ECON 626: Applied Microeconomics

Lecture 7:

Power and Clustering

Professors: Pamela Jakiela and Owen Ozier

Department of Economics
University of Maryland, College Park

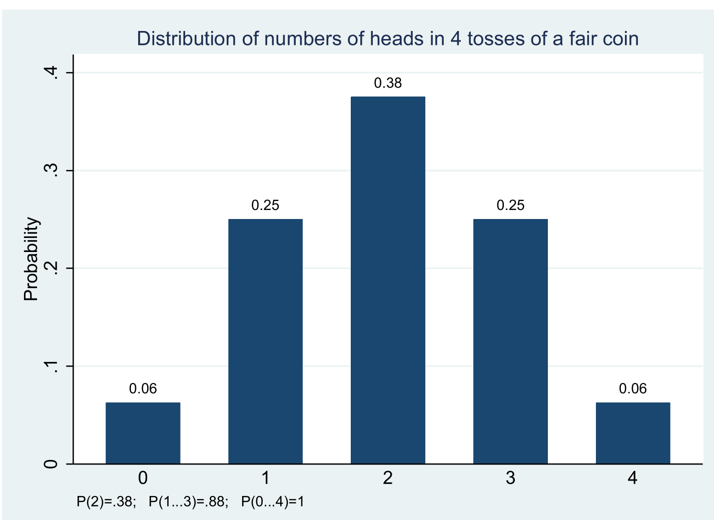
Power

- **Power**:
  probability of rejecting... the null, when... the alternative is true.
- In randomized trials:
  probability of having a statistically significant coefficient on treatment when there is, in fact, an effect of treatment.
- A "power calculation" is... a sample size calculation.
  This means predicting... the standard error.

Coin toss example

- "Null" Hypothesis: the coin is fair
  50% chance of heads, 50% chance of tails.
- Structure of the data:
  Toss the coin a number of times, count heads.
- The test:
  "Fail to reject" null if within some distance of mean under the null;
  "Reject" otherwise.
- If we only had 4 tosses of the coin, what cutoffs could we use?
  Could fail to reject under any of these conditions:
  - (A) never
  - (B) when exactly the mean (2 heads)
  - (C) when within 1 (1, 2, or 3 heads)
  - or (D) always.
- We don't want to reject the null when it is true, though;
  How much accidental rejection would each possible cutoff give us?

Distribution of possible results



Types of error

	Test result	
	"Reject Null," Find an effect!	"Fail to Reject Null," Conclude no effect.
Truth: There is an effect	Great!	"Type II Error" (low power)
Truth: There is NO effect	"Type I Error" (test size)	Great!

The probability of Type I error (given the null) is the "size" of the test. By convention, we are usually interested in tests of "size" 0.05.

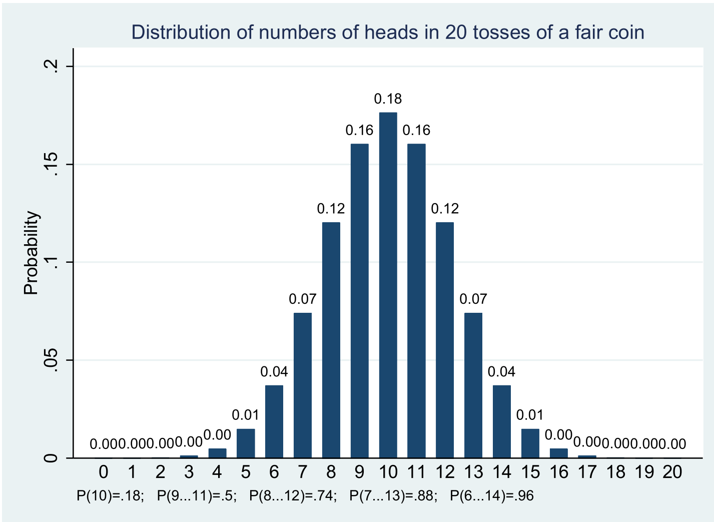
The probability of Type II error is also very important;
If P(failure to detect an effect|there is an effect) $= 1 - \kappa$,
then the power of the test is κ.

Power depends on anticipated effect size; we typically want power $\geq$ 80%.

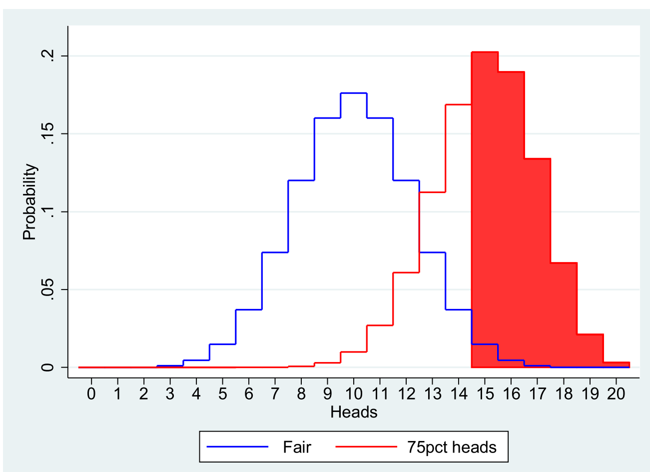
Not enough data even for meaningful test size

- There is no way* to create such a test with four coin tosses so that the chance of accidental rejection under the "null" hypothesis (sometimes written H_0) is less than 5%, a standard in social science.
  ** (Except the "never reject, no matter what" rule. Not very useful.)*
- What about 20 coin tosses?

Distribution of possible results

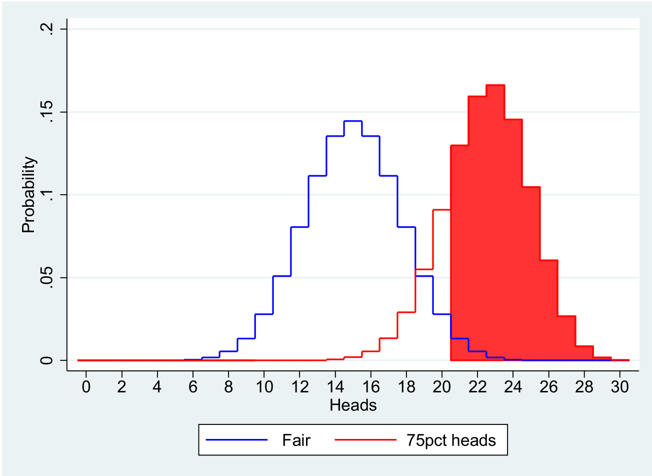


Power with 20 tosses



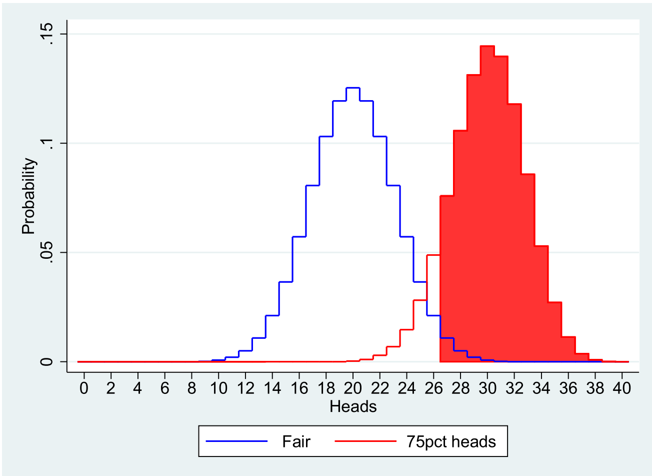
Power: about 0.62

Power with 30 tosses



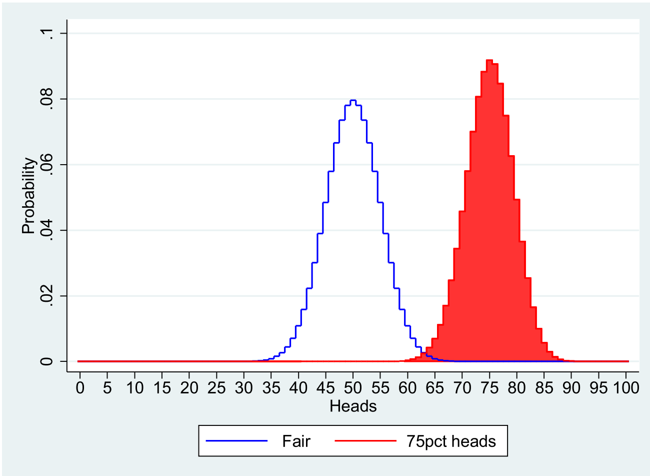
Power: about 0.80

Power with 40 tosses



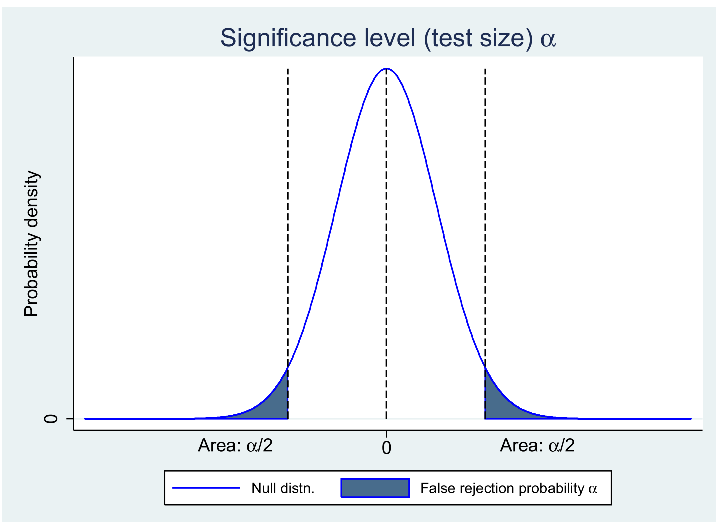
Power: about 0.90

Power with 100 tosses

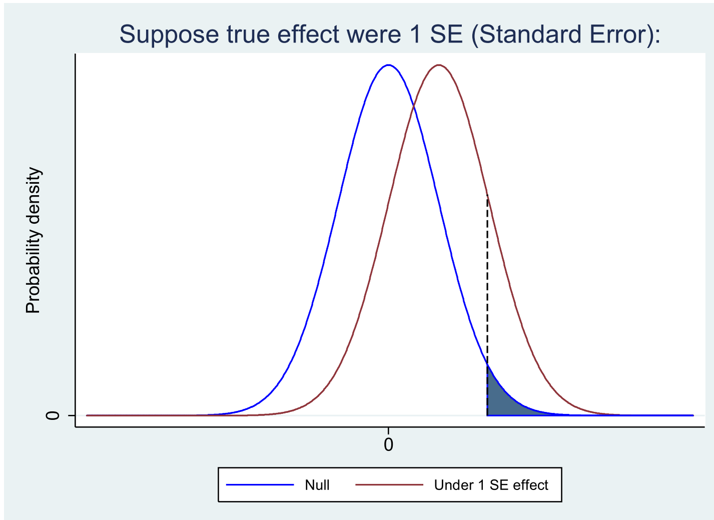


Power: about 0.9997

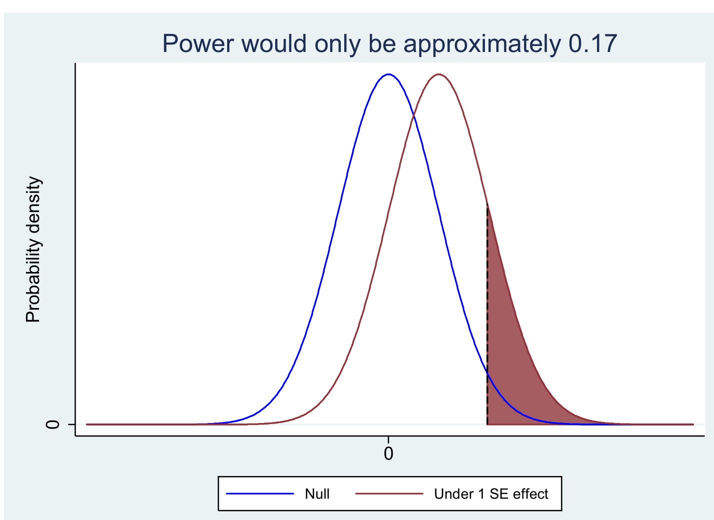
Rejecting H_0 in critical region



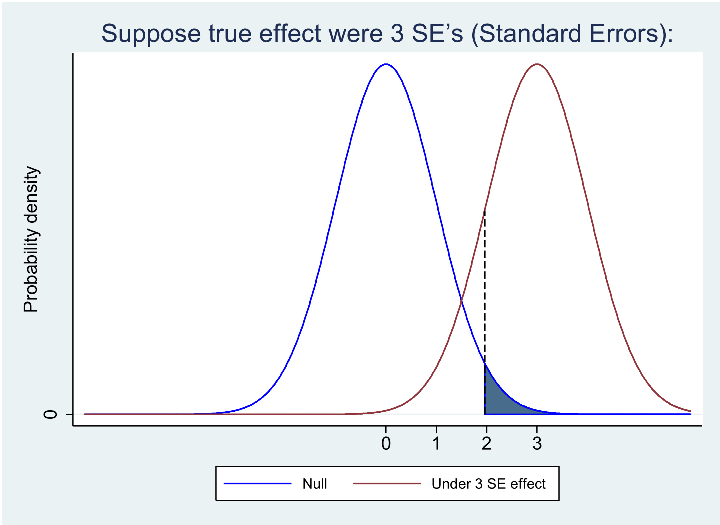
Under an alternative:



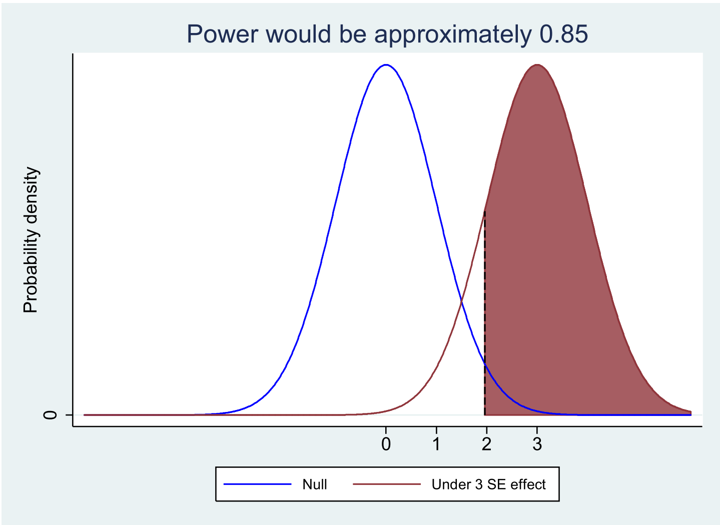
Under an alternative:



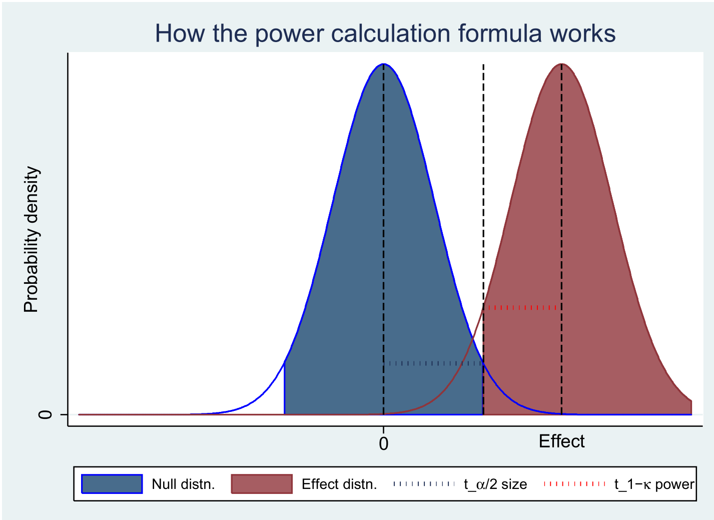
Under an alternative:



Under an alternative:



Power calculation, visually



Note: see the related figure in the *Toolkit* paper.

The formula: for power κ and size α,

$Effect > (t_{1-\kappa}+t_{\alpha/2})SE(\hat{\beta})$ Notation: $t_{1-p} = p^{th}$ percentile of the t dist'n.

Note that the formula above works no matter the design.
Usually: $\alpha = 0.05$, $\kappa = 0.80$, N is large, so:

$$\text{Minimum Detectable Effect} \approx (0.84 + 1.96)SE(\hat{\beta}) \approx 2.8SE(\hat{\beta})$$

We focus on sample size. But how would imperfect compliance or baseline data affect this? Below, I continue for the standard RCT case.

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

In practice (Stata): **sampsi**
Note: Stata uses z rather than t distribution (skirting D.O.F. issue).
We could also flip this equation around:

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

The formula: for power κ and size α,

Where do these numbers come from, σ^2 and the effect size?
Two basic options:

- Consider standardized effect sizes in terms of standard deviations
- Draw on existing data: What is available that could inform your project?

What if treatment is assigned by groups?

We have been thinking here of randomizing at the individual level.
But in practice, we often randomize larger units.
Examples:

- **Entire schools** are assigned to treatment or comparison; we observe outcomes at the level of the individual pupil
- **Classes within a school** are assigned to treatment or comparison; we observe outcomes at the level of the individual pupil
- **Households** are assigned to treatment or comparison; we observe outcomes at the level of the individual family member
- **Sub-district locations** are assigned to treatment or comparison; we observe outcomes at the level of the individual road
- **Bank branch offices** are assigned to treatment or comparison; we observe outcomes at the level of the individual borrower

What does this do?
It depends on how much variation is explained by the group each individual is in.

What happens to the variance of the estimator?

Suppose $y_i = \beta t_i + \epsilon_i$. We compare the means of those with $t_i = 1$ to those with $t_i = 0$. Departure point: iid ϵ_i having variance σ_ϵ^2, and equal numbers of observations in treatment and control ($N/2$ in each):

$$\hat{\beta} = \frac{1}{N/2}\sum_T y_i - \frac{1}{N/2}\sum_C y_i$$

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_T \epsilon_i - \frac{1}{N/2}\sum_C \epsilon_i$$

$$Var(\hat{\beta}) = \frac{1}{N/2}\sigma_\epsilon^2 + \frac{1}{N/2}\sigma_\epsilon^2 = \frac{4}{N}\sigma_\epsilon^2$$

$$SE(\hat{\beta}) = \sqrt{4}\sqrt{\frac{\sigma_\epsilon^2}{N}}$$

This is the formula from before, with $P = 1/2$:

$$\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}$$

What happens to the variance of the estimator?

Now suppose $y_i = \beta t_i + \epsilon_i$, but $\epsilon_i = \nu_g + \eta_{ig}$ for groups g of fixed size n_g. We still compare the means of those with $t_i = 1$ to those with $t_i = 0$. Departure point: within a group, treatment is either 1 or 0; iid ν_g having variance σ_ν^2, iid η_{ig} having variance σ_η^2, so that $\sigma_\epsilon^2 = \sigma_\nu^2 + \sigma_\eta^2$, and equal numbers of observations in treatment and control (still $N/2$ in each). Define the "the intra-cluster correlation," ρ:

$$\rho_\epsilon = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\eta^2} = \frac{\sigma_\nu^2}{\sigma_\epsilon^2}$$

Two other ways of writing this will be convenient:

$$\sigma_\nu^2 = \rho_\epsilon \sigma_\epsilon^2$$

$$\sigma_\eta^2 = (1 - \rho_\epsilon)\sigma_\epsilon^2$$

What happens to the variance of the estimator?

As before,

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_T \epsilon_i - \frac{1}{N/2}\sum_C \epsilon_i$$

$$\hat{\beta} = \beta + \frac{1}{N/2}\sum_T \nu_g + \frac{1}{N/2}\sum_T \eta_{ig} - \frac{1}{N/2}\sum_C \nu_g - \frac{1}{N/2}\sum_C \eta_{ig}$$

$$\begin{aligned} Var(\hat{\beta}) =& \frac{1}{N/(2n_g)}\sigma_\nu^2 + \frac{1}{N/2}\sigma_\eta^2 + \frac{1}{N/(2n_g)}\sigma_\nu^2 + \frac{1}{N/2}\sigma_\eta^2 \\ =& \frac{4n_g}{N}\sigma_\nu^2 + \frac{4}{N}\sigma_\eta^2 \\ =& \frac{4}{N}\left(n_g\sigma_\nu^2 + \sigma_\eta^2\right) \\ =& \frac{4}{N}\left(n_g\rho_\epsilon\sigma_\epsilon^2 + (1-\rho_\epsilon)\sigma_\epsilon^2\right) = \frac{4}{N}\sigma_\epsilon^2\left((n_g-1)\rho_\epsilon + 1\right) \end{aligned}$$

$$SE(\hat{\beta}) = \sqrt{4}\sqrt{\frac{\sigma_\epsilon^2}{N}}\sqrt{(n_g-1)\rho_\epsilon+1} = \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}\sqrt{(n_g-1)\rho_\epsilon+1}$$

The formula

Scale the effective standard error by:

$$\textit{Design Effect (``Moulton factor'')} = \sqrt{1 + (n_{groupsize} - 1)\rho}$$

ρ ("rho") is the intra-class correlation.
In practice (Stata): **loneway** and **sampclus**

Recall earlier formula:

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

We could also flip this equation around:

$$\iff N = (z_{1-\kappa} + z_{\alpha/2})^2 \cdot \left(\frac{1}{P(1-P)}\right) \cdot \left(\frac{\sigma^2}{MDE^2}\right) \cdot (1 + (n_{groupsize} - 1)\rho)$$

Estimation example: clustered standard errors

Stata:

$$V_{cluster} = (X'X)^{-1} \sum_{j=1}^{n_c} u_j' u_j (X'X)^{-1}$$

where

$$u_j = \sum_{j_{cluster}} e_i x_i$$

Angrist and Pischke 8.2.6:

$$\hat{\Omega}_{cl} = (X'X)^{-1} \left(\sum_g X_g' \hat{\Psi}_g X_g \right) (X'X)^{-1}$$

where

$$\hat{\Psi}_g = a\hat{e}_g \hat{e}_g' = a \begin{bmatrix} \hat{e}_{1g}^2 & \hat{e}_{1g}\hat{e}_{2g} & \dots & \hat{e}_{1g}\hat{e}_{n_g g} \\ \hat{e}_{2g}\hat{e}_{1g} & \hat{e}_{2g}^2 & \dots & \hat{e}_{2g}\hat{e}_{n_g g} \\ \dots & \dots & \dots & \dots \\ \hat{e}_{n_g g}\hat{e}_{1g} & \hat{e}_{n_g g}\hat{e}_{2g} & \dots & \hat{e}_{n_g g}^2 \end{bmatrix}$$

Estimation example: clustered standard errors

But remember, in the simplest case, X_g' is either:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{bmatrix} \; or \; \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

So

$$X_g' \begin{bmatrix} \hat{e}_{1g}^2 & \hat{e}_{1g}\hat{e}_{2g} & \dots & \hat{e}_{1g}\hat{e}_{n_g g} \\ \hat{e}_{2g}\hat{e}_{1g} & \hat{e}_{2g}^2 & \dots & \hat{e}_{2g}\hat{e}_{n_g g} \\ \dots & \dots & \dots & \dots \\ \hat{e}_{n_g g}\hat{e}_{1g} & \hat{e}_{n_g g}\hat{e}_{2g} & \dots & \hat{e}_{n_g g}^2 \end{bmatrix} X_g$$

Count the terms. diagonal: n_g; off-diagonal: $n_g(n_g - 1)$.
Diagonal terms have expectation σ_ϵ^2,
while off-diagonal terms have expectation $\sigma_\nu^2 = \rho\sigma_\epsilon^2$.
The matrix product then has expectation:

$$\sigma_\epsilon^2 n_g(1 + (n_g - 1)\rho) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \; or \; \sigma_\epsilon^2 n_g(1 + (n_g - 1)\rho) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Estimation example: clustered standard errors

So:

$$E\left[\left(\sum_g X_g' \hat{\Psi}_g X_g\right)\right] = \sigma_\epsilon^2 (1 + (n_g - 1)\rho) \begin{bmatrix} \frac{N}{2} & \frac{N}{2} \\ \frac{N}{2} & N \end{bmatrix}$$

and thus

$$E\left[\hat{\Omega}_{cl}\right] = E\left[(X'X)^{-1}\left(\sum_g X_g' \hat{\Psi}_g X_g\right)(X'X)^{-1}\right]$$

$$= (1 + (n_g - 1)\rho)\ (X'X)^{-1}\ \sigma_\epsilon^2$$

Intra-cluster correlation ρ (greek letter "rho")

But where does this ρ number come from? Two basic options:

- Consider what might be reasonable assumptions
- Draw on existing data (again):
  What is available that could inform your project?

Intra-class correlations we have known

Data source	ICC (ρ)
Madagascar Math + Language	0.5
Busia, Kenya Math + Language	0.22
Udaipur, India Math + Language	0.23
Mumbai, India Math + Language	0.29
Vadodara, India Math + Language	0.28
Busia, Kenya Math	0.62
Busia, Kenya Language	0.43
Busia, Kenya Science	0.35

Duflo, Glennerster, and Kremer (2006) Using Randomization in Development Economics Research: A Toolkit

Data source	ICC (ρ)
US Elementary Math, unconditional	0.22
US Elementary Math, rural only, unconditional	0.15
US Elementary Math, rural only, conditional on previous scores	0.12

Hedges & Hedberg (2007), Intraclass correlations for planning group randomized experiments in rural education.

More variations for another time

- Imperfect compliance with treatment
- Multiple treatments, multiple testing
- Actual mechanics of randomization
- Covariates, stratification
- Small numbers of groups
- Bayes' rule and power
- Attrition
- Alternative tests
- *"A first comment is that, despite all the precision of these formulas, power calculations involve substantial guess work in practice."*