your advisor and committee. You cannot just change the objectives and scope and expect your committee to accept the changes.
 - For example, suppose in the proposal the research scope was to generate experimental designs of size $N \leq 40$. However, your computer can only handle a maximum of $N = 35$. You should state the original scope was for $N \leq 40$ but due to computational limitations the scope had to be revised to $N \leq 35$.
- After writing the other chapters, you should review the Introduction again. Carefully read the Introduction and ask “Does the Introduction provide motivation regarding the relevance and originality of the research found in the Discussion chapter?”

7.2.2 Writing the Literature Review.

- It is common for a Ph.D. student to want to know how long (in terms of pages) a literature review should be. There is not a specific or required length. It will depend on how much related statistical literature has been published in your research area

and, specifically, on your research topic. Your advisor should be able to indicate the approximate length of your literature review once you have a list of publications to read.

- In general, you should not worry about the length of the Literature Review. The goal is for your the contents of the literature review to be thorough with respect to the important statistical literature related to the problem.
- It is better to have a shorter literature review that is a thorough exploration of the research topic, than to have a longer literature review that contains material that is irrelevant to the research topic. Therefore, you only want to cite publications that are related to your research.
- You want to discuss how the cited literature is related to your research. That is, you want to refer to published research to motivate the originality and relevance of your research.
 - For example, suppose values in the research scope are based on the recommendations appearing in prior studies that are cited references. These values in these references should be stated in the Literature Review and also should be discussed in the Methods chapter as the basis for your research scope.
- You should also include examples related to important definitions and concepts.
 - For example, after you give definition of a ‘Latin hypercube design (LHD)’, you should give a specific example of generating a LHD.
- The literature review in the dissertation will not be exactly the same as the literature review from your proposal. As you write the dissertation, you will find new published literature that is relevant to your research and should be included in this chapter. You can and should add references to your literature review when you find them.

7.2.3 Writing the Methods.

- When writing the Methods section you must clearly identify the statistical methods you will use to generate the results needed to meet the research objectives and the research scope.
- You must also demonstrate that you understand all assumptions associated with the methods you used and justify why those methods are appropriate for your research.
- For example, you cannot just state “Bootstrapping methods will be used to generate confidence intervals for the population correlation coefficient.” You need to show that your study meets the necessary assumptions to use bootstrapping methods to generate confidence intervals for the population correlation coefficient.
- Remember that ‘methods’ are different than ‘results’. Therefore, you should not be presenting summaries (such as tables and figures of results) in the Methods chapter.

- Do not assume that your committee members or anyone who will read the dissertation understands the methods you used. Therefore, in the Methods chapter, you should
 - Clearly describe the methods. You should provide enough details so that anyone reading the dissertation would know how to replicate your research.
 - Justify why the statistical methods are appropriate.
 - Include simple examples that demonstrate an application of methods.

7.2.4 Writing the Results.

- In the Results chapter you describe and summarize the results you found.
- Remember to include comments that remind your committee that your results satisfy your research objectives and research scope.
- Whenever possible, use graphical methods to summarize results. For example, it is difficult to see patterns with tables. Results in tables should be supplemented with graphs, plots, figures, ...
- When presenting specific results, you should mention the research objective that is related to those results.
- By the end of the Results chapters, you want your committee to know that you have satisfied all research objectives and research scope.

7.2.5 Writing the Discussion.

- The Discussion is much more than a restatement of the research results. Your goal is to help your committee understand what all of the results mean (“see the big picture”). You have the opportunity to review your work as a whole.
- You need to relate the contents of the Results chapter to the Literature Review chapter. That is, you want to compare and contrast your results to the contents of the important references that you cited.
- You want to show to your committee that you can *interpret* your research results (and not just summarize the results). That is, you want to convince your committee that your results are truly analyzed (and not only described).
- You should also indicate any limitations of your research (such as reviewing the research scope) that researchers need to address in the future.
- In the Discussion, it must be clear if a statement you make is a direct result of your research results (within the scope) **or** if you are generalizing beyond the research scope.

- For example, if you found a pattern within the scope of design sizes in the dissertation (e.g., $N \leq 40$), then you cannot conclude that the pattern will also be true for design sizes outside the scope (e.g., $N > 40$). You can, however, make a recommendation for larger design sizes based on the research results. You just have state you are making a recommendation, and that future research is needed to validate any generalizations beyond the scope of the dissertation.
- Thus, in the Discussion chapter, you want to be clear regarding how you present conclusions and implications of your results.

7.2.6 Acknowledgements.

There will be a page in the dissertation for ‘acknowledging’ (giving credit) to others who ‘helped’ during the dissertation process. These ‘helpers’ include

- Any collaborators on this work, such as a person who provided technical or administrative assistance (such as people who helped with computing problems or helped you to collect data).
- Any organization (such as a government agency or university) that provided financial support for you to do the research.
- Other people who have helped you personally or professionally over your graduate student career. For example, some students thank their advisor or family members for their support.

7.2.7 Archiving Supplementary Materials.

- If your dissertation involves a lot of computation, then you will generate a lot of computer output. However, you only ‘extract’ the important results from the output.
- All of the computer code and output that does not appear in the dissertation still needs to be available in the future.
- Therefore, all computer code and output needs to be ‘archived’ (saved for the future) on a CD or flash drive. Several copies should be made with one copy given to your advisor and a copy given to any committee member who wants one.

7.3 Making and submitting revisions.

- After you have begun to generate research results, you should consider writing the first draft of a chapter of the dissertation. This could be either the Introduction, the Literature Review, or the Methods chapter. For example, there may be additional comments you want to include in the Methods chapter that you did not mention in the dissertation proposal.
- Thoroughly read each chapter at least twice. Check that the writing is clear and flows logically.

- When you have completed a first draft of a chapter, do not immediately submit it to your advisor to read. Wait a few days and read the chapter again. Always keep track of the consistency of the notation and if all terms have been defined before using them. After you make changes, then give it to your advisor for review.
- If possible, you should schedule a meeting to receive your advisor’s comments and suggestions to be included in the next revision.
- For an international advisor than you cannot meet, you need to make arrangements for the best way to handle revisions.
- For example, your advisor may want you to email the LaTeX code with a pdf copy of the chapter. Then your advisor:
 - May want to type comments directly in the LaTeX code file, **OR**
 - May add comments directly onto the pdf file, **OR**
 - May send you a scanned copy with hand-written comments

for you to use for a revision.

- After you have received comments from your advisor on a draft of a chapter, you should *revise that chapter immediately*. If you make the revisions while the comments of your advisor are ‘fresh in your mind’, it will be easier and it will save time.
- An immediate revision will also prevent the accumulation (or ‘backlog’) of work that you will need to do later. It becomes more and more discouraging to see the list of items remaining (the ‘to-do list’) grow.
- In addition to checking if the content of the chapter is directly related to your research, your advisor also needs time to correct grammar and writing style in your dissertation. Therefore, expect for your advisor to need at least a week (and possibly longer for a long chapter) to read and return the chapter with comments and suggestions.
 - For example, it usually takes me around one hour per page with the first draft of a chapter. So if the chapter has 20 pages, the student should expect me to spend at least 20 hours reviewing it and making comments. These 20+ hours have to be found around my work schedule at Montana State University (which includes teaching, grading, office hours, and administrative responsibilities).
 - Your advisor must have time to read, to think, and then provide you with useful comments and suggestions (‘feedback’).
 - Therefore, you need to be patient with an advisor. Sometimes, your advisor may not be able to complete a review of a chapter for at least two weeks. This is more likely, however, to occur for the first draft of any chapter. Reviewing a revision should take less time.
 - This is also true when your committee reads the chapters of your dissertation. They also need to find enough time around their work schedules to carefully read the dissertation. As a minimum, your committee should have at least two weeks to read the final draft of the dissertation.

- As a Ph.D. advisor at MSU, I do not get any reduction in my work load and do not receive any additional money. I advise Ph.D. students because I find it personally rewarding (and enjoy it most of the time).
- If you *have carefully revised* the first and second drafts of a chapter, then you should be able to make one more revision that is a ‘nearly final’ form. That is, very little will need to be done later with that chapter.
- However, if you *have not carefully revised* the first and second drafts of a chapter, then expect to have to make more and more revisions. There is no limit on the number of revisions your advisor may require.
- Remember again, *it is your dissertation*. Your advisor is not a co-writer or collaborator. If your advisor thinks you are relying too much on him or her to fix problems or are not making sufficient progress, then your advisor can decide to terminate being your advisor. There is no promise that your advisor will indefinitely tolerate lack of independent work.
- So you should revise immediately after receiving comments, but do not be in a hurry to resubmit a chapter. Be careful with each revision before submitting it.
- Even after each chapter has been revised and there are no problems with the grammar or writing style, you can *expect at least one more revision before all of the chapters are submitted as the final draft* of the dissertation.
- The final draft is the dissertation as a ‘whole document’ and not just a collection of chapters. Your advisor (and possibly your committee) may want some additional changes that will connect the chapters to each other. That is, you want the final draft to ensure that:
 - Your research objectives and research scope were met. This may require additional comments in the Results chapter and Discussion chapter.
 - You did not miss discussing important references cited in the Literature Review chapter and how they relate to your results.
 - Your spelling and punctuation is correct. You cannot rely on a spellchecker to find all mistakes.
 - The writing style is of a professional quality.
 - You have not have defined the same terminology in more than one chapter. For example, if you define ‘bootstrapping’ in both the Methods chapter and the Results chapter, you can remove the definition in the Results chapter.
 - Your Bibliography is correct. That is, **verify** that
 - * All cited references appear in the Bibliography.
 - * All references in the Bibliography were cited in the dissertation.
 - * Each citation in the dissertation matches the correct references in the Bibliography. For example, you may have references to ‘Box, G.E.P.’ for years 1960, 1962, 1970, 1972a, 1972b, ... Be sure you have the correct year for a cited reference.

- The title you originally selected still reflects the focus of the dissertation. For example, if any of the objectives in the proposal were changed, you may need to change the title of the dissertation.

My personal experience

- As an advisor, I have to schedule my time for reviewing materials submitted to me by a Ph.D. student. I am not happy when a student promises to submit something on a certain date and then fails to do so without informing me.
- Occasionally, deadlines cannot be met. Advisors understand that. But, you should inform your advisor if you cannot meet a deadline and then state when you will be able to submit it.
- What is unacceptable is when a student repeatedly misses deadlines and especially when there is no reasonable excuse for missing a deadline. As an advisor, I assume that if a student repeatedly misses deadlines then the student is not working as much as he or she should be working.
- Repeated ('chronic') missing of deadlines is an indicator that a student is not making acceptable progress toward completion of the dissertation. This can potentially lead to a Ph.D. advisor to terminate being the advisor.
- To avoid any serious problems I recommend that:
 - You take the initiative in approaching your advisor if you need additional time to complete an assigned task.
 - You keep copies of all your work (including copies of any electronic files).
 - Propose deadlines for drafting sections and subsections in a chapter to help motivate you (and check with your advisor if these deadlines are reasonable). When needed, update these deadlines.

7.4 Plagiarism

- Many students do not understand what plagiarism is. Plagiarism is a topic that all students (especially students writing a Ph.D. dissertation or a MS degree thesis) should understand.
- Plagiarism is a form of academic misconduct that can have very serious penalties for you and your degree. In fact, plagiarism can lead to being expelled from a Ph.D. program.
- Plagiarism covers a range of mistakes that researchers make by failing to properly cite references or collaborators in a professional document (such as a dissertation, journal article, textbook, ...).
- Sometimes plagiarism is unintentional (accidental) and sometimes plagiarism is intentional (deliberate).

- If it is intentional, the plagiarism is considered a very serious form of professional misconduct.
- Sometimes a lack of understanding of what plagiarism is can lead to unintentional or accidental plagiarism.
- So what is plagiarism? In this Chapter, plagiarism will be defined and examples will be given to help you to avoid plagiarism.
- **Plagiarism** is a failure to properly acknowledge the use of the work of another person or group of people in a professional presentation.
- The most common form of plagiarism in a Ph.D. proposal or dissertation in statistics is using the words and ideas of other statisticians *without identifying* that the words and ideas are not yours. In essence, **you are stealing the hard work of others and taking credit for it.**
- In a dissertation, plagiarism can usually be classified into one of three types:
 1. **Verbatim** plagiarism occurs when the exact words of another author are copied and placed into the dissertation without citing the source of the words.
 2. **Near-verbatim** plagiarism occurs when the student only makes very minor changes to another author's words without citing the author. This typically occurs when the student reorders the words in a sentence or changes only a few words.
 3. **Unauthorized collaboration** is a type of plagiarism that occurs when some of the work of in a dissertation was part of a collaborative effort and the collaborator is not acknowledged.
- All three of these types are serious and unprofessional because
 - Each type is an intentional act of the student. That is, the Ph.D. student knows that these are the exact or nearly the exact words of another author or they are the result of collaborative work with another and *still uses them in the hope that the advisor and the committee will not discover the 'theft'.*
 - All three types of plagiarism are unethical. That is, the student is trying to cheat and take credit for something he or she did not do.
- Do not expect an advisor or committee member to be forgiving if you are caught intentionally plagiarizing. If you are lucky, they may only warn you the first time you are caught. But, if you get caught plagiarizing a second time, do not be surprised if you no longer have an Ph.D. advisor.
- Your advisor (and your committee) are very good at identifying plagiarism. Remember that you probably had the advisor and one or more of the committee members as a teacher for one or more courses. They have a good idea of how you write and an overall assessment of your writing ability. Also, your advisor has read most (if not all) of your cited references as well as other publications that you did not cite. They can determine (with a high probability) what has been plagiarized.

- I expect there will be some errors with English when an international Ph.D. student submits a draft for me to read. However, if I am reading and find no problems with the English for several pages (or sometimes even one page), then I suspect the student has plagiarized.
 - For example, suppose a Thai student submits a Literature Review and in it I see:

If the blending of the components is strictly additive, then the first-order model is appropriate. When there is a curvature in the mixture response surface attributable to nonlinear (often referred to as synergistic or antagonistic) blending between pairs of components, the β_{ij} parameters represent deviations of the surface from linear blending. In this case, the quadratic model may be appropriate.

I am immediately suspicious because of the use of the words ‘synergistic’ and ‘antagonistic’. Very few Americans even know what these words mean.
 - Suppose that in another place in the Literature Review, I see

When $q=3$, placing lower bounds on x_1, x_2 and x_3 yields $L_1 \leq x_1, L_2 \leq x_2, L_3 \leq x_3$ and forces the restricted design region to be the shaded triangle in the figure below. To determine the setting of design variables when the design region is a restricted simplex, we set up an alternative coordinate system of pseudo-components.

In this passage, the Ph.D. student carelessly refers to a figure that never appears in the chapter. This again would suggest that the student plagiarized.
- Note that the reasons for the proper citation of references are more than just avoiding plagiarism. Proper citation of references also:
 - Supports the relevance and originality of your research by referring to other experts in statistics.
 - Demonstrates you have thoroughly explored prior research which convinces your committee that you know how independent research is performed.

7.4.1 Examples of verbatim plagiarism

- Suppose a student writes in his dissertation:

Intuitively, systematic sampling seems likely to be more precise than simple random sampling. In effect, it stratifies the population into n strata, which consist of the first k units, the second k units, and so on. We might therefore expect the systematic sample to be about as precise as the corresponding stratified random sample with one unit per stratum. The difference is that with the systematic sample the units all occur at the same relative position in the stratum, whereas with the stratified random sample the position in the stratum is determined separately by randomization within each stratum. The systematic sample is spread more evenly over the population, and this fact has sometimes made systematic sampling considerably more precise than stratified random sampling.

This is an exact quote from *Sampling Techniques* by Cochran (1953), and the student does not cite this textbook. The student is using Cochran's words and is giving the impression that these are his own words. This student has stolen Cochran's words.

- Suppose another student writes in her dissertation:

Consider a Kriging model

$$y(\mathbf{x}) = \sum_{j=1}^m \beta_j h_j(\mathbf{x}) + z(\mathbf{x})$$

where $h_j(\mathbf{x})$ s are known functions, the β_i s are unknown coefficients to be estimated, and $z(\mathbf{x})$ is a stationary Gaussian random process with mean $E(z(\mathbf{x})) = \mathbf{0}$ and covariance $Cov(z(\mathbf{x}_i), z(\mathbf{x}_j)) = \sigma^2 R(\mathbf{x}_i - \mathbf{x}_j)$ where σ^2 is the unknown variance of the random error and the correlation function R is given. The function R may have many choices, for example, a Gaussian correlation function is given by $R(d) = \exp(-\theta d^2)$, when θ is unknown. In this case, we have

$$Cov(z(\mathbf{x}_i), z(\mathbf{x}_j)) = \sigma^2 \exp[-\theta(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)] \quad (3)$$

Consider the linear predictor $\hat{y} = \mathbf{c}'(\mathbf{x})\mathbf{y}_D$, at an untried \mathbf{x} , where $\mathbf{y}_D = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))'$ is the response column vector collected according to the design $D_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. The column vector \mathbf{y}_D can be regarded as an observation taken from $\mathbf{y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))'$. The best linear unbiased predictor (BLUP) is obtained by choosing the vector $\mathbf{c}(\mathbf{x})$ to minimize the mean square error

$$MSE(\hat{y}(\mathbf{x})) = E[\mathbf{c}'(\mathbf{x})\mathbf{y}_D - Y(\mathbf{x})]^2 \quad (4)$$

subject to the unbiased constraint

$$E(\mathbf{c}'(\mathbf{x})\mathbf{y}_D) = E(Y(\mathbf{x})). \quad (5)$$

This passage was taken directly from *Design and Modeling for Computer Experiments* (2006) by Fang, Li, and Sudjianto. It was very easy for me to detect the plagiarism because I gave the textbook to the student.

7.4.2 Examples of near-verbatim plagiarism

- Reconsider the first several sentences of the quote from Cochran's *Sampling Techniques* (1953). Suppose the student writes in the dissertation:

Intuitively, it seems likely that systematic sampling will be more precise than simple random sampling. Effectively, systematic sampling stratifies the population into N strata consisting of the first L units, the second L units, and so on. Therefore, the systematic sample can be expected to be about as precise as the corresponding stratified random sample having one unit in each stratum.

Note that only a few words are changed and n and k are changed to N and L . This is near-verbatim plagiarism. It is better and acceptable to cite the exact quote from Cochran than to just change a few words and not provide a citation (which is unacceptable).

- Here is a short passage from a draft of a dissertation chapter.

In 1978, a Chinese industrial agency proposed a problem of experimental designs, in which six factors with at least 12 levels for each factor should be considered. But the experimental runs could not exceed 50 because the experiments are quite expensive. When uniform design was applied to the problem, only 31 experiments were arranged, with each factor having 31 levels, and a satisfactory result was achieved. From then, uniform design has been gradually popularized in China, particularly in agriculture, textile industry, watch industry, science researches, military sciences, chemistry and chemical engineering (Liang, Fang and Xu ((2001))).

Here is what appeared in *Number-theoretic Methods in Statistics* (1993) by Fang and Wang:

In 1978, a Chinese industrial agency proposed a problem of experimental designs to the authors, where six factors with at least 12 levels for each factors should be considered. Since the experiment was quite expensive, we were asked to provide an efficient experimental design with 50 experiments... Applying the uniform design to our problem, 31 experiments were arranged only, each factor has 31 levels, and a satisfactory result was achieved. From then on, uniform design was gradually popularized inside China in agriculture, the textile industry, the watch industry, science researches, military sciences, and so on.

Again this is near-verbatim plagiarism. Most, but not all, of the Fang and Wang's words were used with only minor changes in many places.

7.4.3 Examples of unauthorized collaboration

- A Ph.D. student pays another person to write computer code to perform a computer simulation, and uses the computer output in the Results chapter of the dissertation. If the Ph.D. student does not give credit to the person who wrote the computer code, it is a form of plagiarism.
- A Ph.D. student uses the work or outline from another statistician to prove a theorem in the dissertation. If the statistician is not acknowledged, then this is plagiarism.
- A Ph.D. student gets another student to write a portion of the Results chapter.

7.4.4 How to avoid plagiarism.

- Most Ph.D. students do not know when or how to make references to other people's work, and often make some plagiarism mistakes due to lack of preparation and experience with professional writing.
- No one can prevent a student from making intentional plagiarism mistakes. The student is knowingly taking a chance of being caught but hopes that he or she does not get caught. That is a matter of conscience and professional ethics.
- However, I hope these notes will prevent you from unintentional plagiarism by making you aware of what is considered unacceptable (such as near-verbatim plagiarism).
- To avoid plagiarism, you must always cite a reference source if you use any of the ideas, concepts, results, examples, or any content in your dissertation. Note: this *is more than just using another person's words*.
- You are allowed to occasionally include an exact or 'verbatim' quote from a reference. Whenever you do include an exact quote, you must cite the reference.
- If the exact quote is short enough to be part of a paragraph (e.g. one or two sentences), then the quote should begin and end with double quotes:
“ (insert quote) ”
- For longer quotes (such as an entire paragraph), the entire quote needs to be indented which separates it from what precedes and follows it. Indented quotes should also be single-spaced. You must, of course, also acknowledge the source of the reference.
- Warning: **The number and length of verbatim quotes should be kept to a minimum.** Remember that your committee wants proof that you understand what you are writing and not just using another statistician's words (even when properly cited).
- Warning: Do not cite one reference again and again (especially if it is a textbook). Repeated references to one source indicates that you are relying too much on one author's work. You want to convince your committee that you have read a wide assortment of original sources from journals and not just the summary presented in one textbook.
- Whenever you are uncertain whether or not your writing can be considered plagiarism, ask for your advisor's opinion. An advisor appreciates when a student is concerned about whether or not the writing is properly referenced. Do not be afraid to ask for advice concerning plagiarism.

7.5 Final Comments on Writing.

- All terminology and notation must be defined. Three common strategies for defining terminology and notation are to:
 1. *Define terms and notation as needed.* That is, present definitions just before the reader needs them for the first time. The reader will now be prepared for the material that follows. However, if you use the terms and notation later in the dissertation, then you will need to refer the reader to the location in the dissertation where the definitions are given.
 2. *Define all terms and notation in one section at the beginning of the dissertation.* For this approach, the reader knows exactly where to find a definition. This, however, is not very common (based on my experience).
 3. *Mix strategies 1. and 2.* Very often, the notation is defined in a section at the beginning of the dissertation, but the terminology is defined as it is needed.
- Whenever possible, *use common notation and terminology.* For example use n or N to represent a sample or design size, and not a letter such as g or κ .
- Verify that you use the same notation throughout the dissertation.
- Consider whether or not portions of a chapter could be moved to an appendix. For example, you may consider including a flowchart or pseudo-code of a computer algorithm in the Methodology chapter, but the actual computer code should be in an appendix at the end of the dissertation.
- Be sure that every table and figure is referenced in the chapter and described in the chapter. Also, make the captions meaningful to read. For example, briefly describe the contents of the figure or table.
 - A caption like “*D-efficiencies*” is poor.
 - You would want a caption like “*D-efficiencies for Central Composite Designs having $3 \leq k \leq 8$ factors and $0 \leq n_0 \leq 5$ Center Points*”.

The goal is to provide guidance on how to understand and interpret the contents of a figure or table.

- Label important equations with numbers making it easy to reference later in the dissertation. This is very helpful to a reader who might lose track of the origin of certain claims.
- You should insert brief comments when presenting a set of equations. These comments often contain the justification of how to go from one step to another. For example, the following was in my dissertation. Note the brief comments between steps.

The Components of \mathbf{A}^{11}

(i) The matrix \mathbf{A}_{22}^{-1} is of the form

$$\begin{aligned}\mathbf{A}_{22}^{-1} &= (a\mathbf{I}_C + b\mathbf{J}_C\mathbf{J}'_C)^{-1} \\ &= \frac{1}{a}(\mathbf{I}_C - \frac{b}{a + Cb}\mathbf{J}_C\mathbf{J}'_C)\end{aligned}$$

where $a = 2\alpha^4$ and $b = F$. Hence,

$$\mathbf{A}_{22}^{-1} = \frac{1}{2\alpha^4}(\mathbf{I}_C - \frac{F}{2\alpha^4 + CF}\mathbf{J}_C\mathbf{J}'_C)$$

(ii) Then pre-multiplying \mathbf{A}_{22}^{-1} by \mathbf{A}_{12} yields

$$\begin{aligned}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} &= \frac{F + 2\alpha^2}{2\alpha^4} \cdot \begin{bmatrix} \mathbf{J}'_C \\ \phi'_2 \end{bmatrix} \cdot \left(\mathbf{I}_C - \frac{F}{2\alpha^4 + CF}\mathbf{J}_C\mathbf{J}'_C \right) \\ &= \frac{F + 2\alpha^2}{2\alpha^4} \cdot \left\{ \begin{bmatrix} \mathbf{J}'_C \\ \phi'_2 \end{bmatrix} - \frac{CF}{2\alpha^4 + CF} \begin{bmatrix} \mathbf{J}_C \\ \phi'_2 \end{bmatrix} \right\} \\ &= \frac{F + 2\alpha^2}{2\alpha^4} \cdot \frac{2\alpha^4}{2\alpha^4 + CF} \begin{bmatrix} \mathbf{J}'_C \\ \phi'_2 \end{bmatrix} \\ &= \frac{F + 2\alpha^2}{2\alpha^4 + CF} \begin{bmatrix} \mathbf{J}'_C \\ \phi'_2 \end{bmatrix}\end{aligned}$$

(iii) Then post-multiplying by \mathbf{A}_{21} yields:

$$\begin{aligned}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} &= \frac{(F + 2\alpha^2)^2}{2\alpha^4 + CF} \begin{bmatrix} \mathbf{J}'_C \\ \phi'_2 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{J}_C & \phi_2 \end{bmatrix} \\ &= \frac{(F + 2\alpha^2)^2}{2\alpha^4 + CF} \cdot \begin{bmatrix} C & \phi'_1 \\ \phi_1 & \phi^* \end{bmatrix}\end{aligned}$$

where ϕ^* is a $K^* \times K^*$ zero matrix. Thus,

$$\begin{aligned}\mathbf{A}^{11} &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \\ &= \begin{bmatrix} N - \frac{C(F+2\alpha^2)^2}{2\alpha^4+CF} & \phi'_1 \\ \phi_1 & \mathbf{Diag}(d_i) \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \frac{CF+2\alpha^4}{2N\alpha^4+CNF-C(F+2\alpha^2)^2} & \phi'_1 \\ \phi_1 & \mathbf{Diag}(\frac{1}{d_i}) \end{bmatrix}\end{aligned}$$

- Avoid long paragraphs. It is easier for the reader to know determine your main points by using multiple paragraphs than to put all of the main points in one long paragraph.
- The same principle also holds for chapters and sections.
 - Divide each chapter into sections that focus on a particular topic.
 - Then divide each section into subsections that focus on particular subtopics.
 - Make the titles of each section and subsection meaningful and helpful to the reader.
 - As you start a chapter or section, briefly state how it will be organized. That is, state what the chapter or section is about and why the committee should read it. Providing a general outline will help you when writing the chapter or section.
 - As you end a section or subsection, remind your committee what was most important.
- For complex concepts, include an example whenever possible. If possible, use a graphical method to convey complex concepts or results. These are particularly useful in the Literature Review and Results chapters.
- Examples make it easier for the committee to understand. Examples keep your committee from getting lost in too much ‘abstraction’.
- Pseudocode, flowcharts, and itemized lists are good ways to describe an algorithm or how to perform a complicated statistical procedure.