Session B:

Introduction to Power Analysis

Kevin J. Grimm
July 8, 2009
University of California, Davis

Overview

- Power Analysis
  - What is it?
  - Why do we care about it?
  - What affects it?
  - How can we calculate it?
- Simple Example
  - Independent Groups t-test
    - Determining Power
    - Determining Sample Size

Power

- Probability of Rejecting the Null Hypothesis when the Null Hypothesis is FALSE
  - Correctly Rejecting the Null Hypothesis
  - Finding the effect when there actually is an effect
  - 1 - \( \beta \)

Review

- Type I Error
  - Rejecting the Null Hypothesis \( (H_0) \) when the Null Hypothesis is TRUE
  - The probability of making a Type I error is \( \alpha \)
- Type II Error
  - Retaining the Null Hypothesis when the Null Hypothesis is FALSE
  - The probability of making a Type II error is \( \beta \)
Graphical Representation of Power

**Power**

- Most applied statistical work has focused on experimental data
  - Primary concern was minimizing the probability of a *Type I error* (*alpha*)
    - *Rejecting the Null Hypothesis* when the Null Hypothesis is TRUE
    - Because we know the sampling distribution of the statistic under the *Null Hypothesis*

- Researchers have increasingly been interested in the probability of making a *Type II error*
  - *Retaining the Null Hypothesis* when the Null Hypothesis is *FALSE*
  - Considering the substantial cost (money, time, effort) that goes into an experiment, we should know the probability of making a Type II error
  - If the probability of making a Type II error is high, then why should we even do the experiment?

- **Challenge**
  - We don’t know the sampling distribution under the *Alternative Hypothesis*
    - We don’t know how the population means differ
    - We don’t know the population correlation

- **Solution**
  - Let’s try a variety of differences and evaluate *beta* (probability of making a *Type II error*) at these different mean differences
What affects Power?

1. Alpha
   Probability of Type I Error – usually held constant at $\alpha = .05$

2. The true alternative hypothesis
   Effect Size
   1. Difference between population means in standardized units (Cohen’s $d$)
   2. Population Correlation
   3. Generally – population parameters (e.g., ICC, factor loadings, etc.)

3. Sample size

Effect Size

- Larger the association, the greater the power

- Affects the distance between the sampling distributions under the Null and Alternative Hypotheses

Effect Size

- How far apart the means actually are in a standardized metric
  - Cohen’s $d$: absolute distance between means in terms of standard deviations
  \[
  d = \frac{\mu_1 - \mu_0}{\sigma}
  \]
**Effect of Effect Size on Power**

- As the effect size (true mean difference) increases, power increases.
- The effect size does not have an effect on *alpha*, the probability of making a *Type I Error*.
  - *Alpha* remains .05.
- Increasing the effect size decreases *beta*, the probability of making a *Type II Error*.

**Cohen’s d Effect Sizes**

<table>
<thead>
<tr>
<th>Effect size</th>
<th>d</th>
<th>% Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>.20</td>
<td>85</td>
</tr>
<tr>
<td>Medium</td>
<td>.50</td>
<td>67</td>
</tr>
<tr>
<td>Large</td>
<td>.80</td>
<td>53</td>
</tr>
</tbody>
</table>

Note: A ‘medium’ effect size is one-half population standard deviation difference.
**Sample Size**

- Changing the sample size has a direct effect on the Sampling Distribution of the Mean (or Mean Difference)

\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]

- Increasing Sample Size decreases the standard error of the mean (or mean difference) resulting in less variance in the sampling distribution of the mean (or mean difference)
- Increasing Sample Size, Increases Power by reducing the Standard Error

**Graphical View of the Effect of Sample Size on Power**

(A) Small Sample Size

(B) Large Sample Size

**Calculating Power**

- Two main methods
  1. Analytically using non-centrality parameter from expected study information (e.g., sample size) and population parameters (e.g., Cohen’s \( d \), population correlation, population covariance matrix)
  2. Monte Carlo Simulation
     1. Simulate a population with desired characteristics (e.g., sample size, population parameters, etc.)
     2. Run statistical model
     3. Output Results
     4. Repeat several (~100 – 10,000) times
     5. Examine significance of parameter of interest, model fit, change in model fit over the replications

**Example #1**

- Experimental Study
  - Study Design – Two groups
  - Effect Size – Mean Difference (Cohen’s \( d \))
  - Expected Effect Size – \( d = .2 \)
  - Sample size = 400 (200 per group)
  - Statistical Model – Independent Sample \( t \)-test
  - What’s our Power?
**Method 1 – Analytical**

- Independent Samples \( t \)-test

\[
\delta = d \sqrt{\frac{n}{2}}
\]

Where \( \delta \) is the *non-centrality* parameter, \( d \) is the effect size (standardized mean difference), and \( n \) is the sample size (in each group for the independent samples \( t \)-test)

---

**Calculations**

- Important Information
  - \( d = .20 \)
  - \( n = 200 \)

\[
\delta = d \sqrt{\frac{n}{2}} = .2 \sqrt{\frac{200}{2}} = .2 \sqrt{100} = 2
\]

---

**Power Table (Available in Most Texts)**

<table>
<thead>
<tr>
<th>( \alpha ) for Two-Tailed Test</th>
<th>( \alpha ) for Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta )</td>
<td>( 0.10 )</td>
</tr>
<tr>
<td>( 1.00 )</td>
<td>0.26</td>
</tr>
<tr>
<td>( 1.10 )</td>
<td>0.29</td>
</tr>
<tr>
<td>( 1.20 )</td>
<td>0.33</td>
</tr>
<tr>
<td>( 1.30 )</td>
<td>0.37</td>
</tr>
<tr>
<td>( 1.40 )</td>
<td>0.40</td>
</tr>
<tr>
<td>( 1.50 )</td>
<td>0.44</td>
</tr>
<tr>
<td>( 1.60 )</td>
<td>0.48</td>
</tr>
<tr>
<td>( 1.70 )</td>
<td>0.52</td>
</tr>
<tr>
<td>( 1.80 )</td>
<td>0.56</td>
</tr>
<tr>
<td>( 1.90 )</td>
<td>0.60</td>
</tr>
<tr>
<td>( 2.00 )</td>
<td>0.64</td>
</tr>
<tr>
<td>( 2.10 )</td>
<td>0.68</td>
</tr>
<tr>
<td>( 2.20 )</td>
<td>0.71</td>
</tr>
<tr>
<td>( 2.30 )</td>
<td>0.74</td>
</tr>
<tr>
<td>( 2.40 )</td>
<td>0.78</td>
</tr>
<tr>
<td>( 2.50 )</td>
<td>0.80</td>
</tr>
<tr>
<td>( 2.60 )</td>
<td>0.83</td>
</tr>
<tr>
<td>( 2.70 )</td>
<td>0.85</td>
</tr>
<tr>
<td>( 2.80 )</td>
<td>0.88</td>
</tr>
<tr>
<td>( 2.90 )</td>
<td>0.90</td>
</tr>
<tr>
<td>( 3.00 )</td>
<td>0.91</td>
</tr>
</tbody>
</table>

---

**Example #1 – Power**

- Power = .52

– 52% of observing a small effect with a sample size of 400 with an independent sample \( t \)-test *IF* this is the effect in the population
Method #2
SAS Script for Generating Group Data

DATA ex1;
  seed = 20090708; *Start of the Random Number Generator;
*Population Parameters;
  N = 200;
  mu_0 = 0;
  mu_1 = .2;
  sigma = 1;
  DO group = 1 to 2;
    IF group = 1 THEN mu = mu_0;
    IF group = 2 THEN mu = mu_1;
    DO i = 1 to N;
      x_i = mu + sigma * RANNOR(seed);
      OUTPUT;
    END;
  END;
RUN;

Simulated Data

<table>
<thead>
<tr>
<th>Obs</th>
<th>group</th>
<th>x_i</th>
<th>Obs</th>
<th>group</th>
<th>x_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.14173</td>
<td>201</td>
<td>2</td>
<td>1.65882</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1.28428</td>
<td>202</td>
<td>2</td>
<td>0.42147</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.44383</td>
<td>203</td>
<td>2</td>
<td>1.71208</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.67989</td>
<td>204</td>
<td>2</td>
<td>-0.51501</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.98533</td>
<td>205</td>
<td>2</td>
<td>0.48204</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-0.85084</td>
<td>206</td>
<td>2</td>
<td>-0.02383</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>-0.09704</td>
<td>207</td>
<td>2</td>
<td>1.19980</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.29904</td>
<td>208</td>
<td>2</td>
<td>0.61976</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.01972</td>
<td>209</td>
<td>2</td>
<td>1.15154</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>-0.81942</td>
<td>210</td>
<td>2</td>
<td>-0.77830</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>-0.40004</td>
<td>211</td>
<td>2</td>
<td>0.77641</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0.73694</td>
<td>212</td>
<td>2</td>
<td>1.55916</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>-0.05544</td>
<td>213</td>
<td>2</td>
<td>-0.85130</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0.28362</td>
<td>214</td>
<td>2</td>
<td>-1.09778</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>-1.47824</td>
<td>215</td>
<td>2</td>
<td>-0.29695</td>
</tr>
</tbody>
</table>

PROC TTEST

*Independent Samples T-Test;
PROC TTEST DATA = ex1;
  CLASS group;
  VAR x_i;
RUN;

PROC t-test Output

The TTEST Procedure

<table>
<thead>
<tr>
<th>group</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Err</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>-0.0473</td>
<td>1.0461</td>
<td>0.0740</td>
<td>-2.3043</td>
<td>2.5463</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>0.2010</td>
<td>0.8519</td>
<td>0.0602</td>
<td>-1.9781</td>
<td>3.0491</td>
</tr>
<tr>
<td>Diff (1-2)</td>
<td>0.2483</td>
<td>0.9540</td>
<td>0.0954</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Method Variances | DF | t Value | Pr > |t| |
|------------------|----|---------|------|---|
| Pooled Equal     | 398| -2.60   | 0.0096 |
| Satterthwaite Unequal | 382.31 | -2.60 | 0.0096 |
Let's run this same example, but for many datasets and find out how many times we have a significant mean difference.

```
DATA ex2;
  seed = 10081023;
  datasets = 10000;
  N = 200;
  mu_0 = 0;
  mu_1 = .2;
  sigma = 1;
  DO d = 1 to datasets;
    DO group = 1 to 2;
      IF group = 1 THEN mu = mu_0;
      IF group = 2 THEN mu = mu_1;
      DO i = 1 to N;
        x_i = mu + sigma * RANNOR(seed);
        OUTPUT;
      END;
    END;
  END;
RUN;
```

Calculate t-statistic for each replication and determine whether it is significant.

```
PROC TTEST DATA = ex2;
  CLASS group;
  VAR x_i;
  ODS OUTPUT ttests = output1;
  BY d;
RUN;
```

```
DATA output2;
  SET output1;
  IF Method ne 'Pooled' THEN DELETE;
  sig = 0;
  IF probt < .05 THEN sig = 1;
RUN;
```

```
PROC FREQ;
  TABLES sig;
RUN;
```

**Results**

The FREQ Procedure

<table>
<thead>
<tr>
<th>sig</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4871</td>
<td>48.71</td>
</tr>
<tr>
<td>1</td>
<td>5129</td>
<td>51.29</td>
</tr>
</tbody>
</table>

This means the *Power* to detect a mean difference of .2 standard deviations with a sample size of 400 for an Independent Samples *t*-test is approximately 51% (.51)

**Example #2**

- Experimental Study
  - Study Design – Two groups
  - Effect Size – Mean Difference (Cohen’s *d*)
    - Expected Effect Size – *d* = .4
    - Desired Power – 80%
  - Statistical Model – Independent Sample *t*-test
  - Sample size – How many participants do we need?
**Method 1 – Analytical**

- Independent Samples \( t \)-test

\[
\delta = d \sqrt{\frac{n}{2}}
\]

Where \( \delta \) is the *non-centrality* parameter, \( d \) is the effect size (standardized mean difference), and \( n \) is the sample size (in each group for the independent samples \( t \)-test)

### Calculations

- Important Information
  - \( d = .40 \)
  - Desired Power = .80 (80%)
    - \( \delta = 2.80 \)

\[
\delta = d \sqrt{\frac{n}{2}} \quad \frac{\delta}{d} = \sqrt{\frac{n}{2}} \quad \left( \frac{\delta}{d} \right)^2 = \frac{n}{2} \quad 2\left( \frac{\delta}{d} \right)^2 = n
\]
Method #2 – Simulation

*Example #2 - Two Groups
Cohen’s d = .4
Desired Power = .80
Don’t know how many subjects we need;

DATA ex2;
seed = 20090708;
datasets = 1000;
mu_0 = 0;
mu_1 = .4;
sigma = 1;
DO d = 1 to datasets;
  DO sample = 10 to 300;
    DO group = 1 to 2;
      IF group = 1 THEN mu = mu_0;
      IF group = 2 THEN mu = mu_1;
      DO i = 1 to sample;
        x_i = mu + sigma * RANNOR(seed);
        OUTPUT;
      END;
    END;
  END;
END;
RUN;

Running Regression and Outputting Parameters

PROC REG DATA = ex2 OUTTEST = output2 OUTSEB NOPRINT;
  MODEL x_i = group;
  BY d sample;
RUN;

PROC PRINT DATA = output2 (where = (d = 1));
  VAR d sample _type_ group;
RUN;

Output2 Dataset

<table>
<thead>
<tr>
<th>Obs</th>
<th>d</th>
<th>sample</th>
<th><em>TYPE</em></th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>PARMS</td>
<td>-0.00658</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10</td>
<td>SEB</td>
<td>0.40538</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>11</td>
<td>PARMS</td>
<td>0.12671</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>11</td>
<td>SEB</td>
<td>0.36651</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>12</td>
<td>PARMS</td>
<td>0.37874</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>12</td>
<td>SEB</td>
<td>0.40312</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>13</td>
<td>PARMS</td>
<td>-0.29313</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>13</td>
<td>SEB</td>
<td>0.38551</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>14</td>
<td>PARMS</td>
<td>0.69658</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>14</td>
<td>SEB</td>
<td>0.48966</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>15</td>
<td>PARMS</td>
<td>0.78011</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>15</td>
<td>SEB</td>
<td>0.45467</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>16</td>
<td>PARMS</td>
<td>1.65829</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>16</td>
<td>SEB</td>
<td>0.33278</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>17</td>
<td>PARMS</td>
<td>0.47955</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>17</td>
<td>SEB</td>
<td>0.27636</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>18</td>
<td>PARMS</td>
<td>0.17450</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>18</td>
<td>SEB</td>
<td>0.31785</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>19</td>
<td>PARMS</td>
<td>-0.10001</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>19</td>
<td>SEB</td>
<td>0.26150</td>
</tr>
</tbody>
</table>

Manipulating Datasets & Determining Significance

DATA output2a;
  SET output2;
  IF _TYPE_ ne 'PARMS' THEN DELETE;
RUN;

DATA output2b;
  SET output2;
  IF _TYPE_ ne 'SEB' THEN DELETE;
  se_group = group;
  KEEP d sample se_group;
RUN;

DATA output2c;
  MERGE output2a output2b;
  BY d sample;
  t_value = group/se_group;
  sig = 0;
  IF t_value > 1.96 THEN sig = 1;
  IF t_value < -1.96 THEN sig = -1;
RUN;

PROC FREQ DATA = output2c NOPRINT;
  TABLES sig/OUT=output_f;
  BY d sample;
RUN;
### Output2c Dataset

<table>
<thead>
<tr>
<th>Obs</th>
<th>t_value</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.01623</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.34571</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.93952</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-0.76036</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1.42258</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1.71578</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>4.98319</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1.73526</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0.54899</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>-0.38244</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1.48212</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>-0.01474</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1.19690</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2.94366</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>0.72588</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1.52944</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0.39413</td>
<td>0</td>
</tr>
</tbody>
</table>

### Calculating Average Number of Significant Effects per Sample Size & Plotting Power Curve

```latex
PROC SORT DATA = output_f;
   BY sample;
RUN;

PROC MEANS DATA = output_f;
   VAR sig;
   BY sample;
   OUTPUT OUT = power;
RUN;

DATA power1;
   SET power;
   IF _STAT_ ne 'MEAN' THEN DELETE;
RUN;
PROC GPLOT DATA = power1;
   SYMBOL
     I = join
     V = dot
     H = 5
     W = 2
     C = black
     R = 100000;
   PLOT sig * sample;
RUN;
```

### Power Curve

- Plot of the relationship between sample size (x-axis) and power (y-axis).
- Visual display to help determine appropriate sample size for a given effect size and statistical test.
**PROC POWER**

*To Generate Power Curve*

```sas
PROC POWER;
  TWOSAMPLEMEANS TEST = DIFF
  MEANDIFF = .4
  STDDEV = 1
  NPERGROUP = 98
  POWER = .;
  PLOT x = n min = 15 max = 500 npoints = 50;
RUN;
```

**Power Curve**

**PROC POWER**

*To Determine Sample Size*

```sas
PROC POWER;
  TWOSAMPLEMEANS TEST = DIFF
  MEANDIFF = .4
  STDDEV = 1
  NPERGROUP = .
  POWER = .80;
RUN;
```

**SAS Output**

The POWER Procedure
Two-sample t Test for Mean Difference

<table>
<thead>
<tr>
<th>Fixed Scenario Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Mean Difference</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Nominal Power</td>
</tr>
<tr>
<td>Number of Sides</td>
</tr>
<tr>
<td>Null Difference</td>
</tr>
<tr>
<td>Alpha</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computed N Per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual N Per Power Group</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>
```
Discussion

Obtaining Adequate Power in a Study

- Power = 0.80 is the standard in psychology
  - 80% ability of rejecting a false $H_0$
  - 80% chance of finding a true effect (if it exists)
- Factors affecting Power
  1. Alpha Level
     Usually fixed at 0.05
  2. Effect Size
     Researcher determines a meaningful Effect Size for study
  3. Number of Participants ($N$)
     Calculated according to algorithms
     Computer Programs

Power

- Analytical methods for calculating power are limited in flexibility, but are quick to calculate
  - Not always a simple non-centrality parameter to calculate
- Monte Carlo methods are more flexible, but time intensive
  - Nested designs
    - Different sample sizes
  - Missing data (attrition)
  - More complex statistical models (SEMs)
Session C:

Power Analysis in Multilevel Models

Kevin J. Grimm
July 8, 2009
University of California, Davis

Overview

1. Multilevel Model
2. Two-Stage Sampling
3. Power Considerations in Multilevel Designs & Models
4. Calculating Power in Multilevel Models
   - Monte Carlo Simulation
   - Optimal Design
5. Cost Considerations
6. Discussion

1. Multilevel Model

- Common statistical model for nested or dependent data
  - Examples of nested data
    - Students from different schools
    - Longitudinal (repeated measures) data

- Model is composed of multiple levels
  - Model for each level of nesting

- Also known as Mixed-Effects, Random Coefficient, & Hierarchical Linear Models
Multilevel Model in Educational Research

- Multilevel and, in particular, two-level designs are used frequently in educational and social research.
- Hierarchical linear models incorporating both random and fixed effects provide a useful statistical paradigm for situations where nesting is an obvious and direct consequence of multistage sampling.

**Basic Two-Level Model**

**Level-1 Model (Within-School Model)**

\[ Y_{ij} = b_{0j} + e_{ij} \]

- Score for student \( i \) in school \( j \)
- Intercept for school \( j \)
- Level-1 Residual

**Level-2 Model (Between-School Model)**

\[ b_{0j} = \gamma_{00} + d_{0j} \]

- Intercept for school \( j \)
- Grand Intercept
- Level-2 Residual

**Basic Two-Level Model – Model Assumptions**

- Level-1 Residual is Normally Distributed with a mean of 0 and a single level-1 variance (Within-School Model)

\[ e_{ij} \sim N\left(0, \sigma_w^2 \right) \]

- Level-2 Residual is Normally Distributed with a mean of 0 and a single level-2 variance

\[ d_{0j} \sim N\left(0, \sigma_b^2 \right) \]

**Basic Two-Level Model**

- Decomposing variance of outcome measure into it’s respective parts
  - Example:
    - Variance of achievement scores is composed of between-school variance (\( \sigma_b^2 \)) and within-school variance (\( \sigma_w^2 \))
      - Students within the school differ from one another
      - Schools have different mean achievement scores
  - Benefit comes when we include predictors of the variability at different levels
Intraclass Correlation

- Measure of the amount of dependency in the data
- Ratio of between-cluster variability to total variability

\[
\rho_{icc} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_w}
\]

Example Data

- Sample
  - 711 children in 210 classrooms (~4 children per class)
- Measures
  - Child outcomes (spring teacher ratings)
    - Child Problem Behaviors
    - Child Social Competence
  - Child-level predictors
    - Boy, Age, Mother’s education, poverty status
  - Classroom/teacher level predictors
    - Years experience, Hours per day

SAS PROC MIXED

```
PROC MIXED DATA = temp COVTEST NOCLPRINT;
   CLASS site;
   MODEL tprobPS = /SOLUTION DDFM = bw;
   RANDOM INTERCEPT / SUBJECT = site TYPE = UN;
RUN;
```

Output

| Cov Parm    | Subject | Estimate | Standard Error | Z    | Pr > |t| |
|-------------|---------|----------|----------------|------|------|---|
| UN(1,1)     | SITE    | 0.06847  | 0.01386        | 4.94 | <.0001 |
| Residual    |         | 0.2275   | 0.01426        | 15.95| <.0001 |

| Effect      | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|-------------|----------|----------------|-----|---------|------|---|
| Intercept   | 1.4659   | 0.02564        | 209 | 57.17   | <.0001 |
**Intraclass Correlation**

\[
\rho_{icc} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}
\]

\[
\rho_{icc} = \frac{.068}{.068 + .228}
\]

\[
\rho_{icc} \approx .23
\]

23% of the variability in behavior problems is at the classroom level.

---

**Two-Level Model with Level-1 Predictor**

\[
Y_{ij} = b_{0j} + b_{1j} \cdot X_{1ij} + b_{2j} \cdot X_{2ij} + \ldots + b_{kj} \cdot X_{kij} + e_{ij}
\]

---

**Two-Level Model with Level-2 Predictors**

\[
b_{0j} = \gamma_{00} + \gamma_{10} \cdot Z_j + d_{0n}
\]

\[
b_{1j} = \gamma_{01} + \gamma_{11} \cdot Z_j + d_{1n}
\]

\[
b_{2j} = \gamma_{02} + \gamma_{12} \cdot Z_j + d_{2n}
\]

\[...
\]

\[
b_{kj} = \gamma_{0k} + \gamma_{1k} \cdot Z_j + d_{kn}
\]

---

**Fixed & Random Effects**

- **Fixed-Effects**
  - Effects (parameters) that do not vary across level-2 units (e.g., schools)

- **Random-Effects**
  - Effects (parameters) that vary across level-2 units
**Example Analysis**

- Child-Level outcome is Teacher-Rated Behavior Problems
- Child-Level predictor is Gender
- Classroom-Level predictor is School Days

---

**SAS PROC MIXED**

Child-Level and Classroom-Level Predictors

```
PROC MIXED DATA = temp COVTEST NOCLPRINT;
CLASS site;
MODEL tprobPS = malep schdayp_mean / SOLUTION DDFM = bw;
RANDOM INTERCEPT / SUBJECT = site TYPE = UN;
RUN;
```

---

**SAS PROC MIXED**

Output

#### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Value</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>SITE</td>
<td>0.07219</td>
<td>0.01371</td>
<td>5.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.2094</td>
<td>0.01316</td>
<td>15.91</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### Solution for Fixed Effects

| Effect       | Estimate | Standard Error | DF   | t Value | Pr > |t|  |
|--------------|----------|----------------|------|---------|-------|
| Intercept    | 1.1981   | 0.06198        | 208  | 19.33   | <.0001|
| MALEP        | 0.2264   | 0.03580        | 500  | 6.32    | <.0001|
| SCHDAYP_mean | 0.03071  | 0.01034        | 208  | 2.97    | 0.0033|

---

**2. Two Stage Sampling**
## Sampling with Nested Designs

- When dealing with nested designs (e.g., children within classrooms) we need to consider the sampling design for each level.

- In a two-level design, the total sample size is broken down into the number of level-2 units and number of level-1 units per level-2 units.

### Example

- Designing classroom intervention
- Dealing with children nested within classrooms
  - Total sample size is $N_T$
  - Number of classrooms is $N_C$
  - Average number of children per classroom is $N_{n|C}$

$$N_T = N_C \cdot N_{n|C}$$

## Two Stage Sampling

- In a two-level design with children nested within classrooms we may ask: How many classrooms and how many students per classroom should we collect data from?
  - Is it better to collect data from many classrooms with few students per classroom or few classrooms with many students per classroom?

## 3. Power Considerations in Multilevel Designs and Models
Question

- How should researchers choose sample sizes at the macro- and microlevel in order to ensure a desired level of power given a relevant (hypothesized) effect size and significance level ($\alpha = .05$)?

Raudenbush (1988)

- In reviewing educational applications of hierarchical linear models, Raudenbush concludes the following: Interval estimates and hypothesis tests for HLM-type models rely on large sample properties of maximum likelihood estimates. Little is known about the small sample behavior of the estimates. Standard errors of macroeffects are least well understood when the number of groups is small. Standard errors of microeffects are even more problematic. Investigators are just beginning to understand the effect of errors of variance component estimation on standard errors of shrinkage estimators.

Issues

- Cost
  - Often costs more to sample at the macro-level (classrooms/schools) than at the micro-level (children)
- Study Constraints
  - May be a limited # of macro- and micro units available to study
  - Total Budget
- Intraclass correlation
  - How much dependency are in the data?
- Research Questions
  - Do we have macro- and/or micro-level research questions? If they are micro effects, are these micro effects varying across macrolevels?

4. Calculating Power in Multilevel Models:

Monte Carlo Simulation
**Example**

- Want to determine the power to detect a medium classroom-level effect with 50 classrooms and 4 children per classroom.
  - Children nested within classrooms
  - Main research interest is a classroom-level (e.g., teacher quality) effect
  - Main outcome is an achievement test

**Study Constants**

1. We expect a medium effect size for quality ($R = .3$)
2. Mean & Standard Deviation of Achievement test is 100 & 15, respectively
   - We expect the intraclass correlation to be .25
3. Mean & Standard Deviation of Quality is 4 & 1, respectively

---

**Monte Carlo Simulation Study**

- Simulate data given the study constants
- Fit appropriate statistical model
- Examine significance of prediction
- Repeat

- Determine number of time parameter of interest is significantly different from zero

**Step 0: Determine Between & Within Variances (& Standard Deviations) for Outcome**

- Given Information:
  - Standard Deviation = 15 (variance is 225)
  - ICC = .25

\[
\sigma_w^2 + \sigma_b^2 = 225 \\
\frac{\sigma_b^2}{\sigma_w^2 + \sigma_b^2} = .25
\]
Step 0: Determine Between & Within Variances (& Standard Deviations) for Outcome

\[
\begin{align*}
\frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2} &= \frac{1}{4} \\
\sigma_w^2 + 3\sigma_w^2 &= 225 \\
4\sigma_b^2 &= \sigma_w^2 + \sigma_b^2 \\
4\sigma_w^2 &= 225 \\
3\sigma_b^2 &= \sigma_w^2 \\
\sigma_w^2 &\sim 55 \\
\sigma_b^2 &\sim 165
\end{align*}
\]

Step 1: Simulate Data (A)

Set Parameters

```plaintext
DATA sim1;
  current = TIME();
  Seed = ROUND(current*10001.5);
  datasets = 1000;
/*Setting Data Conditions*/
  classrooms = 50;
  students = 4;
/*Parameters*/
  mu_ach = 100;  *mean of achievement;  
  icc = .25;    
  sigma_ach1 = 13;  *Note: standard deviations - variance is ~165;  
  sigma_ach2 = 7.5;  *variance is ~55;  
  mu_quality = 4;  *mean of quality;  
  sigma_quality = 1;  *variance is 1;  
  r = .3;  *Standardized Effect Size;  
RUN;
```

Step 1: Simulate Data (B)

Generate data

```plaintext
/*Classroom & Individual Level Data*/
DO d = 1 to datasets;
  data_num = d;
  DO j = 1 to classrooms;
    Zj = RANNOR(seed);
    b0j = r * Zj + (SQRT(1 - r**2) * RANNOR(seed));
    Zj = Zj*sigma_quality + mu_quality;  
    b0j = b0j*sigma_ach2 + mu_ach;
    DO i = 1 to students;
      yij = b0j + sigma_ach1 * RANNOR(seed);
      OUTPUT;
    END;
  END;
END;
RUN;
```

Preliminary Look at Simulated Data

<table>
<thead>
<tr>
<th>Obs</th>
<th>data_num</th>
<th>j</th>
<th>i</th>
<th>b0j</th>
<th>Zj</th>
<th>Yij</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>85.233</td>
<td>3.76478</td>
<td>82.133</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>85.233</td>
<td>3.76478</td>
<td>89.411</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>85.233</td>
<td>3.76478</td>
<td>88.707</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>85.233</td>
<td>3.76478</td>
<td>85.197</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>116.830</td>
<td>3.96750</td>
<td>109.422</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>116.830</td>
<td>3.96750</td>
<td>121.239</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>116.830</td>
<td>3.96750</td>
<td>117.975</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>116.830</td>
<td>3.96750</td>
<td>107.470</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>91.137</td>
<td>2.91848</td>
<td>87.792</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>91.137</td>
<td>2.91848</td>
<td>73.508</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>91.137</td>
<td>2.91848</td>
<td>90.458</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>91.137</td>
<td>2.91848</td>
<td>88.467</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>87.477</td>
<td>3.82530</td>
<td>83.508</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>87.477</td>
<td>3.82530</td>
<td>94.643</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>87.477</td>
<td>3.82530</td>
<td>110.798</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>87.477</td>
<td>3.82530</td>
<td>83.508</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset #</th>
<th>School ID</th>
<th>Child ID</th>
<th>School-level Outcome</th>
<th>School-level Predictor</th>
<th>Child-level Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>2</td>
<td>1</td>
<td>103.564</td>
<td>5.29750</td>
<td>97.160</td>
</tr>
<tr>
<td>202</td>
<td>2</td>
<td>1</td>
<td>103.564</td>
<td>5.29750</td>
<td>93.089</td>
</tr>
<tr>
<td>203</td>
<td>2</td>
<td>1</td>
<td>103.564</td>
<td>5.29750</td>
<td>97.071</td>
</tr>
<tr>
<td>204</td>
<td>2</td>
<td>1</td>
<td>103.564</td>
<td>5.29750</td>
<td>97.160</td>
</tr>
</tbody>
</table>
Preliminary Data Checks
Means & Correlation

PROC MEANS DATA = sim1;
   CLASS data_num;
   VAR yij zj;
RUN;

PROC CORR DATA = sim1;
   VAR b0j Zj;
RUN;

Preliminary Data Checks

The CORR Procedure
2 Variables: b0j  Zj

Simple Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0j</td>
<td>200000</td>
<td>99.99</td>
<td>7.46</td>
<td>69.96</td>
<td>133.39</td>
</tr>
<tr>
<td>Zj</td>
<td>200000</td>
<td>4.00</td>
<td>1.00</td>
<td>-0.034</td>
<td>8.46</td>
</tr>
</tbody>
</table>

Pearson Correlation Coefficients, N = 200000
Prob > |r| under H0: Rho=0

<table>
<thead>
<tr>
<th></th>
<th>b0j</th>
<th>Zj</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0j</td>
<td>1.00000</td>
<td>0.29881</td>
</tr>
<tr>
<td>Zj</td>
<td>0.29881</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Preliminary Data Checks
Intraclass Correlation

PROC MIXED DATA = sim1 COVTEST NOCLPRINT;
   CLASS j;
   MODEL Yij = /SOLUTION DDFM = bw;
   RANDOM INTERCEPT / SUBJECT = j TYPE = UN;
   BY data_num;
   ODS OUTPUT SOLUTIONF = parms_fix1 COVPARMS = parms_random1;
RUN;

PROC PRINT DATA = parms_random1 (where = (data_num < 11));
RUN;
### Preliminary Data Checks

#### Intraclass Correlation

The MEANS Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>level_1_var</td>
<td>1000</td>
<td>168.53</td>
<td>18.54</td>
<td>120.97</td>
<td>235.08</td>
</tr>
<tr>
<td>level_2_varA</td>
<td>1000</td>
<td>55.71</td>
<td>20.90</td>
<td>3.68</td>
<td>136.67</td>
</tr>
<tr>
<td>icc</td>
<td>1000</td>
<td>0.24</td>
<td>0.08</td>
<td>0.02</td>
<td>0.48</td>
</tr>
</tbody>
</table>

**Step 2: Fit Appropriate Statistical Model**

```sql
PROC MIXED DATA = sim1 COVTEST NOCLPRINT;
CLASS j;
MODEL Yij = Zj/SOLUTION DDFM = bw;
RANDOM INTERCEPT / SUBJECT = j TYPE = UN;
BY data_num;
ODS OUTPUT SOLUTIONF=parms_fix2 COVPARMS=parms_random2;
RUN;
```
###Parms_Fix2 Dataset

<table>
<thead>
<tr>
<th>data_num</th>
<th>Effect</th>
<th>Estimate</th>
<th>StdErr</th>
<th>DF</th>
<th>tValue</th>
<th>Probt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intercept</td>
<td>82.8025</td>
<td>7.5530</td>
<td>48</td>
<td>10.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>Intercept</td>
<td>92.2774</td>
<td>4.8190</td>
<td>48</td>
<td>19.15</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3</td>
<td>Intercept</td>
<td>97.7606</td>
<td>5.4685</td>
<td>48</td>
<td>17.88</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>4</td>
<td>Intercept</td>
<td>91.4190</td>
<td>6.1999</td>
<td>48</td>
<td>14.75</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>5</td>
<td>Intercept</td>
<td>94.6666</td>
<td>5.4596</td>
<td>48</td>
<td>17.34</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>6</td>
<td>Intercept</td>
<td>79.6312</td>
<td>6.0816</td>
<td>48</td>
<td>13.09</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>7</td>
<td>Intercept</td>
<td>84.3507</td>
<td>5.2403</td>
<td>48</td>
<td>16.10</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>8</td>
<td>Intercept</td>
<td>91.1435</td>
<td>6.1160</td>
<td>48</td>
<td>14.90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>9</td>
<td>Intercept</td>
<td>82.0001</td>
<td>3.9954</td>
<td>48</td>
<td>20.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>10</td>
<td>Intercept</td>
<td>98.2773</td>
<td>4.6231</td>
<td>48</td>
<td>21.26</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

###Step 3: Examine Significance of Effect

```sas
DATA parms_fix2a;
  SET parms_fix2;
  IF effect ne 'Zj' THEN DELETE;
  IF probt < .05 THEN sig = 1;
  IF probt > .05 THEN sig = 0;
RUN;

PROC FREQ DATA = parms_fix2a;
  TABLES sig;
RUN;
```

The FREQ Procedure

<table>
<thead>
<tr>
<th>sig</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>644</td>
<td>64.40</td>
<td>644</td>
<td>64.40</td>
</tr>
<tr>
<td>1</td>
<td>356</td>
<td>35.60</td>
<td>1000</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Power to detect effect is .36

###What affects the power to detect this effect?

- Increase to 6 students per classroom
  - Total # of students increased by 100
  - Power is .42
- Increase to 75 classrooms (4 per classroom)
  - Total # of students increased by 100
  - Power is .52
- Intraclass correlation is .5
  - Power is .50
- Effect is $r = .50$
  - Power is .80
4. Calculating Power in Multilevel Models:

Optimal Design

- Software used to help design multilevel intervention studies where randomization occurs at the cluster level
- Freely available on the web
- May not be useful for examining studies where randomized is not the cluster level and where no randomization occurs

Non-Centrality Parameter in Multilevel Models

\[
\lambda = \frac{\gamma_{10} (N_C)}{4 \left( \sigma^2_b + \frac{\sigma^2_w}{N_{n|C}} \right)}
\]

- \(\lambda\) = non-centrality parameter
- \(\gamma_{10}\) = School-level effect (mean difference) (note: unstandardized)
- \(N_C\) = Number of clusters
- \(N_{n|C}\) = Number of children per cluster
- \(\sigma^2_b\) = between-cluster variance
- \(\sigma^2_w\) = within-cluster variance

What affects Power?

- Cluster size (Number of children per cluster \((N_{n|C})\))
  - As \(N_{n|C}\) increases, the denominator decreases, \(\lambda\) increases, and power increases

- Number of Clusters \((N_C)\)
  - As \(N_C\) increases, then \(\lambda\) increases, and power increases
What affects Power?

- Intraclass Correlation ($\rho_{icc}$)

\[
\rho_{icc} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}
\]

- As $\rho_{icc}$ increases, then the between-cluster variance increases (holding total variance equal) and power decreases

What affects Power?

- Standardized Effect Size ($\delta$)

\[
\delta = \frac{\gamma_{10}}{\sqrt{\sigma_b^2 + \sigma_w^2}}
\]

- As $\delta$ increases, then power increases

Example

- Developing a new school-level intervention to increase mathematics achievement.
  - Past studies indicate
    - $\rho_{icc} = .1$
    - $\delta = .2$
    - $N_{n|C} = 50$
  - Want to achieve a power of .80
  - How many clusters do we need?

Optimal Design Software

(http://sitemaker.umich.edu/group-based/optimal_design_software)

- Click Cluster Randomized Trial
  - Power vs. Number of Clusters ($J$)
    - Select Study Parameters
      - $\rho_{icc} = .1$
      - $\delta = .2$
      - $N_{n|C} = 50$
  - Change Axes
**Resulting Power Curve**

![Graph showing the power curve with a peak at approximately 0.8 when the number of clusters is around 95.](image)

**Cost Considerations**

- The total variable cost of data collection can often be reasonably approximated by the formula

\[
Total\ Cost = N_C \left( C_1 \cdot N_{n|C} + C_2 \right)
\]

- where
  - \( N_C \) = number of clusters
  - \( N_{n|C} \) = number of participants within a cluster
  - \( C_1 \) = cost per participant
  - \( C_2 \) = cost per cluster

**Optimal Sample Sizes**

- To calculate the optimal sample size
  - Find the optimal \( N_{n|C} \)
  - Find the optimal \( N_C \)
  - To find optimal \( N_{n|C} \) use the following formula

\[
\text{optimal } N_{n|C} = \frac{\sigma_w}{\sigma_b} \sqrt{\frac{C_2}{C_1}}
\]

- Determine \( N_C \)

\[
N_C = \frac{Total\ Cost}{N_{n|C} \cdot C_1 + C_2}
\]

**Example**

- What are optimal sample sizes if
  - Total Budget = $10,000
  - Cost per cluster (\( C_2 \)) = $400
  - Cost per person (\( C_1 \)) = $20
  - Intraclass Correlation (\( \rho_{icc} \)) is .05
Step 1: Determine within and between variances

- Assume total variation in outcome is 1
- Calculate within & between variances
  \[ \rho_{cc} = .05 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \]
  \[ \sigma_b^2 + \sigma_w^2 = 1 \]
- Calculate within and between standard deviations
  \[ \sqrt{\sigma_b^2} = \sqrt{.05} = .2236 \]
  \[ \sqrt{\sigma_w^2} = \sqrt{.95} = .9747 \]

Step 2: Determine optimal \( N_{n|C} \)

- Determine Optimal \( N_{n|C} \)
  \[ \text{optimal } N_{n|C} = \frac{\sigma_w}{\sigma_b} \sqrt{\frac{C_2}{C_1}} = \frac{.9747}{.2236} \sqrt{\frac{400}{20}} \approx 19.5 \approx 20 \]

Step 3: Determine optimal \( N_C \)

- Determine Optimal \( N_C \)
  \[ N_C = \frac{10000}{20(20) + 400} = 12.5 \approx 12 \]

Discussion
**Discussion**

- Standard error of the treatment contrast typically depend more heavily on the number of clusters than on the number of participants per cluster.

- However, studies with large numbers of clusters tend to be expensive.

**Macro-level Power**

- Note:
  - When one wishes to have a sampling design that is optimal with respect to estimating school effects, it is best to make \( N_C \) as large as possible and \( N_{n|C} \) equal to 1.
  - However, we can’t estimate within school variability or between-school variability in level-1 regression coefficients, so we would want to make \( N_C \) as large as possible and \( N_{n|C} \) equal to 2.

**Optimal Design**

- Useful software when designing randomized control trials with multistage sampling.

- If you wish to examine non-random effects and/or child-level effects, it’s useful to conduct Monte Carlo Simulations.

- Monte Carlo simulations can also build in other expectancies (e.g., incomplete data, attrition).
**Session D:**

*Power Analysis in Structural Equation Models*

Kevin J. Grimm  
July 9, 2009  
University of California, Davis

---

**Overview**

1. Structural Equation Model (SEM) Concepts  
2. Path Diagrams  
3. Structural Expectations  
4. Background to SEM Power Analysis  
5. Approaches to Power Analysis in SEM  
6. Example  
   - One vs. Two Factor Model  
7. Discussion

---

1. **SEM Concepts**

   **Various Definitions for SEM**

   - Formal statistical statement about the relations among chosen variables  
   - Hypothesized pattern of (linear) relationships among a set of variables  
   - Collection of statistical techniques that allow the examination of a set of relationships between one or more IV’s (continuous or discrete) and one or more DV’s (continuous or discrete)
Some Advantages of SEM

- Explicit representation of theory (no default model)
- Representation of complex multivariate theories involving latent entities are possible
- Direct and indirect effects can be teased apart
- Analysis of multiple groups
- Variables can be outcomes (DV) and predictors (IV)
- Missing Data

Some Disadvantages of SEM

- Need for strong substantive theory regarding relationship of variables
- Sample Size
- Need for some basic understanding of statistics
- Yield to temptation, “when the tail wags the dog”
- Equivalent models
- Yields to statements of causality easily

Special SEM

- General linear model (t-test, regression, multiple regression, ANOVA, etc.)
- Path model
- Confirmatory factor analysis
- Latent variable path model
- Latent growth curve analysis

Some SEM Software

- LISREL
- Mplus
- AMOS
- SAS Proc Calis
- Mx
- COSAN
- Sepath
- LisComp
- Systat Ramona
**Underlying Principles**

- Because SEM concerns relationships among variables, the emphasis is mainly on covariance matrices.
  - Longitudinal data additionally concern the mean vector.
- The hypothesized model attempts to explain the structure of the data, with *parsimony* and *accuracy*.
- Some statistical statement is needed about the match between the observed data and the hypothesized model (e.g., Fit Statistics).

**Some SEM Terminology**

- Manifest variables: Observed (i.e., measured) variables defining the structure we wish to model.
- Latent variables: Unobserved (i.e., unmeasured) variables implied by the covariance among two or more manifest variables.
- Specification: Exercise of formally stating a model.

**Some SEM Terminology (2)**

- **Association**: Non- (or bi-) directional (reciprocal) relation between 2 variables.
- **Direct effect**: (uni-) Directional (non-reciprocal) relation between 2 variables (IV and DV).
- **Indirect effect**: Effect of an IV on a DV through one or more intervening or mediating variables.
- **Total effect**: Sum of direct and indirect effects of an IV on a DV.

**2. Path Diagrams**
**Path Diagrams**

- In Structural Equation Modeling, we often are dealing with several variables, complex relationships, and latent entities.
- To help explain, describe, and understand SEMs, we often use path diagrams to visually represent our model.
- Some SEM programs use Path Diagrams for programming.
- However, path diagrams may not be very useful for very large models (lots of variables and relationships).
  - Matrix algebra may be more helpful.

---

**SEM Path Diagrams**

**A Key**
- Squares = Observed Variables
- Circles = Latent Variables
- Double-Headed Arrows = Variances/Covariances (Association)
- Single-Headed Arrows = Regressions (Direct Effect)
- Triangle = Assigned variable = Constant (=1.0) for modeling Means

---

**Regression & Structural Regression**

\[
Y_n = \beta_0 \cdot 1_n + \beta_1 \cdot X_n + 1 \cdot e_n
\]

- \(\beta_0\) = intercept, predicted value of Y when predictors (X) are zero.
- \(\beta_1\) = slope coefficient, predicted amount of change in Y for a 1 unit change in X.
- \(e\) = residual, part of Y not predicted by X, uncorrelated with X.

Variance of Y is decomposed into variance explained by X and unexplained variance (e).

**Structural Regression**

\[
Y_n = \beta_0 \cdot 1_n + \beta_1 \cdot X_n + 1 \cdot e_n
\]
3. Structural Expectations

- Every Structural Model has a set of Structural Expectations.
  - Variance/Covariance Expectations
  - Mean Expectations
- Used to estimate parameters
- Difference between Structural Expectations (with estimated parameters) and Observed Statistics (Covariance/Mean) is the model misfit
- Expectations can be calculated based on a path diagram or computed algebraically.
Calculating Covariance Expectations

\[ E[X_1, X_1'] = \sigma_{x_1}^2 \]
\[ E[X_2, X_2'] = \beta_1 \cdot \sigma_{x_1}^2 \cdot \beta_1 + \beta_3 \cdot \sigma_{v_1}^2 \cdot \beta_3 + \beta_1 \cdot \sigma_{x_1v_1} \cdot \beta_3 + \beta_3 \cdot \sigma_{x_1v_1} \cdot \beta_1 + \sigma_{e_{x_2}}^2 \]

Covariance Expectations (Select)

\[ E[Y_1, Y_1'] = \sigma_{y_1}^2 \]
\[ E[Y_1, Y_2'] = \beta_2 \cdot \sigma_{y_1}^2 \cdot \beta_2 + \beta_3 \cdot \sigma_{y_1}^2 \cdot \beta_3 \]
\[ E[Y_2, Y_2'] = \beta_2 \cdot \sigma_{y_1}^2 \cdot \beta_2 + \beta_3 \cdot \sigma_{y_1}^2 \cdot \beta_3 \]

Calculating Mean Expectations

\[ (E[X_1])^2 = \mu_{x_1} \cdot 1 \cdot \mu_{x_1} \]
\[ E[X_1] = \mu_{x_1} \]

\[ (E[X_2])^2 = \beta_1 \cdot \mu_{x_1} \cdot 1 \cdot \mu_{x_1} \cdot \beta_1 + \beta_1 \cdot \mu_{x_1} \cdot 1 \cdot \mu_{y_1} \cdot \beta_3 + \beta_3 \cdot \mu_{y_1} \cdot 1 \cdot \mu_{y_1} \cdot \beta_3 + \beta_3 \cdot \mu_{y_1} \cdot 1 \cdot \mu_{x_1} \cdot \beta_1 \]
\[ E[X_2] = \beta_1 \cdot \mu_{x_1} + \beta_3 \cdot \mu_{y_1} \]
4. Background to SEM

Power Analysis

- Power analysis always compares two models
  - Null Model
  - Alternative Model (the one we’re interested in)

- In SEM, there’s no generic Default model to make comparisons against
  - Still need to compare two models
  - Think of these models as competing hypotheses
  - Models need to be nested

**SEM and Statistical Power**

- The power of the chi-square statistic is well known (since 1920) for contingency table analysis

- Statistical power was not a highlight of most SEM studies until the mid 1980s

- The important paper by Saris & Satorra (1985) altered the way people considered power in the SEM context.

- The ideas are similar to standard SEM tests but are most useful for planning a study.

Expected Variances

\[
E[XX'] = \sigma_x^2 \\
E[YY'] = \beta_1 \cdot \sigma_x^2 \cdot \beta_1 + 1 \cdot \sigma_e^2 \cdot 1
\]

Expected Covariance

\[
E[XY'] = \beta_1 \cdot \sigma_x^2
\]

Expected Means

\[
E[X] = \mu_x \\
E[Y] = \beta_1 \cdot \mu_x + \beta_0
\]
Power Analysis in SEM

- As with basic power analysis and power analysis in multilevel models, power can be calculated using
  - Monte Carlo simulations
- Non-centrality parameter
  - Based on the chi-square fit statistic

Satorra & Saris (1985) discussed how power analyses can be conducted in SEM
- Remains a common way to evaluate power

MacCallum, Browne, & Sugawara (1997) discussed how power calculations for SEM can be based on other fit statistics, namely the RMSEA

Muthen & Muthen (2002) provide examples of how to conduct power analyses using Monte Carlo simulations in Mplus

5. Approaches to Power Analysis in SEM

Satorra & Saris (1985)

- Method to examine power analysis that is based on a “perfect” fitting model and a constrained alternate model
- Important to understand model expectations to generate covariance (or correlation) matrix based on the “perfect” model
- Based on chi-square fit statistic for alternate model and difference in degrees of freedom (perfect and alternate model)
### Steps

- **Step 1** Create a population model with appropriate parameter values
- **Step 2** Calculate Expected Covariance Matrix & Mean Vector
- **Step 3** Create an Alternative and more restricted model
  - This can be easily done by eliminating terms from the a full path model in Step 1
  - Best if it is a specific variation of interest
- **Step 4** Use any SEM computer program to fit the alternative model to the initial population model covariance matrix
  - By stating that N=101, we obtain the Likelihood value as $-2LL=\chi^2/100$
- **Step 5** Using the non-central distribution of the $\chi^2$ examine the power curve for the test(s) of the parameters

### Monte Carlo Simulation

- Simulate data based on chosen population model
- Fit population model
- Fit alternate model
- Examine chi-square difference between population model and alternate model for difference in degrees of freedom

### Example

- Initial Population Model is a two factor model with a .7 correlation between factors
  - Factor loadings are .6

- Alternate and Restrictive Model is a one factor model
  - Two factor model with a perfect correlation between factors
Calculate Covariance Matrix for Initial Population Model

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>.36</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>.36</td>
<td>.36</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y1</td>
<td>.252</td>
<td>.252</td>
<td>.252</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y2</td>
<td>.252</td>
<td>.252</td>
<td>.252</td>
<td>.36</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Y3</td>
<td>.252</td>
<td>.252</td>
<td>.252</td>
<td>.36</td>
<td>.36</td>
<td>1.00</td>
</tr>
</tbody>
</table>
**Mplus Output – Tests of Model Fit**

Chi-Square Test of Model Fit

<table>
<thead>
<tr>
<th>Value</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of Freedom</td>
<td>8</td>
</tr>
<tr>
<td>P-Value</td>
<td>1.000</td>
</tr>
</tbody>
</table>

CFI/TLI

<table>
<thead>
<tr>
<th>CFI</th>
<th>1.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLI</td>
<td>1.194</td>
</tr>
</tbody>
</table>

Loglikelihood

<table>
<thead>
<tr>
<th>H0 Value</th>
<th>-810.646</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Value</td>
<td>-810.646</td>
</tr>
</tbody>
</table>

Information Criteria

<table>
<thead>
<tr>
<th>Number of Free Parameters</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akaike (AIC)</td>
<td>1647.292</td>
</tr>
<tr>
<td>Bayesian (BIC)</td>
<td>1681.288</td>
</tr>
<tr>
<td>Sample-Size Adjusted BIC</td>
<td>1640.228</td>
</tr>
</tbody>
</table>

RMSEA (Root Mean Square Error Of Approximation)

<table>
<thead>
<tr>
<th>Estimate</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 Percent C.I.</td>
<td>0.000 0.000</td>
</tr>
<tr>
<td>Probability RMSEA &lt;= .05</td>
<td>1.000</td>
</tr>
</tbody>
</table>

SRMR (Standardized Root Mean Square Residual)

| Value | 0.000 |

**Fitting Alternate Model**

TITLE: One Factor Model;
DATA: FILE = two_fac.dat;
TYPE = FULLCOV;
NGROUPS = 1;
NOBSERVATIONS = 101;

VARIABLE:

<table>
<thead>
<tr>
<th>NAMES</th>
<th>x1 x2 x3 y1 y2 y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>USEARIABLES</td>
<td>x1 x2 x3 y1 y2 y3</td>
</tr>
</tbody>
</table>

ANALYSIS: TYPE = GENERAL; ESTIMATOR = ML;

MODEL:

| factor BY x1* x2 x3 y1 y2 y3; |
| factor@1; |

OUTPUT: SAMPSTAT RESIDUAL STANDARDIZED TECH1;

**Calculating Power Based on Fit of Alternative Model**

DATA temp1;
*Information from Results;
N = 101;
chi_square = 7.333;
df = 1; *Change in chi-square from reference model;
*Calculating Non-centrality parameter;
Function = Chi_square/(N-1);
alpha=.05;
Central = CINV(1-alpha,df);
ATTRIB Size LABEL = 'Sample Size';
Size=2;
DO WHILE (Power < .999);
Noncent = (Size-1) * Function;
Power = 1 - PROBCHI(Central, DF, Noncent);
OUTPUT;
Size = Size + 1;
END;
RUN;

AXIS I=JOIN;
PROC GGRAPH; PLOT Power*Size / VREF = .8 VREF=.95; RUN;
To Achieve .80 power, we need ~ 110 subjects

Monte Carlo Simulation

- Step 1: Generate multiple datasets according to initial (population) model
- Step 2: Fit reference and alternative model
- Step 3: Examine number of times change in chi-square is significant (for the change in degrees of freedom)
  - Can also use the Satorra & Saris method based on average change in chi-square
  - Can also look at specific effect if the effect is forced to be zero in constrained model

Simulating Two Factor Data

SAS MACRO LANGUAGE

```sas
%MACRO SIMULATION;
DATA sim_mc;
  current = TIME();
  Seed = ROUND(current*10001.5);
  r = .7;
  loading = .6;
  DO I = 1 to 100;
    * Generate Factor Scores with .7 Correlation;
    x = RANNOR(seed);
    y = r * x + SQRT(1 - r**2) * RANNOR(seed);
    * Generate Observed Scores;
    x1 = loading * x + SQRT(1 - loading**2)*RANNOR(seed);
    x2 = loading * x + SQRT(1 - loading**2)*RANNOR(seed);
    x3 = loading * x + SQRT(1 - loading**2)*RANNOR(seed);
    y1 = loading * y + SQRT(1 - loading**2)*RANNOR(seed);
    y2 = loading * y + SQRT(1 - loading**2)*RANNOR(seed);
    y3 = loading * y + SQRT(1 - loading**2)*RANNOR(seed);
  OUTPUT;
END;
RUN;%MEND SIMULATION;
```

Fitting One & Two Factor Models (A)

SAS MACRO LANGUAGE

```sas
%MACRO ANALYSIS;
******************
TITLE2 'One Factor Model';
*************************
PROC CALIS DATA = sim_mc OUTRAM = model1;
  VAR x1 x2 x3 y1 y2 y3;
  RAM
    1 1 7 L1,      2 1 1 R1,
    1 2 7 L2,      2 2 2 R2,
    1 3 7 L3,      2 3 3 R3,
    1 4 7 L4,      2 4 4 R4,
    1 5 7 L5,      2 5 5 R5,
    1 6 7 L6,      2 6 6 R6,
    2 7 7 1;
RUN;

DATA modella;
  SET model1;
  IF _NAME_ ne 'CHISQUAR' THEN DELETE;
  one_fac_c2 = _ESTIM_;  KEEP _NAME_ one_fac_c2;
RUN;
```
Fitting One & Two Factor Models (B)
SAS MACRO LANGUAGE

**************************;
TITLE2 'Two Factor Model';
**************************;
PROC CALIS DATA = sim_mc OUTRAM = model2;
VAR x1 x2 x3 y1 y2 y3;
RAM
  1 1 7 L1, 2 1 1 R1,
  1 2 7 L2, 2 2 2 R2,
  1 3 7 L3, 2 3 3 R3,
  1 4 8 L4, 2 4 4 R4,
  1 5 8 L5, 2 5 5 R5,
  1 6 8 L6, 2 6 6 R6,
  2 7 7 1, 2 8 8 1,
  2 8 7 c;
RUN;
DATA model2a;
  SET model2;
  IF _NAME_ ne 'CHISQUAR' THEN DELETE;
  two_fac_c2 = _ESTIM_; 
  KEEP _NAME_ two_fac_c2;
RUN;

Fitting One & Two Factor Models (C)
SAS MACRO LANGUAGE

DATA temp2;
  MERGE model1a model2a;
  diff = one_fac_c2 - two_fac_c2;
RUN;
DATA _NULL_; 
  SET temp2;
  FILE 'E:\Memory Key 20090331\UVa_20090709\fits.dat' MOD;
  PUT (one_fac_c2 two_fac_c2 diff) (15.5);
RUN;
%MEND

Run MACROS & Evaluate Results

%MACRO ALLJOB;
  %LOCAL count;
  %LOCAL stop;
  %LET stop = 1100;
  %LET count= 1001;
  %DO %WHILE (&count <= &stop);
    SIMULATION; *Recall local macro - SIMULATION;
    ANALYSIS;
    %LET count = %EVAL(&count+1); *Increment count by 1;
  %END;
%MEND ALLJOB;

DATA temp1;
  INFILE 'E:\Memory Key 20090331\UVa_20090709\fits.dat'; 
  INPUT one_fac_c2 two_fac_c2 diff;
  IF diff > 3.84 THEN sig = 1; 
  IF diff < 3.84 THEN sig = 0;
RUN;

Results

The FREQ Procedure

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
<td>21.00</td>
<td>21</td>
<td>21.00</td>
</tr>
<tr>
<td>1</td>
<td>79</td>
<td>79.00</td>
<td>100</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Power is .79 with N = 100
7. Discussion

SEM Power Analysis

- Flexibility of approach is a bonus
  - Can test a variety of models and different competing models (not just whether a coefficient is zero or not)
- Power of latent variable effect depends on the size of the effect, sample size, AND other parameters (e.g., factor loadings) of the model
- Necessary to have an understanding of what to expect for various parameters

SEM Power Analysis

- Monte Carlo SEM Power Analyses can now be conducted in Mplus
  - Is capable of simulating data and fitting model(s)
  - RUNALL Utility (downloadable from www.statmodel.com) can be used in conjunction with Mplus when data are simulated outside of Mplus
  - Useful, but can be difficult to have it work properly