Power Analysis and Sample Size Determination

Basic Ideas, Tools, and Examples

Dan Hall, Director of the SCC

Department of Statistics
Franklin College of Arts and Sciences
Statistical Consulting Center
UNIVERSITY OF GEORGIA

Table of Contents

Related Resources

- Two companion videos for this talk can be found here kaltura.uga.edu/media/t/1_scl60mqx and here kaltura.uga.edu/media/t/1_r5574d7h.
- An accompanying R script, powerExamps.R, is available as an attachment to each of the videos linked above. Follow the link to either video and click on attachments.
- The videos shows how to calculate power and sample size in G*Power, Russ Lenth's power applets, and R for the examples featured in this talk. Video 1 covers examples 1 and 2 (the MOOC and SHS examples), and Video 2 covers examples 3 and 4 (the Koi and COVID-19 Assay examples).

Related Resources

- Two companion videos for this talk can be found here kaltura.uga.edu/media/t/1_scl60mqx and here kaltura.uga.edu/media/t/1_r5574d7h.
- An accompanying R script, powerExamps.R, is available as an attachment to each of the videos linked above. Follow the link to either video and click on attachments.
- The videos shows how to calculate power and sample size in G*Power, Russ Lenth's power applets, and R for the examples featured in this talk. Video 1 covers examples 1 and 2 (the MOOC and SHS examples), and Video 2 covers examples 3 and 4 (the Koi and COVID-19 Assay examples).

Related Resources

- Two companion videos for this talk can be found here kaltura.uga.edu/media/t/1_scl60mqx and here kaltura.uga.edu/media/t/1_r5574d7h.
- An accompanying R script, powerExamps.R, is available as an attachment to each of the videos linked above. Follow the link to either video and click on attachments.
- The videos shows how to calculate power and sample size in G*Power, Russ Lenth's power applets, and R for the examples featured in this talk. Video 1 covers examples 1 and 2 (the MOOC and SHS examples), and Video 2 covers examples 3 and 4 (the Koi and COVID-19 Assay examples).

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- When planning a study, a fundamental design question is, *How many subjects?*
  - Too small and it will not generate enough information to learn anything definitive (low power).
  - Too large and we waste resources (money, time, opportunity).
  - Typically, we calculate the sample size necessary to have enough power to find the effect we are looking for.
- Sometimes the sample size is fixed. In that case, we may wish to calculate the power that our fixed sample size will achieve.
  - May tell us whether our study is worth doing at all.
  - May tell us that we need to modify our study design or analysis plan so that we can achieve more power.

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- There are many good reasons to do a power/sample size analysis:
  - Required by a funding agency or sponsor of the research.
  - Helps avoid wasting resources.
  - Forces you to think carefully about the design and the analysis.
  - Forces you to operationalize and prioritize your research questions.
- Power analysis can be **hard** and it can be **unpleasant**.
  - It requires assumptions about what we are researching, but we do the research precisely because we don't fully understand that subject.
  - Textbook examples are the easy cases. In practice, most studies require more assumptions and more difficult calculations.
  - Usually, consulting a statistician is necessary. But he/she can't do the analysis without help (you still have hard work to do).

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

- Power/sample size determination is inexact.
  - Often must make very rough estimates of inputs.
  - Often based on calculations for a simplified design and/or analysis than the one to be done in reality.
  - Given its imprecision and dependence on unknown inputs, often best to
    - get results over a range of inputs;
    - compute a conservative estimate (e.g., a lower bound on sample size needed to achieve 80% power);
    - explain the rationale for the analysis (reviewer wants due diligence, not a perfect result);
    - **collect preliminary data (conduct a pilot study)!!**

Introduction

The statistician's job:

- to help operationalize and prioritize the research questions,
- to help with other design aspects that sample size/power will depend on,
- to figure out how to approach the power analysis and implement it,
- to elicit the input to the power analysis in ways that make this as easy as possible for the investigator,
- to write up the sample size/power analysis and (often) the broader statisticial analysis plan.

Introduction

The statistician's job:

- to help operationalize and prioritize the research questions,
- to help with other design aspects that sample size/power will depend on,
- to figure out how to approach the power analysis and implement it,
- to elicit the input to the power analysis in ways that make this as easy as possible for the investigator,
- to write up the sample size/power analysis and (often) the broader statisticial analysis plan.

Introduction

The statistician's job:

- to help operationalize and prioritize the research questions,
- to help with other design aspects that sample size/power will depend on,
- to figure out how to approach the power analysis and implement it,
- to elicit the input to the power analysis in ways that make this as easy as possible for the investigator,
- to write up the sample size/power analysis and (often) the broader statisticial analysis plan.

Introduction

The statistician's job:

- to help operationalize and prioritize the research questions,
- to help with other design aspects that sample size/power will depend on,
- to figure out how to approach the power analysis and implement it,
- to elicit the input to the power analysis in ways that make this as easy as possible for the investigator,
- to write up the sample size/power analysis and (often) the broader statisticial analysis plan.

Introduction

The statistician's job:

- to help operationalize and prioritize the research questions,
- to help with other design aspects that sample size/power will depend on,
- to figure out how to approach the power analysis and implement it,
- to elicit the input to the power analysis in ways that make this as easy as possible for the investigator,
- to write up the sample size/power analysis and (often) the broader statisticial analysis plan.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Introduction

The investigator's job:

- sharpen focus:
  - prioritize research questions and hypotheses;
  - distinguish exploratory goals from hypothesis driven ones;
  - select variables (one or very few) with which to address hypotheses;
  - select populations (i.e., domains) within which to assess hypotheses.
- obtain data/results from which assumptions can be made:
  - conduct a pilot study;
  - find papers that measured same responses on similar populations under comparable conditions;
  - not sufficient to send the statistician several papers that you hope have what is needed;
  - think carefully about the minimum magnitude of effect that (a) is plausible in your study, *and* (b) is clinically significant;
- conduct power analysis early and be prepared to change your research questions and/or design.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - R is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - R is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - It is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - It is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - It is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - It is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - R is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Tools for Power Analysis

- G*Power is an excellent, free power and sample size program for Mac and Windows available here.
- Dr. Russell Lenth has some fantastic Java Applets for power analysis here: http://homepage.divms.uiowa.edu/~rlenth/Power/.
  - These used to run in a browser running Java, but now its best to download the single file (piface.jar) that runs these applets and run it after installing Java Runtime Environment on your computer.
  - His webpage also has some excellent advice about what to do and not do when approaching the power/sample size issue.
- Most general purpose commercial packages do power/sample size calculations.
  - E.g., SAS has PROC POWER for simple problems and PROC GLMPOWER for more complex analyses involving linear models. It also has point and click interfaces in both SAS for Windows and SAS Studio.
- R has some power functions in the `stats` package (e.g., `power.t.test()` as well as many specialized packages (`pwr` being the most useful).
  - R is not as friendly and easy-to-use for simple problems, but is the best tool for doing power analysis via simulation, which is the best option for complex/advanced problems.

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

"Power" has a specific meaning tied to statistical hypothesis testing. So, let's review.

- Hypothesis testing:
  - Assume a null hypothesis H_0 is true,
  - gather evidence (data),
  - summarize evidence against H_0 (test statistic),
  - quantify strength of the evidence (p-value), and
  - make a decision (reject, fail to reject).
- Possible outcomes:

	The Truth	
Decision	H_0 is true	H_0 is false
Don't reject	Correct	Type II Error
Reject	Type I Error	Correct

$$\alpha = \Pr(\text{Type I Error}) \quad \beta = \Pr(\text{Type II Error})$$

$$\text{Power} = 1 - \beta = \Pr(\text{reject } H_0 \text{ given that it is false})$$

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t-test.
    - A measure of (experimental) error variability. σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t-test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t-test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparison adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\mathrm{Trt}} - \mu_{\mathrm{Ctrl}}|$ for a two-sample t-test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\mathrm{Trt}} - \mu_{\mathrm{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Basic Ideas

Power depends on some things we can control, some we can't:

- The sample size n, and how it is distributed over the study design.
  - Power increases with n.
- The Type I error rate (α) and # of tails in H_A.
  - Power increases with α, higher for one-tailed alternative.
- How false the null hypothesis is.
  - Power increases with falseness of H_0.
  - This requires assumptions. At a minimum, we need
    - An absolute effect size. E.g., $|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|$ for a two-sample t test.
    - A measure of (experimental) error variability: σ^2, say.
  - These quantities can be combined into a *standardized* effect size (e.g., $\frac{|\mu_{\text{Trt}} - \mu_{\text{Ctrl}}|}{\sigma}$), but best to consider them separately.
  - Other quantities may affect power. For example:
    - the within-subject correlation and degree of non-sphericity in a repeated measures design;
    - the variance components for random effects in the model;
    - multiple comparisons adjustments;
    - for regression: variability of covariate of interest and its (multiple) correlation with other covariates.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Power/Sample Size for a Two-sample t Test

Typical power analysis for a two-sample t test:

1. Fix $\alpha = 0.05$ (or 0.01, or ...) and select 1- or 2-tailed alternative.
2. Assume a value of $|\mu_1 - \mu_2|$ based on clinical/practical significance and plausible magnitude of effect.
   - In this approach, μ_1 can be fixed at an arbitrary value.
   - Sometimes easier to assume a value for $\frac{\mu_2}{\mu_1}$ (percentage increase/decrease due to treatment) and a value for μ_1 from which $|\mu_1 - \mu_2|$ can be calculated.
3. Select minimum power you'd like to achieve (e.g., 80%).
   - Sometimes fixed by funding agency, but there may be wiggle room in 70–90% range.
   - 100 − power represents your tolerance of the risk of missing a real effect.
4. Assume a balanced design (which has best power) or, rarely, some degree of imbalance.
5. Compute power for range of n. Pick smallest n that achieves power target.

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_D - \mu_A|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study (Two-sample t Test and ANOVA)

An educational researcher want to improve MOOCs (massive open online courses) for students in developing countries.

- Planning 4 treatments, but for now, assume only A and D:

  $A =$ Standard MOOC in English

  $B =$ English MOOC with native language subtitles

  $C =$ English MOOC with native language dubbing and subtitles

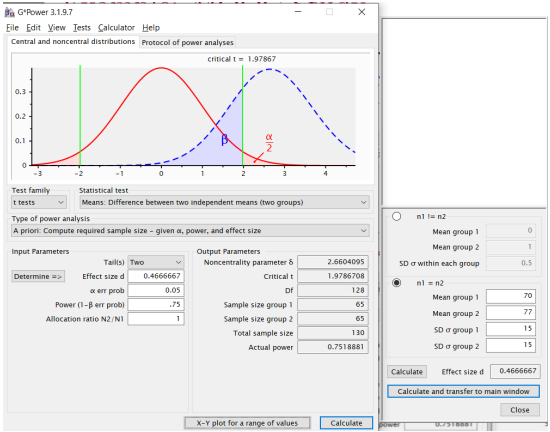
  $D =$ Native language MOOC adapted from original English MOOC

- $Y =$final exam score.
- He decides to use $\alpha = 0.05$, a balanced design, and wants 75% power.
- Basis of assumed effect size:
  - English-speaking past enrollees in A had $\bar{Y} = 78$, $SD(Y) = 10$.
  - Researcher believes non-native English speakers will score worse with more variability, so assumes $\mu_A = 70$, and $\sigma = 15$ in each treatment.
  - He thinks treatment D would be worthwhile if it increases scores by $\geq 10\%$ and believes such an increase is plausible. Thus,

$$\mu_D = 1.1\mu_A = 77.0 \quad \Rightarrow \quad \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{7.0}{15} = 0.47.$$

Example: MOOC Study

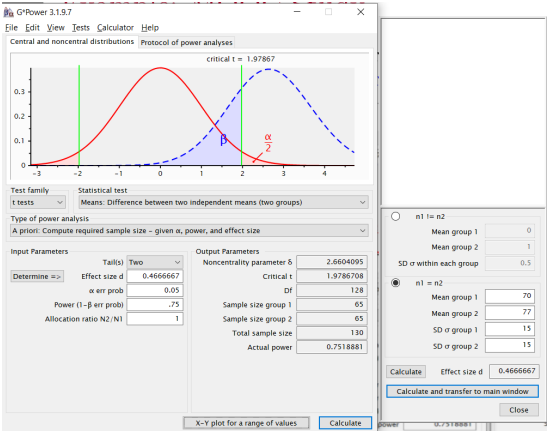
- G*Power dialog and results:



- $n = 130$ subject (65 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

- G*Power dialog and results:



- $n = 130$ subject (65 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: MOOC Study

What if the design includes all 4 treatments A–D?

- Now analysis is an ANOVA F test, not a two-sample t test.
- Effect size is determined by spacing of all four means. In fact, it is a function of sample variance of $\mu_A, \mu_B, \mu_C, \mu_D$.
- Conservative approach: assume a value for

$$\Delta \equiv |\mu_{\text{Max}} - \mu_{\text{Min}}|,$$

  and assume variance of all 4 treatment means is smallest (least favorable) subject to the assumed Δ.
- In MOOC example, we could assume $\Delta = 7$. Least favorable values for other means are half-way between best and worst. So means would be 70, 73.5, 73.5, 77.
- G*Power allows us to specify $\mu_A, \mu_B, \mu_C, \mu_D$ or their (population!) variance in order to determine effect size and the resulting power. Note: pop. variance of 70,73.5,73.5,77 is 6.125.
- Now $n = 364$ subjects (91 per treatment) needed for 75% power (actual power=75.2%).

Example: Secondhand Smoke Study (Repl. Latin Square)

St. Helen et al. (2012) designed an experiment to study secondhand smoke exposure in three outdoor environments:

- A: a **restaurant** patio,
- B: a **bar** beer garden,
- C: and a **control** setting (a park).

Design: Latin square replicated in two directions:

	Subject Number									
Week	1	2	3	4	5	6	...	n-2	n-1	n
1	A	B	C	A	B	C	...	A	B	C
2	B	C	A	C	A	B	...	B	C	A
3	C	A	B	B	C	A	...	C	A	B
4	A	B	C	A	B	C	...	A	B	C
5	C	A	B	B	C	A	...	C	A	B
6	B	C	A	C	A	B	...	B	C	A
	.	.	.	.	.	.	...	.	.	.
m-2	A	B	C	A	B	C	...	A	B	C
m-1	B	C	A	C	A	B	...	B	C	A
m	C	A	B	B	C	A	...	C	A	B

Response variables obtained as post-test minus pretest gain in two biomarkers for tobacco smoke exposure:

- Salivary cotinine
- Log NNAL/creatinine ratio

Example: Secondhand Smoke Study (Repl. Latin Square)

St. Helen et al. (2012) designed an experiment to study secondhand smoke exposure in three outdoor environments:

- A: a **restaurant** patio,
- B: a **bar** beer garden,
- C: and a **control** setting (a park).

Design: Latin square replicated in two directions:

	Subject Number									
Week	1	2	3	4	5	6	...	n-2	n-1	n
1	A	B	C	A	B	C	...	A	B	C
2	B	C	A	C	A	B	...	B	C	A
3	C	A	B	B	C	A	...	C	A	B
4	A	B	C	A	B	C	...	A	B	C
5	C	A	B	B	C	A	...	C	A	B
6	B	C	A	C	A	B	...	B	C	A
	.	.	.	.	.	.	...	.	.	.
m-2	A	B	C	A	B	C	...	A	B	C
m-1	B	C	A	C	A	B	...	B	C	A
m	C	A	B	B	C	A	...	C	A	B

Response variables obtained as post-test minus pretest gain in two biomarkers for tobacco smoke exposure:

- Salivary cotinine
- Log NNAL/creatinine ratio

Example: Secondhand Smoke Study (Repl. Latin Square)

St. Helen et al. (2012) designed an experiment to study secondhand smoke exposure in three outdoor environments:

- A: a **restaurant** patio,
- B: a **bar** beer garden,
- C: and a **control** setting (a park).

Design: Latin square replicated in two directions:

	Subject Number									
Week	1	2	3	4	5	6	...	n-2	n-1	n
1	A	B	C	A	B	C	...	A	B	C
2	B	C	A	C	A	B	...	B	C	A
3	C	A	B	B	C	A	...	C	A	B
4	A	B	C	A	B	C	...	A	B	C
5	C	A	B	B	C	A	...	C	A	B
6	B	C	A	C	A	B	...	B	C	A
	.	.	.	.	.	.	...	.	.	.
m-2	A	B	C	A	B	C	...	A	B	C
m-1	B	C	A	C	A	B	...	B	C	A
m	C	A	B	B	C	A	...	C	A	B

Response variables obtained as post-test minus pretest gain in two biomarkers for tobacco smoke exposure:

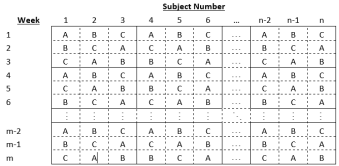
- Salivary cotinine
- Log NNAL/creatinine ratio

Example: Secondhand Smoke Study (Repl. Latin Square)

St. Helen et al. (2012) designed an experiment to study secondhand smoke exposure in three outdoor environments:

- A: a **restaurant** patio,
- B: a **bar** beer garden,
- C: and a **control** setting (a park).

Design: Latin square replicated in two directions:



Response variables obtained as post-test minus pretest gain in two biomarkers for tobacco smoke exposure:

- Salivary cotinine
- Log NNAL/creatinine ratio

Example: Secondhand Smoke Study (Repl. Latin Square)

St. Helen et al. (2012) designed an experiment to study secondhand smoke exposure in three outdoor environments:

- A: a **restaurant** patio,
- B: a **bar** beer garden,
- C: and a **control** setting (a park).

Design: Latin square replicated in two directions:

	Subject Number									
Week	1	2	3	4	5	6	...	n-2	n-1	n
1	A	B	C	A	B	C	...	A	B	C
2	B	C	A	C	A	B	...	B	C	A
3	C	A	B	B	C	A	...	C	A	B
4	A	B	C	A	B	C	...	A	B	C
5	C	A	B	B	C	A	...	C	A	B
6	B	C	A	C	A	B	...	B	C	A
	.	.	.	.	.	.	...	.	.	.
m-2	A	B	C	A	B	C	...	A	B	C
m-1	B	C	A	C	A	B	...	B	C	A
m	C	A	B	B	C	A	...	C	A	B

Response variables obtained as post-test minus pretest gain in two biomarkers for tobacco smoke exposure:

- Salivary cotinine
- Log NNAL/creatinine ratio

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_B - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_B - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_B - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_R - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_R - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_R - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

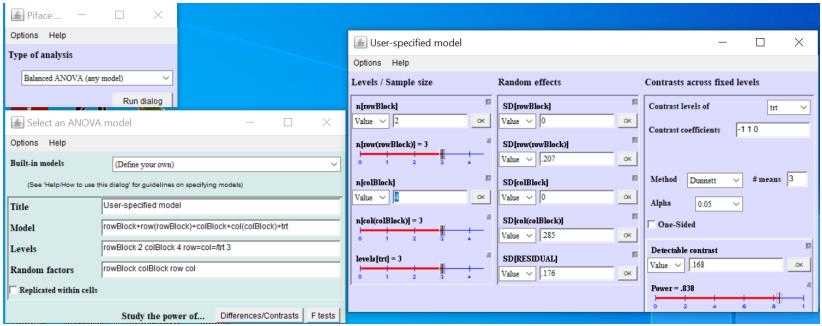
Pilot study:

Parameter	Cotinine	log(NNAL/Creat)
Bar mean	1.432	0.638
Rest mean	0.766	0.134
Ctrl mean	0.141	0.130
SD: Occasions	0.344	0.207
SD: Subjects	0.582	0.285
SD: Error	0.356	0.176

- Analysis based on linear model for Latin square with fixed treatment effects, random subject and measurement occasion effects.
- Larger effects for cotinine, so we powered based on log(NNAL/creat).
- Large diff b/w Bar and Ctrl, so powered for Rest vs Ctrl contrast.
- Tiny difference in pilot data b/w Rest & Ctrl for log(NNAL/creat). But shouldn't always assume effect size that someone else obtained.
  - Pilot study small; much uncertainty in Rest vs Ctrl difference.
  - Should assume effect size that is plausible and clinically significant.
  - Cotinine: Rest vs Ctrl difference about half of Bar vs Ctrl diff. So, for log(NNAL/creat), assume Rest vs Ctrl diff at least one third of Bar vs Ctrl diff: $\mu_R - \mu_C = (0.638 - 0.130)/3 = 0.168$.

Example: Secondhand Smoke Study

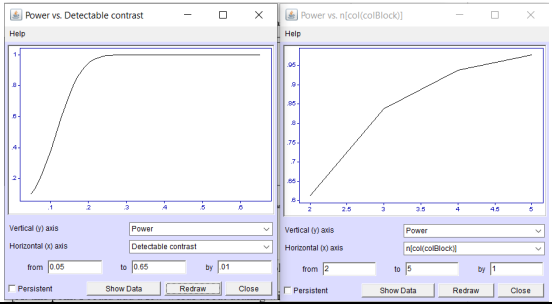
Russ Lenth's power applets are more useful for this example.



- Tried different values of m and n in multiples of 3. Power $\geq .80$ (0.838) for $m = 6$ measurement occasions, $n = 12$ subjects.

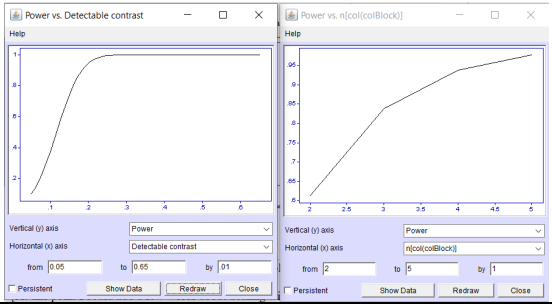
Example: Secondhand Smoke Study

- When an input is especially speculative, its often advisable to compute power/sample size over a range for that input.
- This can also be useful just to visualize how the input affects power.
- E.g., in the SHS example, we assumed Rest vs. Ctrl difference was 0.168, one third the Bar vs. Ctrl difference. What if it were higher or lower?
- Below, we plot power versus the detectable contrast and versus # of replicated Latin squares in the column direction, which determines # subjects needed for the experiment.



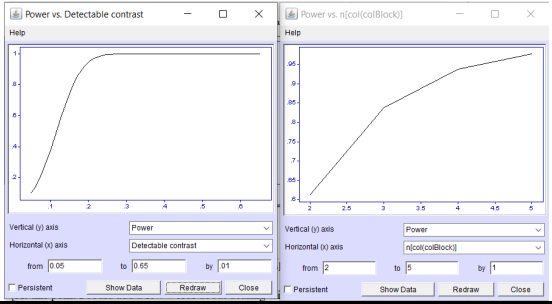
Example: Secondhand Smoke Study

- When an input is especially speculative, its often advisable to compute power/sample size over a range for that input.
- This can also be useful just to visualize how the input affects power.
- E.g., in the SHS example, we assumed Rest vs. Ctrl difference was 0.168, one third the Bar vs. Ctrl difference. What if it were higher or lower?
- Below, we plot power versus the detectable contrast and versus # of replicated Latin squares in the column direction, which determines # subjects needed for the experiment.



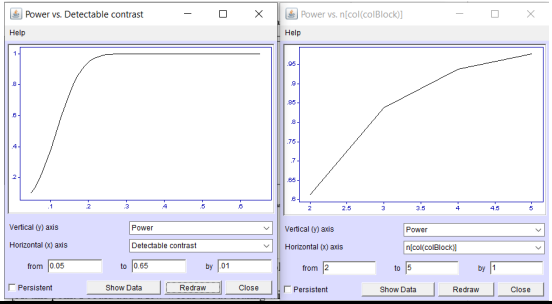
Example: Secondhand Smoke Study

- When an input is especially speculative, its often advisable to compute power/sample size over a range for that input.
- This can also be useful just to visualize how the input affects power.
- E.g., in the SHS example, we assumed Rest vs. Ctrl difference was 0.168, one third the Bar vs. Ctrl difference. What if it were higher or lower?
- Below, we plot power versus the detectable contrast and versus # of replicated Latin squares in the column direction, which determines # subjects needed for the experiment.



Example: Secondhand Smoke Study

- When an input is especially speculative, its often advisable to compute power/sample size over a range for that input.
- This can also be useful just to visualize how the input affects power.
- E.g., in the SHS example, we assumed Rest vs. Ctrl difference was 0.168, one third the Bar vs. Ctrl difference. What if it were higher or lower?
- Below, we plot power versus the detectable contrast and versus # of replicated Latin squares in the column direction, which determines # subjects needed for the experiment.



Power via Simulation

Instead of using power/sample size software, simulation can be used to calculate power.

- E.g., in the SHS problem, we can generate many data sets at random with the assumed treatment means and variance components.
- For each data set, fit the model on which the analysis will be based, and test the hypothesis of interest.
- The proportion of the data sets for which the test is statistically significant is the power.
- See powerExamps.R for the code to estimate power via simulation in the SHS example.
- Simulation result agrees closely (Power=0.839) with that obtained with Lenth's software.

Power via Simulation

Instead of using power/sample size software, simulation can be used to calculate power.

- E.g., in the SHS problem, we can generate many data sets at random with the assumed treatment means and variance components.
- For each data set, fit the model on which the analysis will be based, and test the hypothesis of interest.
- The proportion of the data sets for which the test is statistically significant is the power.
- See powerExamps.R for the code to estimate power via simulation in the SHS example.
- Simulation result agrees closely (Power=0.839) with that obtained with Lenth's software.

Power via Simulation

Instead of using power/sample size software, simulation can be used to calculate power.

- E.g., in the SHS problem, we can generate many data sets at random with the assumed treatment means and variance components.
- For each data set, fit the model on which the analysis will be based, and test the hypothesis of interest.
- The proportion of the data sets for which the test is statistically significant is the power.
- See powerExamps.R for the code to estimate power via simulation in the SHS example.
- Simulation result agrees closely (Power=0.839) with that obtained with Lenth's software.

Power via Simulation

Instead of using power/sample size software, simulation can be used to calculate power.

- E.g., in the SHS problem, we can generate many data sets at random with the assumed treatment means and variance components.
- For each data set, fit the model on which the analysis will be based, and test the hypothesis of interest.
- The proportion of the data sets for which the test is statistically significant is the power.
- See powerExamps.R for the code to estimate power via simulation in the SHS example.
- Simulation result agrees closely (Power=0.839) with that obtained with Lenth's software.

Power via Simulation

Instead of using power/sample size software, simulation can be used to calculate power.

- E.g., in the SHS problem, we can generate many data sets at random with the assumed treatment means and variance components.
- For each data set, fit the model on which the analysis will be based, and test the hypothesis of interest.
- The proportion of the data sets for which the test is statistically significant is the power.
- See powerExamps.R for the code to estimate power via simulation in the SHS example.
- Simulation result agrees closely (Power=0.839) with that obtained with Lenth's software.

Nonstandard Problems: Handle by Simulation and/or Simplification

Many problems are not implemented in power software. In those cases;

- try to simplify the analysis so that the problem is more tractable via simulation or via available software;
- use simulation when software isn't adequate;
- consider how simplifications will affect power and adjust (approximately) the power target or assumptions to compensate.

Nonstandard Problems: Handle by Simulation and/or Simplification

Many problems are not implemented in power software. In those cases;

- try to simplify the analysis so that the problem is more tractable via simulation or via available software;
- use simulation when software isn't adequate;
- consider how simplifications will affect power and adjust (approximately) the power target or assumptions to compensate.

Nonstandard Problems: Handle by Simulation and/or Simplification

Many problems are not implemented in power software. In those cases;

- try to simplify the analysis so that the problem is more tractable via simulation or via available software;
- use simulation when software isn't adequate;
- consider how simplifications will affect power and adjust (approximately) the power target or assumptions to compensate.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study (Paired t Test)

- Partyka et al. interested in comparing echocardiographic measurements taken on koi fish under two conditions:
  - **A=Anaesthesia**, or
  - **B=manual restraint**.
- Planned a crossover experiments where each fish measured twice: once under each treatment in balanced order (AB or BA).
- Investigators...
  - assume anaesthesia will reduce mean response, but not sure by how much.
  - chose heart rate, ejection fraction, fraction shortening as primary responses among several to be obtained.

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(Y_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{ [\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv \rho} \mathrm{SD}(Y_A)\mathrm{SD}(Y_B) \right\}^{1/2}$$

  and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(Y_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv \rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(Y_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv \rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(Y_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{=\rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\text{SE}(Y_B)$ for treatment B. We need $\text{SD}(Y_A - Y_B)$.
  - Get $\text{SD}(Y_B) = \sqrt{n}\text{SE}(Y_B)$.
  - Assume $\text{SD}(Y_A) = \text{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\text{SD}(Y_A - Y_B) = \left\{[\text{SD}(Y_A)]^2 + [\text{SD}(Y_B)]^2 - 2\underbrace{\text{corr}(Y_A, Y_B)}_{=\rho}\text{SD}(Y_A)\text{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\text{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(Y_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{ [\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2 \underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv \rho} \mathrm{SD}(Y_A)\mathrm{SD}(Y_B) \right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(\bar{Y}_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv\rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(\bar{Y}_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

$$\mathrm{SD}(Y_A - Y_B) = \left\{ [\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2 \underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv \rho} \mathrm{SD}(Y_A)\mathrm{SD}(Y_B) \right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(\bar{Y}_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

    $$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv\rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Assumptions, simplifications, and useful results:

- Power on paired t test, a simplification of the full crossover analysis.
  - Unless no order effect, crossover analysis will be more powerful.
- Need difference in means (=mean difference), and SD of difference b/w treatments.
- No literature on this topic. But some results giving summary stats on echo measurements in other non-anaesthetized fish of different species.
- Calculated power for range of % reduction under anaesthesia.
- Literature gives $\bar{Y}_B$, $\mathrm{SE}(\bar{Y}_B)$ for treatment B. We need $\mathrm{SD}(Y_A - Y_B)$.
  - Get $\mathrm{SD}(Y_B) = \sqrt{n}\mathrm{SE}(\bar{Y}_B)$.
  - Assume $\mathrm{SD}(Y_A) = \mathrm{SD}(Y_B) = S$ (equal variance).
  - Use fact that

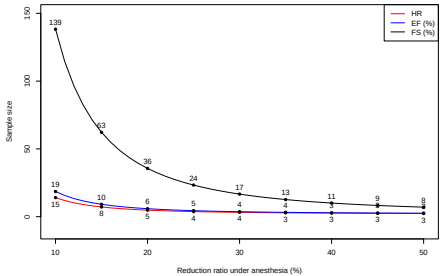
    $$\mathrm{SD}(Y_A - Y_B) = \left\{[\mathrm{SD}(Y_A)]^2 + [\mathrm{SD}(Y_B)]^2 - 2\underbrace{\mathrm{corr}(Y_A, Y_B)}_{\equiv\rho}\mathrm{SD}(Y_A)\mathrm{SD}(Y_B)\right\}^{1/2}$$

    and equal variance assumption to get $\mathrm{SD}(Y_A - Y_B) = S\sqrt{2(1-\rho)}$.
  - Power of test increases with ρ. Assume $\rho = 0.5$ (lowish).

Example: Koi Echocardiography Study

Calculations can be done in G*Power or the Lenth applets. But we used the `power.t.test()` function in R.

```
# means and SD for the 3 responses in trt B:
meanB <- c(70,30,17); SD <- c(10,5,8)
# assume range of effect size (rr=reduction ratio)
rr <- seq(from=0.1,to=0.5,by = 0.01); effSize <- meanB%*%t(rr)
rho <- 0.5 ; sdDiff <- SD*sqrt(2*(1-rho))
# calculate sample sizes with power.t.test() function
sSize <- matrix(NA,ncol=3,nrow= length(rr))
for (j in 1:3){ for(i in 1:length(rr)){
    sSize[i,j]=power.t.test(delta = effSize[j,i],sd=sdDiff[j],
                            power = 0.8,sig.level = 0.05,
                            type="paired",alternative = "one.sided")$n } }
```



Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

- David Blum of UGA Bioexpression and Fermentation Facility contacted us for help planning a study of assays they planned to develop for detection of SARS-CoV-2 antibodies in serum.
- New assay to use yeast-based spike protein, cheaper than existing CDC assay that uses human-derived spike protein.
- Need to buy n_1 serum samples from COVID+ subjects (cases), n_0 samples from COVID- subjects (controls). Each sample to be tested with CDC and UGA assay.
- Samples are expensive. Find n_0, n_1.
- Want to prove that UGA assay not inferior to CDC assay.
  - For α-level test for non-inferiority with respect to a parameter θ, can be done by checking if $100(1-2\alpha)\%$ CI for $\theta_{UGA} - \theta_{CDC}$ has lower limit less than $-\delta$ where δ is a non-inferiority margin.
- Could use AUC for an ROC curve as θ (hard), but decided to use sensitivity and specificity.

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

Complications:

- Non-inferiority trial.
- Paired design.

Inputs:

- Non-inferiority margin (decided on $\delta = 0.05$).
- Significance level ($\alpha = 0.05$).
- Power (80%).
- Correlation between results from the two assays (0.9).
- True sensitivity of each test (assumed 0.96, published value for CDC assay).
- True specificity of each test (assumed 0.993, published value for CDC assay).
- Simulation size (5000).

Example: COVID-19 Assay (Sensitivity and Specificity)

```
library(mvtnorm); library(PropCIs)
# The following function computes a multinomial prob vector for a 2x2 table assuming dichotomized bivariate
# normal random variables generated the table. rho is corr b/w the bivariate normals, and sens is the cut-point
# for dichotomization (i.e, the positivity rate of each assay).
mnorm.pi <- function(rho,sens){
  zsens <- qnorm(sens)
  sigma <- diag(2); sigma[1,2] <- sigma[2,1] <- rho
  p11 <- pmvnorm(lower=c(-Inf,-Inf),upper=c(zsens,zsens),mean=c(0,0),sigma=sigma)
  p01 <- pnorm(zsens)-p11; p10 <- pnorm(zsens)-p11; p00 <- 1-p11-p01-p10
  pi <- c(p00,p01,p10,p11); pi
}
# Now a function to compute the reject rate based on nsim simulated data sets, each of size n,
# where underlying continuous responses are bivariate normal with correlation rho, and where each
# dichotomous response (each test's results) has prob of a positive response (i.e., sensitivity
# when used on cases) equal to sens, and where the non-inferiority margin is delta.
simPowerNoninfSens <- function(nsim=2000,n,rho,sens,delta=0.05,level=.05){

  rcounts <- rmultinom(nsim,n,mnorm.pi(rho=rho,sens=sens))
  rejectVec <- numeric(nsim)
  for(i in 1:nsim){
    rejectVec[i] <- as.numeric(
      scoreci.mp(b=rcounts[2,i],c=rcounts[3,i],n=n,
                 conf.level=1-2*level)$conf.int[1] > -1*delta)}
mean(rejectVec)
}
```

Example: COVID-19 Assay (Sensitivity and Specificity)

- Functions above were used to calculate power

```
set.seed(149689)
# Now compute power at various choices of n_1 assuming sensitivity of 0.96.
# (n_1=117 gives power <.8, n_1=118 gives power>.8)
simPowerNoninfSens(nsim=5000,n=117,rho=.9,sens=.96,delta=0.05)
```

```
[1] 0.7898
```

```
simPowerNoninfSens(nsim=5000,n=118,rho=.9,sens=.96,delta=0.05)
```

```
[1] 0.8342
```

```
# Now compute power at various choices of n_0 assuming  specificity of 0.993.
# Same function can be used, just use specificity value in sens argument.
# (n_0=67 gives power <.8, n_0=68 gives power >.8)
simPowerNoninfSens(nsim=5000,n=67,rho=.9,sens=.993,delta=0.05,level=.05)
```

```
[1] 0.7976
```

```
simPowerNoninfSens(nsim=5000,n=68,rho=.9,sens=.993,delta=0.05,level=.05)
```

```
[1] 0.8008
```

- $n_0 = 68$, $n_1 = 118$.

Example: COVID-19 Assay (Sensitivity and Specificity)

- Functions above were used to calculate power

```
set.seed(149689)
# Now compute power at various choices of n_1 assuming sensitivity of 0.96.
# (n_1=117 gives power <.8, n_1=118 gives power>.8)
simPowerNoninfSens(nsim=5000,n=117,rho=.9,sens=.96,delta=0.05)
```

```
[1] 0.7898
```

```
simPowerNoninfSens(nsim=5000,n=118,rho=.9,sens=.96,delta=0.05)
```

```
[1] 0.8342
```

```
# Now compute power at various choices of n_0 assuming  specificity of 0.993.
# Same function can be used, just use specificity value in sens argument.
# (n_0=67 gives power <.8, n_0=68 gives power >.8)
simPowerNoninfSens(nsim=5000,n=67,rho=.9,sens=.993,delta=0.05,level=.05)
```

```
[1] 0.7976
```

```
simPowerNoninfSens(nsim=5000,n=68,rho=.9,sens=.993,delta=0.05,level=.05)
```

```
[1] 0.8008
```

- $n_0 = 68$, $n_1 = 118$.

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - Small ($f = 0.10 \Rightarrow n = 393$)
    - Medium ($f = 0.25 \Rightarrow n = 64$)
    - Large ($f = 0.40 \Rightarrow n = 25$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - Small ($f = 0.10 \Rightarrow n = 393$)
    - Medium ($f = 0.25 \Rightarrow n = 64$)
    - Large ($f = 0.40 \Rightarrow n = 25$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - Small ($f = 0.10$, $n = 393$)
    - Medium ($f = 0.25$, $n = 64$)
    - Large ($f = 0.40$, $n = 25$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - small ($f = 0.40 \Rightarrow n = 76$)
    - medium ($f = 0.25 \Rightarrow n = 180$)
    - large ($f = 0.10 \Rightarrow n = 1096$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - small ($f = 0.40 \Rightarrow n = 76$)
    - medium ($f = 0.25 \Rightarrow n = 180$)
    - large ($f = 0.10 \Rightarrow n = 1096$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - small ($f = 0.40 \Rightarrow n = 76$)
    - medium ($f = 0.25 \Rightarrow n = 180$)
    - large ($f = 0.10 \Rightarrow n = 1096$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - small ($f = 0.40 \Rightarrow n = 76$)
    - medium ($f = 0.25 \Rightarrow n = 180$)
    - large ($f = 0.10 \Rightarrow n = 1096$)

Final Comments

Cohen's Effect Sizes:

- Based on the social science literature, Cohen (1988) published guidelines for small, medium, and large standardized effect size in various types of power analyses.
  - E.g., for a one-way ANOVA, Cohen's small, medium and large standardized effect sizes ($f = \sigma_W/\sigma_B$) are 0.10, 0.25, and 0.40.
- A deeply flawed strategy is to assume one of these "T-shirt" effect sizes, which allows sample size to be determined without separate consideration of the magnitude of effect on the scale of measurement to be used, the error variability, and the experimental design.
- Essentially, assuming a canned standardized effect size without consideration of the factors that determine it is *pretending to do a power/sample size analysis.*
  - E.g., for 80% power, all balanced one-way layouts come in one of 3 sizes:
    - small ($f = 0.40 \Rightarrow n = 76$)
    - medium ($f = 0.25 \Rightarrow n = 180$)
    - large ($f = 0.10 \Rightarrow n = 1096$)

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-significant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-significant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Post-Hoc Power Analysis

- When a completed study fails to find the significant result it was designed to reveal, the question often arises:
  - *Was the null result because my study was underpowered?*
- To answer this, many investigators embark on retrospective power analysis. This is a fool's errand.
  - If the result was non-significant, then the test necessarily had low power for the *observed effect*. No further calculation is needed.
  - But that does not imply the study was underpowered. It may be that the true effect was trivial or null.
  - And if retrospective power for a *clinically significant effect* is low, this doesn't imply that the null result is less valid or that p would reach significance if study had been larger.
  - The reverse claim is more common: if a non-signifciant p-value was obtained and the study had high power for a clinically significant effect, then it is "more valid" to conclude H_0 is true.
  - These claims are bogus. There is simply no logical basis to strengthen or weaken conclusions based on a post hoc power analysis.
  - For more, see Hoenig and Heisey (2001).

Final Comments

Check your work!

- Often, we have little intuition about the results of a power/sample size calculation.
- That is, it difficult to know if the result "seems about right".
- And there are many opportunities for error when setting up the calculation.
- So, whether your a statistician or an applied researcher, double-check your work, preferably with a second method/piece of software.

Final Comments

Check your work!

- Often, we have little intuition about the results of a power/sample size calculation.
- That is, it difficult to know if the result "seems about right".
- And there are many opportunities for error when setting up the calculation.
- So, whether your a statistician or an applied researcher, double-check your work, preferably with a second method/piece of software.

Final Comments

Check your work!

- Often, we have little intuition about the results of a power/sample size calculation.
- That is, it difficult to know if the result "seems about right".
- And there are many opportunities for error when setting up the calculation.
- So, whether your a statistician or an applied researcher, double-check your work, preferably with a second method/piece of software.

Final Comments

Check your work!

- Often, we have little intuition about the results of a power/sample size calculation.
- That is, it difficult to know if the result "seems about right".
- And there are many opportunities for error when setting up the calculation.
- So, whether your a statistician or an applied researcher, double-check your work, preferably with a second method/piece of software.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/~rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/~rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/∼rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/∼rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/∼rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- G*Power: www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- UCLA's Advanced Research Computing center has illustrations of G*Power for standard several power/sample size problems here: stats.oarc.ucla.edu/other/gpower/.
- Russ Lenth Power Webpage (with downloadable Java program): homepage.divms.uiowa.edu/∼rlenth/Power/
- Papers on Power and Sample Size:
  - Hoenig, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
  - Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193.

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

References & Resources

- Web Tutorials and Presentations on Sample Size/Power Calculations:
  - Williamson, Mark, Sample Size Calculation with R med.und.edu/daccota/_files/pdfs/berdc_resource_pdfs/sample_size_r_module.
  - Schweinberger, Martin, Power Analysis in R ladal.edu.au/pwr.html
  - Introduction to Power Analysis from UCLA's ARC: stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/
  - Liu, Honghu, Study Design: Sample Size Calculation & Power Analysis. ctsi.ucla.edu/education/files/view/rcmar-seminars/2014_Apr_Liu.pdf.
- A funny video: www.youtube.com/watch?v=PbODigCZqL8

Thanks

- If you need assistance with power analysis/sample size determination or with any statistical design or analysis task, please contact the SCC.
  - www.stat.uga/consulting
- We can help!

Thank you!

Thanks

- If you need assistance with power analysis/sample size determination or with any statistical design or analysis task, please contact the SCC.
  - www.stat.uga/consulting
- We can help!

Thank you!

Thanks

- If you need assistance with power analysis/sample size determination or with any statistical design or analysis task, please contact the SCC.
  - www.stat.uga/consulting
- We can help!

Thank you!

Questions?

Questions?