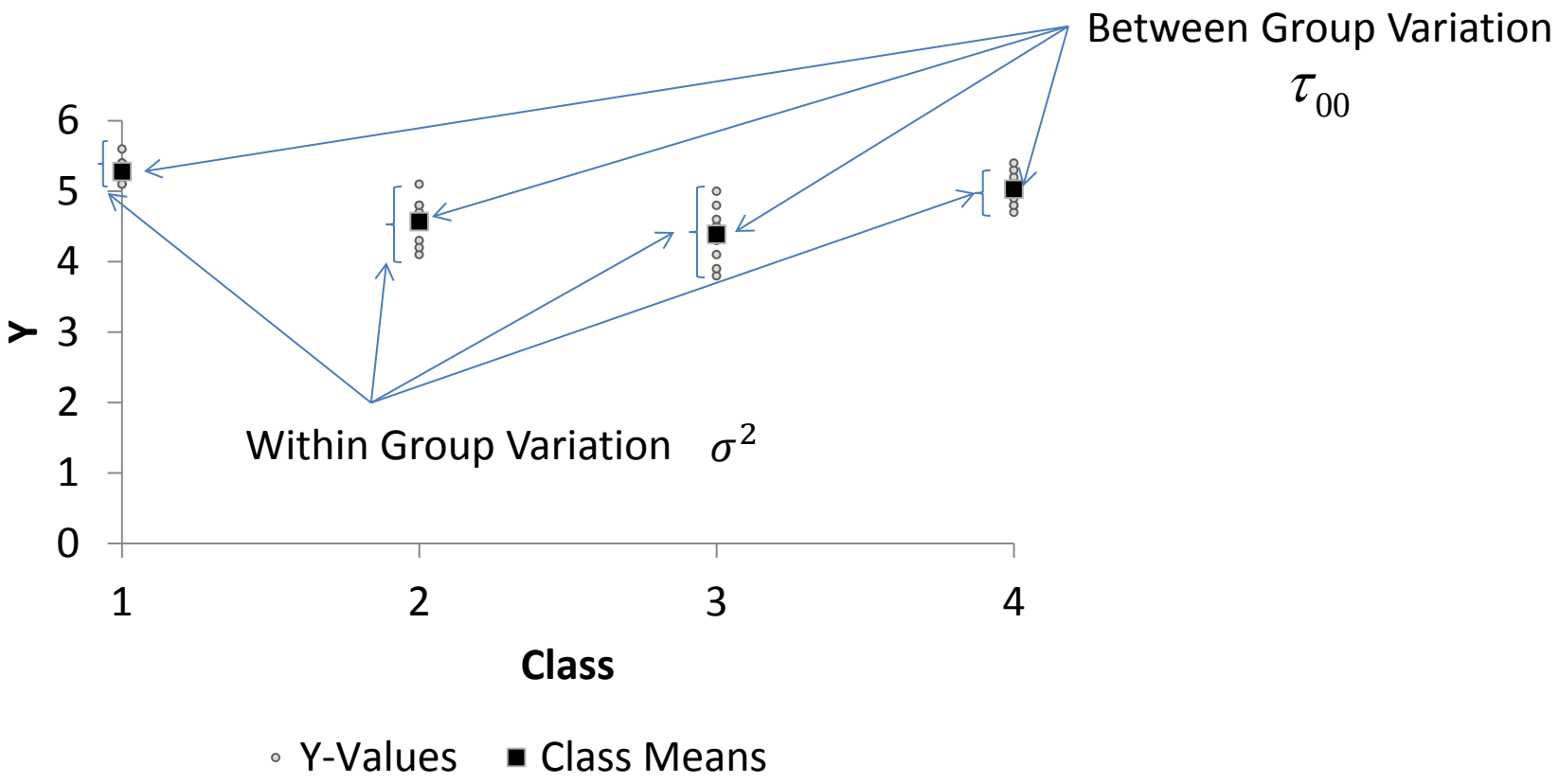
Between and Within Group Variation



Intra Class Correlation (ICC or ρ)

Random Effect ANOVA

$$\rho = \frac{(MS_B - MS_W)}{[MS_B + (n_{cluster} - 1)MS_W]}$$

Unconditional Means or Intercept-only MLM model

$$\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)}$$

Unconditional Means or Intercept-only MLM model

Level 1:

$$y_{ij} = \beta_{0j} + e_{ij}$$

where β_{0j} is the mean for group *j and* e_{ij} (others use r_{ij}) represents residual individual differences from the mean of group *j*.

Level 2:

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

where γ_{00} is the grand mean and μ_{0j} is the deviation of the group mean from the grand mean.

With substitution:

$$y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$$

Unconditional Means or Intercept-only MLM model

$$y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$$

$$\rho = \frac{var(\mu_{0j})}{var(\mu_{0j}) + var(e_{ij})}$$

$$\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)}$$

ICC and Design Effect

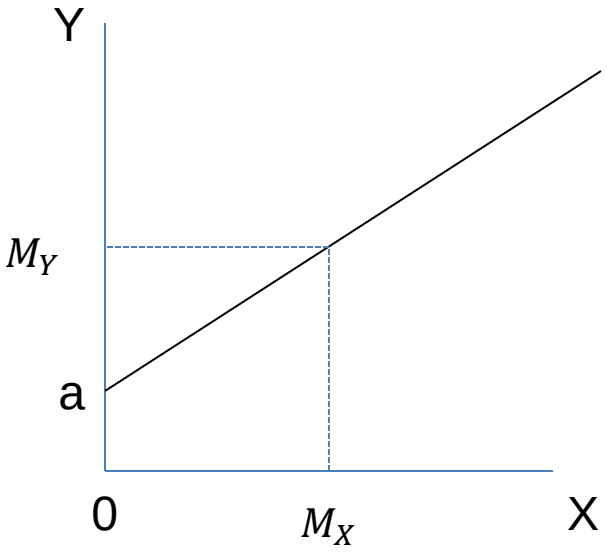
$$Design\ effect = 1 + (n_c - 1)\rho$$

n_c= number of observations per cluster
ρ = intraclass correlation

$$var_{eff} = var[1 + (n_c - 1)\rho]$$

$$n_{eff} = \frac{n_{total}}{[1 + (n_c - 1)\rho]}$$

Intercepts and Centering



$$Y' = a + bX$$

$$a = Y' - bX$$

$$a = M_Y - bM_X$$

If M_X is 0 (centered), then a = M_Y

Group mean vs. grand mean centering

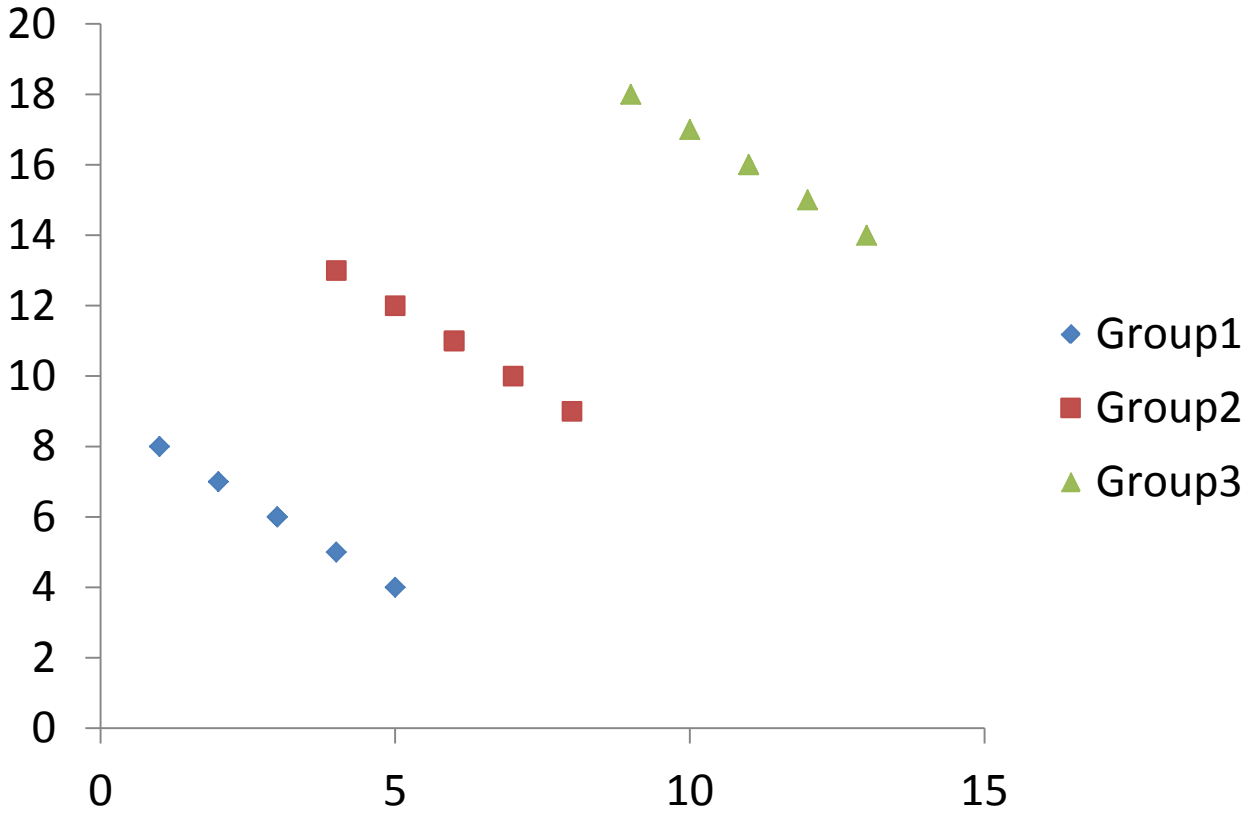
Ignoring Multilevel Structure

Aggregation

Disaggregation

Ecological Fallacy

Between and Within Groups Correlations



Example (from Joop Hox book on Multilevel Analysis, 2010)

DATA

Level 1: pupils (2000 in total)

Level 2: Classes (100 classes of approx 20 pupils each)

Outcome variable:

pupil popularity (scale 0-10)

Predictor variables at pupil level (level 1)

pupil gender (1=girl 0=boy)

pupil extraversion (scale 1-10)

Predictor variable(s) at class level (level 2)

teacher experience (scale 2-25)

Data Structure and Syntax for Intercept-Only Model

```
TITLE: Popularity data using Mplus;
DATA:
  FILE IS "popular2.dat";
VARIABLE:
  NAMES ARE class pupil cons extrav sex texp
  popular popteach zextrav zsex ztexp zpopular zpoptch;
  USEVARIABLES are popular;
  CLUSTER IS class;
ANALYSIS:
  TYPE IS TWOLEVEL random;
  ESTIMATOR IS ML; !default is MLR
MODEL:
  %within%
  popular;
  %between%
  popular;
OUTPUT: SAMPSTAT;
```

```
1 11 1 5 1 24 5.7 5 -.17031 .98881 1.48615 .45102 -.04308
1 12 1 5 1 24 4.8 5 -.17031 .98881 1.48615 -.19996 -.04308
1 13 1 5 0 24 5 5 -.17031 -1.01081 1.48615 -.0553 -.04308
1 14 1 5 1 24 5.5 6 -.17031 .98881 1.48615 .30636 .66906
1 15 1 5 1 24 6 5 -.17031 .98881 1.48615 .66802 -.04308
1 16 1 6 1 24 5.7 5 .62185 .98881 1.48615 .45102 -.04308
1 17 1 4 0 24 3.2 2 -.96248 -1.01081 1.48615 -1.35727 -2.17951
1 18 1 4 0 24 3.1 3 -.96248 -1.01081 1.48615 -1.4296 -1.46737
1 19 1 7 1 24 6.6 7 1.41401 .98881 1.48615 1.10201 1.3812
1 20 1 4 0 24 4.8 4 -.96248 -1.01081 1.48615 -.19996 -.75523
2 21 1 8 1 14 6.4 6 2.20617 .98881 -.04014 .95734 .66906
2 22 1 4 0 14 2.4 3 -.96248 -1.01081 -.04014 -1.93592 -1.46737
2 23 1 6 0 14 3.7 4 .62185 -1.01081 -.04014 -.99561 -.75523
2 24 1 5 1 14 4.4 4 -.17031 .98881 -.04014 -.48929 -.75523
2 25 1 5 1 14 4.3 4 -.17031 .98881 -.04014 -.56162 -.75523
2 26 1 5 0 14 4 4 -.17031 -1.01081 -.04014 -.77861 -.75523
2 27 1 4 1 14 3.8 5 -.96248 .98881 -.04014 -.92328 -.04308
2 28 1 5 0 14 4.2 5 -.17031 -1.01081 -.04014 -.63395 -.04308
2 29 1 6 1 14 5.1 4 .62185 .98881 -.04014 .01703 -.75523
2 30 1 4 1 14 4.1 4 -.96248 .98881 -.04014 -.70628 -.75523
2 31 1 6 1 14 4.6 5 .62185 .98881 -.04014 -.34462 -.04308
```

Intercept Only Model

SUMMARY OF DATA

Estimated Intraclass Correlations for the Y Variables

Variable	Intraclass Correlation
POPULAR	0.362

0.695/(0.695 + 1.222) = .362

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
POPULAR	1.222	0.047	26.199	0.000
Between Level				
Means				
POPULAR	5.078	0.087	58.394	0.000
Variances				
POPULAR	0.695	0.108	6.421	0.000

Influence of Design on Standard Error

Syntax including clustering

```
TITLE: Popularity data using Mplus;
DATA:
  FILE IS "popular2.dat";
VARIABLE:
  NAMES ARE class pupil cons extrav sex texp
  popular popteach zextrav zsex ztexp zpopular zpoptch;
  USEVARIABLES are popular;
  CLUSTER IS class;
ANALYSIS:
  TYPE IS TWOLEVEL random;
  ESTIMATOR IS ML; !default is MLR
MODEL:
  %within%
  popular;
  %between%
  popular;
OUTPUT: SAMPSTAT CINTERVAL;
```

Syntax ignoring clustering

```
TITLE: Popularity data using Mplus;
DATA:
  FILE IS "popular2.dat";
VARIABLE:
  NAMES ARE class pupil cons extrav sex texp
  popular popteach zextrav zsex ztexp zpopular zpoptch;
  USEVARIABLES are popular;
  !CLUSTER IS class;
ANALYSIS:
  !TYPE IS TWOLEVEL random;
  ESTIMATOR IS ML; !default is MLR
MODEL:
  !%within%
  popular;
  !%between%
  !popular;
OUTPUT: SAMPSTAT CINTERVAL;
```

Influence of Design on Standard Error

MODEL RESULTS including clustering

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
POPULAR	1.222	0.040	30.822	0.000
Between Level				
Means				
POPULAR	5.078	0.087	58.390	0.000
Variances				
POPULAR	0.695	0.107	6.489	0.000

MODEL RESULTS ignoring clustering

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
POPULAR	5.076	0.031	164.252	0.000
Variances				
POPULAR	1.910	0.060	31.623	0.000

Standard error lower than it should be

Influence of Design on Standard Error

including clustering

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
Within Level							
Variances							
POPULAR	1.120	1.144	1.157	1.222	1.287	1.299	1.324
Between Level							
Means							
POPULAR	4.854	4.907	4.935	5.078	5.221	5.248	5.302
Variances							
POPULAR	0.419	0.485	0.519	0.695	0.871	0.904	0.970

ignoring clustering

CONFIDENCE INTERVALS OF MODEL RESULTS

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
Means							
POPULAR	4.997	5.016	5.026	5.076	5.127	5.137	5.156
Variances							
POPULAR	1.755	1.792	1.811	1.910	2.010	2.029	2.066

Confidence intervals narrower than should be