EPE - Lecture 4
Power Analysis: Choosing Sample Size and Gauging Sampling Noise Ex Ante

Sylvain Chabé-Ferret

Toulouse School of Economics, Inra

January 2017

In a nutshell

In this lecture, we are going to study how to choose the size of our sample (mainly in a RCT) and how to gauge the size of sampling noise before conducting a study.

Two Conceptually Distinct Approaches

- Using test statistics (power study per se)
- Gauging sampling noise or choosing sample size to reach a given sampling noise

Outline

The Traditional Approach to Power Analysis Using Test Statistics

Power Analysis for RCT Designs

An Alternative Approach: Gauging Sampling Noise

Exercises

What is Power?

Definition (Power)

Power κ is the probability of rejecting the null assumption of a negative or null (resp. null) average treatment effect when the true effect is of at least β_A applying a test of size α to an estimator $\hat{E}$ with a sample of size N. β_A is called the Minimum Detectable Effect (MDE).

One-Sided Test: H_0: $E \leq 0$ *vs* H_A: $E = \beta_A > 0$

Two-Sided Test: H_0: $E = 0$ *vs* H_A: $E = \beta_A \neq 0$

Assumption: Asymptotically Normal Estimator

Assumption (Asymptotically Normal Estimator)

We assume that the estimator $\hat{E}$ is such that there exists a constant (independent of N) $C(\hat{E})$ such that:

$$\lim_{N\to\infty} \Pr\left(\frac{\hat{E}-E}{\sqrt{\mathbb{V}[\hat{E}]}} \leq u\right) = \Phi\left(u\right),$$

with $\mathbb{V}[\hat{E}] = \frac{C(\hat{E})}{N}$.

Power Formula

Theorem (Power with a Normal Estimator)

With $\hat{E}$ an asymptotically normal estimator, we have:

One-Sided Test: H_0: $E \leq 0$ *vs* H_A: $E = \beta_A > 0$

$$\kappa \approx \Phi\left(\frac{\beta_A}{\sqrt{\mathbb{V}[\hat{E}]}} - \Phi^{-1}(1-\alpha)\right),$$

Two-Sided Test: H_0: $E = 0$ *vs* H_A: $E = \beta_A \neq 0$

$$\kappa \approx \Phi\left(\frac{\beta_A}{\sqrt{\mathbb{V}[\hat{E}]}} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right).$$

Proof

Let's start with a one-sided test first. We want to build a test statistic t such that, under H_0, $\Pr(\hat{E} \geq t) = \alpha$. Under H_0 and using asymptotic normality, we have:

$$\Pr(\hat{E} \geq t) = \Pr\left(\frac{\hat{E} - 0}{\sqrt{\mathbb{V}[\hat{E}]}} \geq \frac{t - 0}{\sqrt{\mathbb{V}[\hat{E}]}}\right) \approx 1 - \Phi\left(\frac{t}{\sqrt{\mathbb{V}[\hat{E}]}}\right)$$

As a consequence of $\Pr(\hat{E} \geq t) = \alpha$, we have $t \approx \Phi^{-1}(1-\alpha)\sqrt{\mathbb{V}[\hat{E}]}$. Power is $\Pr(\hat{E} \geq t)$ under H_A. Using asymptotic normality again:

$$\Pr(\hat{E} \geq t) = \Pr\left(\frac{\hat{E} - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}} \geq \frac{t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right) \approx 1 - \Phi\left(\frac{t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right) = \Phi\left(\frac{\beta_A - t}{\sqrt{\mathbb{V}[\hat{E}]}}\right),$$

which proves the first part of the result.

Proof (cont'd)

With a two-sided test, we want a test statistic t such that, under H_0, $\Pr(\hat{E} \leq -t \vee \hat{E} \geq t) = \alpha$. Because the events are disjoint, under H_0 and using asymptotic normality, we have:

$$\begin{aligned}
\Pr(\hat{E} \leq -t \vee \hat{E} \geq t) &= \Pr(\hat{E} \leq -t) + \Pr(\hat{E} \geq t) \\
&= \Pr\left(\frac{\hat{E} - 0}{\sqrt{\mathbb{V}[\hat{E}]}} \leq \frac{-t - 0}{\sqrt{\mathbb{V}[\hat{E}]}}\right) + \Pr\left(\frac{\hat{E} - 0}{\sqrt{\mathbb{V}[\hat{E}]}} \geq \frac{t - 0}{\sqrt{\mathbb{V}[\hat{E}]}}\right) \\
&\approx \Phi\left(-\frac{t}{\sqrt{\mathbb{V}[\hat{E}]}}\right) + 1 - \Phi\left(\frac{t}{\sqrt{\mathbb{V}[\hat{E}]}}\right) \\
&= 2\left(1 - \Phi\left(\frac{t}{\sqrt{\mathbb{V}[\hat{E}]}}\right)\right),
\end{aligned}$$

where the last equality uses the symmetry of the normal distribution. As a consequence of $\Pr(\hat{E} \leq -t \vee \hat{E} \geq t) = \alpha$, we have $t \approx \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\mathbb{V}[\hat{E}]}$. Power is $\Pr(\hat{E} \leq -t \vee \hat{E} \geq t)$ under H_A. Using asymptotic normality and the fact that the two events are disjoint again:

$$\begin{aligned}
\Pr(\hat{E} \leq -t \vee \hat{E} \geq t) &= \Pr(\hat{E} \leq -t) + \Pr(\hat{E} \geq t) \\
&= \Pr\left(\frac{\hat{E} - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}} \leq \frac{-t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right) + \Pr\left(\frac{\hat{E} - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}} \geq \frac{t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right) \\
&\approx \Phi\left(\frac{-t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right) + \Phi\left(\frac{t - \beta_A}{\sqrt{\mathbb{V}[\hat{E}]}}\right).
\end{aligned}$$

When β_A is positive, the first part of the power formula is negligible with respect to the second part. Hence the result.

Power: Illustration

I am going to use the example of a brute force RCT from lecture 3. I am going to assume first that I know the actual distribution of the WW estimator. I am going to use the Monte Carlo simulations to approximate the true distribution of the estimator. What is cool is that if I keep the same name for the code chunk, knitr uses the results of the simulations from lecture 3 and does not run the Monte Carlo again. Pretty awesome.

Power: Illustration

- $\alpha = 0.05$
- $\beta_A = 0.2$
- $\mathbb{V}[\hat{E}] =$0.0030572

Power: Illustration

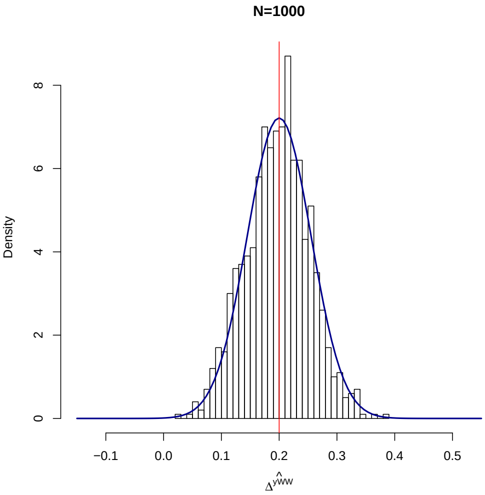

Figure: Power

Power: Illustration

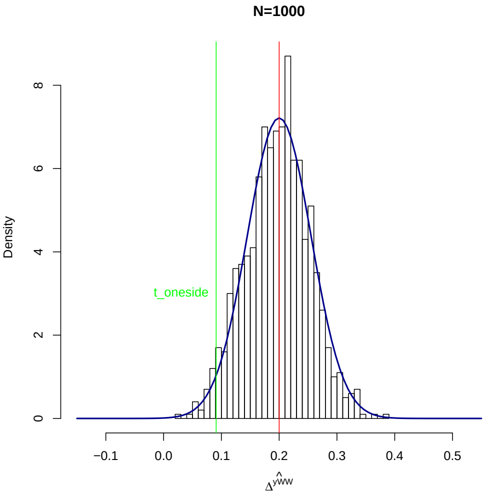

Figure: Power

Power: Illustration

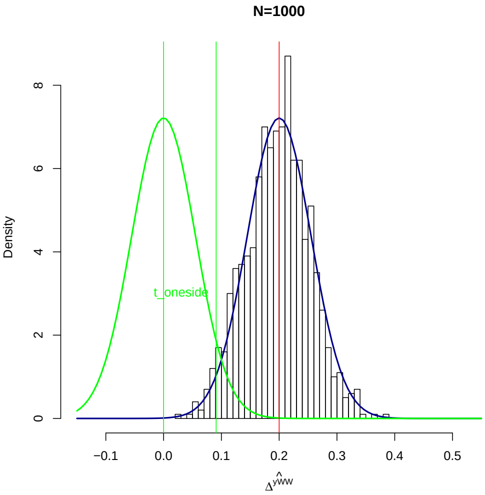

Figure: Power

Power: Illustration

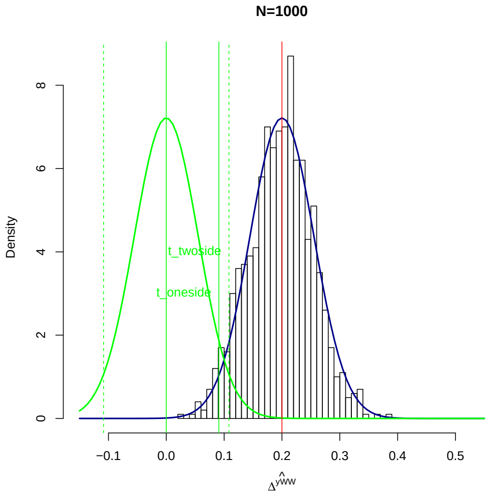

Figure: Power

Power: Illustration

With $\alpha =$0.05, $\beta_A =$0.2 and $\mathbb{V}[\hat{E}] =$0.0030572 (from $N = 1000$)

- One-sided power: 0.9757141
- Two-sided power: 0.9512626
- Approximate two-sided power: 0.9512627

Power: Illustration

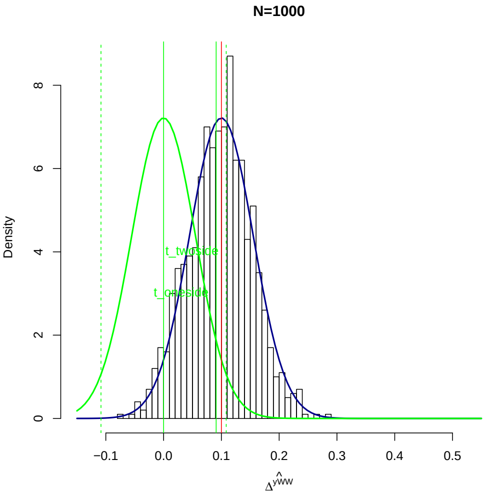

Figure: Power

Power: Illustration

With $\alpha =$0.05, $\beta_A =$0.1 and $\mathbb{V}[\hat{E}] =$0.0030572 (from $N = 1000$)

- One-sided power: 0.5650317
- Two-sided power: 0.4398414
- Approximate two-sided power: 0.4399235

Power: Illustration

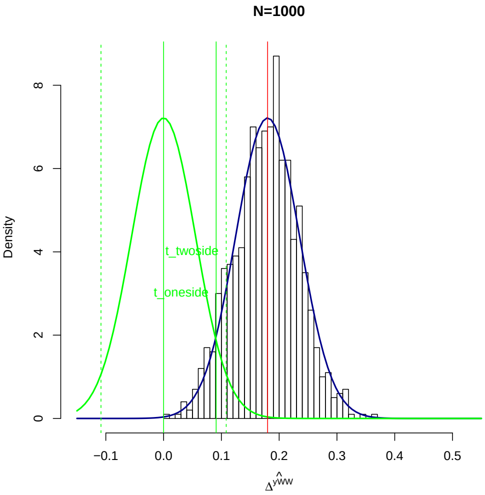

Figure: Power

Power: Illustration

With α =0.05, β_A =0.18 and $\mathbb{V}[\hat{E}]$ =0.0030572 (from $N = 1000$)

- One-sided power: 0.946368
- Two-sided power: 0.9024266
- Approximate two-sided power: 0.9024267

Power: Illustration

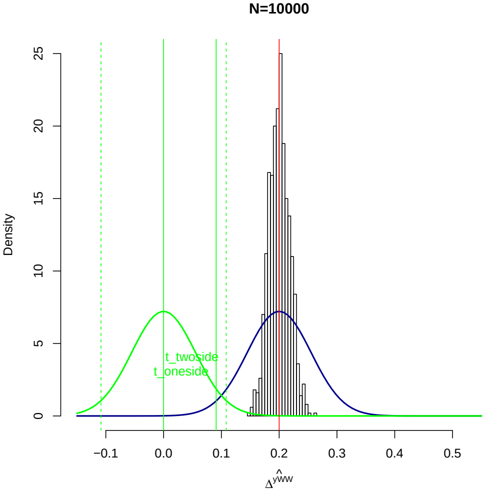

Figure: Power

Power: Illustration

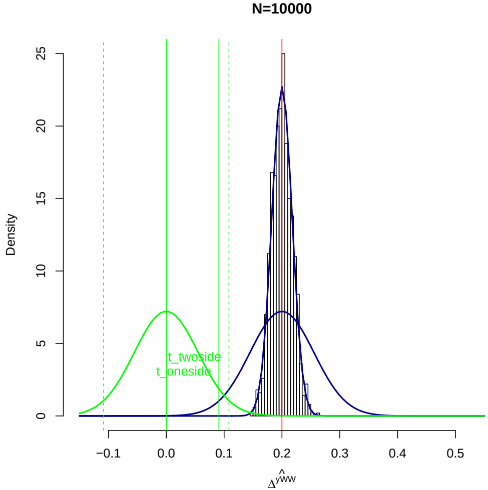

Figure: Power

Power: Illustration

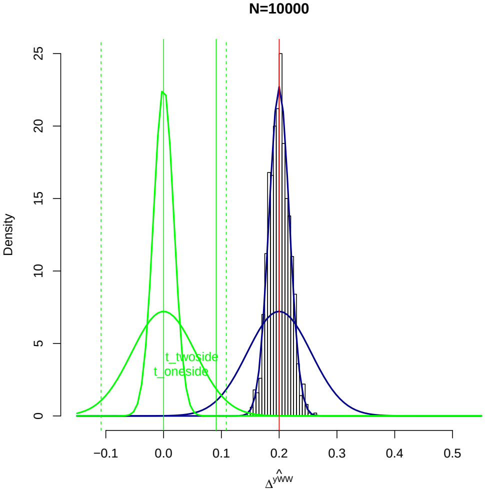

Figure: Power

Power: Illustration

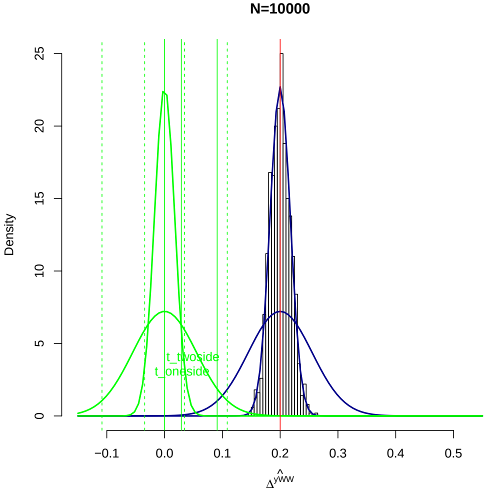

Figure: Power

Power: Illustration

With $\alpha =$0.05, $\beta_A =$0.2 and $\mathbb{V}[\hat{E}] =$3.0857367 $\times 10^{-4}$ (from $N = 10000$)

- One-sided power: 1
- Two-sided power: 1
- Approximate two-sided power: 1

Power: Illustration

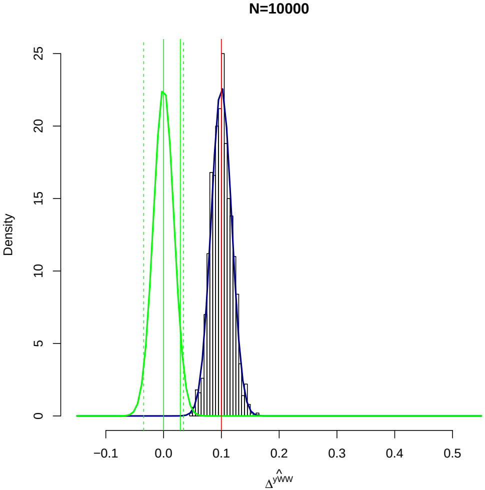

Figure: Power

Power: Illustration

With $\alpha =$0.05, $\beta_A =$0.1 and $\mathbb{V}[\hat{E}] =$3.0857367 $\times 10^{-4}$ (from $N = 10000$)

- One-sided power: 0.9999742
- Two-sided power: 0.9999053
- Approximate two-sided power: 0.9999053

Power: Illustration

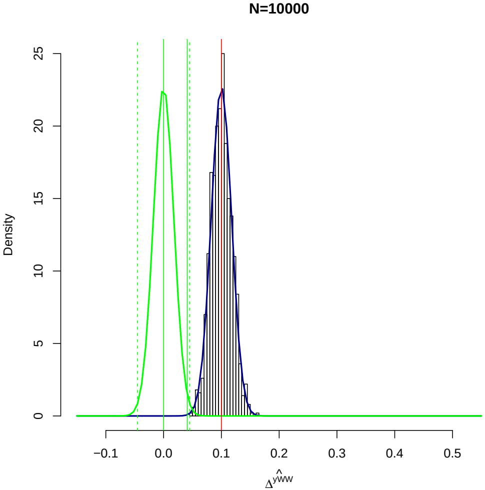

Figure: Power

Power: Illustration

With $\alpha =$0.01, $\beta_A =$0.1 and $\mathbb{V}[\hat{E}] =$3.0857367 $\times 10^{-4}$ (from $N = 10000$)

- One-sided power: 0.9996192
- Two-sided power: 0.9990862
- Approximate two-sided power: 0.9990862

Power and Sample Size: Illustration

With α =0.05 and β_A =0.2 and variances from the Monte Carlo simulations

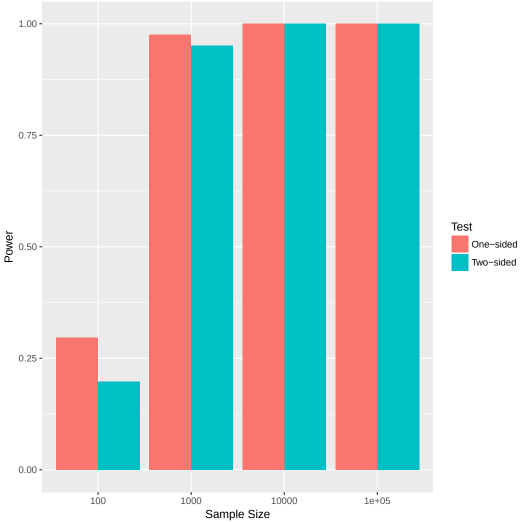

Figure: Power and Sample Size

Minimum Detectable Effect and Sample Size

Theorem (Minimum Detectable Effect and Sample Size)

With an asymptotically normal estimator and a one-sided test, we have:

$$\beta_A \approx \left(\Phi^{-1}(\kappa) + \Phi^{-1}(1-\alpha)\right)\sqrt{\mathbb{V}[\hat{E}]}$$

$$N \approx \left(\Phi^{-1}(\kappa) + \Phi^{-1}(1-\alpha)\right)^2 \frac{C(\hat{E})}{\beta_A^2},$$

with $\mathbb{V}[\hat{E}] = \frac{C(\hat{E})}{N}$.

- For the two-sided result, just replace α by $\alpha/2$.
- The proof is straightforward using the power formula.

MDE: Illustration

With α =0.05 and κ =0.8 and variances from the Monte Carlo simulations

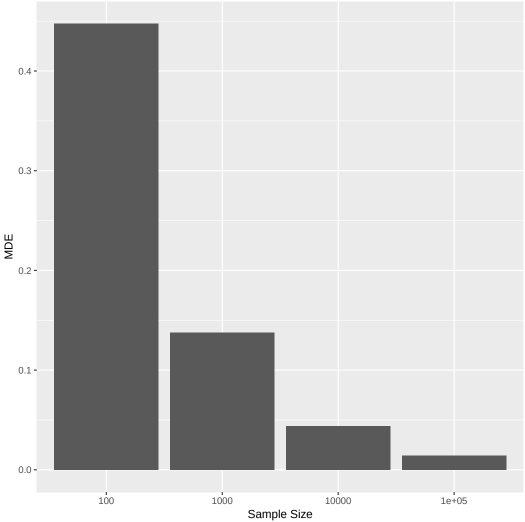

Figure: MDE and Sample Size

Outline

The Traditional Approach to Power Analysis Using Test Statistics

Power Analysis for RCT Designs

An Alternative Approach: Gauging Sampling Noise

Exercises

How Do We Use the Formulas In Practice?

- Problem, ex-ante we do not know $\mathbb{V}[\hat{E}]$ nor $C(\hat{E})$
- We have to come with an idea about the size of $C(\hat{E})$ based on pre-treatment data
- In general, we use observations of the same outcomes from a survey or census conducted prior to our experiment
- There are two ways to use these observations:
    1. We have closed-form formulas for $C(\hat{E})$ based on features of the distribution of the covariates
    2. We do not have these formulas, but we can implement the estimator for the variance on pre-treatment data

Using Closed Form Formulas for $C(\hat{E})$

Example of the Brute Force Design with the With/Without estimator:

$$C(\Delta^{\hat{Y}}_{WW}) = \frac{\mathbb{V}[Y_i^1|R_i=1]}{\Pr(R_i=1)} + \frac{\mathbb{V}[Y_i^0|R_i=0]}{1-\Pr(R_i=1)}.$$

- We do not know $\mathbb{V}[Y_i^1|\hat{R}_i=1]=$0.9359936 nor $\mathbb{V}[Y_i^0|\hat{R}_i=0]=$0.6933382 because these outcomes are not observed yet
- Let's approximate them with $\mathbb{V}[\hat{Y}_i^B]=$0.8083786
- $\Pr(R_i=1)$ we choose by design (let's say 0.5)

With $\alpha=$0.05, $\kappa=$0.8 and $\beta_A=$0.2, we have $N=$500
With $\alpha=$0.05, $\kappa=$0.9 and $\beta_A=$0.2, we have $N=$692

MDE and Sample Size: Illustration

With α =0.05 and κ =0.8

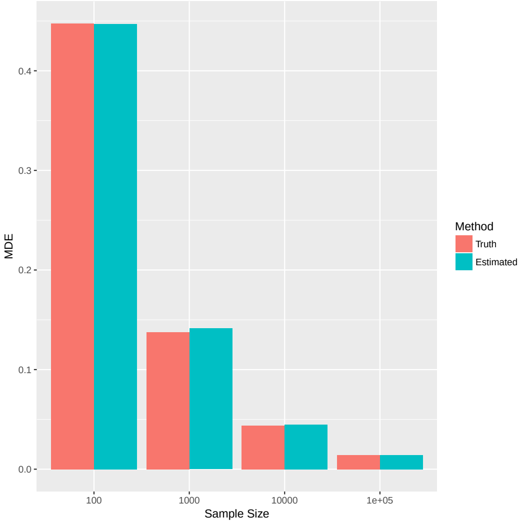

Figure: MDE and Sample Size

Problems with Closed Form Formulas

1. We do not have closed form formulas for all the estimators seen in class
2. Sometimes, components of the formulas are pretty cumbersome to compute

A Simpler Approach: Directly Implement the Estimators on Pre-Treatment Data

1. Resample to choose an adequate sample size
2. Distribute a placebo treatment
3. Estimate the treatment effect (should be zero) and its standard error: this is an estimate of $\sqrt{\mathbb{V}[\hat{E}]}$
4. Compute the MDE
5. Repeat for as many sample sizes as you want to consider

```
N.simul <- 10000
set.seed <- 1234
resample <- sample(N,N.simul,replace=TRUE)
y.simul <- y[resample]
# randomized allocation of 50% of individuals
Rs.simul <- runif(N.simul)
R.simul <- ifelse(Rs.simul<=.5,1,0)
reg.ols.ww.bf.simul <- lm(y.simul~R.simul)
varE.simul <- vcovHC(reg.ols.ww.bf.simul,type='HC2')[2,2]
MDE.simul <- MDE.var(alpha=alpha,kappa=kappa,varE=varE.simu
```

MDE and Sample Size: Illustration

With $\alpha=$0.05 and $\kappa=$0.8, $N = 10^4$, we estimate $\mathbb{V}[\hat{E}]=$3.3443981 $\times 10^{-4}$ and thus $MDE=$0.0454719.

MDE and Sample Size: Illustration

With α =0.05 and κ =0.8

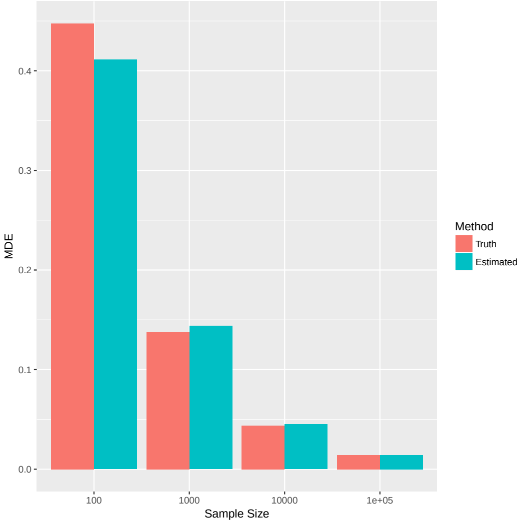

Figure: MDE and Sample Size

Advantages of the Alternative Approach

1. Very intuitive: simulate ex ante what is going to happen ex post
2. Accomodate various designs and estimators easily
3. Accomodate stratification and conditioning on covariates
4. When using covariates, imposes that you list them ex ante (very good to prevent specification search, if you make the list of covariates public ex ante)

MDE and Sample Size Conditioning on X: Illustration

With α =0.05 and κ =0.8, Brute Force Design

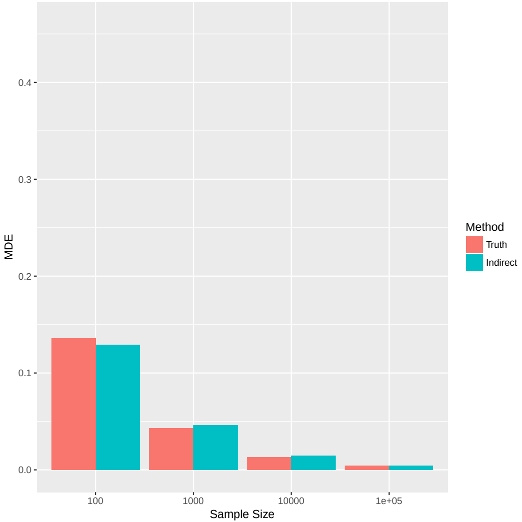

Figure: MDE and Sample Size Conditioning on Covariates

Randomization After Self-Selection

There are additional things we do not know

1. How many people will be eligible for the treatment
2. How many people will apply for the offered treatment among the eligibles
3. How different will applicants be from the eligible population

Randomization After Self-Selection

We have to make assumptions:

1. Try to compute eligibility using pre-treatment data if feasible
2. Guess a proportion of applicants
3. Assume that applicants are the same as the overall eligible population

```
N.simul <- 10000
set.seed <- 1234
resample <- sample(N,N.simul,replace=TRUE)
y.simul <- y[resample]
E.simul <- ifelse(y.simul<=log(param["barY"]),1,0)
Dstar.simul <- runif(N.simul)
Ds.simul <- ifelse(Dstar.simul<=.5 & E.simul==1,1,0)
#random allocation among "self-selected"
Rs.simul <- runif(N.simul)
R.simul <- ifelse(Rs.simul<=.5 & Ds.simul==1,1,0)
reg.ols.ww.rass.simul <- lm(y.simul[Ds.simul==1]~R.simul[Ds
varE.simul <- vcovHC(reg.ols.ww.rass.simul,type='HC2')[2,2]
MDE.simul <- MDE.var(alpha=alpha,kappa=kappa,varE=varE.simu
```

MDE and Sample Size in RCT: Illustration

With $\alpha = 0.05$ and $\kappa = 0.8$, Randomization After Self-Selection

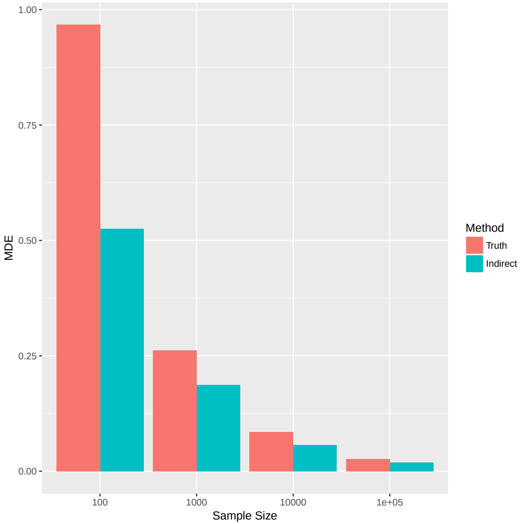

Figure: MDE and Sample Size

Randomization After Eligibility

We have to make the same assumptions as before:

1. Try to compute eligibility using pre-treatment data if feasible
2. Guess a proportion of applicants
3. Assume that applicants are the same as the overall eligible population

```r
N.simul <- 10000
set.seed <- 1234
resample <- sample(N,N.simul,replace=TRUE)
y.simul <- y[resample]
E.simul <- ifelse(y.simul<=log(param["barY"]),1,0)
#random allocation among eligibles
Rs.simul <- runif(N.simul)
R.simul <- ifelse(Rs.simul<=.5 & E.simul==1,1,0)
Dstar.simul <- runif(N.simul)
Ds.simul <- ifelse(Dstar.simul<=.5 & E.simul==1 & R.simul==

reg.2sls.ww.elig.simul <- ivreg(y.simul[E.simul==1]~Ds.simu
varE.simul <- vcovHC(reg.2sls.ww.elig.simul,type='HC2')[2,2
MDE.simul <- MDE.var(alpha=alpha,kappa=kappa,varE=varE.simu
```

MDE and Sample Size in RCT: Illustration

With α =0.05 and κ =0.8, Randomization After Eligibility

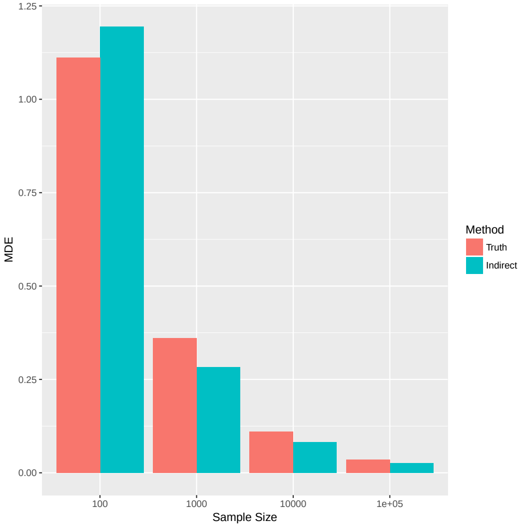

Figure: MDE and Sample Size

Encouragement Design

We have to make the same assumptions as before plus a new one:

1. Try to compute eligibility using pre-treatment data if feasible
2. Guess a proportion of applicants with and without encouragement
3. Assume that applicants are the same as the overall eligible population

```
N.simul <- 10000
set.seed <- 1234
resample <- sample(N,N.simul,replace=TRUE)
y.simul <- y[resample]
E.simul <- ifelse(y.simul<=log(param["barY"]),1,0)
#random allocation among eligibles
Rs.simul <- runif(N.simul)
R.simul <- ifelse(Rs.simul<=.5 & E.simul==1 | Rs.simul<=.75
Dstar.simul <- runif(N.simul)
Ds.simul <- ifelse(Dstar.simul<=.5 & E.simul==1 & R.simul==

reg.2sls.ww.encourage.simul <- ivreg(y.simul[E.simul==1]~Ds
varE.simul <- vcovHC(reg.2sls.ww.encourage.simul,type='HC2
MDE.simul <- MDE.var(alpha=alpha,kappa=kappa,varE=varE.simu
```

MDE and Sample Size in RCT: Illustration

With α =0.05 and κ =0.8, Encouragement Design

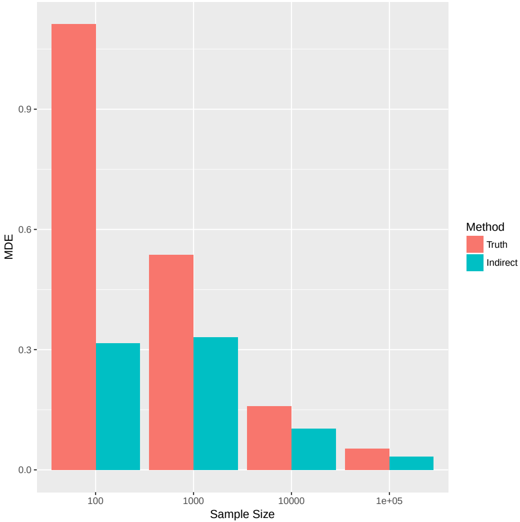

Figure: MDE and Sample Size

Outline

The Traditional Approach to Power Analysis Using Test Statistics

Power Analysis for RCT Designs

An Alternative Approach: Gauging Sampling Noise

Exercises

Pitfalls of the Traditional Approach to Power Analysis

1. We should NOT use tests statistics (cf Lecture 0) but estimate sampling noise
2. It is extremely hard to make sense of the test statistic approach: there are no repeated samples in real life
3. There are A LOT of parameters values to choose: α, κ, β_A
4. Power analysis with usual parameter values ($\alpha = 0.05$, $\kappa = 0.8$) has signal to noise ratio well below one

Signal to Noise Ratio in the Traditional Approach

$$
\begin{aligned}
\frac{\beta_A}{2\epsilon} &\approx \frac{\left(\Phi^{-1}\left(\kappa\right) + \Phi^{-1}\left(1-\alpha\right)\right)\sqrt{\mathbb{V}[\hat{E}]}}{2\Phi^{-1}\left(\frac{\delta+1}{2}\right)\sqrt{\mathbb{V}[\hat{E}]}} \\
&= \frac{\left(\Phi^{-1}\left(\kappa\right) + \Phi^{-1}\left(1-\alpha\right)\right)}{2\Phi^{-1}\left(\frac{\delta+1}{2}\right)}
\end{aligned}
$$

- With $\alpha =$0.05, $\kappa =$0.8, $\delta =$0.95 and one-sided test, the signal to noise ratio is 0.6343165
- With $\alpha =$0.05, $\kappa =$0.8, $\delta =$0.95 and two-sided test, the signal to noise ratio is 0.7147032
- With $\alpha =$0.05, $\kappa =$0.8, $\delta = 0.99$ and two-sided test, the signal to noise ratio is 0.543822
- With $\alpha = 0.01$, $\kappa =$0.8, $\delta = 0.99$ and two-sided test, the signal to noise ratio is 0.663369
- With $\alpha = 0.01$, $\kappa = 0.95$, $\delta = 0.99$ and two-sided test, the signal to noise ratio is 0.8192862

The Proposed Approach

1. Pre-specify the size of sampling noise you are willing to tolerate
2. Implement power formulas to compute sampling noise for various sample sizes
3. Choose sample size accordingly

Advantages of Using Sampling Noise for Power analysis

1. You do not use test statistics
2. It is extremely easy to make sense of sampling noise: it is the amount of variation in the treatment effect estimate that comes from sampling variability
3. There is only ONE parameter value to choose: sampling noise (actually, you also get to choose δ, but its precise value does not make much of a difference on the order of magnitude of the sample size).
4. You do NOT pre-specify the signal to noise ratio (neither do you have to pre-specify a MDE, but if you do, then you can assess the signal to noise ratio ex ante)

Overall, power analysis is simply a way to try to gauge sampling noise ex ante, and it should be kept as simple and direct as that.

Sampling Noise and Sample Size

Theorem (Sampling Noise and Sample Size)

With an asymptotically normal estimator, we have:

$$2\epsilon \approx 2\Phi^{-1}\left(\frac{\delta+1}{2}\right)\sqrt{\mathbb{V}[\hat{E}]}$$
$$N \approx 4\Phi^{-1}\left(\frac{\delta+1}{2}\right)^2 \frac{C(\hat{E})}{(2\epsilon)^2}.$$

The proof is straightforward using the result from Lecture 0.

Sampling Noise and Sample Size In Practice

1. Use closed form formula for $C(\hat{E})$
2. Apply directly estimator to pre-treatment data with pre-selected sample size to recover $\mathbb{V}[\hat{E}]$ and deduce sampling noise

Sampling Noise and Sample Size: Illustration

With δ =0.95 and Brute Force Design

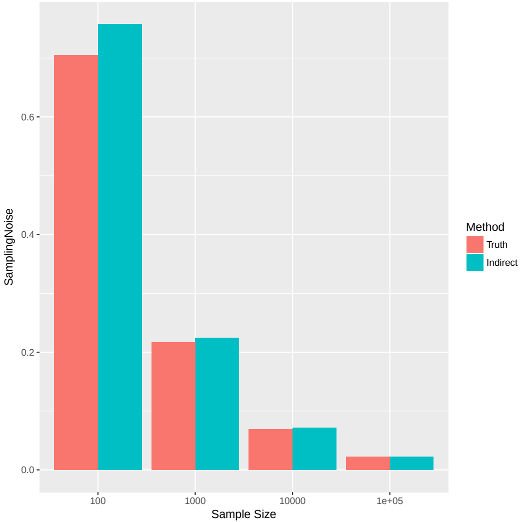

Figure: Sampling Noise and Sample Size

Sampling Noise and Sample Size: Illustration

With $\delta = 0.99$ and Brute Force Design

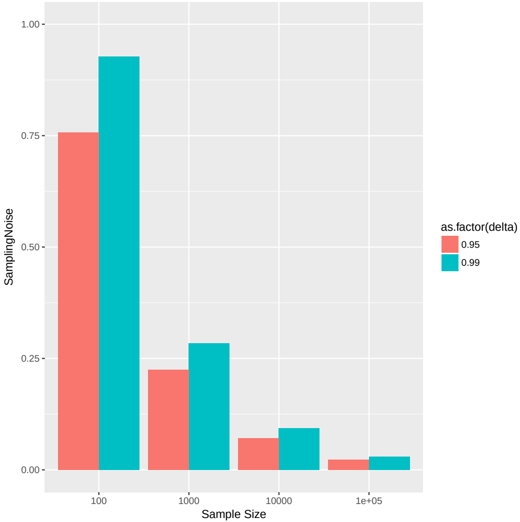

Figure: Sampling Noise and Sample Size

Outline

The Traditional Approach to Power Analysis Using Test Statistics

Power Analysis for RCT Designs

An Alternative Approach: Gauging Sampling Noise

Exercises

Exercises on generated data

1. Using the pre-treatment observations of the generated data, compute the required sample size for MDEs of 0.2 and 0.1 for power 0.8 and size 0.05 for a one and two-sided test for the brute force RCT using the CLT based formulae.
2. How does the sample size change when power is 0.9 for the same MDEs?
3. Do the same thing with the resampling approach suggested in the lecture notes.
4. Compute the required sample size for the same MDEs for the other RCT designs using the resampling method.
5. Advanced: derive the CLT-based formulae for TT (resp. LATE) for the Bloom (and Wald) estimator and implement the formulae to compute sample size for the same MDEs.

Exercises on the RCT data of your choice

1. Use the data from the RCT of your choice to compute the sample size required for MDEs of 0.1σ and 0.2σ for TT or LATE, where σ is the standard deviation of outcomes in the control group. (Treatment effects expressed in units of standard deviation are called effect sizes.) Use the resampling method (and the closed form formulae if you have done the advanced question). If you do not have access to pre-treatment outcomes, use the control group as a source of information. Use $\kappa =$0.8 and 0.9 and $\alpha = 0.05$ and consider one and two-sided tests.
2. For the same MDEs, compute the sample size required to reach a signal to noise ratio of 0.5, 1 and 2. How different are these results from the previous ones?
3. Redo all power calculations with the actual estimate of $C(\hat{E})$ in the sample. How does that change things?