Practical Statistical Power Analysis for Simple and Complex Models

Zhiyong Johnny Zhang

University of Notre Dame

Outline

- 35 minutes lecture + 15 minutes practice + 10 minutes questions & break
- 8:00-8:50: Basic ideas and methods of statistical power analysis
- 9:00-9:50: Power analysis for ANOVA and regression
- 10:00-10:50: Power analysis for mediation analysis and structural equation modeling
- 11:00-11:50: Power analysis for multilevel modeling and a general Monte Carlo method for power calculation

Basic Ideas and Methods

Why statistical power?

- Performing statistical power analysis and sample size estimation is an important aspect of experimental design.
- Without power analysis, sample size may be too high or too low.
- If sample size is too low, an experiment will lack the precision to provide reliable answers to the questions under investigation.
- If sample size is too large, time and resources will be wasted, often for minimal gain.
- Statistical power analysis allows us to decide how large a sample is needed to enable accurate and reliable data analysis and how likely a statistical test can detect effects of a given size in a particular situation.

What is statistical power?

- The power of a statistical test is the probability that the test will reject a false null hypothesis.

	Fail to reject H_0	Reject H_0
H_0 is true	Right decision	Type I error
H_1 is true	Type II error	Power

- Statistical power = 1 – Type II error. As power increases, the chances of a Type II error decrease.

How to calculate statistical power?

- Choose a test statistic
- Derive the distribution of the test statistic under the null and alternative hypothesis
- Calculate power based on the definition

An example

Suppose a researcher is interested in whether training can improve mathematical ability. She plans to conduct a study to get the math test scores from a group of students before and after training. The null hypothesis here is the change is 0. She believes that average change would be 0.1 unit and the standard deviation of data will be 1. She wants to know how many participants she needs to recruit in her study.

- Null hypothesis: $H_0 : \mu = \mu_0 = 0$
- Alternative hypothesis: $H_1 : \mu = \mu_1 = 0.1$

Step 1: Choose a test statistic

- The design is a pre- and post-test design.
- Two-sample paired t-test is often used.
- t statistic can be used here

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}.$$

where $\bar{y}$ and s are sample mean and standard deviation with the sample size n.

Step 2: Distributions under null and alternative hypothesis

- Under the null hypothesis ($\mu = \mu_0$), the statistic

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{\bar{y} - \mu + \mu - \mu_0}{s/\sqrt{n}}$$
$$= \frac{[\bar{y} - \mu + \mu - \mu_0]/\sqrt{\sigma^2/n}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}}$$

  follows a t distribution with degree of freedom $n-1$.
- Under the alternative hypothesis ($\mu = \mu_1$), the statistic

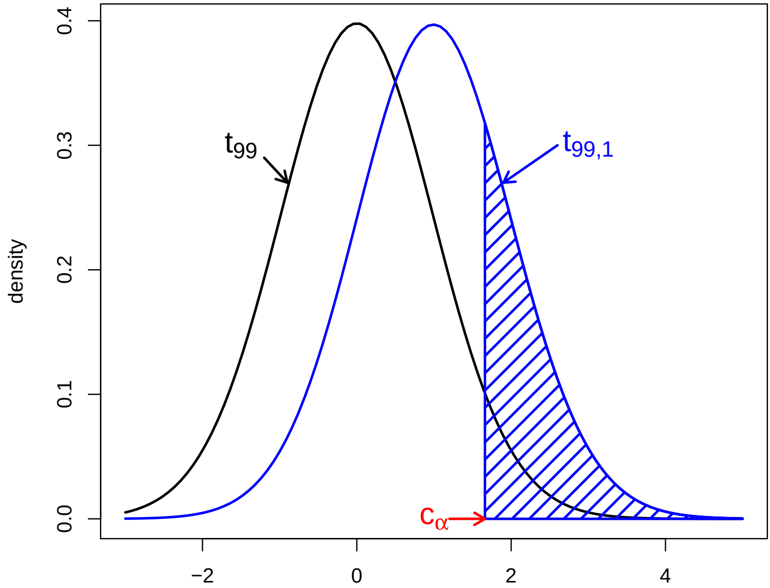
$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}} = \frac{\bar{y} - \mu + \mu_1 - \mu_0}{s/\sqrt{n}}$$

  follows a non-central t distribution with degree of freedom $n-1$ and the non-centrality parameter $\lambda = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$.

Step 3: Calculate power

$$
\begin{aligned}
\text{Power} = \pi &= \Pr(\text{reject } H_0|\mu = \mu_1) \\
&= \Pr(\bar{y} \text{ is larger than the critical value under } H_0|\mu = \mu_1) \\
&= \Pr(\bar{y} > C_{1-\alpha}|\mu = \mu_1) \\
&= \Pr(\bar{y} > \mu_0 + c_{1-\alpha} s/\sqrt{n}|\mu = \mu_1) \\
&= \Pr\left(\frac{\bar{y} - \mu}{s/\sqrt{n}} + \frac{\mu - \mu_0}{s/\sqrt{n}} > c_{1-\alpha}|\mu = \mu_1\right) \\
&= 1 - t_{n-1,\lambda}^{-1}(c_{1-\alpha})
\end{aligned}
$$

where c_α is the 100αth percentile of a central t distribution and $t_{n-1,\lambda}^{-1}$ is the probability of non-central t distribution with the non-centrality parameter $\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$.



density
0.0
0.1
0.2
0.3
0.4
−2
0
2
4
t_{99}
$t_{99,1}$
c_α

How much power?

It depends!

- Power = 1 - Type II error
- Larger power $\Longrightarrow$ smaller type II error $\Longrightarrow$ larger type I error
- Which one is more important? Reject or not to reject null
  - Cognitive training
  - Cancer diagnostics
- Commonly accept a power of 0.8, ratio of 4 for type II error to type I error.

Factors influencing power I

- Sample size. Larger sample size $\Longrightarrow$ larger power
- Significance level, type I error (α):

$$\alpha = \Pr(\text{reject } H_0 | H_0 \text{ is true})$$

  - Increasing α also increases power, reduces type II error.
  - Increasing α increases the risk of obtaining a statistically significant result when the null hypothesis is true.
- Effect size. Larger effect size $\Longrightarrow$ larger power
  - Independent of sample size
  - Standardized effect size
    - might be comparable across studies

Factors influencing power II

- for mean comparison, Cohen's d

$$d = \frac{\mu_1 - \mu_0}{\sigma}$$

  which can be estimated using sample effect size

$$\hat{d} = \frac{\bar{y}_1 - \bar{y}_0}{s}.$$

  - often is sufficient to determine the power
- Unstandardized effect size
  - might not be compared across studies
  - often reserve its own scale
  - for example, the mean itself, μ, which can be estimated by $\bar{y}$
  - might not be sufficient to determine the power because of lacking of variability in the measurement.
- Size of effect
  - no universal way to define small, medium, large effects
  - usually domain dependent
  - Cohen's d: small (0.2), medium (0.5) and large (0.8)
  - "this is an operation fraught with many dangers" (Cohen, 1977)

Factors influencing power III

- Other factors
  - Reliability of data. Power can often be improved by reducing the measurement error in the data.
  - Optimal design. The design of an experiment or observational study often influences the power.
    - often a balanced design has larger power than an unbalanced design
  - Measurement occasions. For longitudinal studies, power increases with the number of measurement occasions. Power may also be related to the measurement intervals.
  - Missing data. Missing data reduce sample size and thus power. Different missing data patterns can have difference power.
  - Non-normal data. Assuming non-normal data to be normal for power analysis might reduce power.

Not only about power

$$\text{Power} = \pi = 1 - t^{-1}_{n-1,\lambda}(c_{1-\alpha})$$

- Power: given sample size n, effect size (used to determine λ), alpha level α
- Sample size planning: given power π, effect size (used to determine λ), alpha level α
- Minimum detectable effect: given sample size n, power π, alpha level α
- Significance level: given sample size n, power π, effect size

Conducting power analysis I

1. Direct calculation
   - Get the critical value $c_{1-\alpha} = c_{0.95} = qt(0.95, 99) = 1.66$.
   - Get the non-centrality parameter $\lambda = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} = \frac{0.1-0}{1/\sqrt{100}} = 1$.
   - Get the power $\pi = 1 - pt(1.66, 99, 1) = 0.257$.
2. Use R function

```
> wp.t(n1=100, d=.1, type="paired",
   alternative="greater")

Paired t test power calculation

      n   d alpha     power
    100 0.1  0.05 0.2573029

NOTE: n is number of *pairs*

WebPower URL: http://w.psychstat.org/ttest
```

Conducting power analysis II

3. Use WebPower

One-sample, paired, two-sample balanced t test

Parameters (Help)	
Sample size	100
Effect size (Calculator)	0.1
Significance level	0.05
Power	
Type of test	Paired ▼
H1	Greater ▼
Power curve	No power curve ▼
Note	t-test

Calculate

Paired t test power calculation

```
  n   d alpha Power
100 0.1  0.05 0.257
```

Note. n is number of *pairs*

4. Other software available such as Gpower, SAS, SPSS, etc.

Software I

- WebPower
  - https://webpower.psychstat.org
  - Use within a web browser on pc, mac, tablets, phone, etc.
  - Needs internet connection
  - Click "New Analysis" to start
  - Current procedures: correlation; one-sample and two-sample proportions; one-sample and two-sample t-test; one-way, two-way, three-way ANOVA; ANCOVA; repeated-measures ANOVA; one-way ANOVA with binary and count data; linear, logistic, and Poisson regression; simple mediation analysis; two level cluster and multisite randomized trials; longitudinal data analysis; structural equation modeling; Monte Carlo based general methods.

Software II

- R package WebPower
  - Still under development
  - Offline use
  - Need to install R on computer
  - To use, copy and paste the content of webpower.R into R
  - or use `source(file.choose())`

Power Analysis for t-test

t-test

- A t test can be used to assess the statistical significance of
  - the difference between population mean and a specific value $\Longrightarrow$ one-sample t-test,
  - the difference between means of matched pairs $\Longrightarrow$ paired two-sample t-test,
  - the difference between two independent population means $\Longrightarrow$ two-sample t-test.

One-sample t-test

In one-sample t test, we are interested in whether the population mean μ is different from a specific value μ_0 (usually $\mu_0 = 0$). The null hypothesis is

$$H_0 : \mu = \mu_0.$$

The alternative hypothesis can be either two-sided or one-sided:

$$H_{11} : \mu = \mu_1 \neq \mu_0,$$

or

$$H_{12} : \mu = \mu_1 > \mu_0 \text{ (greater)},$$

or

$$H_{13} : \mu = \mu_1 < \mu_0 \text{ (less)}.$$

The effect size is defined as $\delta = (\mu_1 - \mu_0)/\sigma$ which can be estimated by $d = (\bar{y} - \mu_0)/s$.

How to get the effect size?

- No easy way!
- Pilot study
- Literature review
- Expert opinions
- Sensitivity analysis

Power for one-sample t-test using WebPower

- Example 1. Calculate power
- Example 2. Generate a power curve
- Example 3. Calculate sample size
- Example 4. Calculate effect size
- Example 5. Determine alpha level

Example 1. Calculate power

A researcher is interested in whether the score on mini-mental state examination of college students is greater than 25. If the effect size is known as 0.2 and there are 150 participants, what is the power to find the significant result?

Parameters (Help)	
Sample size	150
Effect size (Calculator)	0.2
Significance level	0.05
Power	
Type of test	One sample ▼
H1	Greater ▼
Power curve	No power curve ▼
Note	t-test

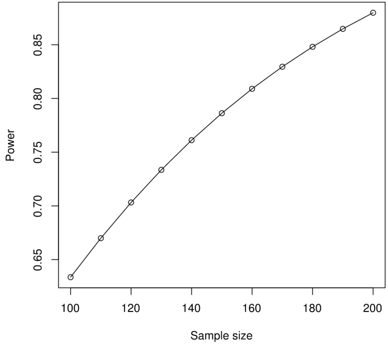
Calculate

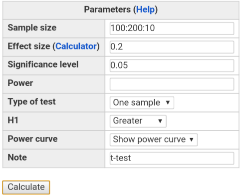
One-sample t test power calculation

```
  n   d alpha Power
150 0.2  0.05 0.786
```

Example 2. Generate a power curve

- Power for sample sizes from 100 to 200 with an interval of 10
- 100:200:10 or 100 110 120 ... 200





One-sample t test power calculation

n	d	alpha	Power
100	0.2	0.05	0.634
110	0.2	0.05	0.670
120	0.2	0.05	0.703
130	0.2	0.05	0.734
140	0.2	0.05	0.761
150	0.2	0.05	0.786
160	0.2	0.05	0.809
170	0.2	0.05	0.830
180	0.2	0.05	0.848
190	0.2	0.05	0.865
200	0.2	0.05	0.880

Example 3. Calculate sample size

- What's the needed sample size to get a power of 0.8?

Parameters (Help)	
Sample size	
Effect size (Calculator)	0.2
Significance level	0.05
Power	0.8
Type of test	One sample
H1	Greater
Power curve	No power curve
Note	t-test

Calculate

One-sample t test power calculation

```
       n   d alpha Power
155.9257 0.2  0.05   0.8
```

Example 3. Calculate sample size

- What's the needed sample size to get a power of 0.8? 156

Parameters (Help)	
Sample size	
Effect size (Calculator)	0.2
Significance level	0.05
Power	0.8
Type of test	One sample
H1	Greater
Power curve	No power curve
Note	t-test

Calculate

One-sample t test power calculation

```
       n   d alpha Power
155.9257 0.2  0.05   0.8
```

Example 4. Calculate effect size

The effect size has to be at least [ ] to get a power of 0.8 with a sample size 150.

Parameters (Help)	
Sample size	150
Effect size (Calculator)	
Significance level	0.05
Power	0.8
Type of test	One sample ▼
H1	Greater ▼
Power curve	No power curve ▼
Note	t-test

Calculate

One-sample t test power calculation

```
  n         d alpha Power
150 0.2039555  0.05   0.8
```

Example 4. Calculate effect size

The effect size has to be at least [0.204] to get a power of 0.8 with a sample size 150.

Parameters (Help)	
Sample size	150
Effect size (Calculator)	
Significance level	0.05
Power	0.8
Type of test	One sample ▼
H1	Greater ▼
Power curve	No power curve ▼
Note	t-test

Calculate

One-sample t test power calculation

```
  n         d alpha Power
150 0.2039555  0.05   0.8
```

Example 5. Determine alpha level

- Although rare, but possible.
- To get a power of 0.8 with a sample size 150 and effect size 0.2, one only needs to slightly increase the alpha level from 0.05 to [ ].

Parameters (Help)	
Sample size	150
Effect size (Calculator)	0.2
Significance level	
Power	0.8
Type of test	One sample
H1	Greater
Power curve	No power curve
Note	t-test

Calculate

One-sample t test power calculation

```
  n   d      alpha Power
150 0.2 0.05509298   0.8
```

Example 5. Determine alpha level

- Although rare, but possible.
- To get a power of 0.8 with a sample size 150 and effect size 0.2, one only needs to slightly increase the alpha level from 0.05 to [0.055].

Parameters (Help)	
Sample size	150
Effect size (Calculator)	0.2
Significance level	
Power	0.8
Type of test	One sample
H1	Greater
Power curve	No power curve
Note	t-test

Calculate

One-sample t test power calculation

```
  n   d      alpha Power
150 0.2 0.05509298   0.8
```

Power analysis using R I

- The same analyses conducted by WebPower can be carried out within R for the same results.
- The R input and output are given below

```
> ## Example 1. Calculate power
> wp.t(150, d=.2, type='one.sample',
   alternative='greater')

One-sample t test power calculation

      n    d alpha      power
    150  0.2  0.05 0.7862539
>
> ## Example 2. Generate a power curve
> res <- wp.t(seq(100,200,10), d=.2, type='
   one.sample', alternative='greater')
```

Power analysis using R II

```
> res

One-sample t test power calculation
      n   d alpha     power
    100 0.2  0.05 0.6336178
    110 0.2  0.05 0.6699290
    120 0.2  0.05 0.7031750
    130 0.2  0.05 0.7335260
    140 0.2  0.05 0.7611590
    150 0.2  0.05 0.7862539
    160 0.2  0.05 0.8089902
    170 0.2  0.05 0.8295443
    180 0.2  0.05 0.8480874
    190 0.2  0.05 0.8647838
    200 0.2  0.05 0.8797900
>
```

Power analysis using R III

```
> plot(res)
>
> ## Example 3. Calculate sample size
> wp.t(NULL, d=.2, power=0.8, type='one.
   sample', alternative='greater')

One-sample t test power calculation

            n   d alpha power
     155.9257 0.2  0.05   0.8
>
> ## Example 4. Calculate effect size
> wp.t(150, d=NULL, power=0.8, type='one.
   sample', alternative='greater')

One-sample t test power calculation
```

Power analysis using R IV

```
      n           d alpha power
    150 0.2039555  0.05   0.8
>
> ## Example 5. Determine alpha level
> wp.t(150, d=0.2, power=0.8, alpha=NULL,
   type='one.sample', alternative='greater')

One-sample t test power calculation

      n   d      alpha power
    150 0.2 0.05509298   0.8
```

Two-sample t-test

Assume there are two samples from $y_{1i} \sim N(\mu_1, \sigma^2)$ and $y_{2i} \sim N(\mu_2, \sigma^2)$. Let $\bar{y_1}$ and $\bar{y_2}$ denote the sample means and ${s_1}^2$ and ${s_2}^2$ denote the sample variances. The t statistic is

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where s_p is the common variance,

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

The t statistic follows a t distribution with degrees of freedom $n_1 + n_2 - 2$ under the null hypothesis. Under the alternative hypothesis, the effect size is $(\mu_1 - \mu_2)/\sigma$ that can be estimated by $(\bar{y}_1 - \bar{y}_2)/s_p$.

- When $n_1 = n_2$, the balanced 2-sample power analysis can be conducted the same as the one-sample one.
- When $n_1 \neq n_2$, the unbalanced 2-sample analysis can be used.

Example: Caclulate effect size

The test scores from two classes with different textbooks are recorded as below. If each class has 25 students, what would be the effect size in this case?

	Class 1	Class 2
Mean	100	125
Variance	900	1225

- The common variance is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 1062.5$$

- The estimated effect size is

$$d = \frac{\bar{y}_1 - \bar{y}_2}{s_p} = -0.767$$

Example: What's the power for $n_1 = n_2 = 25$?

Parameters (Help)	
Sample size	25
Effect size (Calculator)	-0.767
Significance level	0.05
Power	
Type of test	Two sample ▼
H1	Two sided ▼
Power curve	No power curve ▼
Note	t-test

Calculate

Two-sample t test power calculation

```
 n     d alpha Power
25 0.767  0.05 0.757
```

```
> wp.t(n1=25,n2=25,d=-.767)
     n      d  alpha     power
    25  0.767   0.05 0.7571582
NOTE: n is number in *each* group
```

Example: Power for $n_1 = 20$ and $n_2 = 30$

Parameters (Help)	
Sample size of group 1	20
Sample size of group 2	30
Effect size (Calculator)	-0.767
Significance level	0.05
Power	
H1	Two sided ▾
Power curve	No power curve ▾
Note	Unbalanced two-sample t-te

Calculate

t test power calculation

```
n1 n2     d alpha Power
20 30 0.767  0.05  0.74
```

```
> wp.t(n1=20,n2=30,d=-.767,type="two.sample.2
  n")
    n1  n2       d  alpha      power
    20  30  0.767   0.05  0.7400586
NOTE: n1 and n2 are number in *each* group
```

Practice

- For the two-sample t-test, given d=0.767, n1=n2=30, what's the power for alpha=0.1?
- For the two-sample t-test, given d=0.767 and alpha=0.05, generate a power curve for n1=n2=20 to 50.
- For the two-sample t-test, given d=0.767 ,alpha=0.05, and n1=20, what's the needed sample size for n2 to get power 0.9?

Power analysis for ANOVA and regression

One-way ANOVA

Suppose there exists a factor A with k levels or groups. The sample size for each group is $n_g, g = 1, \ldots k$. The total sample size is $n = \sum_{g=1}^{k} n_g$. Let y_{ig} denotes the datum for the ith individual in the gth group. Assume that $y_{ig} \sim N(\mu_g, \sigma^2)$ where μ_g is the group mean of the gth group.

ANOVA usually concerns the overall test of equality of the means across groups with

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

indicating all groups are the same vs.

$$H_1 : \exists j, l; \mu_j \neq \mu_l,$$

existing at least two groups with different means.

For ANOVA, F-test is often used. Under the null hypothesis, a central F distribution is used and under the alternative hypothesis, a non-central F distribution is used.

Effect size for one-way ANOVA

Cohen (1988, p.275) used the statistic f as the measure of effect size for one-way ANOVA. The f is the ratio between the standard deviation of the effect to be tested σ_b (or the standard deviation of the group means, or between-group standard deviation) and the common standard deviation within the populations (or the standard deviation within each group, or within-group standard deviation) σ_w such that

$$f = \frac{\sigma_b}{\sigma_w}.$$

Given the two quantities σ_m and σ_w, the effect size can be determined. Cohen defined the size of effect as: small 0.1, medium 0.25, and large 0.4.

Effect size calculation

The effect size can be determined based the group information.

Group	1	2	...	G
Group size	n_1	n_2	...	n_G
Mean	m_1	m_2	...	m_G
Variance	s_1^2	s_2^2	...	s_G^2

The between group standard deviation σ_b can be calculated by $\sigma_b = \sqrt{\sum_{g=1}^{G} w_g(m_g - \bar{m})^2}$ with $\bar{m} = \sum_{g=1}^{G} w_g m_g$ where w_g is the weight $w_g = \frac{n_g}{\sum_{i=1}^{G} n_g}$. For the within-group standard deviation, it is calculated as $\sigma_w = \sqrt{\sum_{g=1}^{G} s_g^2/G}$. Therefore, the effect size is $f = \sigma_b/\sigma_w$.

Effect size calculator in WebPower

A student hypothesizes that freshman, sophomore, junior and senior college students have different attitude towards obtaining arts degrees.

	n	mean	var
Freshman	25	2	9
Sophomore	25	3	9
Junior	25	3.6	9
Senior	25	4	9

Method 2: Use group mean information

Number of groups: 4 Update

Group	Sample size	Mean	Variance
1	25	2	9
2	25	3	9
3	25	3.6	9
4	25	4	9

Effect size output

The overall effect size f = **0.2511**
The effect size for Group 1 vs Group 2 is f = **0.1179**
The effect size for Group 1 vs Group 3 is f = **0.1886**
The effect size for Group 1 vs Group 4 is f = **0.2357**
The effect size for Group 2 vs Group 3 is f = **0.0707**
The effect size for Group 2 vs Group 4 is f = **0.1179**
The effect size for Group 3 vs Group 4 is f = **0.0471**

Effect size calculation from data

The data file has to be in text format where the first column of the data is the outcome variable and the second is the grouping variable. The first line of the data should be the variable names.

y	group
22.48831	1
15.48998	1
18.97749	1
16.32764	1
22.64907	1
19.14727	1

Method 3: From empirical data analysis

Upload data file:

Choose File anovadata1.txt Calculate

Effect size output

The overall effect size f = **1.7613**
The effect size for Group 1 vs 2 is f= **0.3269**
The effect size for Group 1 vs 3 is f= **0.7833**
The effect size for Group 1 vs 4 is f= **1.6567**
The effect size for Group 2 vs 3 is f= **0.4564**
The effect size for Group 2 vs 4 is f= **1.3298**
The effect size for Group 3 vs 4 is f= **0.8734**

Output from ANOVA

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        3 1166.2   388.7   78.59 <2e-16 ***
Residuals   76  375.9     4.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example. Calculate power

A student hypothesizes that freshman, sophomore, junior and senior college students have different attitude towards obtaining arts degrees. Based on his prior knowledge, he expects that the effect size is about 0.25. If he plans to interview 25 students on their attitude in each student group, what is the power for him to find the significance difference among the four groups?

Statistical Power for One-way ANOVA

Parameters (Help)	
Number of groups	4
Sample size	100
Effect size (Calculator)	0.25
Significance level	0.05
Power	
Type of analysis	Overall
Power curve	No power curve
Note	Example 1 - 1-way ANOVA

Calculate

Power for One-way ANOVA

(Equal sample in each group)

```
    n # of groups Effect size alpha Power
  100           4        0.25  0.05 0.518
```

Note. n is the total sample size adding all groups (overall)

Example. Calculate sample size

In practice, a power 0.8 is often desired. Given the power, the sample size can also be calculated.

One-way ANOVA

Parameters (Help)	
Number of groups	4
Sample size	
Effect size (Calculator)	0.25
Significance level	0.05
Power	0.8
Type of analysis	Overall
Power curve	No power curve
Note	Power analysis for one-way

Calculate

Power for One-way ANOVA

(Equal sample in each group)

```
       n # of groups Effect size alpha Power
178.3971           4        0.25  0.05   0.8
```

Example. Minimum detectable effect

Statistical Power for One-way ANOVA

Parameters (Help)	
Number of groups	4
Sample size	100
Effect size (Calculator)	
Significance level	0.05
Power	0.8
Type of analysis	Overall ▾
Power curve	No power curve ▾
Note	Example 1 - 1-way ANOVA

Calculate

Power for One-way ANOVA

(Equal sample in each group)

```
   n # of groups Effect size alpha Power
 100           4   0.3369881  0.05   0.8
```

Note. n is the total sample size adding all groups (overall)

Power analysis using R - One-way ANOVA I

- The same analyses conducted by WebPower can be carried out within R for the same results.
- The R input and output are given below

```
> #### One-way ANOVA
> # 1. power
> wp.anova(f=0.25,k=4,n=100,alpha=0.05)

Power for One-way ANOVA

    k   n    f alpha     power
    4 100 0.25  0.05 0.5181755

NOTE: n is the total sample size adding all
   groups (overall)
>
```

Power analysis using R - One-way ANOVA II

```
> # 2. power curve
> example <- wp.anova(f=0.25,k=4,n=seq
   (100,200,10),alpha=0.05)
> example

Power for One-way ANOVA

    k   n    f  alpha     power
    4 100 0.25   0.05 0.5181755
    4 110 0.25   0.05 0.5636701
    4 120 0.25   0.05 0.6065228
    4 130 0.25   0.05 0.6465721
    4 140 0.25   0.05 0.6837365
    4 150 0.25   0.05 0.7180010
    4 160 0.25   0.05 0.7494045
    4 170 0.25   0.05 0.7780286
```

Power analysis using R - One-way ANOVA III

```
     4  180  0.25   0.05  0.8039869
     4  190  0.25   0.05  0.8274169
     4  200  0.25   0.05  0.8484718

NOTE: n is the total sample size adding all
   groups (overall)
>
> plot(example, type='b')
>
> # 3. sample size
> wp.anova(f=0.25,k=4,n=NULL,alpha=0.05,power
   =0.8)

Power for One-way ANOVA

     k        n     f  alpha  power
```

Power analysis using R - One-way ANOVA IV

```
    4 178.3971 0.25  0.05   0.8

NOTE: n is the total sample size adding all
   groups (overall)
>
> # 4. effect size
> wp.anova(f=NULL,k=4,n=100,alpha=0.05,power
   =0.8)

Power for One-way ANOVA

    k   n         f alpha power
    4 100 0.3369881  0.05   0.8

NOTE: n is the total sample size adding all
   groups (overall)
```

Two-way ANOVA

Two-way analysis of variance (two-way ANOVA) is a generalization of one-way ANOVA in which two main effects and their interaction effect can be studied. The WebPower interface for power analysis for two-way ANOVA is shown below.

Two-Way, three-Way, and k-Way ANOVA

Parameters (Help)	
Number of groups	4
Total sample size	100
Numerator df	2
Effect size (f) (Calculator)	0.5
Significance level	0.05
Power	
Power curve	No power curve ▼
Note	Power analysis for k-way Al

Calculate

Example

Maxwell & Delaney (2000, p299): A counseling psychologist is interested in three types of therapy for modifying snake phobia. She believes that the best type may depend on degree (i.e., severity) of phobia. She collected data shown below. (M: moderate; S: severe)

Desensitization			Implosion			Insight		
Mild	M	S	Mild	M	S	Mild	M	S
14	15	12	10	12	10	8	9	6
17	11	10	16	14	3	10	6	10
10	12	10	19	10	6	12	7	8
13	10	9	20	11	8	14	12	9
12	9	11	19	13	2	11	11	7

- Illustrate how to obtain the necessary information for power analysis: Number of groups, Total sample size, Numerator df, Effect size.

Needed information for power calculation I

Assume an experiment with two factors A and B. A has J levels and B has K levels. For the snake phobia example, $J = K = 3$.

- Number of groups: the total number of group in the experiment $J \times K$. For the current example, the number of groups is $3 \times 3 = 9$.
- Total sample size: the sample size by adding all participants in all groups. For example, if each group has a sample size 10, the total sample size is $90 = 9 \times 10$ in the snake phobia example.
- Numerator df. The power is calculated based on F distribution which requires the numerator and denominator degrees of freedom. The numerator df depends on the effect to be analyzed. For the main effect, it is the number of levels - 1. For example, if power is calculated for the main effect of A, then the numerator df is $J - 1 = 3 - 1 = 2$. for the

Needed information for power calculation II

interaction between A and B, the numerator df is $(J-1)\times(K-1)=(3-1)\times(3-1)=4.$

- Effect size. Cohen's f used in one-way ANOVA is used here, which is the ratio between the standard deviation of the effect to be tested σ_m and the common standard deviation of the populations σ such that

$$f=\frac{\sigma_m}{\sigma}.$$

Effect size calculation I

From the snake phobia data, we can get the following information

	Mild	Moderate	Severe	Average ($\mu_{j.}$)	α_j
Desensitization	13.2(2.6)	11.4(2.3)	10.4(1.1)	11.67	0.82
Implosion	16.8(4.1)	12(1.6)	5.8(3.3)	11.53	0.69
Insight	11(2.2)	9(2.5)	8(1.6)	9.33	-1.51
Average ($\mu_{.k}$)	13.67	10.8	8.07	10.84	
β_k	2.82	-0.04	-2.78		

Based on the information in the table, we calculate the effect size. First, we calculate the common standard deviation σ,

$$\sigma = \sqrt{\frac{\sum s_{jk}^2}{JK}} = \sqrt{\frac{2.6^2 + 2.3^2 + \ldots + 1.6^2}{9}} = 2.53$$

Effect size calculation II

Effect size for main effect. Suppose we are interested in the main effect of severity. The effect size is the difference among the different level of severity, which can be determined based on the marginal means of severity. Specifically,

$$\begin{aligned}\sigma_m &= \sqrt{\frac{\sum \beta_k^2}{3}} \\ &= \sqrt{\frac{1}{3}[(13.67 - 10.84)^2 + (10.8 - 10.84)^2 + (8.07 - 10.84)^2]} \\ &= 2.29\end{aligned}$$

Then the effect size f is

$$f = \frac{\sigma_m}{\sigma} = \frac{2.29}{2.53} = 0.9$$

Effect size calculation III

Effect size for the interaction effect. For the interaction effect, we first calculate

$$(ab)_{jk} = \mu_{jk} - (\mu_{..} + \alpha_j + \beta_k)$$

and then the standard deviation is

$$\begin{aligned}\sigma_m &= \sqrt{\frac{\sum_j \sum_k (ab)^2_{jk}}{JK}} \\ &= \sqrt{\frac{(13.2 - .82 - 2.82 - 10.84)^2 + \cdots + (8 + 1.51 - 8.07 - 10.84)^2}{3 \times 3}} \\ &= 1.58\end{aligned}$$

The effect size for the interaction is

$$f = \frac{1.58}{2.53} = 0.62.$$

Example. Power for main effect of severity

- Number of groups
  $3 \times 3 = 9$
- Total sample size
  $9 \times 5 = 45$
- Numerator df $3 - 1 = 2$
- Effect size 0.9

Two-Way, three-Way, and k-Way ANOVA

Parameters (Help)	
Number of groups	9
Total sample size	45
Numerator df	2
Effect size (f) (Calculator)	0.9
Significance level	0.05
Power	
Power curve	No power curve
Note	Power analysis for k-way AI

Calculate

Power for multiple ANOVA

```
  n ndf ddf   f ng alpha Power
 45   2  36 0.9  9  0.05     1
```

Note. n is the total sample size

Example. Power for interaction between severity and type

- Number of groups
  $3 \times 3 = 9$
- Total sample size
  $9 \times 5 = 45$
- Numerator df
  $(3-1) \times (3-1) = 4$
- Effect size 0.62

Two-Way, three-Way, and k-Way ANOVA

Parameters (Help)	
Number of groups	9
Total sample size	45
Numerator df	4
Effect size (f) (Calculator)	0.62
Significance level	0.05
Power	
Power curve	No power curve ▼
Note	Power analysis for k-way AI

Calculate

Power for multiple ANOVA

```
 n ndf ddf    f ng alpha Power
45   4  36 0.62  9  0.05 0.895
```

Note. n is the total sample size

Example. ANCOVA

Suppose a continuous variable is related to snake phobia.In power analysis, the number of covariate changes the number of groups. For example, with 1 covariate, the number of groups is now $3 \times 3 + 1 = 10$.

- Number of groups $3 \times 3 + 1 = 9 + 1 = 10$
- Total sample size $9 \times 5 = 45$
- Numerator df $(3 - 1) = 2$
- Effect size 0.9

Parameters (Help)	
Number of groups	10
Total sample size	45
Numerator df	2
Effect size (f) (Calculator)	0.9
Significance level	0.05
Power	
Power curve	No power curve
Note	Power analysis for k-way AI

Calculate

Power for multiple ANOVA

```
 n ndf ddf   f ng alpha Power
45   2  35 0.9 10  0.05     1
```

Power analysis for ANOVA in R I

```
> ## Two-way ANOVA
> # 1. Main effect
> wp.kanova(n=45,ndf=2,f=.9,ng=9)

Multiple way ANOVA analysis

     n ndf ddf   f ng sig.level     power
    45   2  36 0.9  9      0.05 0.9997334


>
> # 2. Interaction effect
> wp.kanova(n=45,ndf=4,f=.62,ng=9)

Multiple way ANOVA analysis

     n ndf ddf    f ng sig.level     power
```

Power analysis for ANOVA in R II

```
    45    4   36 0.62   9        0.05 0.8947855

NOTE: Sample size is the total sample size

WebPower URL: http://w.psychstat.org/kanova

>
> # 3. ANCOVA
> wp.kanova(n=45,ndf=4,f=.9,ng=10)

Multiple way ANOVA analysis

     n ndf ddf   f ng sig.level     power
    45   4  35 0.9 10      0.05 0.9982011
```

Practice

- What's the power for testing the main effect of type of phobia?
- Generate a power for the main effect of type of phobia.
- Try out the online effect size calculator @ https://webpower.psychstat.org/models/means04/effectsize.php

Linear regression

Let y_i denote the measure of a dependent variable for the ith individual, x_{ij} denote the measurement of jth independent variable for the ith individual, and β_j denote the coefficient representing the effect of jth independent variable on the dependent variable. The regression model can be expressed as follows:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i.$$

In the case of omnibus/overall test, the null hypothesis states that all regression coefficients are zero:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0.$$

The alternative hypothesis states that at least one coefficient is not equal to zero:

$$H_1 : \exists j; \beta_j \neq 0, \quad j = 1, 2, 3, \ldots, p$$

Regression hypothesis testing

The hypothesis testing for regression can be viewed as the comparison of two models: a full model and a reduced model. The reduced model can be derived by setting certain parameters in the full model to zero.

For the overall test,

$$\text{Full model: } y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i$$

$$\text{Reduced model: } y_i = \beta_0 + e_i$$

We can also test whether a subset of predictors, e.g., z from (x, z) are jointly related to the outcome variable:

$$\text{Full model: } y_i = \beta_0 + \sum \boldsymbol{\beta}_A x_{ij} + \sum \boldsymbol{\beta}_B z_{ij} + e_i$$

$$\text{Reduced model: } y_i = \beta_0 + \sum \boldsymbol{\beta}_A x_{ij} + e_i$$

F test is used in such hypothesis testing.

Effect size for regression

We use the effect size measure f^2 proposed by Cohen (1988, p.410) as the measure of the regression effect size. Using the idea of full model and reduced model, the f^2 is defined as

$$f^2 = \frac{R^2_{Full} - R^2_{Reduced}}{1 - R^2_{Full}}$$

where R^2_{Full} and $R^2_{Reduced}$ are R-squared for the full and reduced models respectively. Note for the overall test, $R^2_{Reduced} = 0$.

WebPower interface for regression

Power for regression is based on F test. It needs the following information

- Sample size
- Number of predictors
- Effect size

Linear Regression

Parameters (Help)		
Sample size	100	
Number of predictors	1	
Effect size Show	0.15	
	Effect size calculation	
	Full model	
	Number of Predictors	1
	R-squared	0.1
	Reduced model	
	Number of Predictors	0
	R-squared	0
		Calculate
Significance level	0.05	
Power		
Power curve	No power curve ▾	
Note	Power analysis for regressi	

Calculate

Example. Power

A researcher believes that a student's high school GPA and SAT score can explain 10% of variance of her/his college GPA. If she/he has a sample of 50 students, what is her/his power to find significant relationship between college GPA and high school GPA and SAT?

Parameters (Help)	
Sample size	50
Number of predictors	2
Effect size Show	0.1111
	Effect size calculation Full model Number of Predictors: 2 R-squared: 0.1 Reduced model Number of Predictors: 0 R-squared: 0 Calculate
Significance level	0.05
Power	
Power curve	No power curve ▾
Note	Linear regression

Calculate

Power for linear regression

```
  n p1 p2     f2 alpha Power
 50  2  0 0.1111  0.05 0.521
```

Example. Power for addition predictors

Another researcher believes in addition to a student's high school GPA and SAT score, the quality of recommendation letter is also important to predict college GPA. The literature shows the quality of letter can explain an addition 5% of variance of college GPA. In order to find significant relationship between college GPA and the quality of recommendation letter above and beyond high school GPA and SAT score with a power of 0.8, what is the required sample size?

Parameters (Help)	
Sample size	
Number of predictors	1
Effect size Show	0.0588
	Effect size calculation
	Full model
	Number of Predictors: 3
	R-squared: 0.15
	Reduced model
	Number of Predictors: 2
	R-squared: 0.1
	Calculate
Significance level	0.05
Power	0.8
Power curve	No power curve
Note	Linear regression

Calculate

Power for linear regression

```
       n p1 p2     f2 alpha Power
137.4318  1  2 0.0588  0.05   0.8
```

Power analysis for linear regression in R I

```
> ## Linear regression
> # 1. Power
> wp.regression(50,p1=2,f2=.1111)

Multiple regression power calculation

     n p1 p2     f2 alpha     power
    50  2  0 0.1111  0.05 0.5212981

WebPower URL: http://w.psychstat.org/
   regression
```

Power analysis for linear regression in R II

```
>
> # 2. Effect size
> wp.regression(n=NULL, p1=3, p2=2, f2=.0588,
    power=0.8)

Multiple regression power calculation

          n p1 p2     f2 alpha power
   137.4318  3  2 0.0588  0.05   0.8

WebPower URL: http://w.psychstat.org/
   regression
```

Practice

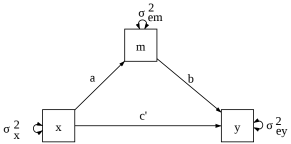
1. A previous study found that years of education and years of working experience could explain 15% of the variance of income. If you need to replicate this research and you select 50 participants for your study. What's your power to find significant results at alpha level 0.1?
2. Suppose that years of education explains 10% and years of working experience explains 5% of the variance of income.
   2.1 Generate a power curve for the predictor years of working experience only with the sample 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150.
   2.2 To get a power 0.8 for the predictor years of working experience, how large the effect size has to be with the sample size 100?

Power analysis for mediation analysis and structural equation modeling

Mediation analysis

Mediation models are useful to investigate the underlying mechanisms related to why an input variable influences an output variable.

- `x`, `m`, and `y` represent the input variable, the mediation variable, and the outcome variable.
- The total effect of `x` on `y`, `c'+a*b`, consists of the direct effect `c'` and the mediation effect θ=`a*b`



Test mediation effects I

Consider a simple mediation model

$$\begin{aligned} m_i &= a_0 + a * x_i + em_i \\ y_i &= b_0 + b * m_i + c * x_i + ey_i \end{aligned}$$

where $em_i \sim N(0, \sigma^2_{em})$ and $ey_i \sim N(0, \sigma^2_{ey})$. The mediation effect is $\theta = ab = a * b$. We have the null and alternative hypothesis

$$H_0 : \theta = 0 \text{ vs. } H_1 : \theta \neq 0.$$

The Sobel test statistic is

$$Z = \frac{\hat{a}\hat{b}}{\hat{\sigma}_{ab}}$$

Test mediation effects II

where $\hat{\sigma}_{ab}^2 = \hat{a}^2 * \hat{\sigma}_b^2 + \hat{b}^2 * \hat{\sigma}_a^2$. From regression analysis, we have

$$\hat{\sigma}_a^2 = \frac{\sigma_{em}^2}{n\sigma_x^2}$$
$$\hat{\sigma}_b^2 = \frac{\sigma_{ey}^2}{n\sigma_m^2(1-\rho_{mx}^2)}$$

where σ_x^2 and σ_m^2 are variance for x and m and ρ_{mx} is the correlation between x and m.
Furthermore, because $\hat{a} = \rho_{xm} * \sigma_m/\sigma_x$, we have

$$\rho_{xm} = \hat{a}\sigma_x/\sigma_m$$
$$\sigma_{em}^2 = \sigma_m^2(1-\rho_{mx}^2) = \sigma_m^2 - a^2\sigma_x^2.$$

Test mediation effects III

Then

$$\hat{\sigma}_a^2 = \frac{\sigma_m^2 - a^2\sigma_x^2}{n\sigma_x^2},$$
$$\hat{\sigma}_b^2 = \frac{\sigma_{ey}^2}{n(\sigma_m^2 - a^2\sigma_x^2)}.$$

Therefore, the Sobel test depends on the sample size, the coefficients a and b, the variances of x and m, and the residual variance of y denoted by $\hat{\sigma}_{ey}^2$ as in

$$Z = \frac{\hat{a}\hat{b}}{\sqrt{\hat{a}^2 * \frac{\sigma_{ey}^2}{n(\sigma_m^2 - a^2\sigma_x^2)} + \hat{b}^2 * \frac{\sigma_m^2 - a^2\sigma_x^2}{n\sigma_x^2}}}$$

WebPower interface for mediation

- No standardized effect size
- Need to provide parameters in the model

Simple Mediation via Sobel Test

Parameters (Help)	
Sample size	100
Path a	0.5
Path b	0.5
Variance of x	1
Variance of m	1
Error variance of y	1
Significance level	0.05
Power	
Power curve	No power curve ▼
Note	Simple mediation via Sobel

Calculate

Example. Power

Suppose we want to investigate whether home environment (m) is a mediator between the relationship of mother's education (x) and child's mathematical ability (y). Furthermore, we know $a = b = 0.5$, $\sigma_x^2 = \sigma_{em}^2 = \sigma_{ey}^2 = 1$. Then, we want to know the statistical power we can achieve with a sample of 100 participants at the significance level 0.05.

Simple Mediation via Sobel Test

Parameters (Help)	
Sample size	100
Path a	0.5
Path b	0.5
Variance of x	1
Variance of m	1
Error variance of y	1
Significance level	0.05
Power	
Power curve	No power curve ▾
Note	Simple mediation via Sobel

Calculate

Power calculation for mediation analysis

```
  N Power   a   b varx varm varey alpha
100 0.934 0.5 0.5    1    1     1  0.05
```

Example. Sample size

What a sample size is needed to get a power of 0.8?

Simple Mediation via Sobel Test

Parameters (Help)	
Sample size	
Path a	0.5
Path b	0.5
Variance of x	1
Variance of m	1
Error variance of y	1
Significance level	0.05
Power	0.8
Power curve	No power curve
Note	Simple mediation via Sobel

Calculate

Power calculation for mediation analysis

```
       N Power   a   b varx varm varey alpha
65.40718   0.8 0.5 0.5    1    1     1  0.05
```

Power for mediation using R I

```
> ## Mediation analysis
> # 1. Power
> wp.mediation(n=100, a=.5, b=.5, varx=1, varm=1, vary
    =1)

Power calculation for simple mediation based on Sobel
    test

      n     power   a   b varx varm vary alpha
    100 0.9337271 0.5 0.5    1    1    1  0.05

WebPower URL: http://w.psychstat.org/mediation
```

Power for mediation using R II

```
>
> wp.mediation(n=NULL, power=.8, a=.5, b=.5, varx=1,
    varm=1, vary=1)

Power calculation for simple mediation based on Sobel
    test

            n power   a   b varx varm vary alpha
     65.40718   0.8 0.5 0.5    1    1    1  0.05

WebPower URL: http://w.psychstat.org/mediation
```

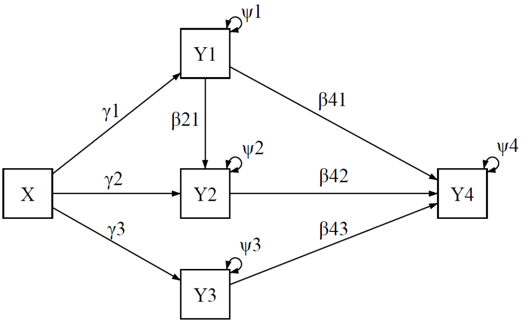
Structural equation modeling

Structural equation modeling (SEM) is one of the most widely used methods in social and behavioral sciences. SEM is a multivariate technique that is used to analyze relationships between observed and latent variables as well as among observed and latent variables. It can be viewed as a combination of factor analysis and multiple regression analysis. Two methods are widely used in power analysis for SEM. The first one is based on the likelihood ratio test proposed by Satorra & Sarris (1985) and the second one is based on RMSEA proposed by MacCallum et al. (1996).

An example

Satorra & Sarris (1985): Y_1, Y_2, Y_3, Y_4, and X satisfy the following population model

$$\begin{aligned} Y_1 &= \gamma_1 X + \zeta_1 \\ Y_2 &= \gamma_2 X + \beta_{21} Y_1 + \zeta_2 \\ Y_3 &= \gamma_3 X + \zeta_3 \\ Y_4 &= \beta_{41} Y_1 + \beta_{42} Y_2 + \beta_{43} Y_3 + \zeta_4 \end{aligned} \tag{1}$$



Satorra & Sarris (1985) method I

Let $\mathbf{S}$ denote an unbiased sample covariance matrix and $\boldsymbol{\theta}$ denote parameters in a SEM model. Let Σ be the covariance matrix defined by the model with parameters $\boldsymbol{\theta}$. From SEM theory, we know that the statistic

$$\hat{W} = (n-1)\left[\log|\Sigma(\hat{\boldsymbol{\theta}})| + \text{tr}(S\Sigma(\hat{\boldsymbol{\theta}})^{-1}) - \log|S| - p\right]$$

follows a chi-squared distribution with degrees of freedom d asymptotically. The purpose is to test the hypothesis that

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$$

vs

$$H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1.$$

Satorra & Sarris (1985) method II

Under H_0, we have $P(\chi^2_d > c_\alpha) = \alpha$ where c_α is the critical value under the chi-squared distribution with degrees of freedom d. Under H_1, $\hat{W}$ follows asymptotically a non-central chi-squared distribution with the non-centrality parameter λ. The statistical power is defined as $Power = P(\hat{W} > c_\alpha | H_1)$. Satorra & Sarris (1985) showed that λ can be approximated by

$$\lambda \approx (n-1)[\log|\hat{\Sigma}_R| + \text{tr}(\Sigma_F \hat{\Sigma}_R^{-1}) - \log|\Sigma_F| - p]$$

where Σ_F and Σ_R are defined under H_1 and H_0, respectively. With this, we can define an effect size independent of sample size as

$$\delta = \lambda/(n-1).$$

WebPower Interface

- Sample size
- Degrees of freedom
- Effect size

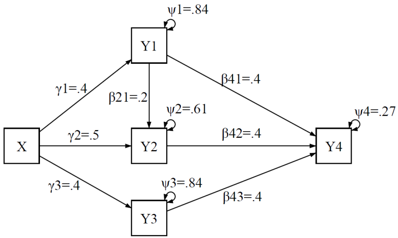
SEM based on Chi-squared test

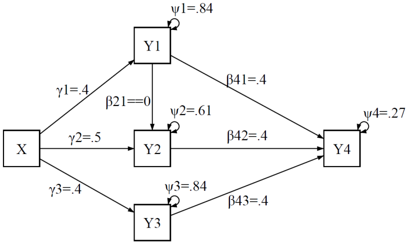
Parameters (Help)	
Sample size	100
Degrees of freedom	8
Effect size (Calculator)	0.1
Significance level	0.05
Power	
Power curve	No power curve ▼
Note	Power for SEM using Satori

Calculate

Effect size I

The effect size is defined as the difference between two SEM models, a full model M_F and a reduced model M_R. The full model includes all the parameters in the population and the reduced model is nested within the full model by setting certain relationship to be null.





Effect size II

The effect size can be calculated in the following way.

1. From the full model, the model implied covariance matrix can be obtained as Σ_F.
2. The reduced model can be fitted to Σ_F.
3. Suppose the estimated covariance matrix for the reduced model is $\hat{\Sigma}_R$. The effect size δ is obtained as

$$\delta = \log|\hat{\Sigma}_R| + \text{tr}(\Sigma_F \hat{\Sigma}_R^{-1}) - \log|\Sigma_F| - p$$

   where p is dimension of Σ_F.

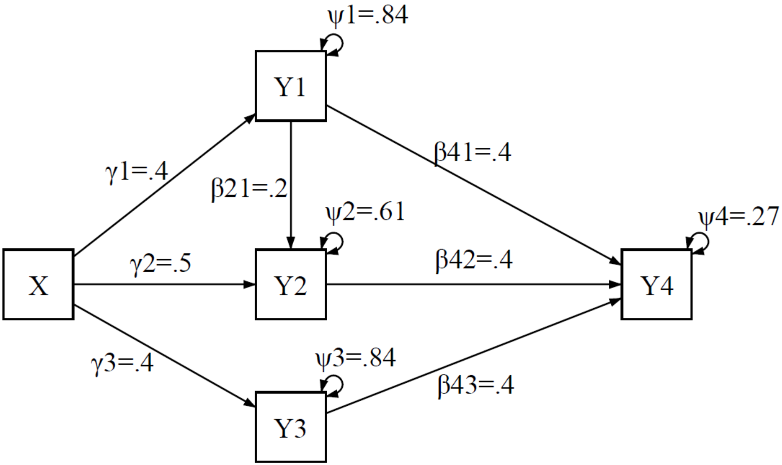
An easy way to get the effect size is to fit the reduced model to Σ_F through SEM software with a predefined sample size n to get the chi-squared statistics λ. Then the effect size is $\delta = \lambda/(n-1)$.

Effect size III

Size of effect: δ and the RMSEA ϵ: $\delta = d\epsilon^2$

	RMSEA (ϵ)	Degrees of freedom (d)	Effect size (δ)
		1	0.0025
		2	0.005
Small	0.05	4	0.01
		8	0.02
		16	0.04
		1	0.0064
		2	0.0128
Medium	0.08	4	0.0256
		8	0.0512
		16	0.1024
		1	0.01
		2	0.02
Large	0.1	4	0.04
		8	0.08
		16	0.16

Effect size calculation I



Effect size calculation II

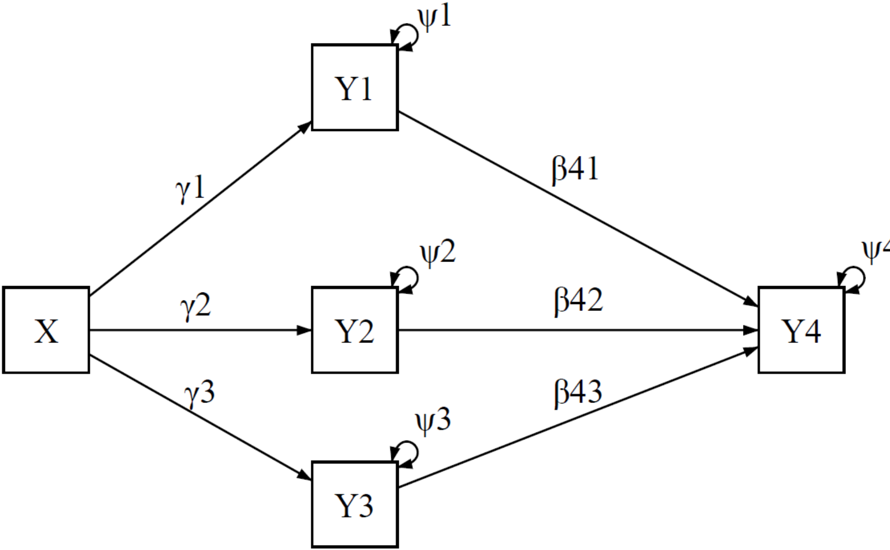
- Using the model parameters, calculate the covariance matrix for the full model

	y1	y2	y3	y4	x
y1	1.000				
y2	0.400	0.980			
y3	0.160	0.232	1.000		
y4	0.624	0.645	0.557	1.000	
x	0.400	0.580	0.400	0.552	1.000

Effect size calculation III

- Fit the reduced model by removing the path from Y1 to Y2 to the covariance matrix from the full model

Effect size calculation IV



Effect size calculation V

- With a sample size 100, the resulted chi-squared statistic is 5.36 with the degrees of freedom 4.
- Therefore, the estimated effect size is 5.36/99=0.054.

Effect size calculation VI

Effect Size Calculator for SEM

1. Effect size from population models

The model with population parameters (Full model)

```
y1 ~ 0.4*x
y2 ~ 0.5*x + 0.2*y1
y3 ~ 0.4*x
y4 ~ 0.4*y1 + 0.4*y2 + 0.4*y3
y1 ~~ 0.84*y1
y2 ~~ 0.61*y2
y3 ~~ 0.84*y3
y4 ~~ 0.27*y4
```

The restricted model fit to the population covariance matrix (Reduced model)

```
y1 ~ x
y2 ~ x
y3 ~ x
y4 ~ y1 + y2 + y3
```

Calculate

Effect size output

The effect size = **0.054**
The degrees of freedom = **4**
RMSEA = **0.05788915**

Example. Power

- The power to detect the path from Y1 to Y2.

Statistical Power for SEM (Satorra & Saris, 1985)

Parameters (Help)	
Sample size	100
Degrees of freedom	4
Effect size (Calculator)	0.054
Significance level	0.05
Power	
Power curve	No power curve ▼
Note	Power for SEM using Satorra

Calculate

Power analysis for SEM (Satorra & Saris, 1985)

```
Sample size Power effect df alpha
        100 0.422  0.054  4  0.05
```

Example. Power with different alpha level

Statistical Power for SEM (Satorra & Saris, 1985)

Parameters (Help)	
Sample size	100
Degrees of freedom	4
Effect size (Calculator)	0.054
Significance level	.001 .01 .025 .05 .1
Power	
Power curve	No power curve ▼
Note	Power for SEM using Satorra

Calculate

Power analysis for SEM (Satorra & Saris, 1985)

```
Sample size Power effect df alpha
        100 0.065  0.054  4 0.001
        100 0.209  0.054  4 0.010
        100 0.316  0.054  4 0.025
        100 0.422  0.054  4 0.050
        100 0.550  0.054  4 0.100
```

Example. Sample size planning

Statistical Power for SEM (Satorra & Saris, 1985)

Parameters (Help)	
Sample size	
Degrees of freedom	4
Effect size (Calculator)	0.054
Significance level	.05
Power	0.8
Power curve	Show power curve ▾
Note	Power for SEM using Satorra

Calculate

Power analysis for SEM (Satorra & Saris, 1985)

```
Sample size Power effect df alpha
   222.0238   0.8  0.054  4  0.05
```

Example. Determine effect size

Statistical Power for SEM (Satorra & Saris, 1985)

Parameters (Help)	
Sample size	100
Degrees of freedom	4
Effect size (Calculator)	
Significance level	.05
Power	0.8
Power curve	No power curve ▼
Note	Power for SEM using Satorra

Calculate

Power analysis for SEM (Satorra & Saris, 1985)

```
Sample size Power effect df alpha
        100   0.8  0.121  4  0.05
```

Using WebPower fro SEM I

```
> ## SEM
>
> # Get the effect size
>
> full.model <-'
+ y1 ~ 0.4*x
+ y2 ~ 0.5*x + 0.2*y1
+ y3 ~ 0.4*x
+ y4 ~ 0.4*y1 + 0.4*y2 + 0.4*y3
+ y1 ~~ 0.84*y1
+ y2 ~~ 0.61*y2
+ y3 ~~ 0.84*y3
+ y4 ~~ 0.27*y4
+ '
>
> full.res<-sem(full.model, do.fit=FALSE)
> sigma.F<-fitted.values(full.res)$cov
>
```

Using WebPower fro SEM II

```
> reduced.model <-'
+ y1 ~ x
+ y2 ~ x
+ y3 ~ x
+ y4 ~ y1 + y2 + y3
+ '
>
> N<-100
> reduced.res<-sem(reduced.model, sample.cov=
   sigma.F, sample.nobs=N)
>
> summary(reduced.res)
  Number of observations
                                       100

  Estimator
                                                      ML
  Minimum Function Test Statistic
                            5.362
```

Using WebPower fro SEM III

```
  Degrees of freedom
                                                4
  P-value (Chi-square)
                                          0.252

>
> delta <- reduced.res@Fit@test[[1]]$stat/N
> df <- reduced.res@Fit@test[[1]]$df
>
> delta
[1] 0.05361846
> df
[1] 4
>
> # 1. Power
> wp.sem.chisq(n = 100, df = 4, effect = .054,
   power = NULL, alpha = 0.05)

Power analysis for SEM (Satorra & Saris, 1985)
```

Using WebPower fro SEM IV

```
      n df effect     power alpha
    100  4  0.054 0.4221152  0.05
>
> # 2. Different alphas
> wp.sem.chisq(n = 100, df = 4, effect = .054,
   power = NULL, alpha = c(.001, .005, .01,
   .025, .05))

Power analysis for SEM (Satorra & Saris, 1985)

      n df effect      power alpha
    100  4  0.054 0.06539478 0.001
    100  4  0.054 0.14952768 0.005
    100  4  0.054 0.20867087 0.010
    100  4  0.054 0.31584011 0.025
    100  4  0.054 0.42211515 0.050
>
> # 3. Sample size
```

Using WebPower fro SEM V

```
> wp.sem.chisq(n = NULL, df = 4, effect = .054,
   power = 0.8, alpha = 0.05)

Power analysis for SEM (Satorra & Saris, 1985)

           n df effect power alpha
    222.0238  4  0.054   0.8  0.05
>
> # 4. Effect size
> wp.sem.chisq(n = 100, df = 4, effect = NULL,
   power = 0.8, alpha = 0.05)

Power analysis for SEM (Satorra & Saris, 1985)

      n df    effect power alpha
    100  4 0.1205597   0.8  0.05
```

Power based on RMSEA

Let ϵ_0 and ϵ_1 be RMSEA under $H_0 : \epsilon = \epsilon_0$ and $H_1 : \epsilon = \epsilon_1$. Under H_0, the test statistic

$$\hat{W} = (n-1)\left[\log|\Sigma(\hat{\boldsymbol{\theta}})| + \mathrm{tr}(S\Sigma(\hat{\boldsymbol{\theta}})^{-1}) - \log|S| - p\right]$$

follows a chi-squared distribution with degree of freedom d and non-centrality parameter $\lambda_0 = nd\epsilon_0^2$. Under H_1, the test statistic follows a chi-squared distribution with degree of freedom d and non-centrality parameter $\lambda_1 = nd\epsilon_1^2$. Power for testing close fit is defined as

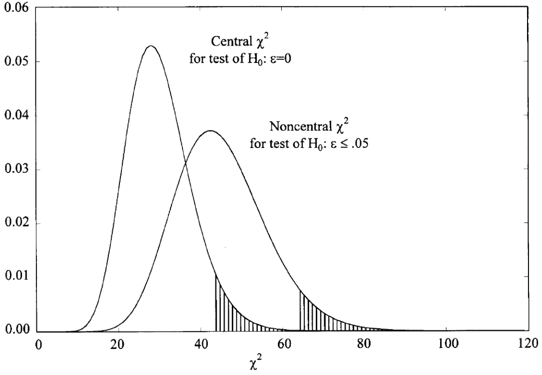
$$\begin{aligned} Power &= P(\hat{W} > c_\alpha | H_1) = 1 - [\chi^2_{d,\lambda_1}(c_\alpha)]^{-1} \\ &= 1 - \chi^2_{d,\lambda_1}[(\chi^2_{d,\lambda_0,\alpha})]^{-1} \end{aligned}$$

where $[\chi^2_{d,\lambda_0,\alpha}]^{-1}$ is the $100\alpha th$ percentile of the chi-squared distribution under H_0. For the not-close fit, the power is

$$\begin{aligned} Power &= P(\hat{W} < c_\alpha | H_1) \\ &= [\chi^2_{d,\lambda_1}(c_\alpha)]^{-1} = \chi^2_{d,\lambda_1}[(\chi^2_{d,\lambda_0,\alpha})]^{-1}. \end{aligned}$$

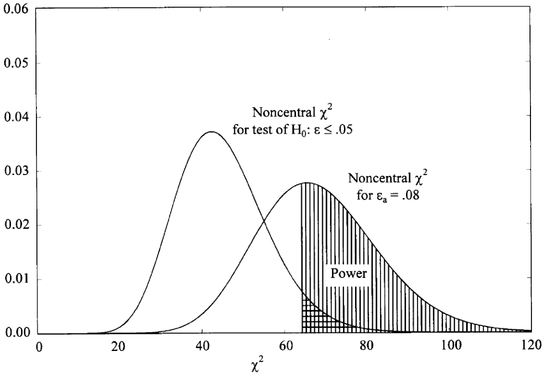
Different tests I

- Chi-square distribution under $\epsilon = 0$ and $\epsilon = .05$
  - If the true is 0.05 and we test 0, what the power of the test to reject H_0
  - If the model fit is close and we test for the exact fit, what is the likelihood to reject the null hypothesis?



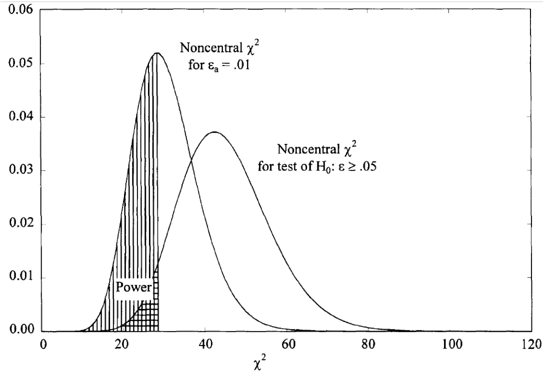
Different tests II

- Test of close fit $H_0 : \epsilon \leq 0.05$ vs. $H_1 : \epsilon = 0.08$
  - If the true is 0.08 and we test 0.05, what the power of the test to reject H_0
  - If the model fit is mediocre and we test for the close fit, what is the likelihood to reject the null hypothesis?

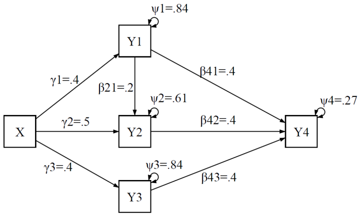


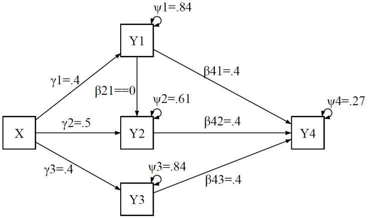
Different tests III

- Test of not-close fit $H_0 : \epsilon \geq 0.05$ vs. $H_1 : \epsilon = 0.01$
  - If the model fits extremely well, what is the likelihood to reject the null hypothesis that the model does not fit well?



Example. Power





SEM based on RMSEA

Parameters (Help)	
Sample size	100
Degrees of freedom	4
RMSEA for H0	0
RMSEA for H1	0.116
Significance level	0.05
Power	
Type of analysis	Close fit ▼
Power curve	No power curve ▼
Note	SEM based on RMSEA

Calculate

SEM based on RMSEA

```
  n Power RMSEA0 RMSEA1 df alpha
100 0.421      0  0.116  4  0.05
```

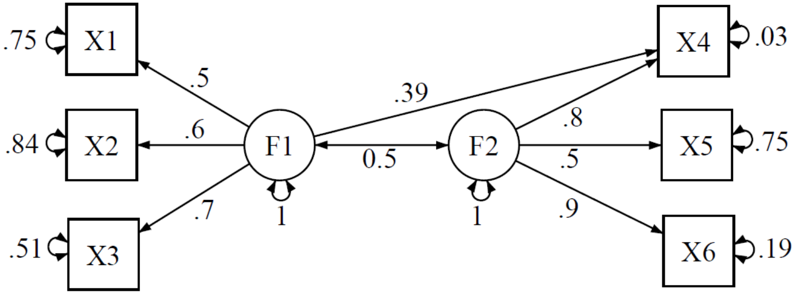
Power analysis using R I

```
> ## SEM RMSEA
> # 1. Power
> wp.sem.rmsea(n = 100, df = 4, rmsea0 = 0,
   rmsea1 = .116, power = NULL, alpha = 0.05)

Power analysis for SEM based on RMSEA

      n df rmsea0 rmsea1     power alpha
    100  4      0  0.116 0.4208173  0.05

>
> # 2. Sample size
> wp.sem.rmsea(n = NULL, df = 4, rmsea0 = 0,
   rmsea1 = 0.116, power = 0.8, alpha = 0.05)

Power analysis for SEM based on RMSEA
```

Power analysis using R II

```
        n df rmsea0 rmsea1 power alpha
 222.7465  4      0  0.116   0.8  0.05

>
> # 3. Effect size
> wp.sem.rmsea(n = 100, df = 4, rmsea0 = 0,
  rmsea1 = NULL, power = 0.8, alpha = 0.05)

Power analysis for SEM based on RMSEA

     n df rmsea0    rmsea1 power alpha
   100  4      0 0.1736082   0.8  0.05
```

Practice



- The full model is given above. The power analysis concerns the cross loading from F1 to X4.
  1. Obtain the effect size for the Satarro & Sarris method and the RMSEA
  2. What's the power for a sample size 100?
  3. To get a power 0.8, what's the needed sample size?

Power analysis for multilevel modeling

Multilevel Modeling — When?

In educational studies, the total sample size is often a combination of students sampled from different classrooms or schools. When data exhibit such nested structure, multilevel modeling can be conducted.

Student (ID)	School (Name)	Verbal Score
1	Potato	88
2	Potato	85
3	Potato	92
4	Tomato	76
5	Tomato	78
6	Tomato	80
⋮	⋮	
60	Sheep	77

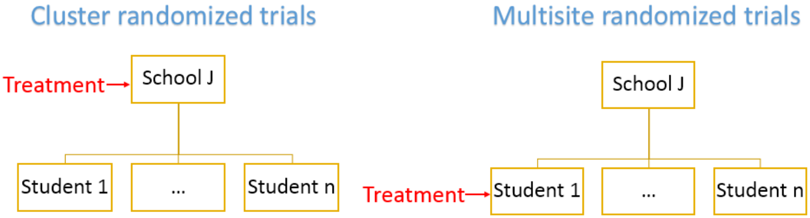
Multilevel Modeling — Why?

When data are nested, it is natural that the individuals within the same cluster (e.g., school) are correlated, which violates one of the assumptions of traditional models such as multiple regression and ANOVA. As a consequence, traditional models will produce biased estimates of parameter standard errors, and thus lead to significance tests with inflated type I error rates (e.g., Hox, 1998).

Advantages of using multilevel modeling:

- Handle nested data
- Allow us to know both individual and cluster differences
- More powerful

Multilevel Modeling (CRT vs MRT)



CRT:

- The entire site (school) is randomly assigned to treatment or control.
- Avoids a possible "spill over" effect within schools.

MRT:

- Students within schools are randomly assigned.
- More convenient and economical because we have a larger pool.
- Easy to manage because each cluster follows the same study design.

Power Analysis for CRT (1 treatment & 1 control)

$$Y_{ij} = \beta_{0j} + e_{ij},\ e_{ij} \sim N(0, \sigma_W^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_j + u_{0j},\ u_{0j} \sim N(0, \sigma_B^2)$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (cluster);

X_j: treatment indicator of cluster j, $X_j = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$

γ_{00} : grand mean;

γ_{01} : treatment main effect (i.e., $\mu_D = \mu_T - \mu_C$)

β_{0j} : cluster mean

σ_W^2 : within-cluster variance; σ_B^2 : between-cluster variance

Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect
$H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma_{01}})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under $H_0 : T \sim t_{J-2}$.

Under $H_1 : T \sim t_{J-2,\lambda}$.

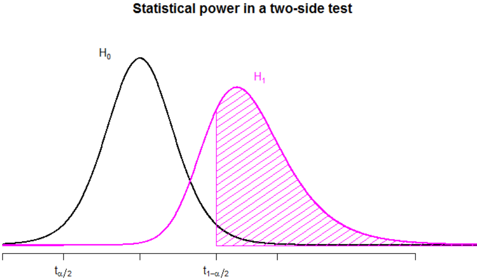
Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect
$H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma_{01}})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under $H_0 : T \sim t_{J-2}$.

Under $H_1 : T \sim t_{J-2,\lambda}$.



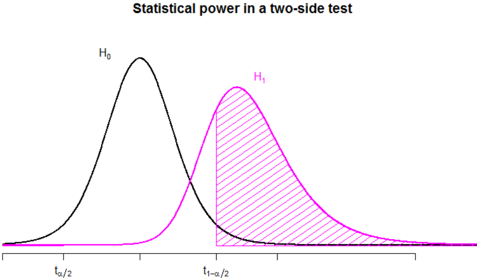
Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect
$H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma_{01}})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under H_0 : $T \sim t_{J-2}$.

Under $H_1 : T \sim t_{J-2,\lambda}$.



$$\text{Power} = P(\text{reject } H_0 | H_1 \text{ true})$$
$$= \begin{cases} 1 - P[T_{J-2,\lambda} < t_0] + P[T_{J-2,\lambda} \leq -t_0] & \text{two}-\text{sided;} \\ 1 - P[T_{J-2,\lambda} < t_0] & \text{one}-\text{sided,} \end{cases}$$

Power Analysis for CRT (1 treatment & 1 control)

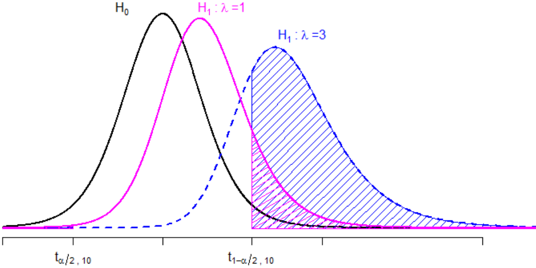
$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}}.$$

- As λ increases, power increases.
- λ is a function of μ_D, n, J, σ_B^2 and σ_W^2.
- To give more meaningful definition, we can reparameterize λ in terms of effect size and intra-class correlation (ICC).

Power Analysis for CRT (1 treatment & 1 control)

$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}}.$$

- As λ increases, power increases.
- λ is a function of μ_D, n, J, σ_B^2 and σ_W^2.
- To give more meaningful definition, we can reparameterize λ in terms of effect size and intra-class correlation (ICC).



ICC in CRT

The intra-class correlation (ICC) quantifies the degree to which two randomly drawn observations within a cluster are correlated. In CRT, the ICC is defined as

$$\rho = corr(Y_{ij}, Y_{i'j}) = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2}.$$

- The proportion of total variance that is accounted for by clustering.
- $\rho = 0$, no between cluster variation.
- As ρ increases, more variation is due to between-cluster variability.
- For school-based data sets, ρ usually ranges between 0.10 to 0.30. (Bloom, Bos & Lee, 1999; Hedges & Hedberg, 2007)

Effect Size in CRT (1 treatment & 1 control)

The effect sizes used in educational and psychological research are typically standardized mean differences. Possible definitions for the effect size in CRT (Hedges, 2007):

- $f = \mu_D/\sigma_W$. This effect size might be of interest in a meta-analysis where the studies being compared are single-site studies.
- $f = \mu_D/\sigma_B$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies that have been analyzed by using cluster means as the unit of analysis.
- $f = \mu_D/\sqrt{\sigma_B^2 + \sigma_W^2}$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies or studies that sample from a broader population but do not include clusters.

Power Analysis for CRT (1 treatment & 1 control)

Redefine λ in standardized notation:

$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}} = \frac{\sqrt{J}f}{\sqrt{4(\rho + \frac{1-\rho}{n})}}.$$

Now, λ is a function of n, J, f and ρ.

- As J or n increases, λ increases and thus power increases.
- As f increases, λ increases and thus power increases.
- As ρ increases, λ decreases and thus power decreases.

Example. Power

A group of educational researchers developed a new teaching method to help students improve their math scores. They plan to randomly assign 5 classrooms to the new method and 5 classrooms to the standard method. Each classroom has 20 students. Based on their prior knowledge, they hypothesize that the effect size is 0.6 and the intra-class correlation is 0.1. What is the power for them to find a significant difference between the standard classrooms and those using the new teaching method?

Parameters (Help)	
Sample size	20
Effect size (Calculator)	0.6
Number of clusters	10
Intra-class correlation	0.1
Power	
Significance level	0.05
Power curve	No power curve ▼
Type of analysis	Two-sided test ▼
Note	Cluster randomized trials wi

Calculate

```
 n  J power   f alpha
20 10  0.59 0.6  0.05
```

Note. n is the number of observations in each cluster.

Power Analysis for CRT (2 treatments & 1 control)

$$Y_{ij} = \beta_{0j} + e_{ij},\ e_{ij} \sim N(0, \sigma_W^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{1j} + \gamma_{02} X_{2j} + u_{0j},\ u_{0j} \sim N(0, \sigma_B^2)$$

$$X_{1j} = \begin{cases} 1/3 & \text{treatment1} \\ 1/3 & \text{treatment2} \\ -2/3 & \text{control} \end{cases};\quad X_{2j} = \begin{cases} 1/2 & \text{treatment1} \\ -1/2 & \text{treatment2} \\ 0 & \text{control} \end{cases}$$

β_{0j} : cluster mean; γ_{00} : grand mean

γ_{01}: mean difference between the average of the two treatments and the control

γ_{02}: mean difference between the two treatments

$$Y_{ij} = \gamma_{00} + \gamma_{01} X_{1j} + \gamma_{02} X_{2j} + u_{0j} + e_{ij}$$

$$\begin{cases} \mu_{T1} = \gamma_{00} + \frac{1}{3}\gamma_{01} + \frac{1}{2}\gamma_{02} \\ \mu_{T2} = \gamma_{00} + \frac{1}{3}\gamma_{01} - \frac{1}{2}\gamma_{02} \\ \mu_C = \gamma_{00} - \gamma_{01} \end{cases} \Longrightarrow \begin{cases} 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = \gamma_{01} \\ \mu_{T1} - \mu_{T2} = \gamma_{02} \end{cases}$$

Power Analysis for CRT (2 treatments & 1 control)

1. Test treatment main effect:

   $H_0: \gamma_{01} = 0 \Leftrightarrow \mu_D = 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = 0$
   Under $H_0: T_1 \sim t_{J-3}$. Under $H_1: T_1 \sim t_{J-3,\lambda_1}$, where

   $$\lambda_1 = \frac{\sqrt{J} f_1}{\sqrt{4.5(\rho + \frac{1-\rho}{n})}} \text{ and } f_1 = \frac{0.5(\mu_{T1} + \mu_{T2}) - \mu_C}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

2. Comparing the two treatments:

   $H_0: \gamma_{02} = 0 \Leftrightarrow \mu_D = \mu_{T1} - \mu_{T2} = 0$
   Under $H_0: T_2 \sim t_{J-3}$. Under $H_1: T_2 \sim t_{J-3,\lambda_2}$, where

   $$\lambda_2 = \frac{\sqrt{J} f_2}{\sqrt{6(\rho + \frac{1-\rho}{n})}} \text{ and } f_2 = \frac{\mu_{T1} - \mu_{T2}}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

3. Ominibus test: $H_0: \gamma_{01} = \gamma_{02} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$
   Under $H_0: F \sim F_{2,J-3}$. Under $H_1: F \sim F_{2,J-3,\lambda}$, where
   $\lambda = \lambda_1^2 + \lambda_2^2$.

   $$f_3 = \sqrt{\frac{\frac{1}{18}(\mu_{D1} + \mu_{D2})^2 + \frac{1}{6}(\mu_{D1} - \mu_{D2})^2}{\sigma_B^2 + \sigma_W^2}}$$

Example. Power

A medical researcher plans to compare two sleep aids and a placebo in helping sleep disorders. The outcome variable is self-reported sleep quality. The researcher plans to conduct the study in 21 clinics, with one-third receiving treatment 1, one-third receiving treatment 2, and the rest receiving placebo. Suppose there are 20 patients in each clinic. Past study reveals that effect size for comparing the two sleep aids to the placebo is 0.5 and the intra-class correlation is 0.1. What is the power for detecting a significance difference between the average treatments and the placebo?

Parameters (Help)	
Sample size	20
Effect size (Calculator)	0.5
Number of clusters	21
Intra-class correlation	0.1
Power	
Significance level	0.05
Power curve	No power curve ▾
H1	Two-sided test ▾
Type of analysis	Average treatment v.s. control ▾
Note	Cluster Randomized Trials \

Calculate

```
 J  n power   f alpha
21 20 0.765 0.5  0.05
```

Using R I

```
> ### Multilevel modeling
> ## Cluster randomized trials with two arms
>
> # 1. Power
> wp.crt2arm(f=0.6,n=20,J=10,icc=.1)

Multilevel model cluster randomized trials
   with two arms

     J  n   f icc     power alpha
    10 20 0.6 0.1 0.5901684  0.05
>
> # 2. Number of cluster
> wp.crt2arm(f=0.6,n=20,J=NULL,icc=.1,power
   =.8)
```

Using R II

```
Multilevel model cluster randomized trials
    with two arms

             J   n    f icc power alpha
     14.83587 20 0.6 0.1   0.8  0.05

>
> ## Cluster randomized trials with three
    arms
>
> # 1. Power
> wp.crt3arm(n=20, f=.5, J=21, icc=.1, power=
    NULL)

Multilevel model cluster randomized trials
    with three arms
```

Using R III

```
   J  n   f icc     power alpha
  21 20 0.5 0.1 0.7650611  0.05

>
> # 2. Sample size
> wp.crt3arm(n=NULL, f=.5, J=21, icc=.1,
   power=.8)

Multilevel model cluster randomized trials
   with three arms

   J        n   f icc power alpha
  21 27.34175 0.5 0.1   0.8  0.05
```

Power Analysis for MRT (1 treatment & 1 control)

For multisite randomized trials with 1 treatment and 1 control:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j}.\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(\mathbf{0}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix})$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (site);

X_{ij}: indicator of treatment assignment with $X_{ij} = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$

β_{0j} : mean at the jth site

β_{1j} : mean difference between treatment and control at the jth site

γ_{00} : grand mean;

γ_{10} : treatment main effect

σ^2 : level-1 error variance

τ_{00} : site variability

τ_{11} : variance of site-specfic treatment effects

Power Analysis for MRT (1 treatment & 1 control)

For multisite randomized trials with 1 treatment and 1 control:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j}.\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(\mathbf{0}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix})$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (site);

X_{ij}: indicator of treatment assignment with $X_{ij} = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$

β_{0j} : mean at the jth site

β_{1j} : mean difference between treatment and control at the jth site

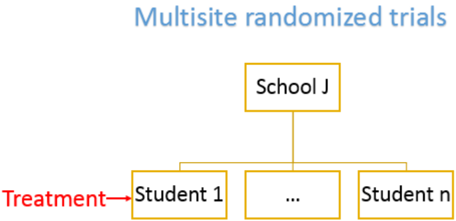
γ_{00} : grand mean;

γ_{10} : treatment main effect

σ^2 : level-1 error variance

τ_{00} : site variability

τ_{11} : variance of site-specfic treatment effects



Power Analysis for MRT (1 treatment & 1 control) I

Power for testing treatment main effect $H_0 : \gamma_{10} = 0$ is calculated by

$$\text{Power} = \begin{cases} 1 - P[T_{J-1,\lambda} < t_0] + P[T_{J-1,\lambda} \leq -t_0] & \text{two} - \text{sided test}, \\ 1 - P[T_{J-1,\lambda} < t_0] & \text{one} - \text{sided test}, \end{cases} \quad (2)$$

where

$$\lambda = \frac{\gamma_{10}}{\sqrt{(\frac{4\sigma^2}{n} + \tau_{11})/J}}, \quad (3)$$

t_0 is the $100(1 - \frac{\alpha}{2})$th percentile for a two-sided test and the $100(1 - \alpha)$th percentile for a one-sided test of the t distribution with $J - 1$ degrees of freedom, and α is the significance level.

Power Analysis for MRT (1 treatment & 1 control) II

The effect size for the main effect is defined as

$$f_1 = \frac{\mu_D}{\sqrt{\sigma^2}}, \tag{4}$$

where μ_D is the mean difference between the treatment and control across all the sites, σ^2 is the level-1 error variance.

Example. Power I

A researcher plans to conduct a multisite randomized trial to evaluate the efficacy of an intervention for alcoholics. Patients will be recruited from 20 sites, and at each site half of the patients will be assigned to the treatment condition and the other half will be assigned to the control condition. The number of patients at each site is expected to be 45. The outcome variable is the reduction in abuse symptoms. Past study reveals that the effect size for the treatment is 0.5. Further, the researcher estimates that the level-1 error is 1.25, the variance of site means is 0.1 and the variance in treatment effect across sites is 0.5. What's the power for testing the treatment main effect?

Example. Power II

Parameters (Help)	
Sample size	45
Effect size (Calculator)	0.5
Number of clusters	20
Variance of site means	
Variance of treatment effects across sites	0.5
Level 1 error variance	1.25
Power	
Significance level	0.05
Power curve	No power curve ▼
H1	Two-sided test ▼
Type of test	Treatment main effect ▼
Note	Multisite randomized trials v

Calculate

```
  J  n power   f alpha
20 45 0.858 0.5  0.05
```

Note. Sample size is sample size per cluster.

Power Analysis for MRT (2 treatments & 1 control) I

The model for a 3-arm (2 treatments & 1 control) MRT can be expressed as

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + e_{ij},$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j},\ \beta_{2j} = \gamma_{20} + u_{2j},$$

$$Var(e_{ij}) = \sigma^2,\ Var(u_{0j}) = \tau_{00},\ Var(u_{1j}) = \tau_{11},\ Var(u_{2j}) = \tau_{22},$$

- Y_{ij} is the ith outcome in jth cluster $(i = 1, 2, ..., N;\ j = 1, 2, ...J)$
- X_{1ij} is used to compare the average outcome of the two treatment arms with that of the control arm (1/3 for the first treatment, 1/3 for the second treatment and -2/3 for the control condition)

Power Analysis for MRT (2 treatments & 1 control) II

- X_{2ij} is used to contrast the average outcome between the two treatment arms (1/2 for the first treatment, -1/2 for the second treatment and 0 for the control condition)
- β_{0j} is mean of the jth site
- β_{1j} is the mean difference between average treatment and control of the jth site
- β_{2j} is the mean difference between the two treatments of the jth site
- γ_{00} is grand mean
- γ_{10} is the contrast between average of the two treatments and control
- γ_{20} is the contrast between the two treatments.

Power Analysis for MRT (2 treatments & 1 control) III

Power for testing the treatment main effect
$H_0: \ \gamma_{10} = 0 \Leftrightarrow \frac{1}{2}(\mu_{T1} + \mu_{T2}) = \mu_C$ is calculated by

$$\text{Power} = \begin{cases} 1 - P[T_{J-1,\lambda_1} < t_0] + P[T_{J-1,\lambda_1} \leq -t_0] & \text{two} - \text{sided test}, \\ 1 - P[T_{J-1,\lambda_1} < t_0] & \text{one} - \text{sided test}, \end{cases} \tag{5}$$

where

$$\lambda_1 = \frac{\gamma_{10}}{\sqrt{(4.5\frac{\sigma^2}{n} + \tau_{11})/J}}, \tag{6}$$

t_0 is the $100(1 - \frac{\alpha}{2})$th percentile for a two-sided test and the $100(1 - \alpha)$th percentile for a one-sided test of the t distribution with $J - 1$ degrees of freedom, and α is the significance level.
Power for comparing the two treatments
$H_0: \ \gamma_{20} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2}$ is calculated by

Power Analysis for MRT (2 treatments & 1 control) IV

$$\text{Power} = \begin{cases} 1 - P[T_{J-1,\lambda_2} < t_0] + P[T_{J-1,\lambda_2} \leq -t_0] & \text{two}-\text{sided test}, \\ 1 - P[T_{J-1,\lambda_2} < t_0] & \text{one}-\text{sided test}, \end{cases} \quad (7)$$

where

$$\lambda_2 = \frac{\gamma_{20}}{\sqrt{(6\frac{\sigma^2}{n} + \tau_{22})/J}}, \quad (8)$$

t_0 is the $100(1-\frac{\alpha}{2})$th percentile for a two-sided test and the $100(1-\alpha)$th percentile for a one-sided test of the t distribution with $J-1$ degrees of freedom, and α is the significance level.
Power for the omnibus test
$H_0: \ \gamma_{10} = \gamma_{20} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$ is calculated by

$$\text{Power} = P(F_{2,2(J-1),\lambda_3} \geq F_0), \quad (9)$$

Power Analysis for MRT (2 treatments & 1 control) V

where

$$\lambda_3 = \lambda_1^2 + \lambda_2^2, \tag{10}$$

F_0 is the $100(1-\alpha)$th percentile of the F distribution with degrees of freedom 2 and $2(J-1)$.

Given the mean difference between the treatment 1 and control (μ_{D1}), mean difference between the treatment 2 and control (μ_{D2}), and the Level-one error variance (σ^2), the effect sizes of the first two tests can be calculated as

$$f_1 = \frac{(\mu_{D1} + \mu_{D2})/2}{\sqrt{\sigma^2}}, \tag{11}$$

$$f_2 = \frac{\mu_{D1} + \mu_{D2}}{\sqrt{\sigma^2}}. \tag{12}$$

Note that effect size is not defined under the omnibus test.

Example. Power I

A researcher plans to collect data from 20 clinics to examine the effect of certain behavioral therapies on recovering from anorexia. At each clinic, 30 anorexic girls will be randomly assigned to therapy 1, therapy 2, or the control group. Previous research suggests that therapy 1 might lead to an increase of 0.5 in BMI and therapy 2 might lead to an increase of 0.8 in BMI. Further, the person-specific error variance is 2.25 and the variance in treatment effects across sites is 0.4. What's the power for testing treatment main effect ? Based on the provided information, the effect size is $(0.8 + 0.5)/2/\sqrt{2.25} = 0.43$.

Example. Power II

Parameters (Help)	
Sample size	30
Effect size of treatment main effect (Calculator)	0.43
Effect size of two treatment difference	
Number of clusters	20
Variance of treatment effects across sites	0.4
Level 1 error variance	2.25
Power	
Significance level	0.05
Power curve	No power curve ▾
H1	Two-sided test ▾
Type of analysis	Test treatment main effect ▾
Note	Multisite randomized trials v

Calculate

```
  J  n power   f1 f2 alpha
 20 30 0.807 0.43 NA  0.05
```

Note. Sample size is sample size per cluster.

Power analysis using R I

```
> ## Multisite randomized trials with two
   arms
> # 1. Power
> wp.mrt2arm(n=45, f=0.5, J=20, tau11=.5, sg2
   =1.25, power=NULL)

Multilevel model multisite randomized trials
   with two arms

     J  n   f tau11  sg2     power alpha
    20 45 0.5   0.5 1.25 0.8583253  0.05

>
> # 2. Sample size
> wp.mrt2arm(n=NULL, f=0.5, J=20, tau11=.5,
   sg2=1.25, power=.8)
```

Power analysis using R II

```
Multilevel model multisite randomized trials
    with two arms

     J          n    f tau11   sg2 power alpha
    20 23.10086 0.5    0.5 1.25   0.8  0.05

>
> ## Multisite randomized trials with three
    arms
>
> # 1. Power
> wp.mrt3arm(n=30, f1=0.43, J=20, tau=.4, sg2
    =2.25, power=NULL)
```

Power analysis using R III

```
Multilevel model multisite randomized trials
    with three arms

      J  n    f1 tau  sg2      power alpha
     20 30 0.43 0.4 2.25 0.8066964  0.05

>
> # 2. Sample size
> wp.mrt3arm(n=NULL, f1=0.43, J=20, tau=.4,
    sg2=2.25, power=0.8)

Multilevel model multisite randomized trials
    with three arms

      J        n   f1 tau  sg2 power alpha
     20 28.61907 0.43 0.4 2.25   0.8  0.05
```

A general Monte Carlo based methods

Monte Carlo method

- By definition, statistical power

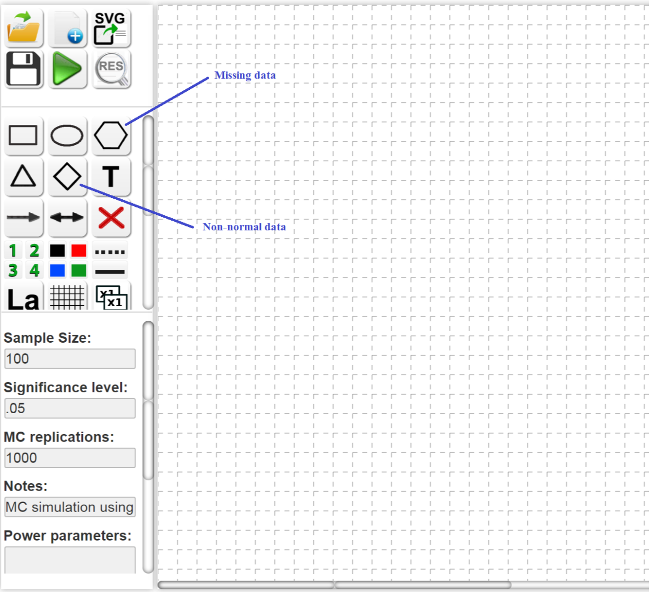
$$\pi = \Pr(\text{reject } H_0 | H_1)$$

- The power can be empirically estimated through a Monte Carlo procedure.
  - Specify a model with population parameter.
  - Generate R, for example $R = 1000$, sets of data with a chosen sample size.
  - For each sets of data, fit the model and test the significance of a parameter based on a test.
  - If for r sets of data, the results are significant, then the power is

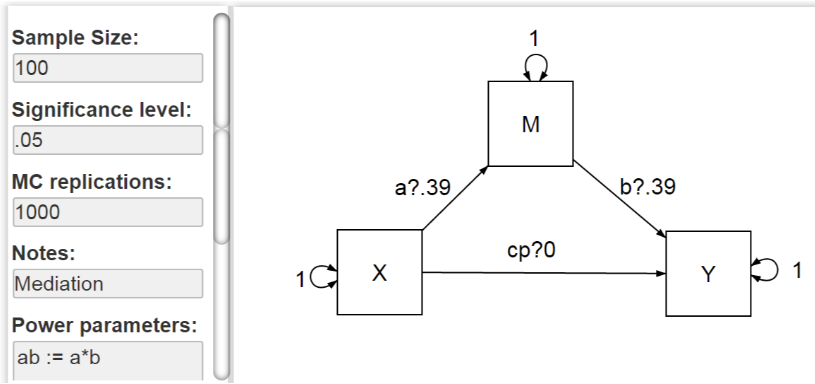
$$power = \frac{r}{R}.$$

- WebPower implements the procedure and also allows non-normal data and missing data

WebPower diagram based Monte Carlo method



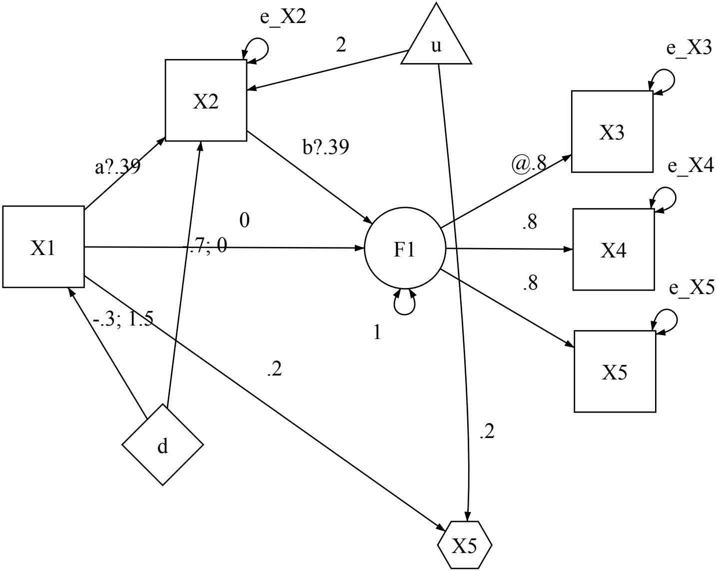
Example. Mediation I



Example. Mediation II

```
Basic information:

  Esimation method                                    ML
  Standard error                                standard
  Number of requested replications                  1000
  Number of successful replications                 1000

                       True  Estimate     MSE      SD     Power Power.se Coverage
Regressions:
  M ~
    X          (a)    0.390     0.379   0.100   0.096     0.968    0.006    0.962
  Y ~
    M          (b)    0.390     0.391   0.100   0.105     0.960    0.006    0.936
    X         (cp)    0.000    -0.001   0.106   0.112     0.059    0.007    0.941

Intercepts:
    M                 0.000     0.003   0.099   0.099     0.051    0.007    0.949
    Y                 0.000     0.003   0.099   0.100     0.064    0.008    0.936
    X                 0.000    -0.006   0.099   0.102     0.052    0.007    0.948

Variances:
    M                 1.000     0.979   0.138   0.146     1.000    0.000    0.918
    Y                 1.000     0.964   0.136   0.138     1.000    0.000    0.906
    X                 1.000     0.992   0.140   0.143     1.000    0.000    0.926

Indirect/Mediation effects:
    ab                0.152     0.148   0.055   0.055     0.863    0.011    0.928
```

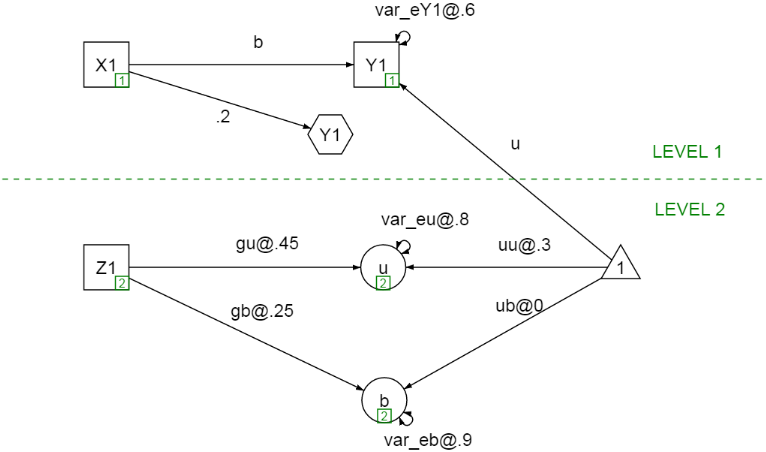
Example. Non-normal and missing data I



Example. Non-normal and missing data II

```
Basic information:

  Esimation method                                  ML
  Standard error                            robust.huber.white
  Number of requested replications                1000
  Number of successful replications               1000

                       True  Estimate     MSE      SD     Power Power.se Coverage
Latent variables:
  F1 =~
    X3                0.800    0.800    0.000    0.000      NaN      NaN    0.000
    X4                0.800    0.851    0.311    0.328    0.908    0.009    0.934
    X5                0.800    0.816    0.285    0.298    0.877    0.010    0.913

Regressions:
  X2 ~
    X1         (a)    0.390    0.395    0.097    0.098    0.971    0.005    0.938
  F1 ~
    X2         (b)    0.390    0.394    0.146    0.153    0.766    0.013    0.922
    X1                0.000   -0.009    0.147    0.145    0.043    0.006    0.957

Intercepts:
    X2                2.000    2.001    0.099    0.098    1.000    0.000    0.946
    X3                0.000   -0.007    0.266    0.283    0.082    0.009    0.918
    X4                0.000    0.000    0.270    0.275    0.069    0.008    0.931
    X5                0.000    0.004    0.318    0.327    0.072    0.008    0.928
    F1                0.000    0.000    0.000    0.000      NaN      NaN    0.000

Variances:
    X2      (e_X2)    1.000    0.977    0.136    0.141    1.000    0.000    0.902
    F1                1.000    1.039    0.483    0.521    0.756    0.014    0.918
    X3      (e_X3)    1.000    0.938    0.312    0.334    0.862    0.011    0.953
    X4      (e_X4)    1.000    0.947    0.295    0.289    0.879    0.010    0.957
    X5      (e_X5)    1.000    0.936    0.316    0.314    0.855    0.011    0.913

Indirect/Mediation effects:
    ab                0.152    0.155    0.071    0.073    0.629    0.015    0.909
```

Example. A two-level model I



Example. A two-level model II

Power Analysis Results

```
  Number of requested replications                1000
  Number of successful replications               1000

Level    Equation             Est       SD       SE     Power
        Fixed effects
    2     u~1               0.296    0.206    0.203    0.319
    2     b~1              -0.008    0.213    0.217    0.055
    2     u~Z1              0.454    0.224    0.210    0.605
    2     b~Z1              0.246    0.234    0.225    0.233
       Random effects
    1    Y1~~Y1            0.601    0.028
    2    u~~u              0.787    0.273
    2    b~~b              0.907    0.308
    2    u~~b             -0.001    0.196
```

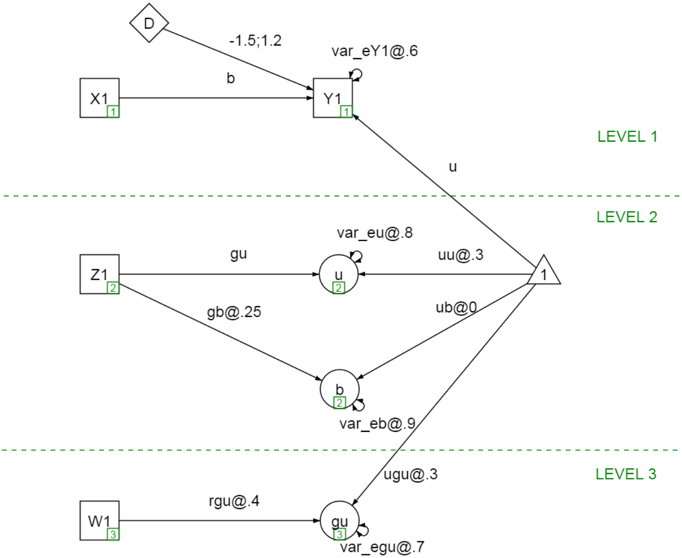
==

WebPower ended at 14:14:44 on Jul 22, 2016

Time spent on the analysis

```
   user  system elapsed
840.682   1.585 841.805
```

Example. A three-level model I



Example. A three-level model II

Power Analysis Results

```
  Number of requested replications              1000
  Number of successful replications              386

Level    Equation             Est      SD       SE     Power
        Fixed effects
    2     u~1               0.300    0.018    0.019    1.000
    2     b~1               0.001    0.020    0.019    0.065
    3     gu~1              0.305    0.120    0.119    0.712
    3     gu~W1             0.391    0.122    0.120    0.894
    2     b~Z1              0.249    0.020    0.019    1.000
       Random effects
    1    Y1~~Y1            0.600    0.002
    2    u~~u              0.796    0.023
    2    b~~b              0.901    0.026
    2    u~~b              0.000    0.017
    3    gu~~gu            0.688    0.142
```

Time spent on the analysis

```
     user   system  elapsed
260982.2    127.9 260938.4
```

Summary

Method	WebPower URL	R function
proportion	http://w.psychstat.org/prop	wp.prop
t-test	http://w.psychstat.org/ttest	wp.t
correlation	http://w.psychstat.org/correlation	wp.correlation
one-way ANOVA	http://w.psychstat.org/anova	wp.anova
two-way ANOVA	http://w.psychstat.org/kanova	wp.kanova
Linear regression	http://w.psychstat.org/regression	wp.regression
Logistic regression	http://w.psychstat.org/logistic	wp.logistic
Poisson regression	http://w.psychstat.org/poisson	wp.poisson
Simple mediation	http://w.psychstat.org/mediation	wp.mediation
SEM Satorra & Saris	http://w.psychstat.org/semchisq	wp.sem.chisq
SEM RMSEA	http://w.psychstat.org/rmsea	wp.sem.rmsea
CRT 2 arms	http://w.psychstat.org/crt2arm	wp.crt2arm
CRT 3 arms	http://w.psychstat.org/crt3arm	wp.crt3arm
MRT 2 arms	http://w.psychstat.org/mrt2arm	wp.mrt2arm
MRT 3 arms	http://w.psychstat.org/mrt3arm	wp.mrt3arm
Path diagram	http://w.psychstat.org/diagram	N/A

Acknowledgment

- The Power team: Ke-Hai Yuan, Yujiao Mai, Meghan Cain, Miao Yang, Ge Jiang, Agung Santoso, Haiyan Liu, Han Du, Lin Xing
- Supported by the Institute of Education Sciences at the Department of Education (R305D140037)
- For more information, contact Zhiyong Zhang (zzhang4@nd.edu).
- Website: `https://WebPower.psychstat.org`

Thank you! Questions and Comments?