Introduction

560 Hierarchical modeling

Peter Hoff

Statistics, University of Washington

Multilevel data

Multilevel data: Data for which there are

- multiple nested levels of sampling, and/or
- multiple nested sources of variability.

Such data are also often called *hierarchical data*.

Examples:

Educational testing: students are nested within classes, classes within schools.

Agricultural experiments: subplots nested within whole plots.

Clinical trials: observations over time nested within patients, patients within hospitals.

Multilevel data

Multilevel data: Data for which there are

- multiple nested levels of sampling, and/or
- multiple nested sources of variability.

Such data are also often called *hierarchical data*.

Examples:

Educational testing: students are nested within classes, classes within schools.

Agricultural experiments: subplots nested within whole plots.

Clinical trials: observations over time nested within patients, patients within hospitals.

Multilevel data

Multilevel data: Data for which there are

- multiple nested levels of sampling, and/or
- multiple nested sources of variability.

Such data are also often called *hierarchical data*.

Examples:

Educational testing: students are nested within classes, classes within schools.

Agricultural experiments: subplots nested within whole plots.

Clinical trials: observations over time nested within patients, patients within hospitals.

Multilevel data

Multilevel data: Data for which there are

- multiple nested levels of sampling, and/or
- multiple nested sources of variability.

Such data are also often called *hierarchical data*.

Examples:

Educational testing: students are nested within classes, classes within schools.

Agricultural experiments: subplots nested within whole plots.

Clinical trials: observations over time nested within patients, patients within hospitals.

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$y_{i,j}$ = response of ith micro unit in jth macro unit

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$$y_{i,j} = \text{response of } i\text{th micro unit in } j\text{th macro unit}$$

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$$y_{i,j} = \text{response of } i\text{th micro unit in } j\text{th macro unit}$$

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$$y_{i,j} = \text{response of } i\text{th micro unit in } j\text{th macro unit}$$

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$y_{i,j}$ = response of ith micro unit in jth macro unit

Terminology

observational unit: an object for which data maybe observed.

macro-level unit: a unit within which other units are nested.

micro-level unit: a unit nested within another unit.

Note that unless there are only two levels, macro versus micro is relative.

Synonyms:

macro units	micro units
top-level units	bottom-level units
primary units	secondary units
clusters	units
groups	units

Notation:

$$y_{i,j} = \text{response of } i\text{th micro unit in } j\text{th macro unit}$$

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling
The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling

The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling

The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling

The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling

The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Why multilevel data?

Consider the costs of administering an in-person survey to:

- 100 randomly sampled public high-school students in Washington state;
- 10 students sampled from 10 randomly sampled high schools.

Cluster sampling

The second sampling scheme is called *cluster sampling* or *two-stage sampling*.

Cluster sampling

- is often cheaper per sampled unit;
- often gives less reliable estimates of population means.

Task: Estimation of a population mean

Task: Estimate the population mean μ from sample data.

Questions:

- How do cluster sampling and SRS compare?
- How do you infer μ from cluster sample data?

Task: Estimation of a population mean

Task: Estimate the population mean μ from sample data.

Questions:

- How do cluster sampling and SRS compare?
- How do you infer μ from cluster sample data?

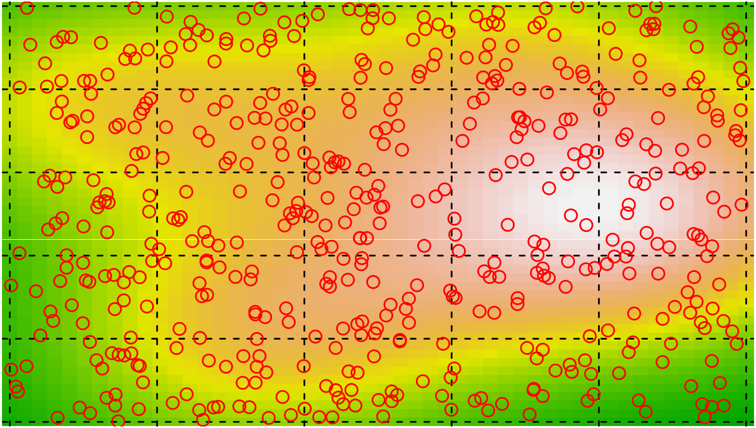
Task: Estimation of a population mean

Task: Estimate the population mean μ from sample data.

Questions:

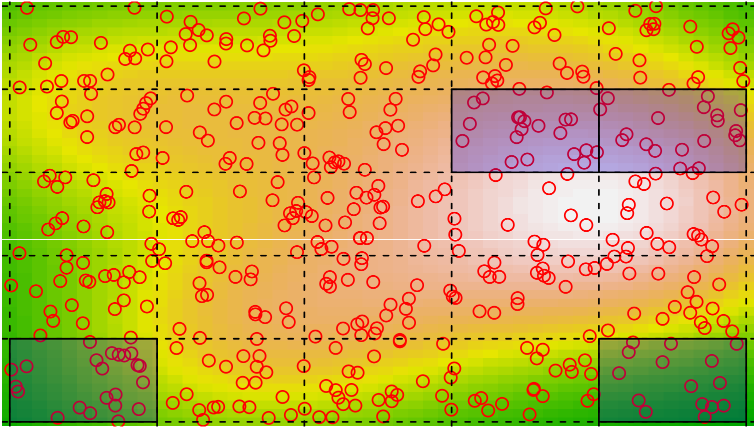
- How do cluster sampling and SRS compare?
- How do you infer μ from cluster sample data?

Two-stage sampling



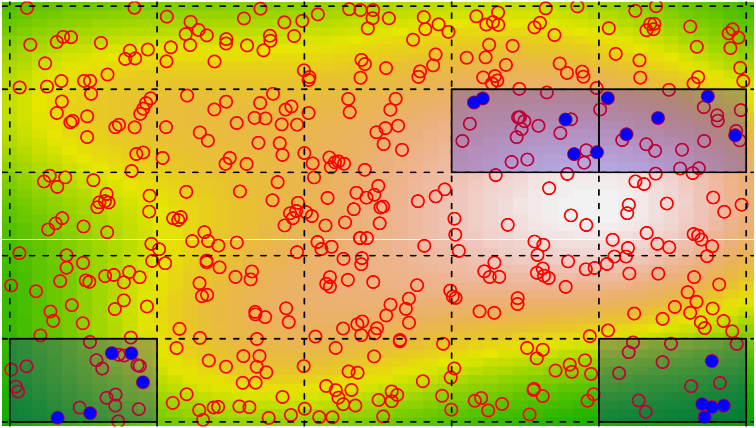
μ=2.1124814

Two-stage sampling



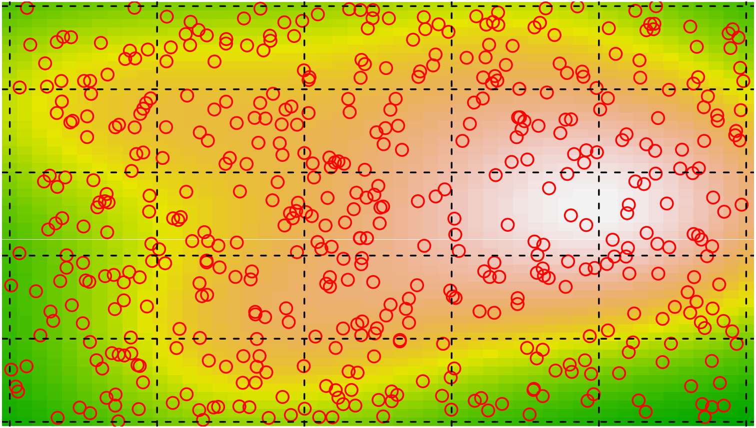
μ=2.1124814

Two-stage sampling



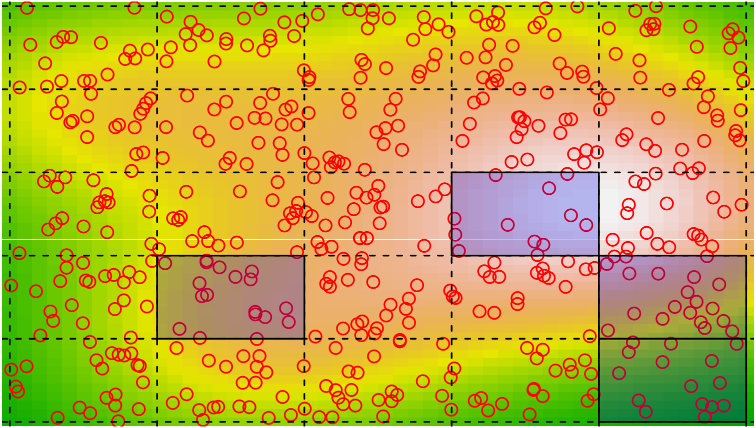
μ=2.1124814 , $\bar{y}$=1.8687553

Two-stage sampling



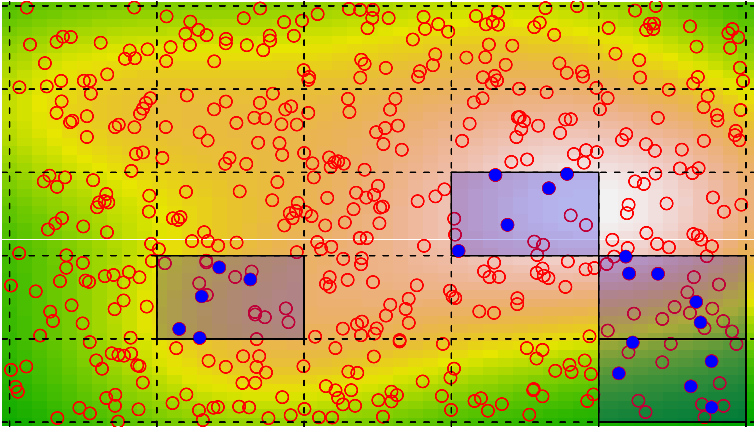
μ=2.1124814

Two-stage sampling



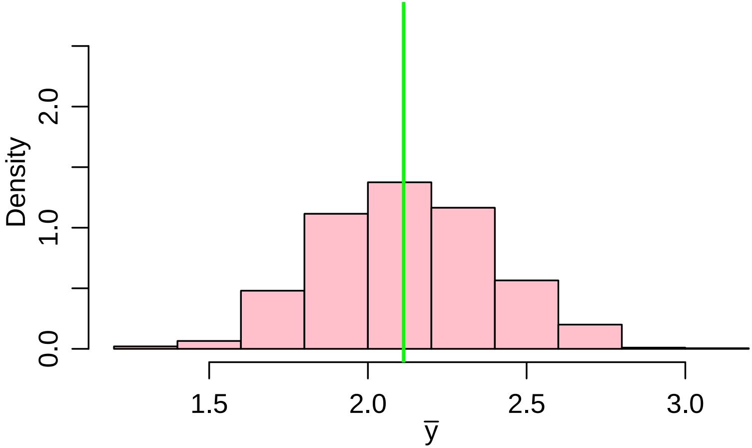
μ=2.1124814

Two-stage sampling

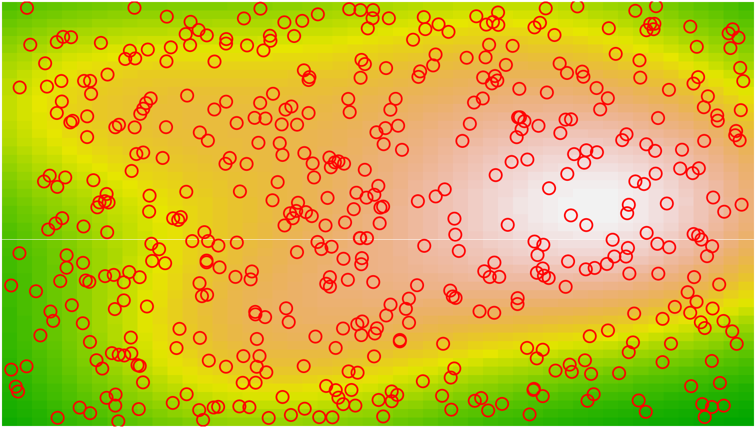


μ=2.1124814 , $\bar{y}$=2.2758576

Variability of sample mean

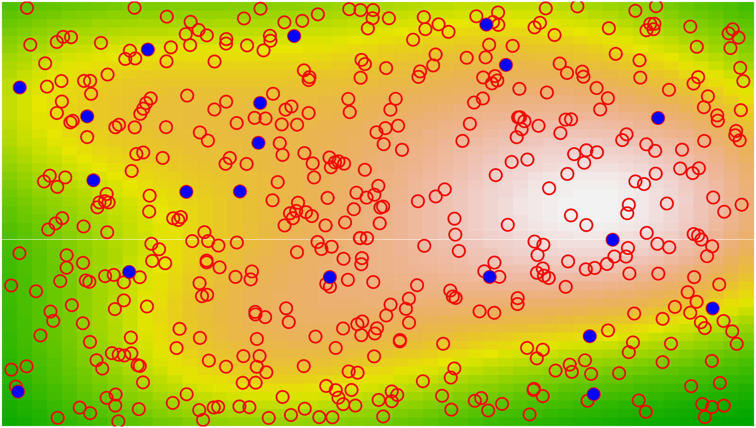


Comparison to SRS



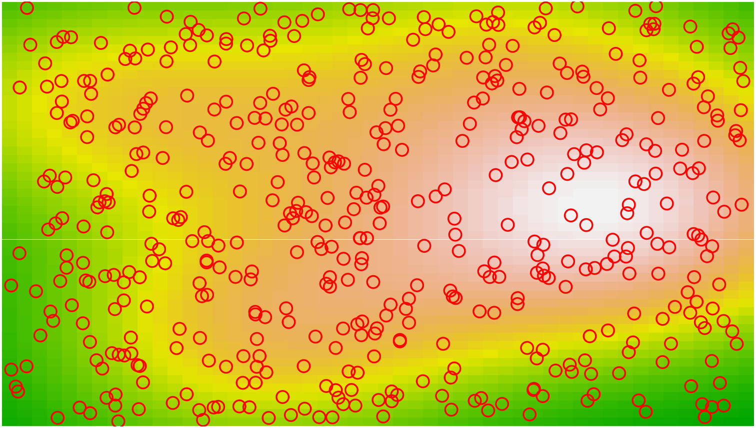
μ=2.1124814

Comparison to SRS



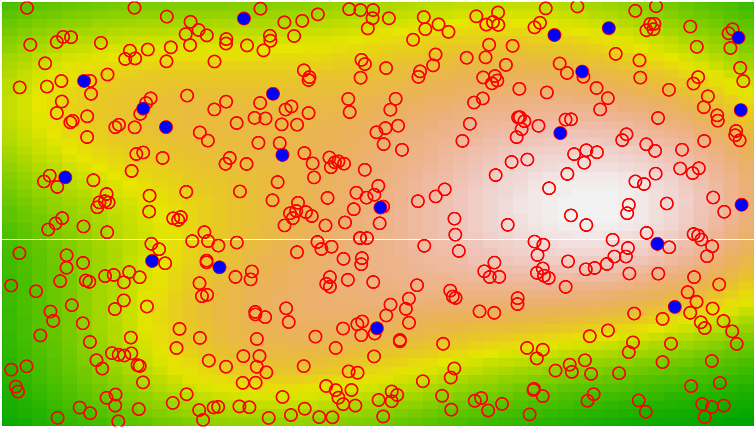
μ=2.1124814 , $\bar{y}$=2.0593259

Comparison to SRS



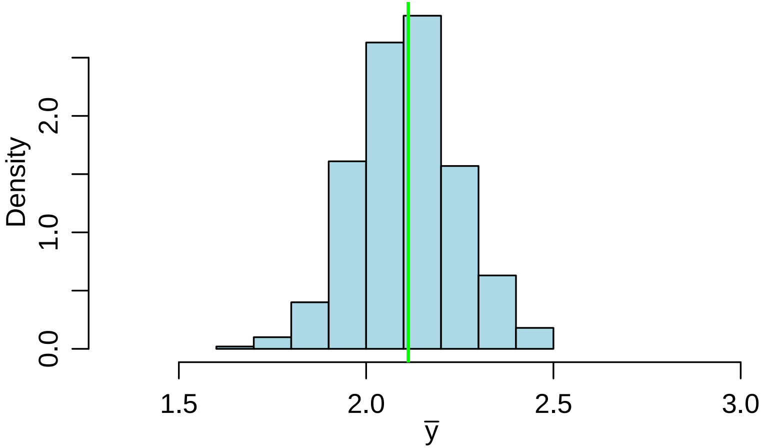
μ=2.1124814

Comparison to SRS

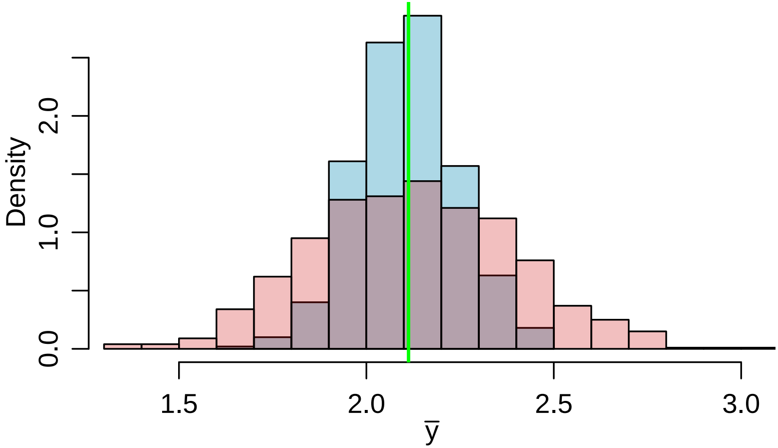


μ=2.1124814 , $\bar{y}$=2.20719

Variability of sample mean



Comparison of sampling variability



Heterogeneity, homogeneity and dependence

As we will show mathematically,

$$\text{across-group heterogeneity} \Leftrightarrow \text{within-group homogeneity}$$
$$\Leftrightarrow \text{within-group correlation or dependence}$$

Equivalently, across-group heterogeneity generally

- increases the variance of the sample;
- reduces the precision of estimates.

$$\mathrm{Var}[\bar{y}_{tss}] \geq \mathrm{Var}[\bar{y}_{srs}]$$

Heterogeneity, homogeneity and dependence

As we will show mathematically,

$$\text{across-group heterogeneity} \Leftrightarrow \text{within-group homogeneity}$$
$$\Leftrightarrow \text{within-group correlation or dependence}$$

Equivalently, across-group heterogeneity generally

- increases the variance of the sample;
- reduces the precision of estimates.

$$\text{Var}[\bar{y}_{tss}] \geq \text{Var}[\bar{y}_{srs}]$$

Heterogeneity, homogeneity and dependence

As we will show mathematically,

$$\text{across-group heterogeneity} \Leftrightarrow \text{within-group homogeneity}$$
$$\Leftrightarrow \text{within-group correlation or dependence}$$

Equivalently, across-group heterogeneity generally

- increases the variance of the sample;
- reduces the precision of estimates.

$$\text{Var}[\bar{y}_{tss}] \geq \text{Var}[\bar{y}_{srs}]$$

Heterogeneity, homogeneity and dependence

As we will show mathematically,

$$\begin{aligned} \text{across-group heterogeneity} &\Leftrightarrow \text{within-group homogeneity} \\ &\Leftrightarrow \text{within-group correlation or dependence} \end{aligned}$$

Equivalently, across-group heterogeneity generally

- increases the variance of the sample;
- reduces the precision of estimates.

$$\text{Var}[\bar{y}_{tss}] \geq \text{Var}[\bar{y}_{srs}]$$

Heterogeneity, homogeneity and dependence

As we will show mathematically,

$$\text{across-group heterogeneity} \Leftrightarrow \text{within-group homogeneity}$$
$$\Leftrightarrow \text{within-group correlation or dependence}$$

Equivalently, across-group heterogeneity generally

- increases the variance of the sample;
- reduces the precision of estimates.

$$\text{Var}[\bar{y}_{tss}] \geq \text{Var}[\bar{y}_{srs}]$$

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

t-interval for SRS:
If $y_1, \ldots, y_n$ is an iid sample with $\mathrm{E}[y_i] = \mu$ and $\mathrm{Var}[y_i] = \sigma^2$,

$$\mathrm{E}[\bar{y}] = \mu \ , \ \mathrm{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot\sim N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot\sim N(0,1).$$

As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot\sim t_{n-1}, \ , \text{where} s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

***t*-interval for SRS:**
If $y_1, \ldots, y_n$ is an iid sample with $\mathrm{E}[y_i] = \mu$ and $\mathrm{Var}[y_i] = \sigma^2$,

$$\mathrm{E}[\bar{y}] = \mu \ , \ \mathrm{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot{\sim} N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1).$$

As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot{\sim} t_{n-1}, \ \text{, where} s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

***t*-interval for SRS:**
If $y_1, \ldots, y_n$ is an iid sample with $\text{E}[y_i] = \mu$ and $\text{Var}[y_i] = \sigma^2$,

$$\text{E}[\bar{y}] = \mu \ , \ \text{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot{\sim} N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1).$$

As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot{\sim} t_{n-1}, \ \text{, where} s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

***t*-interval for SRS:**
If $y_1, \dots, y_n$ is an iid sample with $\mathrm{E}[y_i] = \mu$ and $\mathrm{Var}[y_i] = \sigma^2$,

$$\mathrm{E}[\bar{y}] = \mu \ , \ \mathrm{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot{\sim} N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1).$$

As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot{\sim} t_{n-1}, \ \text{, where} s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity

Task: Construct a 95% CI for the population mean.

***t*-interval for SRS:**
If $y_1, \ldots, y_n$ is an iid sample with $\mathrm{E}[y_i] = \mu$ and $\mathrm{Var}[y_i] = \sigma^2$,

$$\mathrm{E}[\bar{y}] = \mu \ , \ \mathrm{Var}[\bar{y}] = \sigma^2/n.$$

By the central limit theorem,

$$\bar{y} \dot{\sim} N(\mu, \sigma^2/n) \ , \ \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0, 1).$$

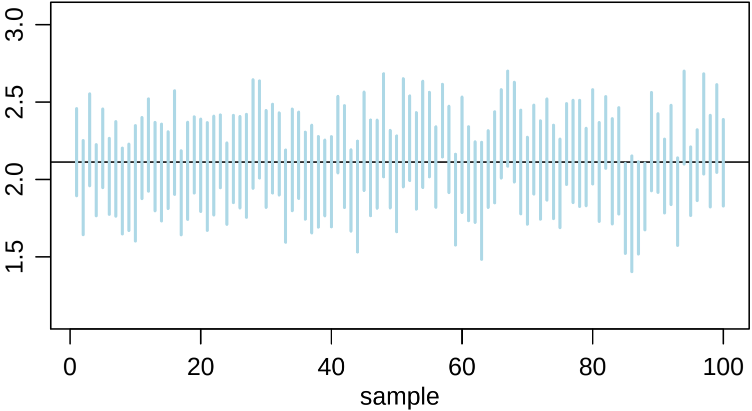
As σ^2 is generally unknown, we use

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \dot{\sim} t_{n-1}, \ \text{, where} s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2.$$

From this, we have

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n} \ \text{ is a 95\% CI for } \mu.$$

Ignoring across-group heterogeneity



Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity

$$\bar{y} \pm t_{n-1,.975} \times s/\sqrt{n}$$

What if we apply the formula to data from a cluster sample?

If $y_1, \ldots, y_n$ are from a SRS, then

$$\text{Var}[\bar{y}] = \sigma^2/n = \text{E}[s^2/n].$$

$s/\sqrt{n}$ provides a good estimate of the sd of $\bar{y}$.

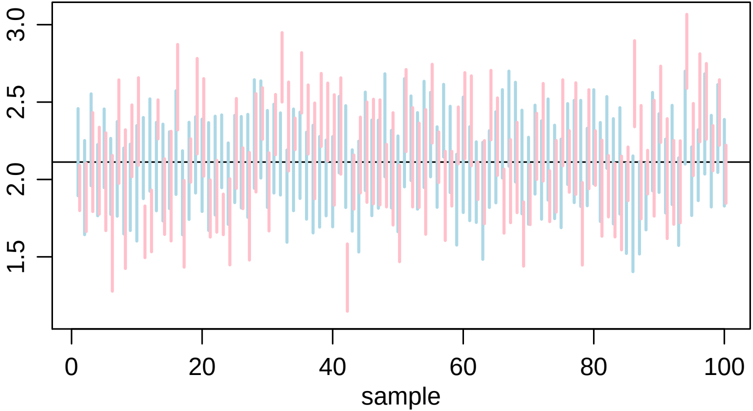
If $y_1, \ldots, y_n$ are from a cluster sample, then generally

$$\text{Var}[\bar{y}] > \sigma^2/n \approx \text{E}[s^2/n].$$

$s/\sqrt{n}$ is generally an underestimate of the sd of $\bar{y}$.

How will the resulting confidence interval behave if $\text{sd}(\bar{y}) > s/\sqrt{n}$?

Ignoring across-group heterogeneity



Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity = within group similarity.
- Within group similarity leads to positively correlated cluster sample data.
- The variance of the sample mean from (positively) correlated data is higher than that of the mean of uncorrelated data.
- Statistical inference ignoring such correlation will be inaccurate.

Moving forward: We will develop techniques to

- evaluate within and across-group heterogeneity;
- provide accurate statistical inference based on cluster samples.

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = \mathrm{E}[y|x = 1]$
- $\mu_0 = \mathrm{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = \text{E}[y|x = 1]$
- $\mu_0 = \text{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = \mathrm{E}[y|x = 1]$
- $\mu_0 = \mathrm{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = \mathrm{E}[y|x = 1]$
- $\mu_0 = \mathrm{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = E[y|x = 1]$
- $\mu_0 = E[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

- $x \in \{0, 1\}$
- $\mu_1 = \mathrm{E}[y|x = 1]$
- $\mu_0 = \mathrm{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Task: Estimation of an effect

Suppose

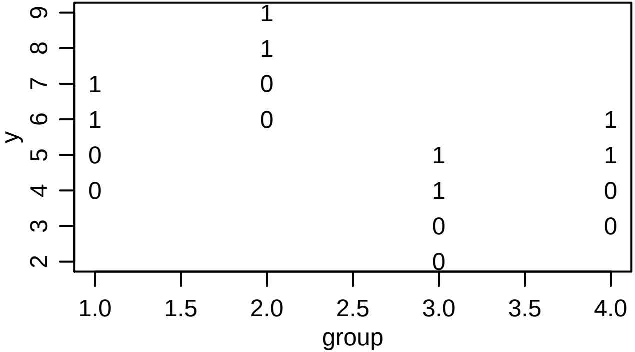
- $x \in \{0, 1\}$
- $\mu_1 = \mathrm{E}[y|x = 1]$
- $\mu_0 = \mathrm{E}[y|x = 0]$

Task: Estimate the difference $\delta = \mu_1 - \mu_0$ based on cluster sample data.

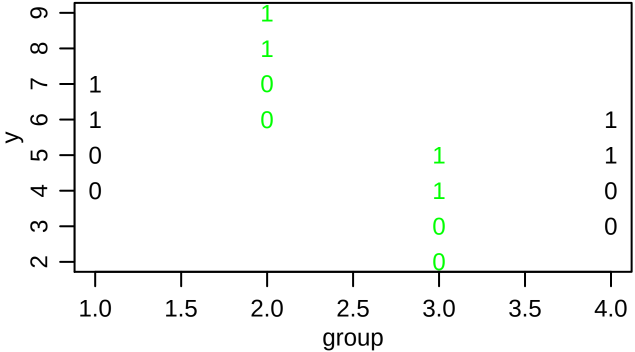
Data: For each group j, we have $(y_{1,j}, x_{1,j}), \ldots, (y_{n,j}, x_{n,j})$.

Question: What could go wrong by ignoring the multilevel nature of the data?

Estimation of an effect

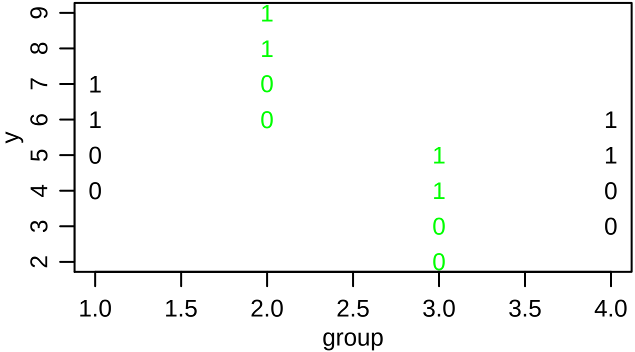


Estimation of an effect



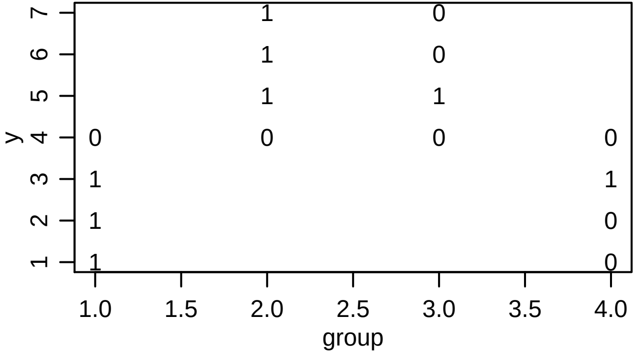
Ignoring group differences can lead to *overconservative inference.*

Estimation of an effect



Ignoring group differences can lead to *overconservative inference.*

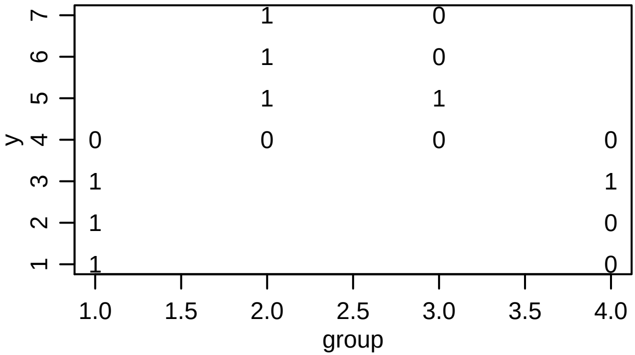
Estimation of an effect



- The population mean difference is zero;
- The sample mean difference based on any two groups is not zero.

Ignoring group differences can lead to *underconservative inference.*

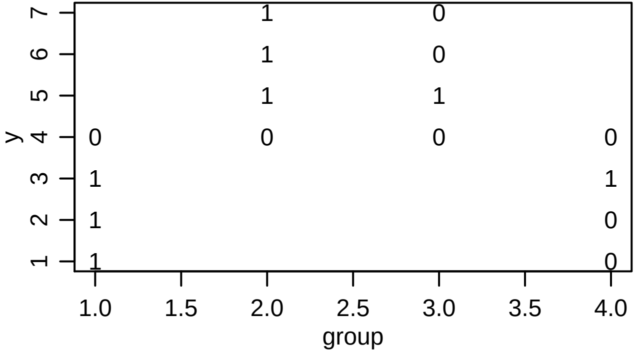
Estimation of an effect



- The population mean difference is zero;
- The sample mean difference based on any two groups is not zero.

Ignoring group differences can lead to *underconservative inference.*

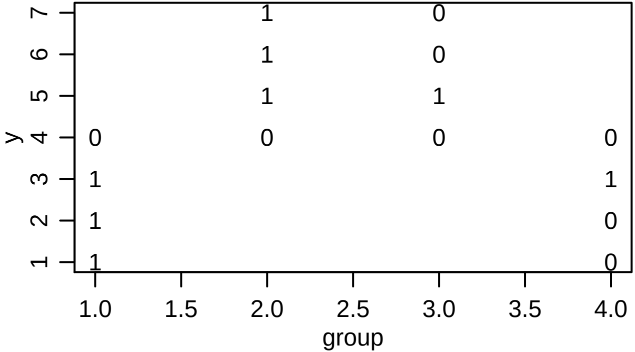
Estimation of an effect



- The population mean difference is zero;
- The sample mean difference based on any two groups is not zero.

Ignoring group differences can lead to *underconservative inference.*

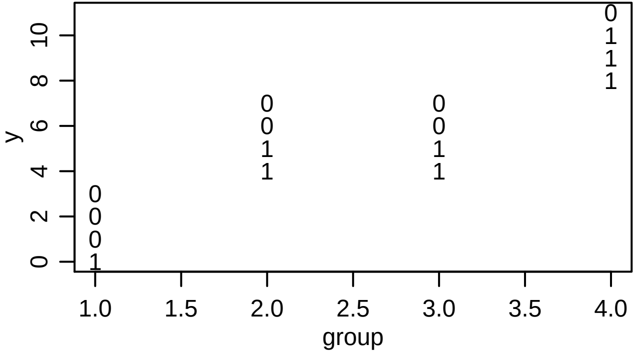
Estimation of an effect



- The population mean difference is zero;
- The sample mean difference based on any two groups is not zero.

Ignoring group differences can lead to *underconservative inference.*

Estimation of an effect

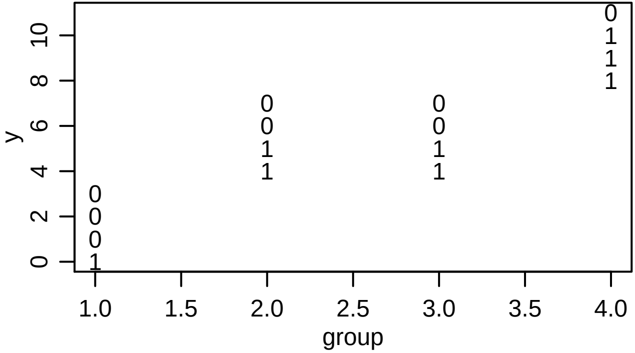


- $\mu_1 - \mu_0 > 0$ in the overall population;
- $\mu_{1,j} - \mu_{0,j} < 0$ in every group.

Micro/group effects may be different from macro/population effects.

This is sometimes called *Simpson's paradox*.

Estimation of an effect

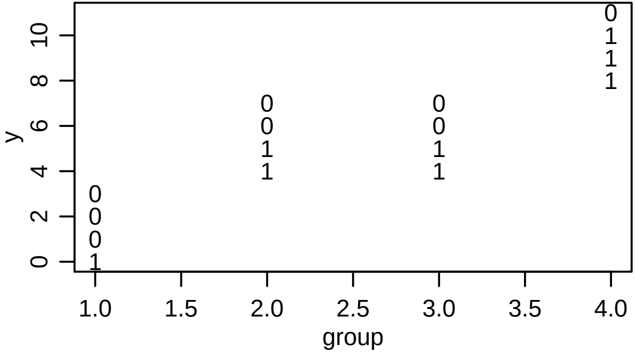


- $\mu_1 - \mu_0 > 0$ in the overall population;
- $\mu_{1,j} - \mu_{0,j} < 0$ in every group.

Micro/group effects may be different from macro/population effects.

This is sometimes called *Simpson's paradox*.

Estimation of an effect

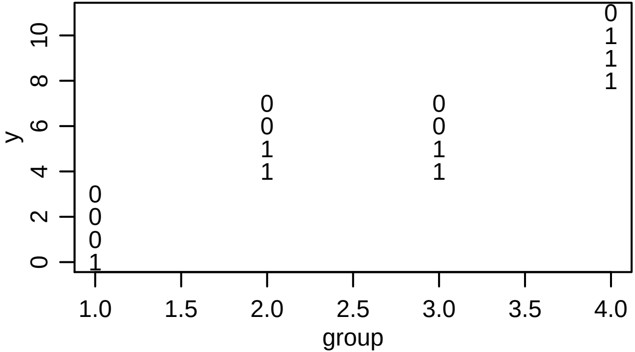


- $\mu_1 - \mu_0 > 0$ in the overall population;
- $\mu_{1,j} - \mu_{0,j} < 0$ in every group.

Micro/group effects may be different from macro/population effects.

This is sometimes called *Simpson's paradox*.

Estimation of an effect

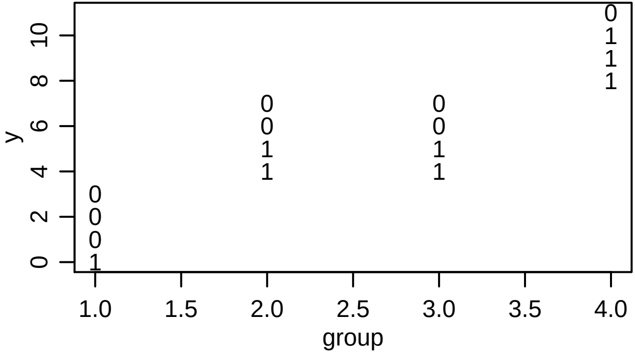


- $\mu_1 - \mu_0 > 0$ in the overall population;
- $\mu_{1,j} - \mu_{0,j} < 0$ in every group.

Micro/group effects may be different from macro/population effects.

This is sometimes called *Simpson's paradox*.

Estimation of an effect



- $\mu_1 - \mu_0 > 0$ in the overall population;
- $\mu_{1,j} - \mu_{0,j} < 0$ in every group.

Micro/group effects may be different from macro/population effects.

This is sometimes called *Simpson's paradox*.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Ignoring across-group heterogeneity

Summary:

- Across group heterogeneity can lead to *over or under* conservative inference.
- Aggregated macro effects may be different from micro effects.
- Statistical inference ignoring groups can be inaccurate in *unpredictable* ways.

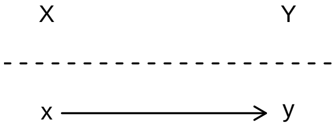
Moving forward: We will develop techniques to

- differentiate between macro and micro level effects;
- appropriately control for within and between-group heterogeneity.

Macro, micro and cross-level effects

X, x are macro and micro level explanatory variables

Y, y are macro and micro level outcome variables



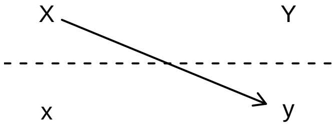
What are the effects of SES (x) on political opinion (y)?

(a *micro effect*)

Macro, micro and cross-level effects

X, x are macro and micro level explanatory variables

Y, y are macro and micro level outcome variables



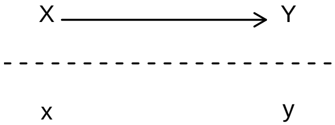
What are the effects of State GDP (X) on political opinion (y) ?

(a *macro-micro effect*)

Macro, micro and cross-level effects

X, x are macro and micro level explanatory variables

Y, y are macro and micro level outcome variables



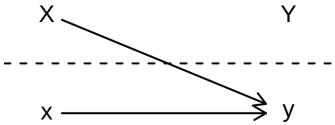
What are the effects of State GDP (X) on statewide political opinion (Y)?

(a *a macro effect*)

Macro, micro and cross-level effects

X, x are macro and micro level explanatory variables

Y, y are macro and micro level outcome variables



What are the effects of State GDP (X) and SES (x) on political opinion (y)?

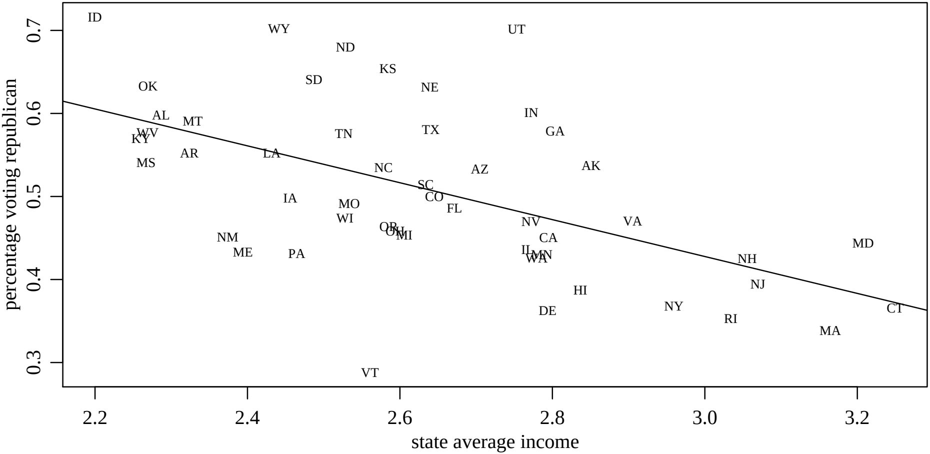
(a *multilevel effects*)

Example: Income and voting patterns

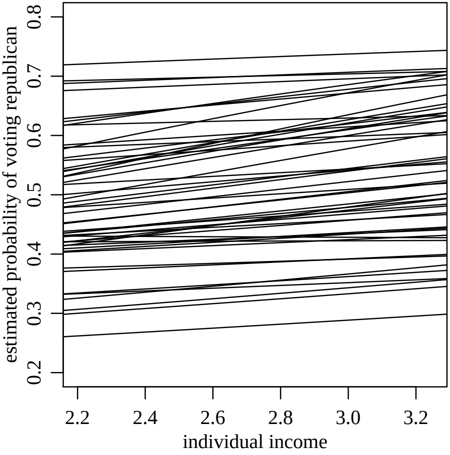
Exit poll data from 2004 presidential election

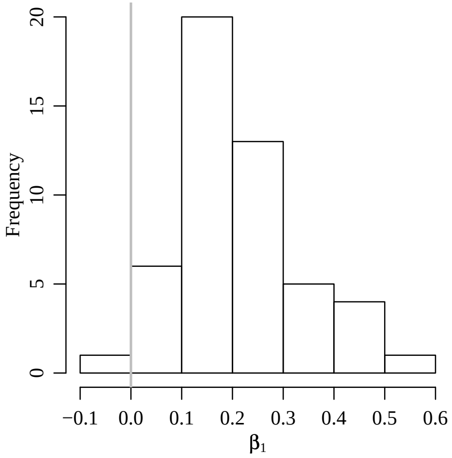
- $j \in \{1, \dots, 50\}$ indexes the states,
- $y_{i,j}$ is the voting variable for person i in state j,
- $x_{i,j}$ is the measure of income for person (i, j).

Macro-level income-voting relationships



Micro-level income-voting relationships





Joint estimation of effects

In general we may be interested in understanding all of the following:

- macro level effects,
- micro level effects,
- heterogeneity of micro effects across groups.

Joint estimation of effects

In general we may be interested in understanding all of the following:

- macro level effects,
- micro level effects,
- heterogeneity of micro effects across groups.

Joint estimation of effects

In general we may be interested in understanding all of the following:

- macro level effects,
- micro level effects,
- heterogeneity of micro effects across groups.