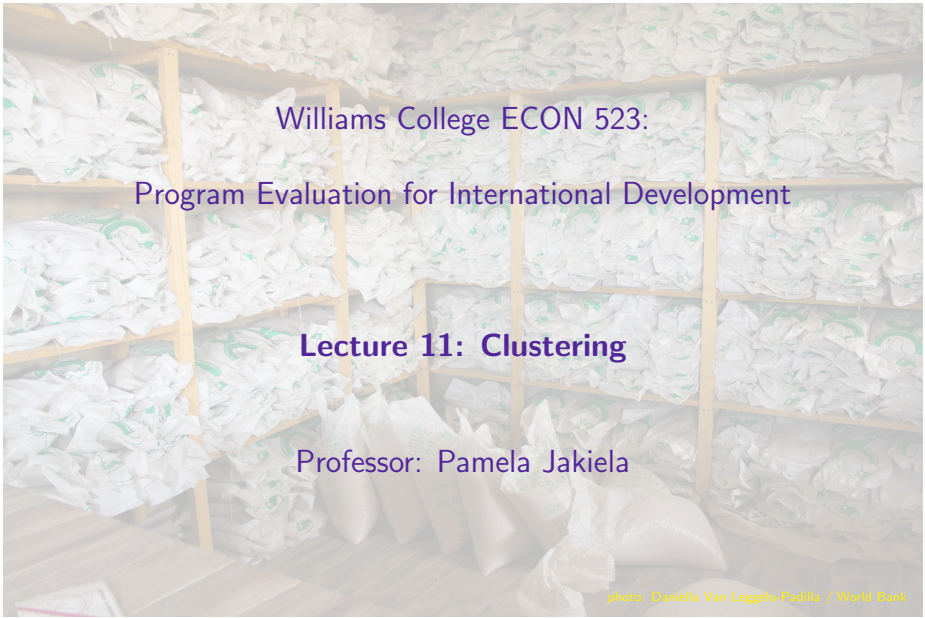


Williams College ECON 523:

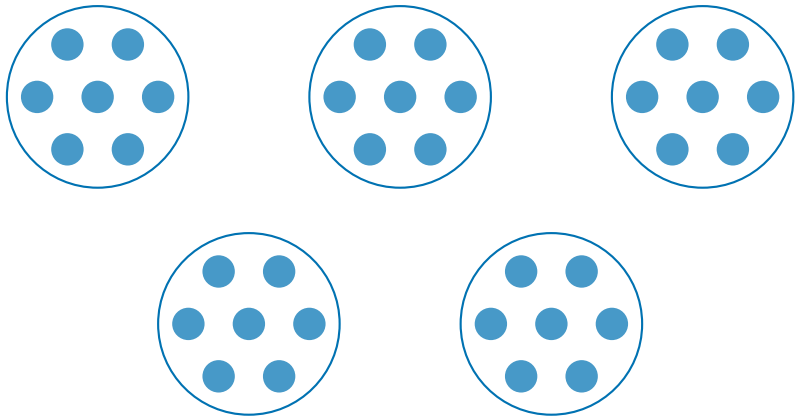
Program Evaluation for International Development

Lecture 11: Clustering

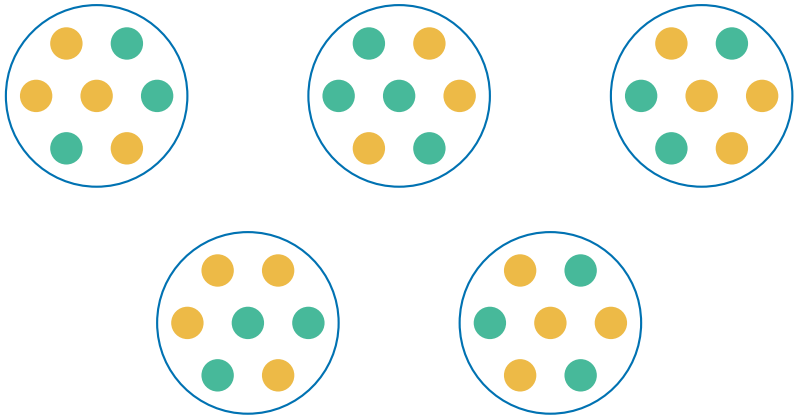
Professor: Pamela Jakiela

photo: Daniella Van Leggelo-Padilla / World Bank

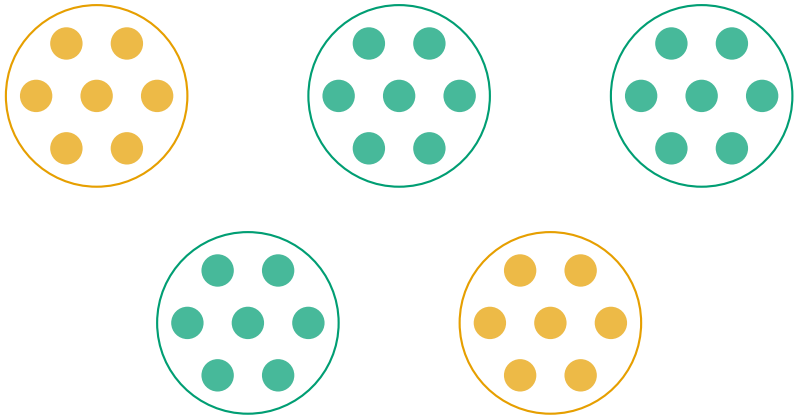
What Happens When We Assign Treatment at the Cluster Level?



Outcomes Are Independent Within Clusters



Outcomes Are Perfectly Correlated Within Clusters



Power Calculations in Cluster-Randomized Trials

$$MDE = \left(t_{1-\kappa} + t_{\alpha/2}\right) \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{N}}$$

size and power ≈ 2.8

$SE(\hat{\beta})$

Power Calculations in Cluster-Randomized Trials

$$MDE = \left(t_{1-\kappa} + t_{\alpha/2}\right) \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{\sigma^2}{N}} \sqrt{1 + (n_g - 1)\,\rho}$$

Moulton factor or **design effect**

n_g = observations per cluster

ρ = intra-class correlation

Power Calculations in Cluster-Randomized Trials

$$
\begin{aligned}
MDE &= \left(t_{1-\kappa} + t_{\alpha/2}\right) \sqrt{\tfrac{1}{P(1-P)}} \sqrt{\tfrac{\sigma^2}{N}} \sqrt{1 + (n_g - 1)\,\rho} \\
&\approx 2.8 \sqrt{\tfrac{1}{P(1-P)}} \sqrt{\tfrac{\sigma^2}{N}} \sqrt{1 + (n_g - 1)\,\rho}
\end{aligned}
$$

Power Calculations in Cluster-Randomized Trials

$$\begin{aligned} MDE &= \left(t_{1-\kappa} + t_{\alpha/2}\right) \sqrt{\tfrac{1}{P(1-P)}} \sqrt{\tfrac{\sigma^2}{N}} \sqrt{1 + (n_g - 1)\,\rho} \\ &\approx 2.8 \sqrt{\tfrac{1}{P(1-P)}} \sqrt{\tfrac{\sigma^2}{N}} \sqrt{1 + (n_g - 1)\,\rho} \end{aligned}$$

When $\rho = 0$, clusters do not matter (from a power perspective)

When $\rho = 1$, effective sample size is the number of clusters N/n_g

What Is ρ?

Outcome	ICC	Context	Cluster	Reference
Microenterprise assets	0.01	Hyderabad, India	Urban neighborhood	Banerjee et al. (2015)
Borrowing from banks	0.04	Hyderabad, India	Urban neighborhood	Banerjee et al. (2015)
Microcredit borrowing	0.11	Hyderabad, India	Urban neighborhood	Banerjee et al. (2015)
Height-for-age z-score	0.02	Kisumu, Kenya	Rural village	Jakiela et al. (2023)
Mother tongue receptive vocabulary	0.02	Kisumu, Kenya	Rural village	Jakiela et al. (2023)
English receptive vocabulary	0.03	Kisumu, Kenya	Rural village	Jakiela et al. (2023)
Expressive vocabulary	0.05	Kisumu, Kenya	Rural village	Jakiela et al. (2023)
Neonatal mortality	0.01	Malawi	DHS clusters	Godlonton and Okeke (2015)
Skilled birth attendant present	0.14	Malawi	DHS clusters	Godlonton and Okeke (2015)
Household has electricity	0.29	Malawi	DHS clusters	Godlonton and Okeke (2015)
Math and language test scores	0.22	Busia, Kenya	Classroom	Miguel and Kremer (2004)
Math and language test scores	0.23	Udaipur, India	Classroom	Duflo and Hanna (2005)
Math test scores	0.62	Busia, Kenya	Classroom	Glewwe et al. (2004)

In-Class Activity

```
clear all
set seed 24601

local numclusters = 1000
local obspercluster = 1
local effect = 0

// create an empty matrix to save results
local loopmax=100
matrix pval=J(`loopmax',1,.)

// create data sets w/ clusters
forvalues i  =1/`loopmax' {
    display "Loop iteration `i'"
    quietly set obs `numclusters'
    quietly gen clustid = _n
    quietly gen treatment=cond(_n>`numclusters'/2,1,0)
    quietly gen clusteffect = rnormal()
    quietly expand `obspercluster'
    quietly gen y = `effect'*treatment + clusteffect + rnormal()
    quietly reg y treatment
    mat V = r(table)
    matrix pval[`i',1]=V[4,1]
    drop clustid treatment clusteffect y
}

```

In-Class Activity

```
7
8  // create an empty matrix to save results
9  local loopmax=100
10 matrix pval=J(`loopmax',1,.)
11
12 // create data sets w/ clusters
13 forvalues i  =1/`loopmax' {
14     display "Loop iteration `i'"
15     quietly set obs `numclusters'
16     quietly gen clustid = _n
17     quietly gen treatment=cond(_n>`numclusters'/2,1,0)
18     quietly gen clusteffect = rnormal()
19     quietly expand `obspercluster'
20     quietly gen y = `effect'*treatment + clusteffect + rnormal()
21     quietly reg y treatment
22     mat V = r(table)
23     matrix pval[`i',1]=V[4,1]
24     drop clustid treatment clusteffect y
25 }
26
27 // store results
28 svmat pval
29 summarize
30 gen significant = pval<0.05
31 tab significant
32
```