All Power Analyses are Wrong, How to be More Right

Christopher David Desjardins

Research Methodology Consulting Center

5/1/19

IES RFA - Goal 1

1. Exploration (Goal One)

a) Purpose

The Exploration goal supports projects that will identify malleable factors associated with student education outcomes and/or the factors and conditions that mediate or moderate that relationship. Exploration projects are intended to build and inform theoretical foundations to support future applied research efforts such as (1) the development of interventions or the evaluation of these interventions or (2) the development and validation of assessments.

Recommendations for a Strong Application: In order to address the above requirements, the Institute recommends that you include the following in your Research Plan section to strengthen the methodological rigor of the proposed exploratory work.

- For all quantitative inferential analyses, demonstrate that the sample provides sufficient power to address your research aims.

IES RFA - Goal 3

3. Efficacy and Follow-up (Goal Three)

b. Research Plan – The purpose of this section is to describe the evaluation of the intervention.

Requirements: In order to be responsive and sent forward for scientific peer review, **all** applications under the Efficacy and Follow-up goal **must** describe

(ii) The power analysis; and

IES RFA - Goal 4

4. Replication: Efficacy and Effectiveness (Goal Four)

b. Research Plan – The purpose of this section is to describe the evaluation of the intervention.

Requirements: In order to be responsive and sent forward for scientific peer review, **all** applications under Goal Four **must** describe:

(i) The research design;

(ii) The power analysis; and

Motivating research question

Do low-income high school students receiving *TutorNow* have higher composite ACT scores than low-income high school students not receiving *TutorNow*?

Possible outcomes

1. Students in *TutorNow* have higher ACT composite scores than students not in *TutorNow*.
2. Students in *TutorNow* have lower ACT composite scores than students not in *TutorNow* (paradoxical effect).
3. There are no differences in ACT composite scores between students in *TutorNow* and students not in *TutorNow*.

How many students do we need to recruit?

What (partially) impacts this?

- Anticipated size of effect

What (partially) impacts this?

- Anticipated size of effect
- Comfort with making a Type I or Type II error

What (partially) impacts this?

- Anticipated size of effect
- Comfort with making a Type I or Type II error
- Statistical test/model

What (partially) impacts this?

- Anticipated size of effect
- Comfort with making a Type I or Type II error
- Statistical test/model
- Measurement error

What (partially) impacts this?

- Anticipated size of effect
- Comfort with making a Type I or Type II error
- Statistical test/model
- Measurement error
- Attrition

Which effect size?

In our *TutorNow* example, there are lots of potential effect sizes one could consider.

1. Cohen's d
2. Point-biserial correlation
3. R^2
4. Unstandardized regression coefficient

Cohen's d

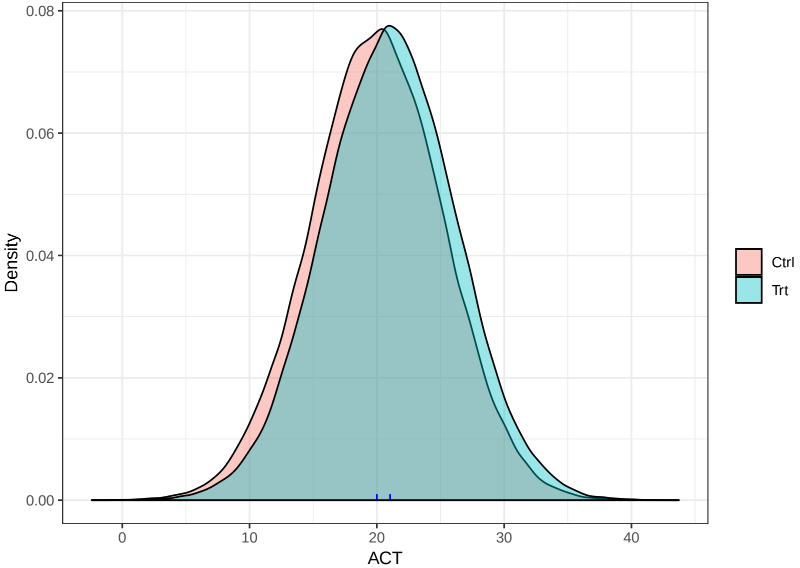
Given that we have two independent samples, most likely consider Cohen's d.

$$d = \frac{\bar{X}_{Trt} - \bar{X}_{Ctrl}}{s}$$

Benefits:
Units are in standard deviations
Criteria for small (0.2), medium (0.5), and large (0.8) effects

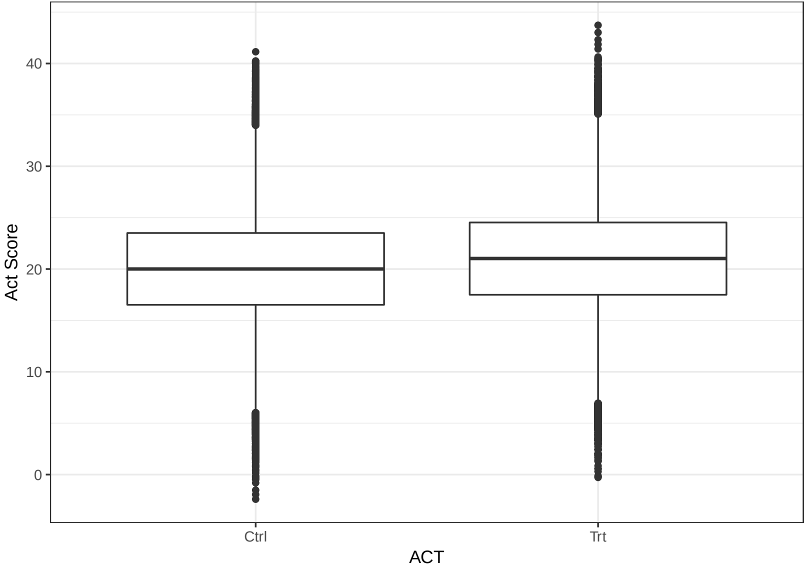
Drawbacks:
Units are in standard deviations
Criteria for small (0.2), medium (0.5), and large (0.8) effects

What might a Cohen's d of 0.2 look like for the ACT?



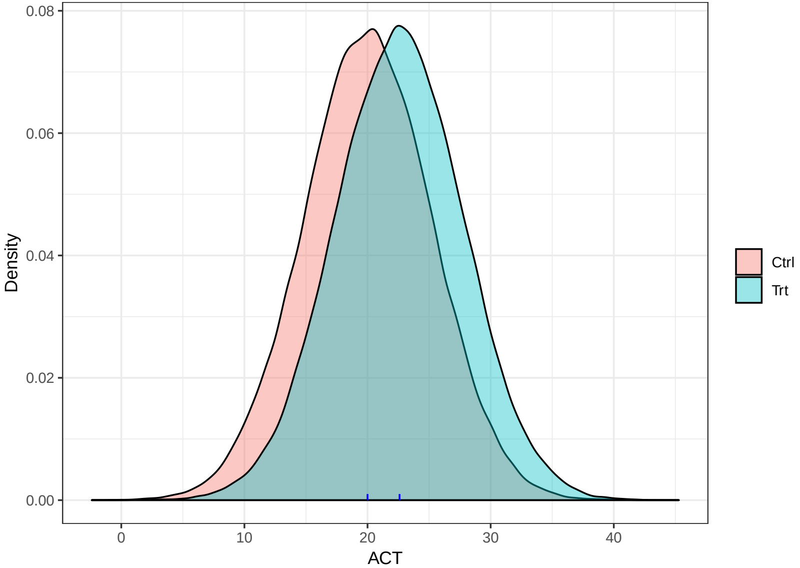
Assuming a standard deviation of 5.2.

What might a Cohen's d of 0.2 look like for the ACT?



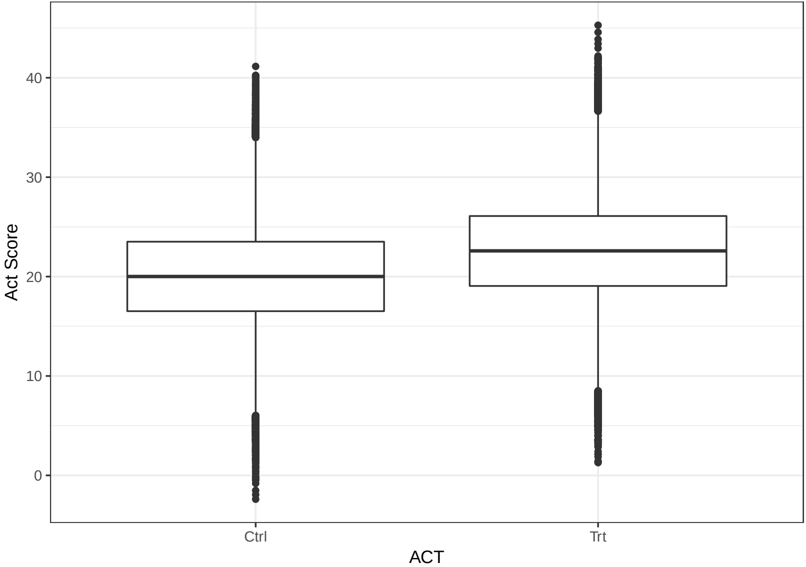
Assuming a standard deviation of 5.2.

What might a Cohen's d of 0.5 look like for the ACT?



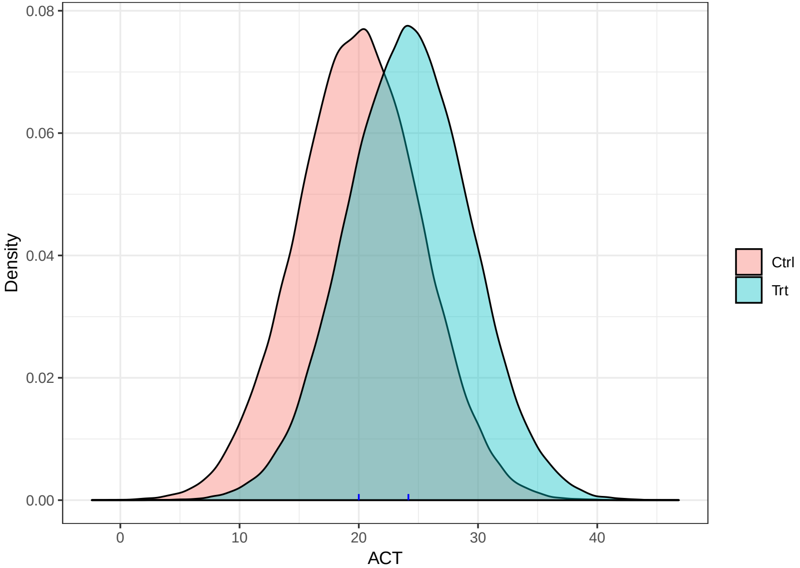
Assuming a standard deviation of 5.2.

What might a Cohen's d of 0.5 look like for the ACT?



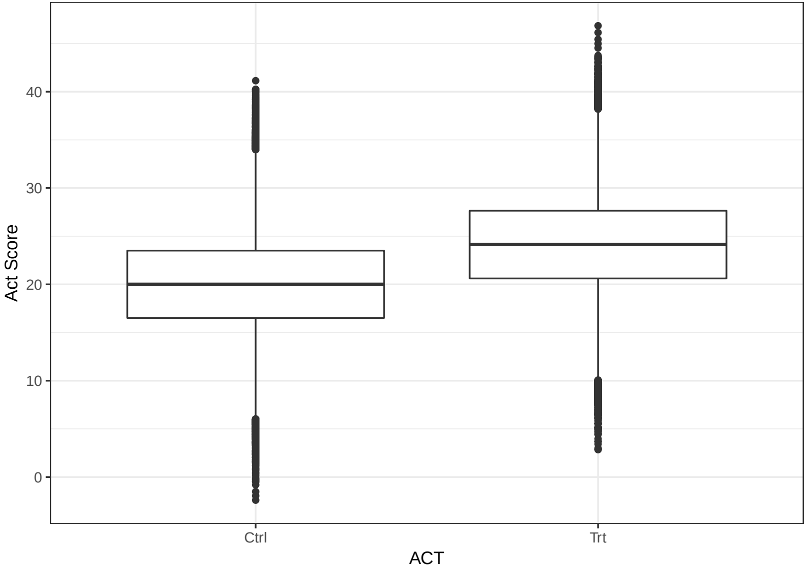
Assuming a standard deviation of 5.2.

What might an a Cohen's d of 0.8 look like for the ACT?



Assuming a standard deviation of 5.2.

What does a Cohen's d of 0.8 look like for the ACT?



Assuming a standard deviation of 5.2.

Interpreting these effect sizes

Assuming a standard deviation of 5.2

- Cohen's d of 0.2 would correspond to a difference in means of 1.04.
- Cohen's d of 0.5 would correspond to a difference in means of 2.6.
- Cohen's d of 0.8 would correspond to a difference in means of 4.16.

The standard error of measurement is 0.93 (according to the ACT Technical Manual), half of the width of a 95% confidence interval would be:

$$(1.96 * 0.93) = 1.8228$$

This "half" width is larger than the mean difference that resulted in a Cohen's d of 0.2, but not a Cohen's d of 0.5. (Apples to Oranges)

A better approach - MMES

- What is a minimum meaningful effect size (MMES)?
  - A theoretically-driven effect size that if detected in the study is clinically/educationally/economically meaningful. Anything smaller, is trivial.
  - IES calls this **practical significance**.
  - Differs from minimum detectable effect size (MDES) because MDES is a statistical artifact.
- How to work with the MMES?
  1. Consult literature and determine effect sizes reported.
  2. Discuss with practitioners what is a meaningful effect size for your application.
  3. Articulate why this is meaningful to colleagues.
  4. Articulate it in a meaningful and understandable unit.
  5. Approach RMCC folks with a MMES and ask them to run a power analysis.

Statistical Inference

Compare the students receiving TutorNow with students not receiving TutorNow using a two-sample t-test:

$$t = \frac{\bar{X}_{Trt} - \bar{X}_{Ctrl}}{\sqrt{\frac{s_{Trt}^2 + s_{Ctrl}^2}{2}}\sqrt{\frac{2}{N}}}$$

Assumes that

1. Samples are from normally distributed populations
2. Sample sizes are equal
3. Groups have the same variance

Hypothesis testing

We're principally interested in testing the difference between the means

$$H_0\colon \mu_{Trt} - \mu_{Ctrl} = 0$$
$$H_1\colon \mu_{Trt} - \mu_{Ctrl} \neq 0$$

But aren't we really interested in?

$$H_0\colon \mu_{Trt} - \mu_{Ctrl} = MMES$$
$$H_1\colon \mu_{Trt} - \mu_{Ctrl} > MMES$$

There's no reason we have to test the first one.

Errors in hypothesis testing

Let H_0 represent a null hypothesis.

	H_0 True	H_0 False
Fail to Reject H_0	Correct Decision (True Negative, $1 - \alpha$)	**Type II Error** (False Negative, β)
Reject H_0	**Type I Error** (False positive, α)	Correct Decision (True Positive, $1 - \beta$)

Errors in hypothesis testing

$$H_0\colon \mu_{Trt} - \mu_{Ctrl} = MMES$$
$$H_1\colon \mu_{Trt} - \mu_{Ctrl} > MMES$$

		Truth	
		Not a meaningfully improvement	A meaningfully improvement
Conclusion	Doesn't work	Correct Decision	Type II
	Works	Type I	Correct Decision

Power is our ability to conclude that *TutorNow* represents a meaningful (and significant) improvement on the ACT for low-income students compared to low-income students not receiving *TutorNow* when in fact *TutorNow* actually is an improvement.

Consequences

- Type I error (α)
  - Conclude that TutorNow has a significant, and meaningful, improvement on the ACT over students not receiving tutoring when it does not.
  - Waste resources (time, money)
  - Program's effectiveness isn't reproducible.
- Type II error (β)
  - Conclude that TutorNow does not have a significant, and meaningful, improvement over students not receiving tutoring when in fact it does.
  - Withhold a treatment that could improve student outcomes
  - Program is no longer developed

Consequences

- Type I error (α)
  - Conclude that TutorNow has a significant, and meaningful, improvement on the ACT over students not receiving tutoring when it does not.
  - Waste resources (time, money)
  - Program's effectiveness isn't reproducible.
- Type II error (β)
  - Conclude that TutorNow does not have a significant, and meaningful, improvement over students not receiving tutoring when in fact it does.
  - Withhold a treatment that could improve student outcomes
  - Program is no longer developed
- **Which one is the worse error?**

Quick digression - Ecological example

- RQ: Does timber harvesting affect the presence of river otters in a watershed?
- H_0: Timber harvesting has no effect on the presence of river otters in a watershed.
- H_1: Timber harvesting negatively affects the presence of river otters in a watershed.

Quick digression - Ecological example

- Type I error (α)
  - Conclude that timber harvesting negatively affects the presence of river otters use of a watershed when it does not.
  - Local economic impact
  - Timber harvested halted regionally, nationally.
- Type II error ($1 - \beta$)
  - Conclude that timber harvesting has no effect on the presence of river otters use of a watershed when it does.
  - River otters go locally extinct (extirpated)
  - River otters go extinct where ever timber harvesting occurs

Quick digression - Ecological example

- Type I error (α)
  - Conclude that timber harvesting negatively affects the presence of river otters use of a watershed when it does not.
  - Local economic impact
  - Timber harvested halted regionally, nationally.
- Type II error ($1 - \beta$)
  - Conclude that timber harvesting has no effect on the presence of river otters use of a watershed when it does.
  - River otters go locally extinct (extirpated)
  - River otters go extinct where ever timber harvesting occurs
- **Which one is the worse error?**

Power Analysis

What do we need?

- **A statistical model/test**
- An effect size estimate
- A Type I error (α)
- A Type II error (β) or power (1 - β)
- Sample size

How do we perform an actual power analysis?

- Identify the statistical test you wish to perform.

How do we perform an actual power analysis?

- Identify the statistical test you wish to perform.
- For basic tests, power calculators exist
  - Must meet assumptions of the test.
  - **This is why many power analyses are wrong.**

How do we perform an actual power analysis?

- Identify the statistical test you wish to perform.
- For basic tests, power calculators exist
  - Must meet assumptions of the test.
  - **This is why many power analyses are wrong.**
- For complex models, power is mostly simulation-based.
  - Every parameter in the model must have a value.

Why would we do a power analysis via simulation?

1. Can reproduce power analysis calculators.

Why would we do a power analysis via simulation?

1. Can reproduce power analysis calculators.
2. Real data rarely conform exactly to modeling assumptions.

Why would we do a power analysis via simulation?

1. Can reproduce power analysis calculators.
2. Real data rarely conform exactly to modeling assumptions.
3. Understand the extent that violations could impact findings.

Why would we do a power analysis via simulation?

1. Can reproduce power analysis calculators.
2. Real data rarely conform exactly to modeling assumptions.
3. Understand the extent that violations could impact findings.
4. Want to play with scenarios.

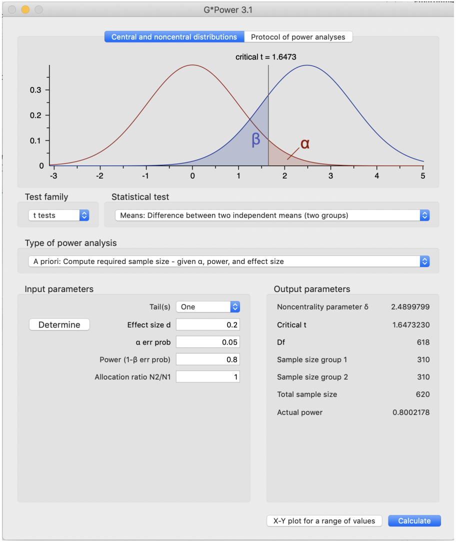
Power analysis for ACT

Given an effect size (Cohen's d) of 0.2 (aka MDES), $\alpha = .05$, power $=$ .80, and two-sample t-test with a one-sided alternative hypothesis.

In R:

```
pwr::pwr.t.test(d = 0.2, sig.level = .05, power = .80,
                type = "two.sample", alternative = "greater")

##
##      Two-sample t test power calculation
##
##               n = 309.8065
##               d = 0.2
##       sig.level = 0.05
##           power = 0.8
##     alternative = greater
##
## NOTE: n is number in *each* group
```

In G-Power:


G*Power 3.1
Central and noncentral distributions
Protocol of power analyses
critical t = 1.6473
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
4
5
β
α
Test family
t tests
Statistical test
Means: Difference between two independent means (two groups)
Type of power analysis
A priori: Compute required sample size - given α, power, and effect size
Input parameters
Tail(s) One
Determine
Effect size d 0.2
α err prob 0.05
Power (1-β err prob) 0.8
Allocation ratio N2/N1 1
Output parameters
Noncentrality parameter δ 2.4899799
Critical t 1.6473230
Df 618
Sample size group 1 310
Sample size group 2 310
Total sample size 620
Actual power 0.8002178
X-Y plot for a range of values
Calculate

How do we do a power analysis via simulation?

- Specify the data-generating mechanism, α, and sample size or effect size.
- Assuming sample size (N) and effect size are provided,
  - Randomly generate data of size N a large number of times (e.g., 2000 or 5000 replicates)
  - Perform the desired statistical test and save the p-value.
  - An empirical estimate of power is the number of times out of the 2000 or 5000 replicates where $p < \alpha$,
- Trying to obtain a desired sample/effect size,
  - Randomly generate data a large number of times (e.g., 2000 or 5000 replicates) but vary either sample size or effect size.
  - Perform the desired statistical test, save the p-value and, record the estimated power.
  - The smallest sample size size/MDES is where power is .80 or greater.

Power analysis of t-test via simulation

$$H_0 : \mu_{Trt} - \mu_{Ctrl} = 0$$

$$H_1 : \mu_{Trt} - \mu_{Ctrl} > 0$$

```r
set.seed(512019)
run_sim <- replicate(5000, expr = {
trt <- rnorm(310, mean = 21.04, sd = 5.2)
ctrl <- rnorm(310, mean = 20, sd = 5.2)
test <- t.test(trt, ctrl, var.equal = FALSE, alternative = "greater")
d <- cohen.d(trt, ctrl)
c(test$p.value, d)
})
run_sim <- data.frame(t(run_sim))
colnames(run_sim) <- c("p.value", "d")
mean(run_sim$d) # effect size

## [1] 0.2004489

mean(run_sim$p.value < .05) # empirical power

## [1] 0.7958
```

Power analysis of simple linear regression via simulation

$$\text{ACT} \mid \text{Trt} \sim N(\beta_0 + \beta_1 \text{Trt}, \sigma^2)$$
$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 > 0$$

```
set.seed(512019)
run_sim <- replicate(5000, expr = {
trt <- rnorm(310, mean = 21.04, sd = 5.2)
ctrl <- rnorm(310, mean = 20, sd = 5.2)
y <- c(trt, ctrl)
x <- rep(c("Trt", "Ctrl"), each = 310)
mod <- lm(y ~ x)
pt(summary(mod)$coef[2, 3], df = 618, lower.tail = FALSE)
})
mean(run_sim < .05) # empirical power

## [1] 0.7958
```

A more interesting power analysis

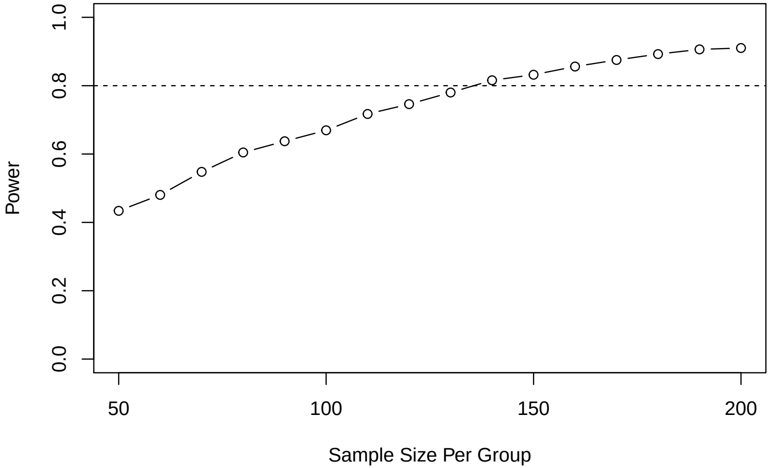
Let's say you've determined the MMES is a Cohen's d of 0.2. Anything less than 0.2 is not of practical interest. You expect a medium size effect of 0.5. You want to know the **minimum sample size** to detect an effect greater than 0.2 assuming you have an effect of 0.5 with a power of 0.8. Also, we'll relax our assumption about equal variance for the t-test.

$$H_0 : \mu_{Trt} - \mu_{Ctrl} = 1.04$$

$$H_1 : \mu_{Trt} - \mu_{Ctrl} > 1.04$$

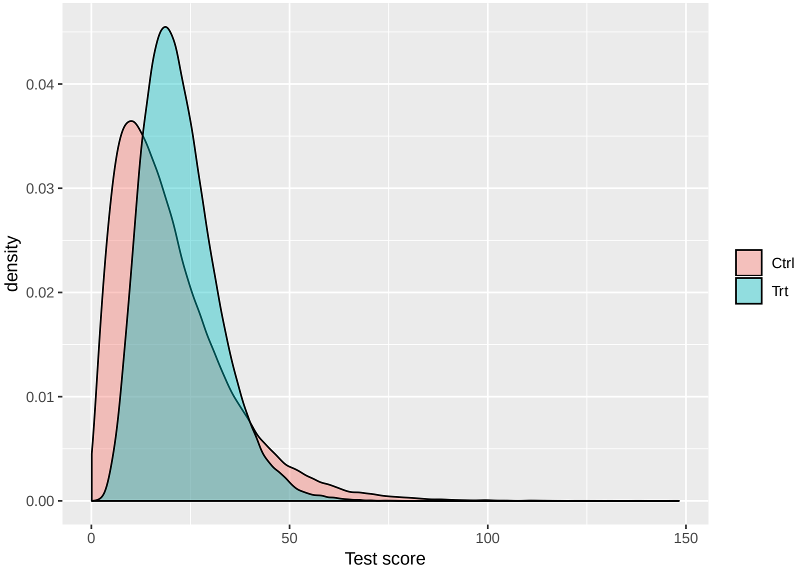
```
n <- seq(50, 200, by = 10)
calc_p <- function(n){
    trt <- rnorm(n, mean = 22.6, sd = 5.2)
    ctrl <- rnorm(n, mean = 20, sd = 5.2)
    test <- t.test(trt, ctrl, mu = 1.04,
                   var.equal = FALSE, alternative = "greater")
    test$p.value
}

set.seed(512019)
power <- NULL
for(samp in n){
    tmp <- replicate(5000, calc_p(samp))
    power <- c(power, mean(tmp < .05))
}
```

Power curve



Let's play with the assumptions now

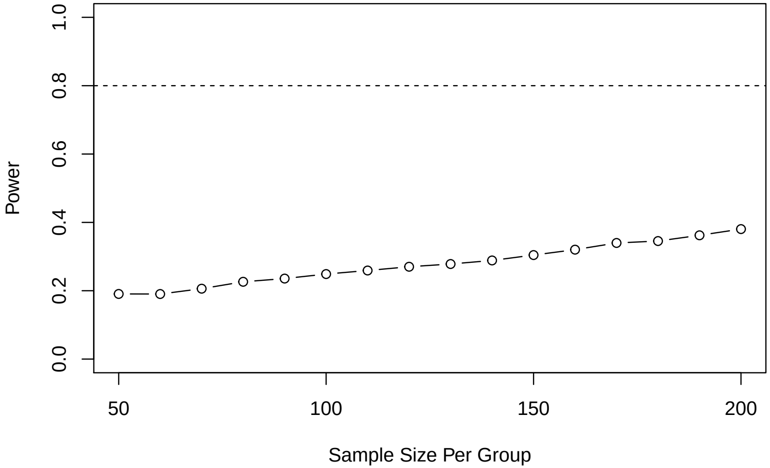
Instead of the two groups coming from two normal populations, what happens if they come from two skewed distributions but still have the *same mean difference* in the population?



Treatment: mean = 22.6, sd = 9.51
Control: mean = 20, sd = 14.14

```
n <- seq(50, 200, by = 10)
calc_p <- function(n){
    trt <- rgamma(n, shape = 5.65, scale = 4)
    ctrl <- rgamma(n, shape = 2, scale = 10)
    test <- t.test(trt, ctrl, mu = 1.04,
                   var.equal = FALSE, alternative = "greater")
    test$p.value
}

set.seed(512019)
power <- NULL
for(samp in n){
    tmp <- replicate(5000, calc_p(samp))
    power <- c(power, mean(tmp < .05))
}
```

Power curve



Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.
   - RMCC can help with #1, #2 but not #3.

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.
   - RMCC can help with #1, #2 but not #3.
4. Complexity increases quickly.

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.
   - RMCC can help with #1, #2 but not #3.
4. Complexity increases quickly.
5. Takes a long time (computationally)

Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.
   - RMCC can help with #1, #2 but not #3.
4. Complexity increases quickly.
5. Takes a long time (computationally)
6. Didn't plan enough time during grant proposal.

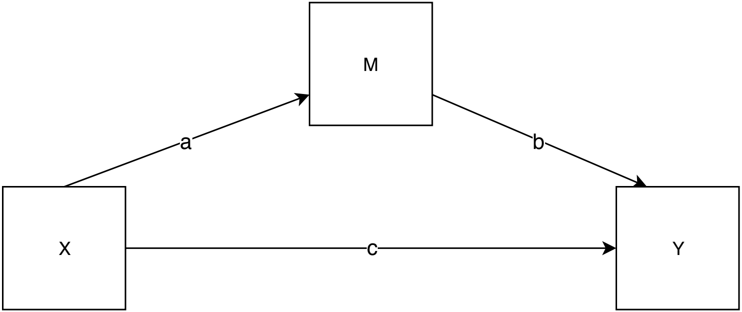
Why aren't we doing simulation-based power analyses

1. Uncertainty regarding how to write your statistical model and what your statistical assumptions are.
2. The power of what? A part of your model? The global fit of your model?
3. Uncertainty about what the parameters in the model should be.
   - RMCC can help with #1, #2 but not #3.
4. Complexity increases quickly.
5. Takes a long time (computationally)
6. Didn't plan enough time during grant proposal.
7. SPSS can't do it.

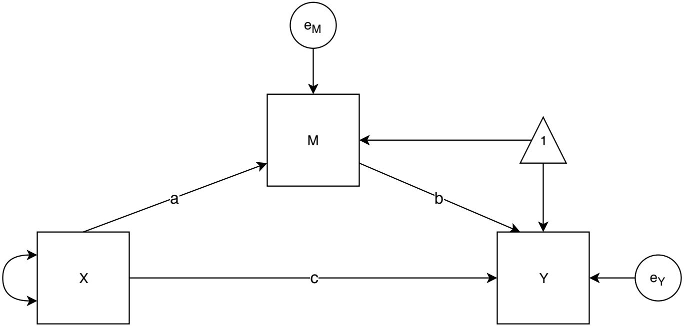
Good news?

- If you can write your model, it is very easy to simulate power in R.
  - Can use base functions to simulate mixed effects model, SEM, regressions, etc.
  - Can use functions in MASS and lavaan for simulating from multivariate normal and SEM models.
- If you use Mplus, https://www.statmodel.com/usersguide/chapter12.shtml

Power analysis - Mediation



Power analysis - Mediation (explicit)



Equations

1. M regressed on X

$$M|X \sim N(\beta_m + aX, \sigma_m^2)$$

2. Y regressed on M and X

$$Y|X, M \sim N(\beta_y + bM + cX, \sigma_y^2)$$

We are interested in the indirect effect, ab.

What we need for the power analysis if interest is in the indirect effect, ab

- β_m and β_y - the intercepts. Could set to zero if variables are mean-centered.
- a, b, and c - the slope terms. The effect of X on M, the effect of M on Y holding X constant, and the effect of X on Y holding M constant.
- σ_m^2 and σ_y^2 - residual variance
- X needs a distribution.
    - Should it be continuous or discrete?
- **Need to provide 8 parameters** and to find power of ab could vary a, b, or both.
- What type of effect size to use? Several proposed ...
    - Standardized indirect effect squared (v) (Lachowicz, Preacher, & Kelley, 2018)
        - The variance in Y accounted for jointly by M and X adjusting for the ordering of variables.
        - *Could* use Cohen's criteria for R^2 (.02, .15, .25)

How to proceed

First, what do we want? A sample size given a MMES or do we know the sample size and need to know our MDES? Let's assume the former and let's assume a MMES of .10 and want to know our sample size given a power of .80 and an α of .05.

Second, we need to set the parameters.

- We'll standardize X and make M and Y standardized, too.
- No need for β_m and β_y.
- We know we need to constrain $ab^2 = .10$, so how do to decide what value to use for a and b?
- Can set each standardized regression weights (beta) to .562 as long as ab = .316.
- If using Baron & Kenny framework, probably the above approach is best.
- We'll set c to .14 (a small effect).

Third, let's assume we can recruit (or afford) between 50 - 120 participants, but would prefer the least amount possible.

Fourth, we'll run 2,000 replications of each condition.

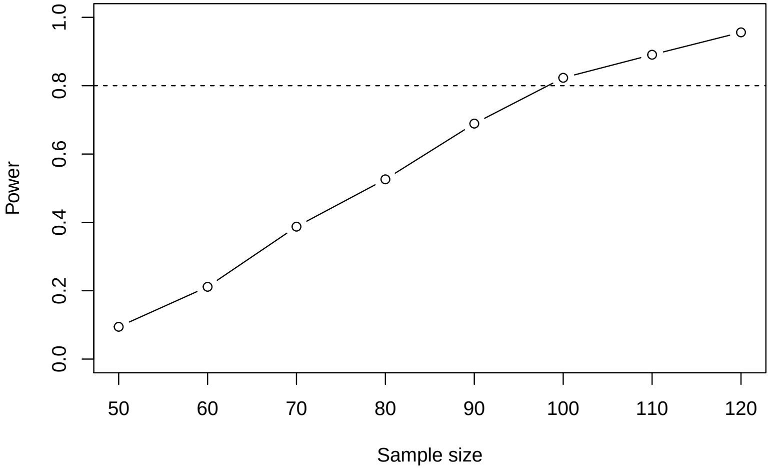
Setting up the model

```
set.seed(05012019)
library(lavaan)
# relationship between X and M, M and Y, and X and Y
a <- .562
b <- .562
c <- .14

# Fit the model
mod <- '
M ~ a*X             # M regressed onto X
Y ~ b*M + c*X       # Y regressed onto M and X
v := (a*b)*(a*b)    # indirect effect squared
'

# How many observations should we generate.
n <- 100

# Generate X to be a random normal variable, M, and Y
X <- rnorm(n = n, mean = 0, sd = 1)
M <- a*X + rnorm(n, mean = 0, sd = sqrt(1 - a^2))
Y <- b*M + c*X + rnorm(n, mean = 0, sd = sqrt(1 - (b^2 + c^2 + 2*b*c*a)))
dat <- data.frame(Y, X, M)

# Fit the model
fit <- sem(model = mod, data = dat)

# Extract the relevant information
param <- parameterEstimates(fit)
subset(param, label == "v", pvalue)
```

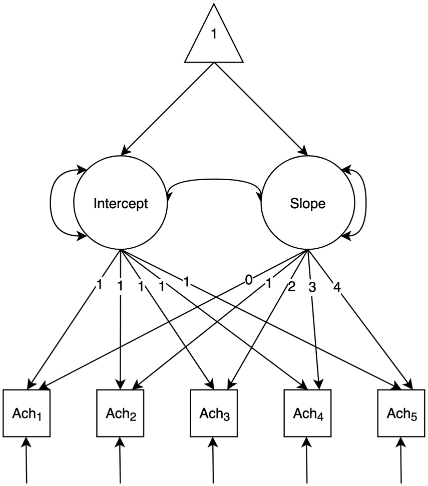
Running the simulation

```r
sample_size <- seq(50, 120, by = 10)
alpha <- .05
run_sim <- function(N){
  # How many observations should we generate.
  n <- N

  # Generate X to be a random normal variable, M, and Y
  X <- rnorm(n = n, mean = 0, sd = 1)
  M <- a*X + rnorm(n, mean = 0, sd = sqrt(1 - a^2))
  Y <- b*M + c*X + rnorm(n, mean = 0, sd = sqrt(1 - (b^2 + c^2 + 2*b*c*a)))
  dat <- data.frame(Y, X, M)

  # Fit the model
  fit <- sem(model = mod, data = dat)

  # Extract the relevant information
  param <- parameterEstimates(fit)
  subset(param, label == "v", pvalue, drop = TRUE)
}
set.seed(05012019)
power <- NULL
for(samp in sample_size){
  p_values <- replicate(2000, run_sim(samp))
  power <- c(power, mean(p_values < alpha))
}
```

Power curve for mediation analysis



Linear (latent) growth curve model



Let's say we're interested in detecting the minimum unstandardized parameter coefficient for the mean of the slope (i.e., change in achievement over time) assuming we have a sample size of 100.

Generating data in R

```
nobs  <- 100
ran_effs <- c(5, .5)
cor_mat <- matrix(c(1, -.2, -.2, 1), nrow = 2)
cov_mat <- cor2cov(cor_mat, ran_effs)
int_mean <- 20
slope_mean <- 1.5 # this is our parameter of interest!
fact_scores <- MASS::mvrnorm(n = nobs, mu = c(int_mean, slope_mean), Sigma = cov_mat)

ach1 <- 1*fact_scores[, 1] + 0*fact_scores[, 2] + rnorm(nobs, sd = .5)
ach2 <- 1*fact_scores[, 1] + 1*fact_scores[, 2] + rnorm(nobs, sd = .5)
ach3 <- 1*fact_scores[, 1] + 2*fact_scores[, 2] + rnorm(nobs, sd = .5)
ach4 <- 1*fact_scores[, 1] + 3*fact_scores[, 2] + rnorm(nobs, sd = .5)
ach5 <- 1*fact_scores[, 1] + 4*fact_scores[, 2] + rnorm(nobs, sd = .5)

dat <- data.frame(id = 1:nobs, ach1, ach2, ach3, ach4, ach5)
dat_l <- reshape(dat, direction = "long", varying = 2:6,
                 timevar = "year", v.names = "score", times = 0:4)
```

Fitting the model in R

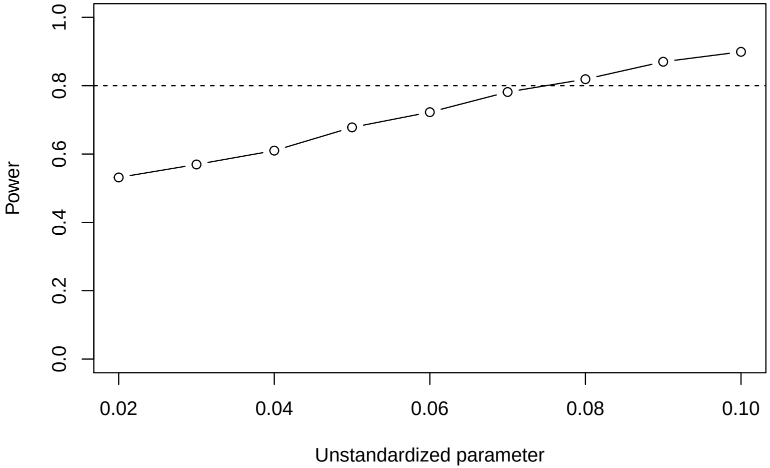
```
mod <- "
int =~ 1*ach1 + 1*ach2 + 1*ach3 + 1*ach4 + 1*ach5
slope =~ 0*ach1 + 1*ach2 + 2*ach3 + 3*ach4 + 4*ach5

int ~ 1
slope ~ eff*1

int ~~ int + slope
slope ~~ slope

ach1 ~~ e*ach1
ach2 ~~ e*ach2
ach3 ~~ e*ach3
ach4 ~~ e*ach4
ach5 ~~ e*ach5
"
```

Equivalent models

```r
# as a latent growth curve model
lgcm.fit <- growth(mod, data = dat)
summary(lgcm.fit)

# as a mixed effects model
me.fit <- lme4::lmer(score ~ 1 + year + (1 + year | id), data = dat_l)
summary(me.fit)
```

```r
alpha <- .5
nobs  <- 100
ran_effs <- c(5, .5)
cor_mat <- matrix(c(1, -.2, -.2, 1), nrow = 2)
cov_mat <- cor2cov(cor_mat, ran_effs)
int_mean <- 20
eff_sizes <- seq(.02, .1, by = .01)
run_sim <- function(param){

  slope_mean <- param # this is our parameter of interest!
  fact_scores <- MASS::mvrnorm(n = nobs, mu = c(int_mean, slope_mean), Sigma = cov_mat)

  ach1 <- 1*fact_scores[, 1] + 0*fact_scores[, 2] + rnorm(nobs, sd = .5)
  ach2 <- 1*fact_scores[, 1] + 1*fact_scores[, 2] + rnorm(nobs, sd = .5)
  ach3 <- 1*fact_scores[, 1] + 2*fact_scores[, 2] + rnorm(nobs, sd = .5)
  ach4 <- 1*fact_scores[, 1] + 3*fact_scores[, 2] + rnorm(nobs, sd = .5)
  ach5 <- 1*fact_scores[, 1] + 4*fact_scores[, 2] + rnorm(nobs, sd = .5)

  dat <- data.frame(id = 1:nobs, ach1, ach2, ach3, ach4, ach5)
  lgcm.fit <- growth(mod, data = dat)
  params <- parameterEstimates(lgcm.fit)
  subset(params, label == "eff", pvalue, drop = TRUE)
}
set.seed(05012019)
power <- NULL
for(param in eff_sizes){
  p_values <- replicate(2000, run_sim(param))
  power <- c(power, mean(p_values < alpha))
}
```

Power curve for LGCM



Why should we promote a culture of simulation-based power analyses?

- All power analyses can be performed via simulation.
- Simulation requires deeper thought about research questions and models that will be examined a priori.
- Deepens our understanding of power and statistical concepts (more generally)

How can we promote a culture of simulation-based power analyses in CEHD?

1. Work with RMCC and involve us early.
2. Make sure students exiting PhD-level statistics courses know how to write their models and know the assumptions.
   - If they know the models and the assumptions, then they can test them and understand what to expect if they don't meet their assumptions.
3. Avoid thinking about power as a binary concept.
4. Involve simulation, as much as possible, in the classroom.
5. Just teach R.