Power and Sample Size

Fiona Callaghan, MA MS PhD

12 February 2015

Motivating example

Hypothesis Testing

Power

Examples

- Example 1: One sample test of proportion
- Example 2: Two-sample test of proportion (Risk difference RD)
- Example 3: Two sample T-test (Normal, continuous outcome measurements)

Conclusion/Comments

Disclaimer

Disclosure: Opinions/statements in this course reflect personal opinions, not policy or opinions of Amgen, Inc. or FDA.

Section 1

Motivating example

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.
- Study Design: We find 80 HIV negative people who live with an HIV+ person and follow them for 6 months. Take HIV tests before and after.

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.
- Study Design: We find 80 HIV negative people who live with an HIV+ person and follow them for 6 months. Take HIV tests before and after.
- Results: 0/80 people contract HIV.

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.
- Study Design: We find 80 HIV negative people who live with an HIV+ person and follow them for 6 months. Take HIV tests before and after.
- Results: 0/80 people contract HIV.
- Claim: The risk of contracting HIV if you live with an HIV+ person is zero.

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.
- Study Design: We find 80 HIV negative people who live with an HIV+ person and follow them for 6 months. Take HIV tests before and after.
- Results: 0/80 people contract HIV.
- Claim: The risk of contracting HIV if you live with an HIV+ person is zero.
- Is this claim true?

Motivating Example

- Hypothesis: Suppose we wanted to estimate the probability of getting HIV/AIDS for people living in a family with at least one HIV+ family member.
- Study Design: We find 80 HIV negative people who live with an HIV+ person and follow them for 6 months. Take HIV tests before and after.
- Results: 0/80 people contract HIV.
- Claim: The risk of contracting HIV if you live with an HIV+ person is zero.
- Is this claim true?
- NB. Based on a real (published) example.

Motivating Example

- The "best" estimate of risk based on the data is 0.

Motivating Example

- The "best" estimate of risk based on the data is 0.
- However, the "95% confidence interval" for this quantity is $(0, 0.045)$.

Motivating Example

- The "best" estimate of risk based on the data is 0.
- However, the "95% confidence interval" for this quantity is $(0, 0.045)$.
- i.e. "plausible" values for the risk of contracting HIV range from 0% to 4.5%!

Motivating Example

- The "best" estimate of risk based on the data is 0.
- However, the "95% confidence interval" for this quantity is $(0, 0.045)$.
- i.e. "plausible" values for the risk of contracting HIV range from 0% to 4.5%!
- How could we fix this?

Questions

1. Is it a sample size issue?

Questions

1. Is it a sample size issue?
2. Can we ever prove the true risk is 0%?

Questions

1. Is it a sample size issue?
2. Can we ever prove the true risk is 0%?
3. Can we ever rule out (disprove) that the true risk is, say, $> 1\%$?

Questions

1. Is it a sample size issue?

Questions

1. Is it a sample size issue?
    - What if they found $0/1000$ people contracted HIV?

Questions

1. Is it a sample size issue?
   - What if they found $0/1000$ people contracted HIV?
   - The confidence interval (CI) is $(0, 0.004)$ i.e. 0% to 0.4%

Questions

1. Is it a sample size issue?
   - What if they found $0/1000$ people contracted HIV?
   - The confidence interval (CI) is $(0, 0.004)$ i.e. 0% to 0.4%
   - The larger the sample size, the smaller the confidence interval.

Questions

1. Is it a sample size issue?
   - What if they found $0/1000$ people contracted HIV?
   - The confidence interval (CI) is $(0, 0.004)$ i.e. 0% to 0.4%
   - The larger the sample size, the smaller the confidence interval.
   - Even if we recruited a billion people, and never saw anyone contract HIV, we would only make our CI smaller and smaller, but we would never "prove" 0%, we would only "exclude" smaller and smaller non-zero values.

Questions

1. Is it a sample size issue?
   - What if they found $0/1000$ people contracted HIV?
   - The confidence interval (CI) is $(0, 0.004)$ i.e. 0% to 0.4%
   - The larger the sample size, the smaller the confidence interval.
   - Even if we recruited a billion people, and never saw anyone contract HIV, we would only make our CI smaller and smaller, but we would never "prove" 0%, we would only "exclude" smaller and smaller non-zero values.
   - On the other hand, a possible risk of 0.4% is certainly more comforting than a risk of 4.5%.

Questions

1. Is it a sample size issue?
   - What if they found $0/1000$ people contracted HIV?
   - The confidence interval (CI) is $(0, 0.004)$ i.e. 0% to 0.4%
   - The larger the sample size, the smaller the confidence interval.
   - Even if we recruited a billion people, and never saw anyone contract HIV, we would only make our CI smaller and smaller, but we would never "prove" 0%, we would only "exclude" smaller and smaller non-zero values.
   - On the other hand, a possible risk of 0.4% is certainly more comforting than a risk of 4.5%.
   - Part of the solution may be to ask: what risk of contracting HIV is considered "acceptable" to not be able to rule out?

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.
     - Formulate an "alternative hypothesis", which is typically defined as *not* the null hypothesis.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.
     - Formulate an "alternative hypothesis", which is typically defined as *not* the null hypothesis.
     - If you find a "positive" result (e.g. p-value < 0.05) you have enough evidence to reject the null hypothesis (but we have not proven the alternative hypothesis).

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.
     - Formulate an "alternative hypothesis", which is typically defined as *not* the null hypothesis.
     - If you find a "positive" result (e.g. p-value < 0.05) you have enough evidence to reject the null hypothesis (but we have not proven the alternative hypothesis).
     - If you find a "negative" result (e.g. p-value > 0.05) you have not proven anything.

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.
     - Formulate an "alternative hypothesis", which is typically defined as *not* the null hypothesis.
     - If you find a "positive" result (e.g. p-value < 0.05) you have enough evidence to reject the null hypothesis (but we have not proven the alternative hypothesis).
     - If you find a "negative" result (e.g. p-value > 0.05) you have not proven anything.
     - A hypothesis is "innocent until proven guilty"

Questions

2. Can we ever prove the true risk is 0%?
   - Short answer: No.
   - Long answer: We never "prove" a hypothesis.
     - Classical stats uses a falsification theory of scientific method i.e. theories can only be disproved.
     - Formulate a "null hypothesis": a conservative position that you want to disprove.
     - Formulate an "alternative hypothesis", which is typically defined as *not* the null hypothesis.
     - If you find a "positive" result (e.g. p-value < 0.05) you have enough evidence to reject the null hypothesis (but we have not proven the alternative hypothesis).
     - If you find a "negative" result (e.g. p-value > 0.05) you have not proven anything.
     - A hypothesis is "innocent until proven guilty"
     - Note: Bayesian stats does not use a falsification theory of scientific method, but that is for another time...

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "
   - i.e. Before we begin our study, we can define what level of proof (i.e. 5%) will be required to say we have disproved something.

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "
   - i.e. Before we begin our study, we can define what level of proof (i.e. 5%) will be required to say we have disproved something.
   - How incompatible/unusual do the results we see have to be, before a skeptic will admit that the null hypothesis is not plausible?

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "
   - i.e. Before we begin our study, we can define what level of proof (i.e. 5%) will be required to say we have disproved something.
   - How incompatible/unusual do the results we see have to be, before a skeptic will admit that the null hypothesis is not plausible?
   - Typically, we say that if the results we see have a 5% or less chance of being observed if the null were true, then we reject the null.

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "
   - i.e. Before we begin our study, we can define what level of proof (i.e. 5%) will be required to say we have disproved something.
   - How incompatible/unusual do the results we see have to be, before a skeptic will admit that the null hypothesis is not plausible?
   - Typically, we say that if the results we see have a 5% or less chance of being observed if the null were true, then we reject the null.
   - This is called the "alpha-level", e.g. $\alpha = 0.05$.

Questions

3. Can we ever rule out (disprove) that the true risk is $> 1\%$?
   - Yes, but what counts as evidence?
   - We can observe enough evidence to "reject the null hypothesis at the 5% level "
   - i.e. Before we begin our study, we can define what level of proof (i.e. 5%) will be required to say we have disproved something.
   - How incompatible/unusual do the results we see have to be, before a skeptic will admit that the null hypothesis is not plausible?
   - Typically, we say that if the results we see have a 5% or less chance of being observed if the null were true, then we reject the null.
   - This is called the "alpha-level", e.g. $\alpha = 0.05$.
   - Another way to think of it is the probability of making a mistake if the null hypothesis is actually true.

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$
- Which, in turn, is equivalent to setting the null hypothesis (the one you want to disprove) to "The risk is 0.01" vs. "The risk is less than 0.01"

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$
- Which, in turn, is equivalent to setting the null hypothesis (the one you want to disprove) to "The risk is 0.01" vs. "The risk is less than 0.01"
- Then calculate a sample size required in order to have a *high probability of seeing results that would lead you to reject the null hypothesis* ...

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$
- Which, in turn, is equivalent to setting the null hypothesis (the one you want to disprove) to "The risk is 0.01" vs. "The risk is less than 0.01"
- Then calculate a sample size required in order to have a *high probability of seeing results that would lead you to reject the null hypothesis* ...
- ...assuming the null hypothesis is false (of course they could be wrong, there are no guarantees).

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$
- Which, in turn, is equivalent to setting the null hypothesis (the one you want to disprove) to "The risk is 0.01" vs. "The risk is less than 0.01"
- Then calculate a sample size required in order to have a *high probability of seeing results that would lead you to reject the null hypothesis* ...
- ...assuming the null hypothesis is false (of course they could be wrong, there are no guarantees).
- So pick a sample size that will get you an $80 - 90\%$ chance of getting a significant result (if you are correct about the general direction of the results).

Summary: How to fix the example

- They should pick the level of risk that they want to rule out, e.g. 1%? 0.1%?
- In this case, equivalent to setting the confidence interval you would like to see, e.g. $(0, 0.01)$
- Which, in turn, is equivalent to setting the null hypothesis (the one you want to disprove) to "The risk is 0.01" vs. "The risk is less than 0.01"
- Then calculate a sample size required in order to have a *high probability of seeing results that would lead you to reject the null hypothesis* ...
- ...assuming the null hypothesis is false (of course they could be wrong, there are no guarantees).
- So pick a sample size that will get you an $80 - 90\%$ chance of getting a significant result (if you are correct about the general direction of the results).
- How do you do that? Read on...

Plan

- Next part of the talk will formalize some of the ideas in the example:

Plan

- Next part of the talk will formalize some of the ideas in the example:
  - Hypothesis Testing

Plan

- Next part of the talk will formalize some of the ideas in the example:
  - Hypothesis Testing
  - Power (probability of getting a significant result, if the null is false)

Plan

- Next part of the talk will formalize some of the ideas in the example:
  - Hypothesis Testing
  - Power (probability of getting a significant result, if the null is false)
  - Sample size examples

Section 2

Hypothesis Testing

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"
2. Quantify the hypothesis, e.g. "The probability of contracting HIV is greater than 1%" $= H_0 : \pi > 0.01$

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"
2. Quantify the hypothesis, e.g. "The probability of contracting HIV is greater than 1%" $= H_0 : \pi > 0.01$
3. What will you be measuring or counting? e.g. The proportion of people who get HIV

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"
2. Quantify the hypothesis, e.g. "The probability of contracting HIV is greater than 1%" $= H_0 : \pi > 0.01$
3. What will you be measuring or counting? e.g. The proportion of people who get HIV
4. How will you design your experiment in order to answer your question? e.g. one group, two groups, compare over time?

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"
2. Quantify the hypothesis, e.g. "The probability of contracting HIV is greater than 1%" $= H_0 : \pi > 0.01$
3. What will you be measuring or counting? e.g. The proportion of people who get HIV
4. How will you design your experiment in order to answer your question? e.g. one group, two groups, compare over time?
5. Decide what type of test you will need (talk to a statistician) e.g. one-sample test of proportion

General Idea

1. Set up a conservative hypothesis you want to disprove, e.g. "The is some risk of contracting HIV"
2. Quantify the hypothesis, e.g. "The probability of contracting HIV is greater than 1%" $= H_0 : \pi > 0.01$
3. What will you be measuring or counting? e.g. The proportion of people who get HIV
4. How will you design your experiment in order to answer your question? e.g. one group, two groups, compare over time?
5. Decide what type of test you will need (talk to a statistician) e.g. one-sample test of proportion
6. Define in advance what counts as "rejecting" the hypothesis (α-level) e.g. p-value < 0.05

General Idea, continued

7. Decide the smallest clinically interesting difference that you would like to detect, i.e. the difference between the null hypothesis value ($p_0 = 0.01$), and the value closest to the null value that you think is interesting to detect (e.g. $p_1 = 0.0001$?).

General Idea, continued

7. Decide the smallest clinically interesting difference that you would like to detect, i.e. the difference between the null hypothesis value ($p_0 = 0.01$), and the value closest to the null value that you think is interesting to detect (e.g. $p_1 = 0.0001$?).
8. Calculate the sample size you will need to have a "reasonable" chance of finding a significant result.

General Idea, continued

7. Decide the smallest clinically interesting difference that you would like to detect, i.e. the difference between the null hypothesis value ($p_0 = 0.01$), and the value closest to the null value that you think is interesting to detect (e.g. $p_1 = 0.0001$?).
8. Calculate the sample size you will need to have a "reasonable" chance of finding a significant result.
9. Gather data, check you met the test assumptions in step 5.

General Idea, continued

7. Decide the smallest clinically interesting difference that you would like to detect, i.e. the difference between the null hypothesis value ($p_0 = 0.01$), and the value closest to the null value that you think is interesting to detect (e.g. $p_1 = 0.0001$?).
8. Calculate the sample size you will need to have a "reasonable" chance of finding a significant result.
9. Gather data, check you met the test assumptions in step 5.
10. Perform test. If $p < 0.05$, then the probability of seeing your results if the null is true is very unlikely, so....the null is rejected: publish!

General Idea, continued

7. Decide the smallest clinically interesting difference that you would like to detect, i.e. the difference between the null hypothesis value ($p_0 = 0.01$), and the value closest to the null value that you think is interesting to detect (e.g. $p_1 = 0.0001$?).
8. Calculate the sample size you will need to have a "reasonable" chance of finding a significant result.
9. Gather data, check you met the test assumptions in step 5.
10. Perform test. If $p < 0.05$, then the probability of seeing your results if the null is true is very unlikely, so....the null is rejected: publish!
11. If $p > 0.05$, results compatible with null: have not proved *or disproved* anything.

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$
- Alternative hypothesis: e.g. $H_1 : \pi \leq 0.01$.

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$
- Alternative hypothesis: e.g. $H_1 : \pi \leq 0.01$.
- Parameter: The truth that you would like to know, distilled to a number (unknown) e.g. π (true proportion).

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$
- Alternative hypothesis: e.g. $H_1 : \pi \leq 0.01$.
- Parameter: The truth that you would like to know, distilled to a number (unknown) e.g. π (true proportion).
- Estimate: How you are going to estimate the parameter using your data, e.g. $\hat{p} = \frac{\#\text{ of new HIV cases}}{\text{Everyone in study}}$

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$
- Alternative hypothesis: e.g. $H_1 : \pi \leq 0.01$.
- Parameter: The truth that you would like to know, distilled to a number (unknown) e.g. π (true proportion).
- Estimate: How you are going to estimate the parameter using your data, e.g. $\hat{p} = \frac{\#\text{ of new HIV cases}}{\text{Everyone in study}}$
- P-value: Assuming the null is true, what is the probability of seeing your results?

Parts of the hypothesis test

- Null hypothesis: e.g. $H_0 : \pi > 0.01$
- Alternative hypothesis: e.g. $H_1 : \pi \leq 0.01$.
- Parameter: The truth that you would like to know, distilled to a number (unknown) e.g. π (true proportion).
- Estimate: How you are going to estimate the parameter using your data, e.g. $\hat{p} = \frac{\#\text{ of new HIV cases}}{\text{Everyone in study}}$
- P-value: Assuming the null is true, what is the probability of seeing your results?
- Decision rule: Q: How "unlikely" do your results have to be (assuming the null is true) before you conclude the null hypothesis is false? A: $\alpha-$level.

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).
- There are also many other things to think of to ensure a successful experiment *before* you start:

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).
- There are also many other things to think of to ensure a successful experiment *before* you start:
  - What is your objective? (come back to this with every other decision)

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).
- There are also many other things to think of to ensure a successful experiment *before* you start:
  - What is your objective? (come back to this with every other decision)
  - Pre-specify your hypotheses.

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).
- There are also many other things to think of to ensure a successful experiment *before* you start:
  - What is your objective? (come back to this with every other decision)
  - Pre-specify your hypotheses.
  - Experimental design (different experimental design will leads different analysis/tests).

How do I make a "successful" experiment more likely?

- Calculate sample size you will need to have a power of $80 - 90\%$ (Power covered next section).
- There are also many other things to think of to ensure a successful experiment *before* you start:
  - What is your objective? (come back to this with every other decision)
  - Pre-specify your hypotheses.
  - Experimental design (different experimental design will leads different analysis/tests).
  - What is the appropriate analysis?

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?
  - Will you have missing data? Systematically excluded? (bias)

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?
  - Will you have missing data? Systematically excluded? (bias)
  - What is your outcome? (leads to different analyses)

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?
  - Will you have missing data? Systematically excluded? (bias)
  - What is your outcome? (leads to different analyses)
  - Analysis assumptions, other assumptions

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?
  - Will you have missing data? Systematically excluded? (bias)
  - What is your outcome? (leads to different analyses)
  - Analysis assumptions, other assumptions
  - ...many other important questions.

How do I make a "successful" experiment more likely? (continued)

- Many other important issues:
  - Do you have a random sample?
  - Who is in your data, who isn't? (bias)
  - How many tests will you perform?
  - Will you have missing data? Systematically excluded? (bias)
  - What is your outcome? (leads to different analyses)
  - Analysis assumptions, other assumptions
  - ...many other important questions.
- Talk to your statistician/epidemiologist and plan ahead!

Section 3

Power

Power

- What is power and why should you care?

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.
- Typically, you want power of $70 - 90\%$.

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.
- Typically, you want power of $70 - 90\%$.
- Power is notated with $1 - \beta$.

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.
- Typically, you want power of $70 - 90\%$.
- Power is notated with $1 - \beta$.
- What happens if you ignore power?

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.
- Typically, you want power of $70 - 90\%$.
- Power is notated with $1 - \beta$.
- What happens if you ignore power?
  - and your study "fails" (not significant)?

Power

- What is power and why should you care?
- Power = probability of rejecting the null hypothesis, if one of the alternative hypotheses is true.
- In other words, the chance you are going to get a significant result, given the alternative hypothesis that you think is true.
- Typically, you want power of $70 - 90\%$.
- Power is notated with $1 - \beta$.
- What happens if you ignore power?
  - and your study "fails" (not significant)?
  - and your study "succeeds" (significant)?

Another way to think of power

Decision	Truth	
	Null hypothesis true	Null hypothesis false
Reject null	Type I error, FP $\Pr(\text{Type I error}\|\text{Null true}) = \alpha$	Correct decision, TP
Fail to reject null (Keep null)	Correct decision, TN	Type II error, FN $\Pr(\text{Type II error}\|\text{alternative true}) = \beta$

Table: Power, alpha-level and types of error in statistical decision rule.

Want both α and β to be small. Want $1 - \beta$ (power) to be large.

Factors that affect power

- In general, power depends on a number of factors:

Factors that affect power

- In general, power depends on a number of factors:
  - Power typically increases with increased sample size.

Factors that affect power

- In general, power depends on a number of factors:
  - Power typically increases with increased sample size.
  - The difference between the null and alternative hypothesis scenarios. If you are trying to find a big difference , then you have a higher probability of finding it; if the signal is very small, then it is more difficult to detect and therefore you have less power.

Factors that affect power

- In general, power depends on a number of factors:
  - Power typically increases with increased sample size.
  - The difference between the null and alternative hypothesis scenarios. If you are trying to find a big difference , then you have a higher probability of finding it; if the signal is very small, then it is more difficult to detect and therefore you have less power.
  - Variability. If the standard deviation is large, then there is more "noise" and it is harder to detect a difference, power goes down. If variance/sd is small, it is easier to detect a difference (power goes up).

Factors that affect power

- In general, power depends on a number of factors:
  - Power typically increases with increased sample size.
  - The difference between the null and alternative hypothesis scenarios. If you are trying to find a big difference , then you have a higher probability of finding it; if the signal is very small, then it is more difficult to detect and therefore you have less power.
  - Variability. If the standard deviation is large, then there is more "noise" and it is harder to detect a difference, power goes down. If variance/sd is small, it is easier to detect a difference (power goes up).
  - The α-level. Bigger α, more power – but, your conclusions are less impressive, i.e. a $\alpha = 0.10$ is easier to achieve, but a less interesting result.

Factors that affect power

- In general, power depends on a number of factors:
  - Power typically increases with increased sample size.
  - The difference between the null and alternative hypothesis scenarios. If you are trying to find a big difference , then you have a higher probability of finding it; if the signal is very small, then it is more difficult to detect and therefore you have less power.
  - Variability. If the standard deviation is large, then there is more "noise" and it is harder to detect a difference, power goes down. If variance/sd is small, it is easier to detect a difference (power goes up).
  - The α-level. Bigger α, more power – but, your conclusions are less impressive, i.e. a $\alpha = 0.10$ is easier to achieve, but a less interesting result.
  - Some tests are more powerful than others because they assume you know more about the data e.g. tests that assume normally distributed data are more powerful than tests that do not. But this only helps you if it is true!

Sample size

- Power and sample size are inter-related.

Sample size

- Power and sample size are inter-related.
- The precise formula for power and sample size depends on the type of test, i.e. there are different formulas for one-sample test of proportion, 2 sample tests, regression, time-series, etc.

Sample size

- Power and sample size are inter-related.
- The precise formula for power and sample size depends on the type of test, i.e. there are different formulas for one-sample test of proportion, 2 sample tests, regression, time-series, etc.
- Sample size can be difficult to estimate because of all the factors that influence power and sample size.

Sample size

- Power and sample size are inter-related.
- The precise formula for power and sample size depends on the type of test, i.e. there are different formulas for one-sample test of proportion, 2 sample tests, regression, time-series, etc.
- Sample size can be difficult to estimate because of all the factors that influence power and sample size.
- The examples focus on estimating sample size for a given amount of power.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.
    - Previous studies? Pilot studies?

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.
    - Previous studies? Pilot studies?
    - Typically, you calculate sample size for a range of sds/means/power and take a (reasonable) worst case scenario.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.
    - Previous studies? Pilot studies?
    - Typically, you calculate sample size for a range of sds/means/power and take a (reasonable) worst case scenario.
  - "Clinically interesting difference"$=\Delta$. Usually easy, because you define it: it is the amount of difference between the two groups that you would find it "interesting" to detect

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.
    - Previous studies? Pilot studies?
    - Typically, you calculate sample size for a range of sds/means/power and take a (reasonable) worst case scenario.
  - "Clinically interesting difference"$=\Delta$. Usually easy, because you define it: it is the amount of difference between the two groups that you would find it "interesting" to detect
  - α level –easy, usually 0.05.

What is often required for sample size

- After deciding on an analysis method but *before you have done the experiment:*
  - Power – easy, because you define it.
  - Estimating the means and the standard deviations.
    - Difficult, especially anticipating what the SDs will be.
    - Previous studies? Pilot studies?
    - Typically, you calculate sample size for a range of sds/means/power and take a (reasonable) worst case scenario.
  - "Clinically interesting difference"$=\Delta$. Usually easy, because you define it: it is the amount of difference between the two groups that you would find it "interesting" to detect
  - α level –easy, usually 0.05.
  - One-tailed vs two-tailed test

Section 4

Examples

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$
  - $q_1 = 1 - p_1$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$
  - $q_1 = 1 - p_1$
  - $\alpha = 0.05$, so $1 - \alpha/2 = 0.975$.

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$
  - $q_1 = 1 - p_1$
  - $\alpha = 0.05$, so $1 - \alpha/2 = 0.975$.
  - $z_{1-\alpha/2} =$ the normal percentile ("z-score") corresponding to $1 - \alpha/2$, e.g. $1 - \alpha/2 = 0.975 \rightarrow z_{0.975} \approx 1.96$ (look-up table).

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$
  - $q_1 = 1 - p_1$
  - $\alpha = 0.05$, so $1 - \alpha/2 = 0.975$.
  - $z_{1-\alpha/2} =$ the normal percentile ("z-score") corresponding to $1 - \alpha/2$, e.g. $1 - \alpha/2 = 0.975 \rightarrow z_{0.975} \approx 1.96$ (look-up table).
  - $1 - \beta =$ power. In practice, calculate a range of values, $1 - \beta = 0.7, 0.8, 0.9, ...$

Example 1: One sample test of proportion (HIV example)

- Formula: $n = \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$
  - $p_0 =$ null hypothesis value, e.g. $H_0 : p \geq p_0 = 0.01$
  - $q_0 = 1 - p_0 = 0.99$
  - $p_1 =$ alternative hypothesis values $p_1 < 0.01$. e.g. $p_1 = 0.001$
  - $q_1 = 1 - p_1$
  - $\alpha = 0.05$, so $1 - \alpha/2 = 0.975$.
  - $z_{1-\alpha/2} =$ the normal percentile ("z-score") corresponding to $1 - \alpha/2$, e.g. $1 - \alpha/2 = 0.975 \rightarrow z_{0.975} \approx 1.96$ (look-up table).
  - $1 - \beta =$ power. In practice, calculate a range of values, $1 - \beta = 0.7, 0.8, 0.9, ...$
  - $z_{1-\beta} =$ the normal percentile corresponding to the power, e.g. for 90% power, $z_{0.9} \approx 1.28$

Sample size calculation

For 90% power, with $\alpha = 0.05$, null value $p_0 = 0.01$ and alternative value $p_1 = 0.001$, the sample size is...

$$
\begin{aligned}
n &= \frac{p_0 q_0 \left(z_{1-\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2} \\
&= \frac{0.01 \times 0.99 \left(1.96 + 1.28 \sqrt{\frac{0.001 \times 0.999}{0.99 \times 0.01}} \right)^2}{(0.001 - 0.01)^2} \\
&= \frac{0.0099 \left(1.96 + 1.28 \times 0.32 \right)^2}{0.000081} \\
&= 684.5 \\
\rightarrow &\quad \text{need at least 685 subjects (always round up).}
\end{aligned}
$$

What power did the original ($N = 80$) example have?

- Formula: $\text{Power} = \Psi\left[\sqrt{\frac{p_0 q_0}{p_1 q_1}}\left(z_{\alpha/2} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_0 q_0}}\right)\right]$
  - $\Psi(z) =$ inverse of the "z-scores"

What power did the original ($N = 80$) example have?

- Formula: $\text{Power} = \Psi\left[\sqrt{\frac{p_0 q_0}{p_1 q_1}}\left(z_{\alpha/2} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_0 q_0}}\right)\right]$
  - $\Psi(z) =$ inverse of the "z-scores"
  - i.e. the cumulative probability of the normal distribution corresponding to z.

What power did the original ($N = 80$) example have?

For $N = 80$, with $\alpha = 0.05$, null value $p_0 = 0.01$ and alternative value $p_1 = 0.001$ the power is...

$$
\begin{aligned}
\text{Power} &= \Psi\left[\sqrt{\frac{p_0 q_0}{p_1 q_1}}\left(z_{\alpha/2} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_0 q_0}}\right)\right] \\
&= \Psi\left[\sqrt{\frac{0.01 \times 0.99}{0.001 \times 0.999}}\left(z_{0.025} + \frac{|0.01 - 0.001|\sqrt{80}}{\sqrt{0.01 \times 0.99}}\right)\right] \\
&= \Psi\Big[3.15\Big(-1.96 + 0.809\Big)\Big] \\
&= \Psi\Big[-3.62\Big] \\
&= 0.00015
\end{aligned}
$$

$\rightarrow$ virtually no chance of ruling out 1% as a possible value.

Another method: specify length of confidence interval

- Sometimes the goal is to get an estimate within a certain precision
- Handy rule of thumb: $N \sim 1000$ required for rate of $\sim 50\%$ with margin of error/width of confidence interval $+/3\%$ (see this on political polling a lot).
- Formula: $ME = z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n}}$
- or, $n = \frac{p_1 q_1}{\left(\frac{ME}{z_{1-\alpha/2}}\right)^2}$

Another method: specify length of confidence interval

- For ME of $+-0.1\%$ when we think the true rate is $p_1 = 0.001$ or 0.1%:

$$
\begin{aligned}
n &= \frac{p_1 q_1}{\left(\frac{ME}{z_{1-\alpha/2}}\right)^2} \\
&= \frac{0.001 * 0.999}{\left(\frac{0.001}{1.96}\right)^2} \\
&= \frac{0.000999}{\left(0.00051\right)^2} \\
&= 3837.8 \\
&\rightarrow 3838
\end{aligned}
$$

Assumptions of 1 sample test of proportion

- Also have to check the assumptions of the test, to ensure conclusions are valid:

Assumptions of 1 sample test of proportion

- Also have to check the assumptions of the test, to ensure conclusions are valid:
- "Large sample" test. How large is large? Rule of thumb: $np_0q_0 > 5$

Assumptions of 1 sample test of proportion

- Also have to check the assumptions of the test, to ensure conclusions are valid:
- "Large sample" test. How large is large? Rule of thumb: $np_0q_0 > 5$
- In our example, $np_0q_0 = 685 \times 0.01 \times 0.99 = 6.8$, so we are ok.

Assumptions of 1 sample test of proportion

- Also have to check the assumptions of the test, to ensure conclusions are valid:
- "Large sample" test. How large is large? Rule of thumb: $np_0q_0 > 5$
- In our example, $np_0q_0 = 685 \times 0.01 \times 0.99 = 6.8$, so we are ok.
- Other assumptions.

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.
- If 700 is impossible, could re-run calculations with different assumptions e.g. power, detect a larger $p_1 - p_0$

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.
- If 700 is impossible, could re-run calculations with different assumptions e.g. power, detect a larger $p_1 - p_0$
- If sample size is fixed for practical reasons, e.g. money, then you can still calculate the power that your study has (what chance do you have of getting a significant result?)

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.
- If 700 is impossible, could re-run calculations with different assumptions e.g. power, detect a larger $p_1 - p_0$
- If sample size is fixed for practical reasons, e.g. money, then you can still calculate the power that your study has (what chance do you have of getting a significant result?)
- Power is important for "negative" studies i.e. when you want evidence to *support* the null hypothesis: "We had 95% power to detect a difference of 1% and we (still) did not find it".

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.
- If 700 is impossible, could re-run calculations with different assumptions e.g. power, detect a larger $p_1 - p_0$
- If sample size is fixed for practical reasons, e.g. money, then you can still calculate the power that your study has (what chance do you have of getting a significant result?)
- Power is important for "negative" studies i.e. when you want evidence to *support* the null hypothesis: "We had 95% power to detect a difference of 1% and we (still) did not find it".
- For "negative" studies, should really use a "non-inferiority" test. (not covered here).

Comments

- Experimenters needed approximately 700 people to say what they wanted to say.
- If 700 is impossible, could re-run calculations with different assumptions e.g. power, detect a larger $p_1 - p_0$
- If sample size is fixed for practical reasons, e.g. money, then you can still calculate the power that your study has (what chance do you have of getting a significant result?)
- Power is important for "negative" studies i.e. when you want evidence to *support* the null hypothesis: "We had 95% power to detect a difference of 1% and we (still) did not find it".
- For "negative" studies, should really use a "non-inferiority" test. (not covered here).
- Remember "Absence of evidence is not evidence of absence"; if you don't have enough evidence, you haven't proven (or disproved) anything

Example 2: Two-sample test of proportion

- Comparing response rate or disease prevalence in two groups

Example 2: Two-sample test of proportion

- Comparing response rate or disease prevalence in two groups
- Group 1: N=40, response rate is 75%.

Example 2: Two-sample test of proportion

- Comparing response rate or disease prevalence in two groups
- Group 1: N=40, response rate is 75%.
- Group 2: N=40, response rate is 85%.

Example 2: Two-sample test of proportion

- Comparing response rate or disease prevalence in two groups
- Group 1: N=40, response rate is 75%.
- Group 2: N=40, response rate is 85%.
- Conclusion: New method improvement over old method by 10%!

Example 2: Two-sample test of proportion

- Comparing response rate or disease prevalence in two groups
- Group 1: N=40, response rate is 75%.
- Group 2: N=40, response rate is 85%.
- Conclusion: New method improvement over old method by 10%!
- Is the new method really better than the old method?

Two-sample test of proportions

- Null: Proportions are the same, $H_0 : p_1 = p_2$

Two-sample test of proportions

- Null: Proportions are the same, $H_0 : p_1 = p_2$
- Alternative: Proportions are different, $H_1 : p_1 \neq p_2$

Two-sample test of proportions

- Null: Proportions are the same, $H_0 : p_1 = p_2$
- Alternative: Proportions are different, $H_1 : p_1 \neq p_2$
- Parameter: True value of difference, $p_1 - p_2$

Two-sample test of proportions

- Null: Proportions are the same, $H_0 : p_1 = p_2$
- Alternative: Proportions are different, $H_1 : p_1 \neq p_2$
- Parameter: True value of difference, $p_1 - p_2$
- Estimate: Difference of estimates in each sample, response rate group 1-response rate for group 2.

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}} z_{1-\alpha/2} + \sqrt{p_1 q_1 + p_2 q_2} z_{1-\beta} \right]^2 / \Delta^2$

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}} z_{1-\alpha/2} + \sqrt{p_1 q_1 + p_2 q_2} z_{1-\beta} \right]^2 / \Delta^2$
- for n in each sample and where:

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}} z_{1-\alpha/2} + \sqrt{p_1 q_1 + p_2 q_2} z_{1-\beta} \right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}z_{1-\beta}\right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.
  - p_2 = the projected response rate in group 2 under alternative hypothesis. $q_2 = 1 - p_2$.

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}} z_{1-\alpha/2} + \sqrt{p_1 q_1 + p_2 q_2} z_{1-\beta} \right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.
  - p_2 = the projected response rate in group 2 under alternative hypothesis. $q_2 = 1 - p_2$.
  - $\Delta = |p_2 - p_1|$

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}z_{1-\beta}\right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.
  - p_2 = the projected response rate in group 2 under alternative hypothesis. $q_2 = 1 - p_2$.
  - $\Delta = |p_2 - p_1|$
  - $\bar{p} = \frac{1}{2}(p_1 + p_2)$. $\bar{q} = 1 - \bar{p}$

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}z_{1-\beta}\right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.
  - p_2 = the projected response rate in group 2 under alternative hypothesis. $q_2 = 1 - p_2$.
  - $\Delta = |p_2 - p_1|$
  - $\bar{p} = \frac{1}{2}(p_1 + p_2)$. $\bar{q} = 1 - \bar{p}$
  - $z_{1-\alpha/2}$ =the z-score associated with the $1 - \alpha/2$-th percentile.

Formula

- $n = \left[\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}z_{1-\beta}\right]^2 / \Delta^2$
- for n in each sample and where:
  - p_1 = the projected response rate in group 1 under alternative hypothesis. $q_1 = 1 - p_1$.
  - p_2 = the projected response rate in group 2 under alternative hypothesis. $q_2 = 1 - p_2$.
  - $\Delta = |p_2 - p_1|$
  - $\bar{p} = \frac{1}{2}(p_1 + p_2)$. $\bar{q} = 1 - \bar{p}$
  - $z_{1-\alpha/2}$ =the z-score associated with the $1 - \alpha/2$-th percentile.
  - $z_{1-\beta}$ = the z-score associated with the $1 - \beta$-th (power) percentile

What sample size is required?

For 90% power $(1-\beta)$, $\alpha = 0.05, p_1 = 0.75, p_2 = 0.85$,
$\rightarrow \bar{p} = \frac{0.75+0.85}{2} = 0.8, \bar{q} = 0.2, \Delta = 0.1$

$$
\begin{aligned}
n &= \frac{\left[\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}z_{1-\beta}\right]^2}{\Delta^2} \\
&= \frac{\left[\sqrt{2 \times 0.8 \times 0.2} \times 1.96 + \sqrt{0.75 \times 0.25 + 0.85 \times 0.15} \times 1.28\right]^2}{0.1^2} \\
&= \frac{\left[1.11 + 0.718\right]^2}{0.01} \\
&= 333.8 \\
\rightarrow &\quad 334 \text{ in each group}
\end{aligned}
$$

Example: 2-sample test of normal outcomes

- Suppose you want to show LDL levels are the same in two populations.

Example: 2-sample test of normal outcomes

- Suppose you want to show LDL levels are the same in two populations.
- An important assumption of this test is that the data is normally distributed.

Example: 2-sample test of normal outcomes

- Suppose you want to show LDL levels are the same in two populations.
- An important assumption of this test is that the data is normally distributed.
- Typically, lab values are not normal, but the log(LDL) might be.

Example: 2-sample test of normal outcomes

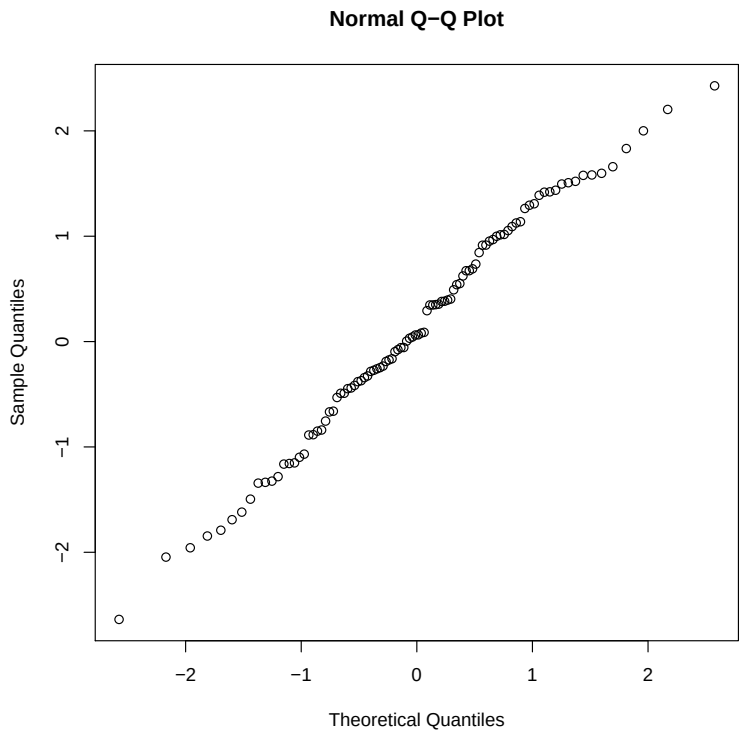
- Suppose you want to show LDL levels are the same in two populations.
- An important assumption of this test is that the data is normally distributed.
- Typically, lab values are not normal, but the log(LDL) might be.
- Suppose you take the log of the LDL: how do you check normality?

Example: 2-sample test of normal outcomes

- Suppose you want to show LDL levels are the same in two populations.
- An important assumption of this test is that the data is normally distributed.
- Typically, lab values are not normal, but the log(LDL) might be.
- Suppose you take the log of the LDL: how do you check normality?
- Many methods, a good way is to plot the data using a "QQ-plot'.

Assumptions: QQ-plot

- A QQ-plot plots the quantiles from the normal distribution (x-axis) against the quantiles from your data (y-axis) and see if they agree (ie. form a straight 45^o line)



- If not normal, may need different test (non-parametric)

2-sample t-test

- Null: Means are the same, $H_0 : \mu_1 = \mu_2$

2-sample t-test

- Null: Means are the same, $H_0 : \mu_1 = \mu_2$
- Alternative: Means are different, $H_1 : \mu_1 \neq \mu_2$

2-sample t-test

- Null: Means are the same, $H_0 : \mu_1 = \mu_2$
- Alternative: Means are different, $H_1 : \mu_1 \neq \mu_2$
- Parameter: True value of difference, $\mu_1 - \mu_2$

2-sample t-test

- Null: Means are the same, $H_0 : \mu_1 = \mu_2$
- Alternative: Means are different, $H_1 : \mu_1 \neq \mu_2$
- Parameter: True value of difference, $\mu_1 - \mu_2$
- Estimate: Difference of estimates in each sample, mean of group1-mean for group 2, $\bar{x}_1 - \bar{x}_2$.

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:
  - $\Delta = |\mu_1 - \mu_2| =$ the absolute difference between your projected (true) mean values, if alternative hypothesis is true.

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:
  - $\Delta = |\mu_1 - \mu_2| =$ the absolute difference between your projected (true) mean values, if alternative hypothesis is true.
  - $\mu_1, \mu_2 =$ the true (projected) means for group 1 and 2, under alternative.

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:
  - $\Delta = |\mu_1 - \mu_2| =$ the absolute difference between your projected (true) mean values, if alternative hypothesis is true.
  - $\mu_1, \mu_2 =$ the true (projected) means for group 1 and 2, under alternative.
  - $\sigma_1, \sigma_2 =$ the true (projected) standard deviations for group 1 and 2.

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:
  - $\Delta = |\mu_1 - \mu_2|$ = the absolute difference between your projected (true) mean values, if alternative hypothesis is true.
  - μ_1, μ_2 = the true (projected) means for group 1 and 2, under alternative.
  - σ_1, σ_2 = the true (projected) standard deviations for group 1 and 2.
  - $z_{1-\alpha/2}$ = the $1 - \alpha/2$-th normal percentile.

Formula for sample size

- $n = \frac{(\sigma_1^2+\sigma_2^2)(z_{1-\alpha/2}+z_{1-\beta})^2}{\Delta^2}$ for each group, where:
  - $\Delta = |\mu_1 - \mu_2|$ = the absolute difference between your projected (true) mean values, if alternative hypothesis is true.
  - μ_1, μ_2 = the true (projected) means for group 1 and 2, under alternative.
  - σ_1, σ_2 = the true (projected) standard deviations for group 1 and 2.
  - $z_{1-\alpha/2}$ = the $1-\alpha/2$-th normal percentile.
  - $z_{1-\beta}$ = the $1-\beta$-th normal percentile (z-score associated with power).

Example 3

Sample size required for 90% power $1-\beta$, $\alpha = 0.05$, $\Delta = \log(2)$,

$$\begin{aligned} n &= \frac{(\sigma_1^2 + \sigma_2^2)(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2} \\ &= \frac{(2^2 + 2^2)(z_{0.975} + z_{0.9})^2}{\log(2)^2} \\ &= \frac{(8)(1.96 + 1.28)^2}{\log(2)^2} \\ &= 174.8 \\ &\rightarrow 175 \text{ items in each sample.} \end{aligned}$$

Assumptions

- Normality

Assumptions

- Normality
- Large enough sample (to be sure that it comes from a normal distribution)

Assumptions

- Normality
- Large enough sample (to be sure that it comes from a normal distribution)
- Other assumptions: Independence, same variance in each sample or not (different test mechanisms).

Section 5

Conclusion/Comments

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.
    - RD, RR, survival analysis, logistic regression, ...

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.
  - RD, RR, survival analysis, logistic regression, ...
- All statistical packages do power/sample size.

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.
  - RD, RR, survival analysis, logistic regression, ...
- All statistical packages do power/sample size.
- Also, lots of good online calculators for sample size/power (make sure from a reputable source!, e.g., biostat/epi departments)

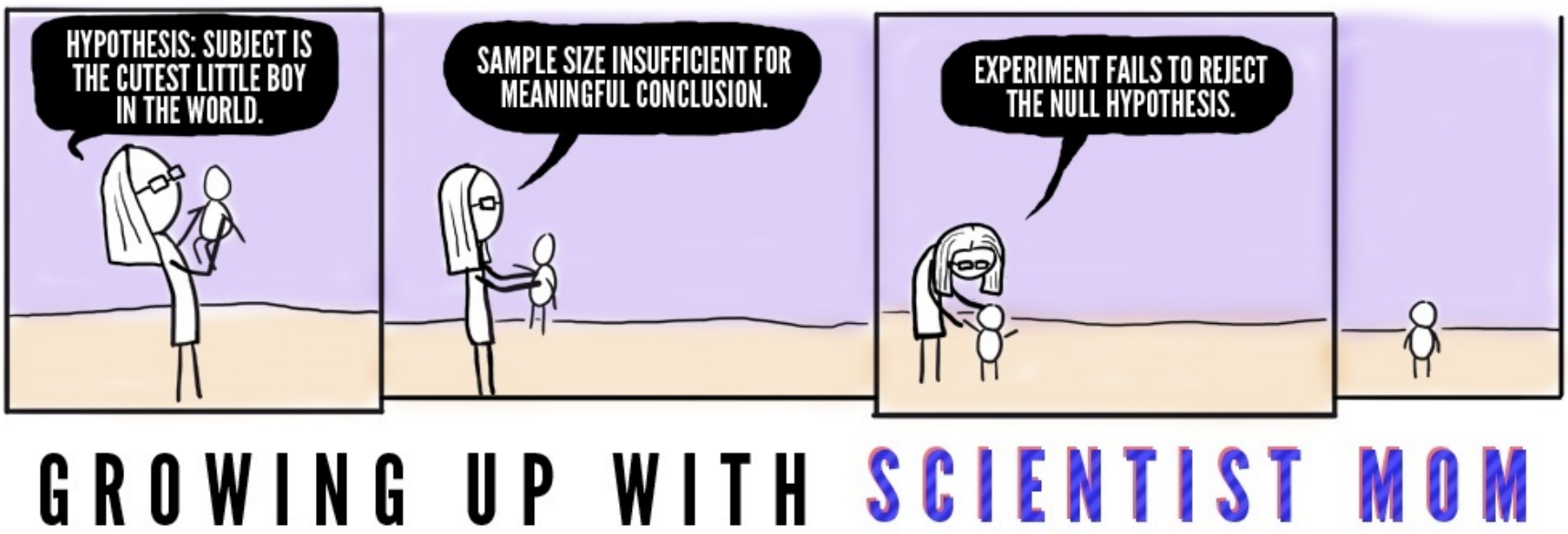
Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.
  - RD, RR, survival analysis, logistic regression, ...
- All statistical packages do power/sample size.
- Also, lots of good online calculators for sample size/power (make sure from a reputable source!, e.g., biostat/epi departments)
- Plan ahead. It is never too early to talk to your statistician/epidemiologist!

Concluding comments

- If the range of values (confidence interval) is really what you are interested in, there are sample size formulas for that.
- Many different analyses possible, all with different sample size formulas.
  - RD, RR, survival analysis, logistic regression, ...
- All statistical packages do power/sample size.
- Also, lots of good online calculators for sample size/power (make sure from a reputable source!, e.g., biostat/epi departments)
- Plan ahead. It is never too early to talk to your statistician/epidemiologist!
- "To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of." – Sir Ronald Fisher

Cartoon


HYPOTHESIS: SUBJECT IS THE CUTEST LITTLE BOY IN THE WORLD.
SAMPLE SIZE INSUFFICIENT FOR MEANINGFUL CONCLUSION.
EXPERIMENT FAILS TO REJECT THE NULL HYPOTHESIS.
GROWING UP WITH SCIENTIST MOM