Workshop on Design and Analysis of Experiments in Education

Spyros Konstantopoulos
spyros@msu.edu

Kim Maier
kmaier@msu.edu

Michigan State University

Prepared for a workshop at the 2010 meeting of
The Society for Research on Educational Effectiveness

Experimental Design

- "Experimental Design" encompasses:
  1. Strategies for organizing data collection
  2. Data analysis procedures *matched* to those data collection strategies.
- Analysis of Variance (ANOVA) is traditional analysis procedure applied to experimental designs, especially RCTs.
- Other appropriate analytic procedures include:
  - Multilevel or hierarchical statistical models.
  - Statistical models applied to aggregrates, such as class or school means.
- Use of analytic procedures should match with research hypotheses.

Why do we need experimental design?

- To identify treatment effects in the presence of ***variability*** of units and/or responses.
- Variability exists because:
  - Units (students, teachers, & schools) are not identical
  - Units respond in different ways to treatments
- We need experimental design to control this variability so that treatment effects can be identified.

Principles of Experimental Design

- Goal: control variability so that the systematic effects of treatments can be identified
  - Create sample groups that are equivalent "in the long run"
    - Measures of traits are similar across groups
    - Groups would have the same response if given the same treatment.
- Methods to achieve this control:

1. ***Randomization*** ← "True" experiments
2. ***Matching***
3. ***Statistical Adjustment***

Quasi-experiments
Observational studies

Control by Randomization

Controls for the effects of ***all*** characteristics:

- observable or non-observable
- known or unknown
  - ⇨ Makes groups equivalent *on average* on ***all*** characteristics

- Differences in outcomes after treatment that are larger than would be expected by chance can be attributed to *treatment effect* and not to preexisting differences between the groups.

Control by Randomization

- Considered to be the gold standard in clinical research.
  - ⇨ Strongest design for causal inference
- May not work well in all contexts, such as policy interventions
- In some cases, randomization may not be:
  - Desirable
  - Ethical
  - Possible

Control by Matching

- *Known* sources of variation may be eliminated by matching
- Example: try to eliminate district, school, or classroom effects before comparing students
- Some important limitations:
  - Matching can only be done on **known and observable** characteristics that are measured
  - Perfect matching is not always possible
  - Limits generalizability by removing possibly informative variation
  - May reduce sample size

Control by Statistical Adjustment

- A form of post-hoc pseudo-matching
- Uses statistical relations between outcomes and controls/covariates to simulate matching
- Increases precision of regression estimates
- Some important limitations:
  - Statistical control is possible using **known and observable** characteristics only
  - Does not necessarily address *all* preexisting biases prior to assignment to treatment conditions

Using Principles of Experimental Design

- When using randomization, we do not have to know a lot to use if effectively
  - Simply conduct random assignment
- When using matching or statistical control, we have to know a lot and think about what variables are important ahead of time
- Thorough thinking when designing the study is necessary in order to measure all variables that are important in the study
- Randomization may not be as efficient (would require larger sample sizes for the same power) as matching or statistical control

Basic Ideas of Design: Independent Variables

- Independent variables are also called ***factors***
- The *values* of independent variables are called levels
- Some independent variables can be manipulated, others cannot:
  - ***Treatments*** are independent variables that can be manipulated
  - ***Blocks*** and ***covariates*** are independent variables that cannot be manipulated
- Units can be randomly assigned to treatment levels, but *<u>not</u>* to blocks

Basic Ideas of Design: Nesting & Crossing

- Crossing refers to relations between independent variables
- Factors (e.g., treatments or blocks) are crossed if every level of one factor occurs with every level of another factor
- *Example:* The Tennessee class size experiment assigned students to one of three class size conditions.
  - All three treatment conditions (factor A) occurred within each of the participating schools (factor B)

⇨ ***treatment is crossed with schools***

Basic Ideas of Design: Nesting & Crossing

- Factor B is nested in factor A if every level of factor B occurs within only one level of factor A
- *Example:* The Tennessee class size experiment assigned students to one of three class size conditions.
  - All three treatment conditions occurred within each of the participating schools
  - Classrooms are randomly assigned to treatment conditions and each classroom occurs in only one treatment condition

⇨ ***classrooms are nested within treatments***

⇨ ***treatment is crossed with schools***

Basic Ideas of Design: Nesting & Crossing

- *Example:* schools are randomly assigned to treatment conditions.
  - ⇨ ***schools are nested within each treatment condition***

Schools

1, 2, … ,m m + 1, … , 2m

Treatments

1 2

Basic Ideas of Design: Nesting & Crossing

- *Example:* classrooms or students are randomly assigned to treatment conditions within schools

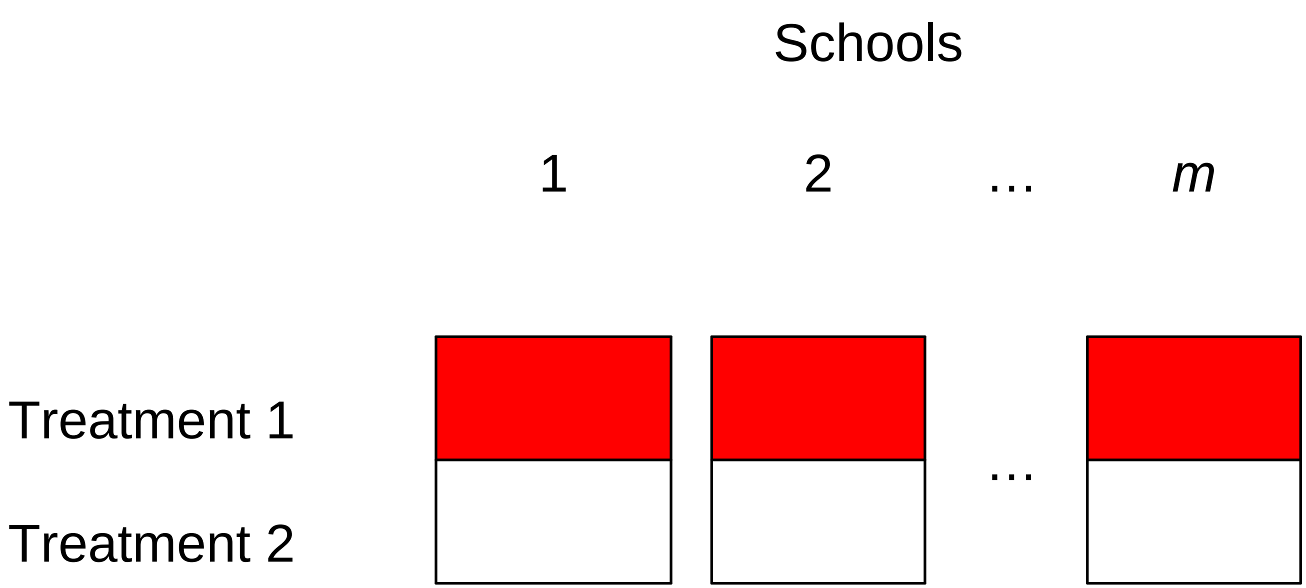
Schools

1	2	…	*m*
T1	T2		T1
T2	T1	…	T2

⇨ ***treatments are crossed with schools***

Basic Ideas of Design: Nesting & Crossing

- *Example:* Again, classrooms or students are randomly assigned to treatment conditions within schools (but notice pattern of treatment)
  - What is wrong with this layout?
  - What are the possible sources of bias?



Three Basic Designs

- Completely Randomized Design
  - Treatments are randomly assigned to students
- Block Randomized Design
  - Students (or classrooms) are assigned randomly to treatments *within* schools
- Cluster (or Group) Randomized Design
  - Also called a Hierarchical Design
  - Schools are assigned randomly to treatments, the same treatment is assigned to all individuals within the school

Completely Randomized Design

- Individuals are randomly assigned to one of two treatments:

Treatment	Control
Individual 1	Individual 1
Individual 2	Individual 2
⋮	⋮
Individual *n*	Individual *n*

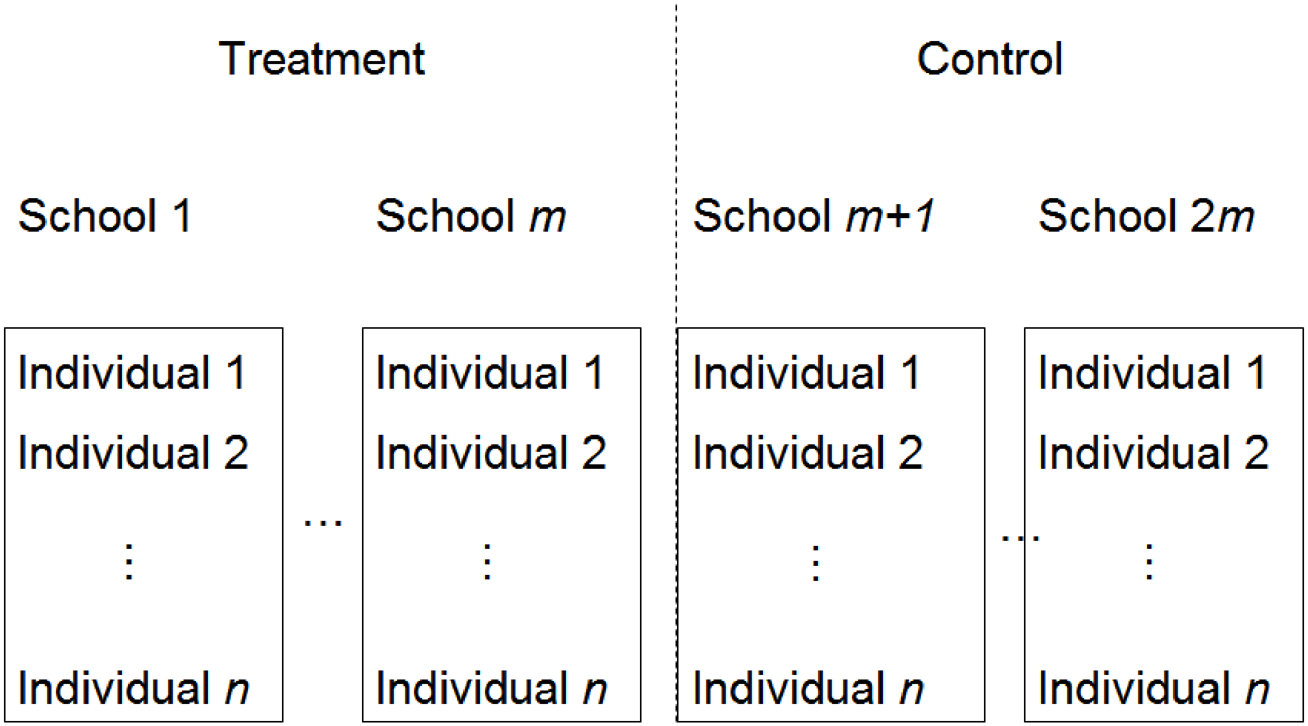
Block Randomized Design

- Individuals are randomly assigned to one of two treatments within their school:

	School 1	…	School *m*
Treatment 1	Individual 1 ⋮ Individual *n*	…	Individual 1 ⋮ Individual *n*
Treatment 2	Individual *n* +1 ⋮ Individual 2*n*	…	Individual *n*+1 ⋮ Individual 2*n*

Cluster Randomized Design

- Schools are randomly assigned to one of two treatments, all students within receive treatment:



Sampling Models

Sampling Models in Educational Research

- Sampling models are often ignored in educational research
- But, sampling is where the randomness comes from in social science research
- Sampling therefore has profound consequences for statistical analysis and research designs

Sampling Models in Educational Research

- In *randomized experiments*, there are two accounts of where randomness comes from:
  - From sampling individuals from treatment and control populations created in the experiment
  - From random assignment of unites to treatments
- Regardless of the perspective, both lead to essentially identical statistical tests

Sampling Models in Educational Research

- Simple random samples are rare in large-scale field research
- Most samples are not probability samples, but the goal is often to make inferences to some population
  - ⇨ sampling design doesn't match goal
- The structure of educational systems is challenging for the application of randomized experiments
  - Educational populations have nested structures
  - Students in classrooms in schools
  - Schools in districts in states

Sampling Models in Educational Research

- Research often exploits this nested structure to sample students by first sampling schools
- This sampling strategy is called *multi-stage (multilevel) cluster sampling* in survey research
- *Example:* Clusters are first sampled (e.g., schools) and then individuals within clusters (e.g., students) are sampled
  - ➪ ***Two-stage (two-level) cluster sample***
- *Example:* Schools are first sampled, then classrooms, then students
  - ➪ ***Three-stage (three-level) cluster sample***

Variance of the Mean of Clustered Sample

- The usual variance calculation is based on a simple random sample
- When clustering is used, the variance must reflect the dependence within a cluster
- The variance of the mean of a cluster sample:

$$\frac{\tau^2}{m}+\frac{\sigma^2}{mn}=\frac{\sigma^2+n\tau^2}{mn}$$

where:

τ^2 = Level-2 variance, m = number of Level-2 units

σ^2 = Level-2 variance, n = number of Level-1 units within a Level-2 unit

Variance of the Mean of Clustered Sample

- The intraclass correlation coefficient (ICC), ρ, is the proportion of total variance at the 2nd level
- If we write $\rho = \tau^2/(\sigma^2 + \tau^2)$, the variance becomes:

$$\frac{\left(\sigma^2+\tau^2\right)}{mn}\left[\left(1-\rho\right)+n\rho\right]=\frac{\left(\sigma^2+\tau^2\right)}{mn}\left[1+\left(n-1\right)\rho\right]$$

Variance of the Mean of Clustered Sample

- This variance can be decomposed:

Total variance

Variance of the mean of
a simple random sample

Design effect

$$\frac{\left(\sigma^2+\tau^2\right)}{mn}\left[n\rho+\left(1-\rho\right)\right]=\frac{\left(\sigma^2+\tau^2\right)}{mn}\left[1+\left(n-1\right)\rho\right]$$

Total sample size

Variance of the Mean of Clustered Sample

- The design effect is a multiplier of the variance of the mean for a simple random sample
- Suppose a sample of size $N = mn$ has a total population variance of σ_T^2
- If the sampling strategy was simple random or stratified, the variance of the mean is:

$$\sigma_T^2 \big/ mn$$

Variance of the Mean of Clustered Sample

- Now, suppose we have a two-stage (two-level cluster design: *n students from each of m schools*
- Assume the same sample size $N = mn$, and same total population variance of σ_T^2
- Since we have clustering, we have an ICC, ρ
- For a clustered sample of n students from each of the m schools, the variance of the mean is now:

$$\frac{\left(\sigma_T^2\right)}{mn}\left[1+(n-1)\rho\right]$$

⇨ The inflation factor $\left[1+(n-1)\rho\right]$ is the *design effect*

Variance of the Mean of Clustered Sample

- Suppose we have n students from p classes in each of m schools
- Assume a sample size $N = mpn$, and same total population variance of σ_T^2
- If the sampling strategy had been *simple* or *stratified* then the variance of the mean would be:

$$\frac{\left(\sigma_T^2\right)}{mpn}$$

Variance of the Mean of Clustered Sample

- Now, suppose we have a three-stage (two-level) cluster design: *n* students from *p* classes in each of *m* schools
- Assume sample size $N = mpn$, and same total population variance of σ_T^2
- We have three levels of clustering, so we have two ICCs, ρ_s *(school level) and* ρ_c *(class level)*
- The variance of the mean is now:

$$\frac{\left(\sigma_T^2\right)}{mpn}\left[1+(n-1)\rho_C+(pn-1)\rho_s\right]$$

⇨ The three level design effect is: $\left[1+(n-1)\rho_c+(pn-1)\rho_s\right]$

Variance of the Mean of Clustered Sample

- Treatment effects in experiments and quasi-experiments are *mean differences*
- The sampling model dictates variance estimates
- Variance impacts:
  - Precision of treatment effect estimates
  - Statistical power

Sampling Models in Educational Research

- The fact that the population is structured in groups or clusters (e.g., students within schools) ***does not*** imply that the sample must be a clustered sample
- Whether a sample is a *clustered sample* depends on:
  - How the sample is drawn
    - ⇨ Are schools sampled first then students randomly in schools?
  - What the inferential population is
    - ⇨ Are inferences about the schools studied or a larger population of schools?

Inferential Population and Inference Models

- The inferential population or the inference model has implications for analysis and therefore for the design of experiments
- Question to consider: *Do we make inferences to the schools in this sample or to a larger population of schools?*
  - Inferences to the ***sampled schools or classes*** in the sample are called **<u>conditional inferences</u>**
  - Inferences to a ***larger population of schools or classes*** are called **<u>unconditional inferences</u>**

  ⇨ *Bottom line:* Inferences are different in **conditional** or **unconditional** models

Inferential Population and Inference Models

- In a conditional inference, we are estimating the mean or treatment effect in the observed schools
- In an unconditional inferences, we are estimating the mean or treatment effect in the population of schools from which the observed schools are a sample
- In both cases, a mean or a treatment effect is estimated, but they are *different* parameters with their own respective variances.

Fixed and Random Effects

- Fixed Effects
  - The levels of a factor in a study constitute the entire inference population
  - The inference model is conditional
    - ⇨ The factor is called fixed, and its effects are called *fixed effects*
- Random Effects
  - The levels of a factor in a study are sampled
  - The inference model is unconditional
    - ⇨ The factor is called random, and its effects are called *random effects*

Power Analysis

Statistical Power

- Power is the probability of detecting the treatment effect when it exists (or accepting the research hypothesis when it is true)
- A critical component in the design of experiments
- In general larger effect sizes increase power
- Every researcher's goal: High power to detect treatment effects
  - Cohen suggests 0.80 as the lower bound

Statistical Power for Two Level Designs

- Larger variances (or ICC) at the second level decreases power
- An increase in the number of units at the second level (i.e., schools) increases power to a greater extent than an increase in the number of students.
- Covariates at the first or second level increase power
- Power is typically higher in block randomized designs

Statistical Power for Two Level Designs

- When the cost for sampling schools is high, then sample more students within schools (but the power will be lower)
- When variances (or intraclass correlations) are larger, then sample more schools and fewer students per school (but the power is lower)
- When effect sizes are large (0.8 or larger) cost and clustering issues are overcome

Statistical Power for Three Level Designs

- Larger variances (or ICC) at the second and third level decrease power
- The number of schools impacts power more than the number of classrooms or students
- The number of classrooms impacts power more than the number of students
- Covariates at first, second, or third level increase power
- The power is typically higher in block randomized designs (especially when treatment is at the student level)

Statistical Power for Three Level Designs

- When the cost for sampling schools is high, then sample more classrooms within schools (but the power will be lower)
- When variances (or intraclass correlations) are larger, then sample more schools and fewer students per classroom (but the power is lower)
- When clustering at the second level is larger than in the third level, then sample more classrooms than students
- When effect sizes are large (0.8 or larger) cost and clustering issues are overcome

Review of Regression

Basic Regression Model

Let Y_{ij} be the outcome of the i^{th} student in the j^{th} school:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \varepsilon_{ij}$$

If T_{ij} is a treatment dummy ($T_{ij} = 1$ in treatment group, $T_{ij} = 0$ in the control group), then

- β_0 is the control group mean,
- β_1 is the treatment effect
- ε_{ij} is a residual for the i^{th} student in the j^{th} school

Coding in Regression

- By defining the independent variables differently, we can change the interpretation of regression coefficients
- For example, use effect coding that assumes balance:

  Define $T_{ij} = \pm \frac{1}{2}$

  Now

  β_0 is the grand mean

  β_1 is half the treatment control mean difference (the treatment effect usually defined in ANOVA)

Centering

- Centering is a transformation applied to the independent variables
- In simple random sample designs, a variable is center by subtracting the mean from each value
- If X_i is the independent variable, the centered variable is:

$$X_i^C = \left(X_i - \bar{X}_{\bullet}\right)$$

where $\bar{X}_{\bullet}$ is the mean of the X_i*'s* in the sample

- The mean of the new centered variable is zero

Centering

- Centering changes the value and the meaning of the intercept (it's the mean of the outcome)
- It ***does not*** change the value or the meaning of the slope
- Consider a regression equation:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- And the regression equation with the centered predictor:

$$Y_i = \beta_0^C + \beta_1^C X_i^C + \varepsilon_i^C = \beta_0^C + \beta_1^C \left(X_i - \bar{X}_{\bullet} \right) + \varepsilon_i^C$$

- Equating the centered and uncentered regression see that:

$$\beta_0 = \beta_0^C - \beta_1^C \bar{X}_{\bullet} \Rightarrow \beta_0^C = \beta_0 + \beta_1^C \bar{X}_{\bullet}$$

- Centering also changes the precision of the intercept (but not of the slope)

Centering: Two-Level Case

- In two-level designs (e.g., students within schools, there two kinds of centering:
  - Grand mean centering of student predictors
  - Group mean centering of student predictors
- Grand mean centering is subtracting the grand mean:

$$X_{ij}^{Grand} = \left(X_{ij} - \bar{X}_{\bullet\bullet} \right)$$

- Group mean centering is subtracting the group/school mean:

$$X_{ij}^{Group} = \left(X_{ij} - \bar{X}_{i\bullet} \right)$$

- These centering methods affect the interpretation of the school intercept

Grand Mean Centering

- Grand mean centering changes the meaning of the intercept in the j^{th} school
- The school intercept is now the mean outcome in the school minus an adjustment due to the student predictors
- With Grand Mean Centering:
  - Student predictors can explain school variance
  - Student predictors are not independent of school predictors
- **Centering changes the precision of the intercept only (as in regression)**

Group Mean Centering

- Group mean centering changes the meaning of the intercept in the i^{th} school
- The school intercept is now the mean outcome in the school *not* adjusted by student predictors
- With Group Mean Centering:
  - Student predictors *cannot* explain school variance
  - Student predictors are *independent* of school predictors
  - Can use aggregate variables at the school level
  - Student predictors are adjusted for school differences (school effects)
- **Centering changes the precision of all estimates**

Specifying Analyses

Know the inference model

- Think through the levels of the design that will be included in the analysis
- Decide on the inference model for each level
  - *Do I want to generalize to a larger universe than just the units in the sample?*
- Different levels may have different inference models
- Levels that exist in the design may be left out of the analysis model

Specifying Analyses

Know the design

- Generally, ***Covariate effects*** should be fixed effects
- ***Treatment effects*** should be random effects when the design permits it (e.g., block randomized designs)
- Sometimes, there are too few units at a level to permit making it a random effect, even when this is theoretically sensible (e.g., 1 – 2 classrooms in a cluster randomized design that assigns schools)

Applications to Experimental Design

- We will look in detail at the two most widely used experimental designs in education:
  - Cluster randomized designs
  - Block randomized designs

Cluster Randomized Design

The Cluster Randomized Design

The Population (sampling frame)

- We wish to compare two treatments
- Assignment of treatments is made to whole schools randomly
- Many schools with *n* students in each school (assume balanced design)
- Assign *all* students in each school to the *same* treatment

The Cluster Randomized Design

The Experiment

- We wish to compare two treatments
- Assignment of treatments is made to whole schools randomly
- Assign *2m* schools with *n* students in each school (assume balanced design)
- There are *m* schools in each treatment condition
- Assign *all* students in each school to the *same* treatment

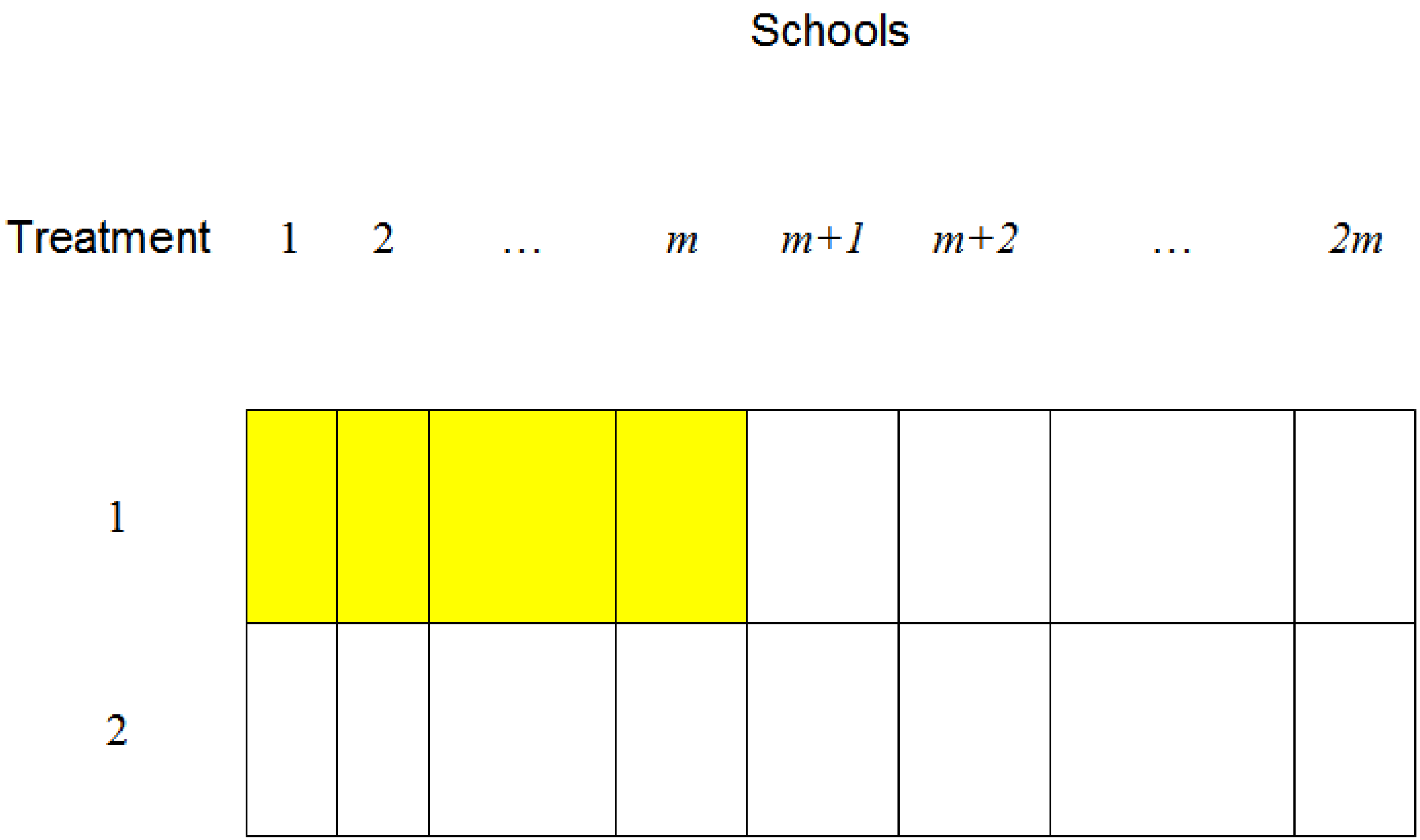
The Cluster Randomized Design

- Diagram of the Experiment:

	Schools							
Treatment	1	2	…	m	$m+1$	$m+2$	…	$2m$
1								
2								

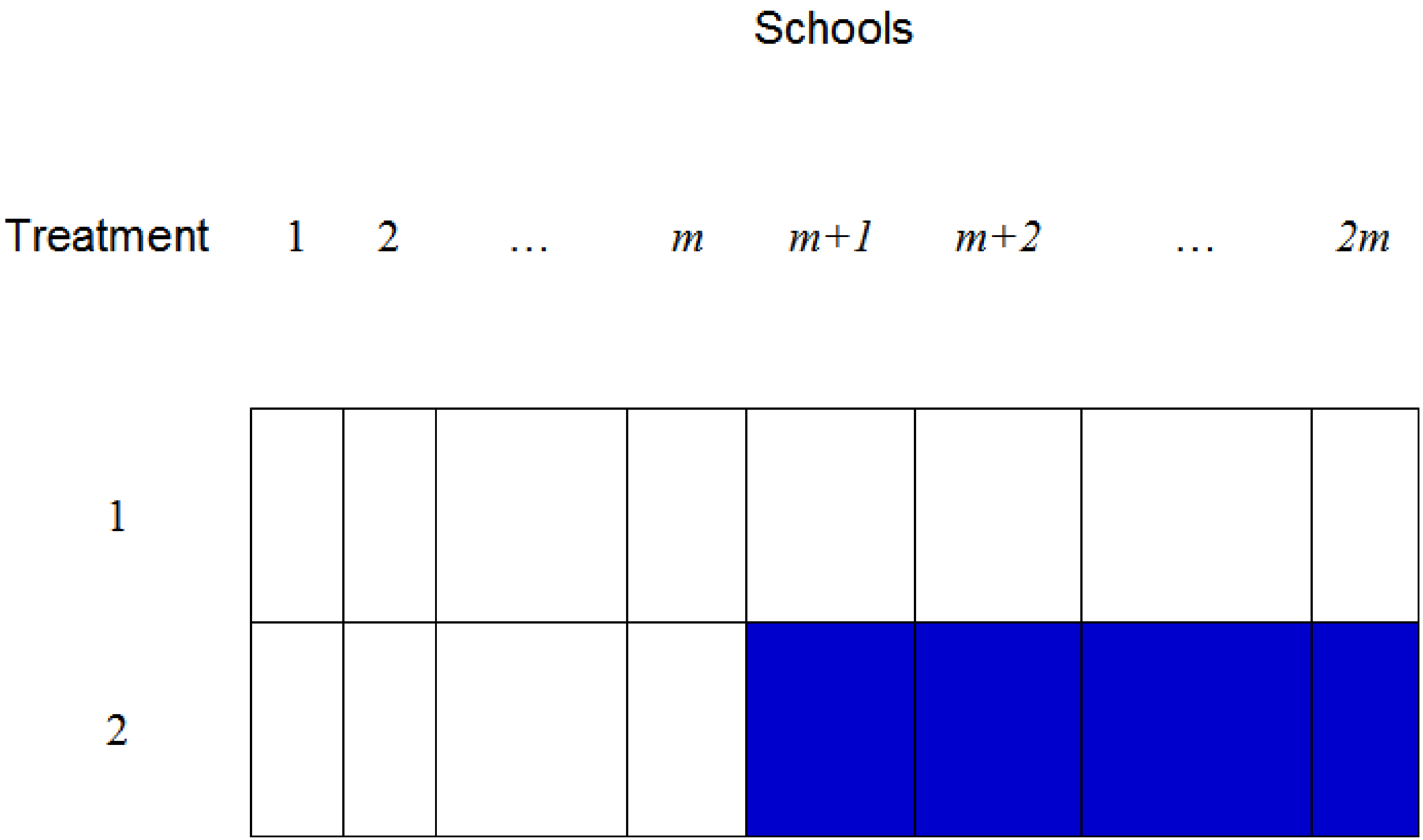
The Cluster Randomized Design

- Treatment 1 Schools:



The Cluster Randomized Design

- Treatment 2 Schools:



Two-Level CRT Design With No Covariates

- With no covariates:

  Level 1 (individual level):

  $$Y_{ij} = \beta_{0j} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_W^2\right)$$

  Level 2 (school level):

  $$\beta_{0j} = \gamma_{00} + \gamma_{01}T_j + \xi_{0j} \qquad \xi_{0j} \sim N\left(0, \sigma_S^2\right)$$

  The intraclass correlation is:

  $$\rho = \sigma_S^2 \Big/ \left(\sigma_S^2 + \sigma_W^2\right) = \sigma_S^2 / \sigma^2$$

Two-Level CRT Design with Covariates

Level 1 (individual level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_W^2\right)$$

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} T_j + \gamma_{02} W_j + \xi_{0j} \qquad \xi_{0j} \sim N\left(0, \sigma_S^2\right)$$

$$\beta_{1j} = \gamma_{10}$$

Note that σ_W^2 and σ_S^2 are adjusted

Covariate effect $\beta_{1j} = \gamma_{10}$ is fixed

Three-Level CRT Design Without Covariates

Level 1 (individual level):

$$Y_{ijk} = \pi_{0jk} + \eta_{ijk} \qquad \eta_{ijk} \sim N\left(0, \sigma_W^2\right)$$

Level 2 (class level):

$$\pi_{0jk} = \beta_{00k} + \varepsilon_{0jk} \qquad \varepsilon_{0jk} \sim N\left(0, \sigma_C^2\right)$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + \gamma_{001} T_k + \xi_{00k} \qquad \xi_{00j} \sim N\left(0, \sigma_S^2\right)$$

There are two intraclass correlation coefficients:

$$\rho_S = \sigma_S^2 \Big/ \left(\sigma_S^2 + \sigma_C^2 + \sigma_W^2\right) = \sigma_S^2 / \sigma^2 \quad \text{(School)}$$

$$\rho_C = \sigma_C^2 \Big/ \left(\sigma_S^2 + \sigma_C^2 + \sigma_W^2\right) = \sigma_C^2 / \sigma^2 \quad \text{(Classroom)}$$

Three-Level CRT Design With Covariates

Level 1 (individual level):

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + \eta_{ijk} \qquad \eta_{ijk} \sim N\left(0, \sigma_W^2\right)$$

Level 2 (class level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} Z_{jk} + \varepsilon_{0jk} \qquad \varepsilon_{0jk} \sim N\left(0, \sigma_C^2\right)$$

$$\pi_{1jk} = \beta_{10k}$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + \gamma_{001} T_k + \gamma_{002} W_k + \xi_{00k} \qquad \xi_{00k} \sim N\left(0, \sigma_S^2\right)$$

$$\beta_{01k} = \gamma_{010}$$

$$\beta_{10k} = \gamma_{100}$$

Covariate effects $\pi_{1jk} = \beta_{10k} = \gamma_{100}$ and $\beta_{01k} = \gamma_{010}$ are fixed

Note that σ_W^2 σ_c^2 and σ_S^2 are adjusted

Standard Errors of Regression Estimates and Clustering

- Appropriate analyses of two and three level data must take into account the multilevel structure (clustering)
- Otherwise the standard errors of regression estimates and statistical tests are incorrect
- The standard errors of treatment effect estimates are typically smaller when clustering is ignored
- This results to higher values of *t*-tests and higher probabilities of finding a significant effect (committing a Type I error)

Standard Errors of Regression Estimates and Clustering

- There are at least three ways of adjusting standard errors for clustered data
  - Conduct the analysis using multilevel models (e.g., SAS proc MIXED, HLM, Stata xtmixed)
  - Post hoc corrections:
    - Use the design effect: multiply the square root of the design effect with the standard error of the estimate.
    - Use the Huber-White robust standard errors (Stata) that adjust for clustering and heterogeneity of variance

Block Randomized Design

The Block Randomized Design

The Population (sampling frame)

- We wish to compare two treatments
- We may assign treatments within schools randomly
- Many schools with 2*n* students in each school (assume balanced design)
- Assign *n* students to each treatment in each of the schools

The Block Randomized Design

- Diagram of the Experiment:

	Schools			
Treatment	1	2	…	m
1			. . .	
2			. . .	

The Block Randomized Design

Schools

Treatment	1	2	…	m
1			. . .	
2			. . .	

Two-Level BRT Design

- The statistical model for the observation on the i^{th} student in the k^{th} treatment in the j^{th} school:

(tradition Design of Experiment notation)

$$Y_{ijk} = \mu + \alpha_k + \beta_j + \alpha\beta_{jk} + \varepsilon_{ijk}$$

where:

$\mu =$ grand mean

$\alpha_k =$ average effect of being in treatment *k*

$\beta_j =$ average effect of being in school *j*

$\alpha\beta_{jk} =$ treatment by school interaction

$\varepsilon_{ijk} =$ residual

Effect of School

School random effect

$$Y_{ijk} = \mu + \alpha_k + \beta_j + \alpha\beta_{jk} + \varepsilon_{ijk}$$

Treatment by school random effect

Two-Level BRT Design With No Covariates

- With no covariates:

Level 1 (individual level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} T_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_W^2\right)$$

Level 2 (school level):

School random effect

$$\beta_{0j} = \gamma_{00} + \xi_{0j} \qquad \xi_{0j} \sim N\left(0, \sigma_S^2\right)$$

$$\beta_{1j} = \gamma_{10} + \xi_{1j} \qquad \xi_{1j} \sim N\left(0, \sigma_{S\times T}^2\right)$$

Treatment by School random effect

Two-Level BRT Design with Covariates

Level 1 (individual level):

$$Y_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma^2_{W,adjusted}\right)$$

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \xi_{0j} \qquad \xi_{0j} \sim N\left(0, \sigma^2_{S,adjusted}\right)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + \xi_{1j} \qquad \xi_{1j} \sim N\left(0, \sigma^2_{T\times S,adjusted}\right)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}W_j$$

Three-Level BRT Design With No Covariates, Treatment at Level 2

Level 1 (individual level):

$$Y_{ijk} = \pi_{0jk} + \eta_{ijk} \qquad \eta_{ij} \sim N\left(0, \sigma_W^2\right)$$

Level 2 (class level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} T_{jk} + \varepsilon_{0jk} \qquad \varepsilon_{0jk} \sim N\left(0, \sigma_C^2\right)$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + \xi_{00k} \qquad \xi_{00j} \sim N\left(0, \sigma_S^2\right)$$

$$\beta_{01k} = \gamma_{010} + \xi_{01k} \qquad \xi_{01j} \sim N\left(0, \sigma_{T \times S}^2\right)$$

Three-Level CRT Design With Covariates Treatment at Level 2

Level 1 (individual level):

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + \eta_{ijk} \qquad \eta_{ijk} \sim N\left(0, \sigma^2_{W,adjusted}\right)$$

Level 2 (class level):

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} T_{jk} + \beta_{02k} Z_{jk} + \varepsilon_{0jk} \qquad \varepsilon_{0jk} \sim N\left(0, \sigma^2_{C,adjusted}\right)$$

$$\pi_{1jk} = \beta_{10k}$$

Level 3 (school level):

$$\beta_{00k} = \gamma_{000} + \gamma_{001} W_k + \xi_{00k} \qquad \xi_{00k} \sim N\left(0, \sigma^2_{S,adjusted}\right)$$

$$\beta_{01k} = \gamma_{010} + \gamma_{011} W_k + \xi_{01k} \qquad \xi_{01k} \sim N\left(0, \sigma^2_{T\times S,adjusted}\right)$$

$$\beta_{02k} = \gamma_{020}$$

$$\beta_{10k} = \gamma_{100}$$

Covariate effects for X and Z are fixed

Moderator/Interaction Effects

Does the Treatment have the Same Effect on ALL Groups?

- School interventions frequently have dual objectives:
  - Increase achievement for all students
  - Decrease the achievement gap
- Note that decreasing the achievement gap suggests that the treatment effect is NOT the same for ALL students
- Some students may benefit more from treatments than others
- That suggests that the treatment interacts with student characteristics or that student characteristics moderate the treatment effect on the outcome (e.g., treatment effect on achievement is more pronounced for disadvantaged students)

Moderator/Interaction

- We are interested in whether small classes increase minority students' achievement more than for other students

  Or
- We are interested in whether small classes affect student achievement more in urban schools than other schools
- In these examples minority status or urban school are the moderating variables

Modeling Statistical Interactions

- In linear models the simplest way to model two-way interactions is by including in the regression equation the two main effects (e.g., small class and minority or urban school) and the two-way interaction (the product of the two main variables)

Two-Level Block Randomized Design

Same level interaction:

Level 1 (individual level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} T_{ij} + \beta_{2j} X_{ij} + \beta_{3j} T_{ij} X_{ij} + \varepsilon_{ij}$$

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \xi_{0j}$$

$$\beta_{1j} = \gamma_{10} + \xi_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

and γ_{30} is the moderator/interaction effect

Two-Level Block Randomized Design

Cross level interaction:

Level 1 (individual level):

$$Y_{ij} = \beta_{0j} + \beta_{1j} T_{ij} + \beta_{2j} T_{ij} + \varepsilon_{ij}$$

Level 2 (school level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \xi_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \xi_{1j}$$

and γ_{11} is the moderator/interaction effect

Effect Sizes

Effect Sizes

- Effect sizes can be defined in more than one way in multilevel designs
- The key difference is what standard deviation is used to standardize
- The easiest one to use is the total standard deviation

Effect Sizes

- In two-level cluster randomized designs, this leads to:

$$\delta = \frac{\gamma_{01}}{\sqrt{\sigma_S^2 + \sigma_W^2}}$$

Treatment effect

Total standard deviation

- In three-level cluster randomized designs, this leads to:

$$\delta = \frac{\gamma_{001}}{\sqrt{\sigma_S^2 + \sigma_C^2 + \sigma_W^2}}$$

Effect Sizes

- In two-level block randomized designs, this leads to:

$$\delta = \frac{\gamma_{10}}{\sqrt{\sigma_S^2 + \sigma_{T\times S}^2 + \sigma_W^2}}$$

Treatment effect

Total standard deviation

- In three-level block randomized designs, this leads to:

$$\delta = \frac{\gamma_{010}}{\sqrt{\sigma_S^2 + \sigma_{T\times S}^2 + \sigma_C^2 + \sigma_W^2}}$$

Questions About Analyses

Questions About Analyses

Q. My schools all come from two districts, but I am randomly assigning the schools. Do I have to take district into account some way?

A. In this case a district dummy can be created and included as a school level variable to capture possible district effects. But districts cannot be modeled as random effects in this case (very few units).

Questions About Analyses

Q. Why can't I use regression to analyze experiments? What's the advantage of multilevel models?

A. Of course regression can be used to analyze experimental data. However, if there is clustering in the data, the standard errors of the estimates need to be adjusted. This can be done using either a design effect or robust standard errors. Multilevel model methods correct the standard errors instantly via the computation. In addition, multilevel models allow analysts to model random effects, may that be treatment effects (e.g., block randomized design), classrooms or schools, or socioeconomic or minority status. That is, effects can differ from school to school. In regression one can model interactions between school and other variables but that suggests many predictors in the model (not a parsimonious model).

Questions About Analyses

Q. Can I use “school fixed effects” to analyze data from a randomized block design?

A. Yes one can use a regression model and include school fixed effects. It is also recommended however, that one includes interactions between the treatment and schools. If there are 81 schools in the sample that suggests 160 dummies in the regression model. In multilevel models one only needs to introduce two random effects (one for schools and one for treatment across schools)

Questions About Analyses

Q. We randomly assigned, but our assignment was corrupted by treatment switchers. What do we do?

A. One possibility is to use in the regression the intention to treat variable instead of the treatment received. One can do the analysis both ways and examine differences in estimates. Another possibility is to use intention to treat as an instrument for received treatment.

Questions About Analyses

Q. We randomly assigned, but our assignment was corrupted by attrition. What do we do?

A. One possibility is to use simple tests (t of F tests) to examine whether dropouts are different on average than stayers within treatment conditions in an outcome such as achievement. Another possibility is to use some kind of imputation scheme, analyze the data with and without imputation and examine differences in the estimates.

Questions About Analyses

Q. We randomly assigned but got a big imbalance on characteristics we care about (gender, race, language, SES, pretest scores). What do we do?

A. One possibility is to use a post hoc adjustment. That is, in the regression or multilevel model one needs to include these variables as statistical controls or covariates to correct for selection. Perhaps interactions between treatment and characteristics should also be included in the model. Another possibility is to use these variables in propensity score methods that aim to create similar groups for the two treatment conditions.

Questions About Analyses

Q. We want to use student covariates to improve precision, but we find that they act somewhat differently in different schools (have different slopes). What do we do?

A. In a multilevel model one can model student covariates as random effects at the school level and compute their variances across schools. In regression one would need to create student covariate by school interactions (that may produce a large number of interaction effects)

Questions About Analyses

Q. We get somewhat different variances in different schools. Should we use robust standard errors?

A. Yes that is a common way to address the problem of variance heterogeneity.