

# Stratified and Cluster Sampling

## Lecture 8 Sections 2.6, 2.8

Robb T. Koether

Hampden-Sydney College

Tue, Sep 8, 2009

# Outline

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- 1 Introduction
- 2 Stratified Random Samples
- 3 Estimating Parameters
- 4 Cluster Samples
- 5 Stratified vs. Cluster
- 6 Assignment

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose we want to do a survey among HSC students that will allow to compare various ethnic groups on campus (Black, White, Hispanic, Orientals, South Asians).
- What might go wrong if we select a simple random sample of 50 HSC students?
- How can we be assured that each ethnic group will be sufficiently represented in our sample?

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose we want to do a survey among HSC students that will allow to compare various ethnic groups on campus (Black, White, Hispanic, Orientals, South Asians).
- What might go wrong if we select a simple random sample of 50 HSC students?
- How can we be assured that each ethnic group will be sufficiently represented in our sample?

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose we want to do a survey among HSC students that will allow to compare various ethnic groups on campus (Black, White, Hispanic, Orientals, South Asians).
- What might go wrong if we select a simple random sample of 50 HSC students?
- How can we be assured that each ethnic group will be sufficiently represented in our sample?

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose that we intentionally select 10 Whites, 10 Blacks, 10 Hispanics, 10 Orientals, and 10 South Asians.
- Clearly, some groups are overrepresented and some are underrepresented in our sample.
- Therefore, if we simply compute an average or a proportion for all 50 students, it will not be representative of the student body.
- So how can we calculate a statistic for the sample that will be representative of the population?

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose that we intentionally select 10 Whites, 10 Blacks, 10 Hispanics, 10 Orientals, and 10 South Asians.
- Clearly, some groups are overrepresented and some are underrepresented in our sample.
- Therefore, if we simply compute an average or a proportion for all 50 students, it will not be representative of the student body.
- So how can we calculate a statistic for the sample that will be representative of the population?

# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose that we intentionally select 10 Whites, 10 Blacks, 10 Hispanics, 10 Orientals, and 10 South Asians.
- Clearly, some groups are overrepresented and some are underrepresented in our sample.
- Therefore, if we simply compute an average or a proportion for all 50 students, it will not be representative of the student body.
- So how can we calculate a statistic for the sample that will be representative of the population?



# Introduction

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose that we intentionally select 10 Whites, 10 Blacks, 10 Hispanics, 10 Orientals, and 10 South Asians.
- Clearly, some groups are overrepresented and some are underrepresented in our sample.
- Therefore, if we simply compute an average or a proportion for all 50 students, it will not be representative of the student body.
- So how can we calculate a statistic for the sample that will be representative of the population?

# Stratified Random Sample

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Definition (Homogeneous)

A group is **homogeneous** if its member all have similar characteristics with regard to a variable of interest.

## Definition (Stratum)

A **stratum** is a homogeneous subset of the population.

## Definition (Stratified random sampling)

**Stratified random sampling** is a sampling method in which the population is first divided into strata. Then a simple random sample is taken from each stratum. The combined results constitute the sample.

# Examples

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Possible strata:
  - Male and female strata.
  - Resident and non-resident strata.
  - White, Black, Hispanic, and Asian strata.
  - Protestant, Catholic, Jewish, Muslim, etc., strata.

# Stratified Random Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

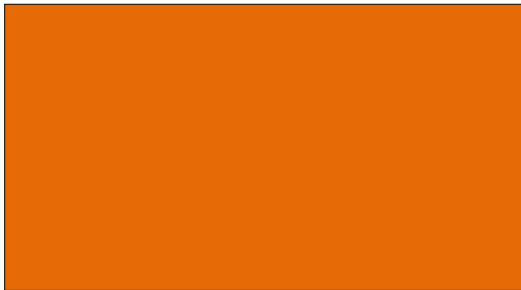
**Stratified  
Random  
Samples**

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment



The population

# Stratified Random Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

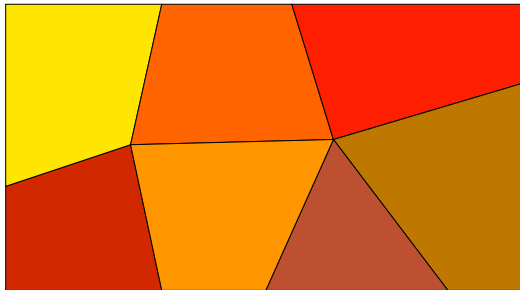
**Stratified  
Random  
Samples**

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment



The strata

# Stratified Random Sampling

## Stratified and Cluster Sampling

Robb T. Koether

Introduction

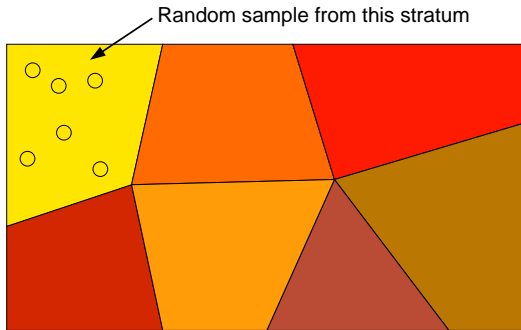
Stratified Random Samples

Estimating Parameters

Cluster Samples

Stratified vs. Cluster

Assignment



The strata

# Stratified Random Sampling

## Stratified and Cluster Sampling

Robb T. Koether

Introduction

Stratified Random Samples

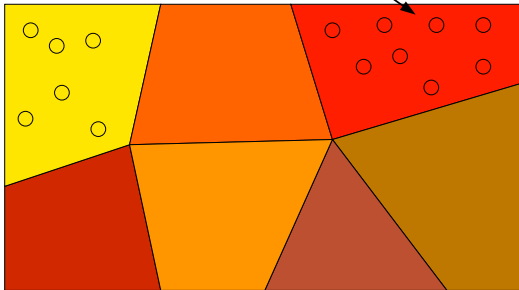
Estimating Parameters

Cluster Samples

Stratified vs. Cluster

Assignment

Random sample from this stratum



The strata

# Stratified Random Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

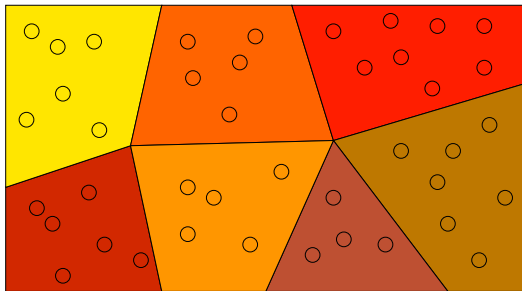
Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment



Random samples from all strata



# Stratified Random Sampling

## Stratified and Cluster Sampling

Robb T.  
Koether

Introduction

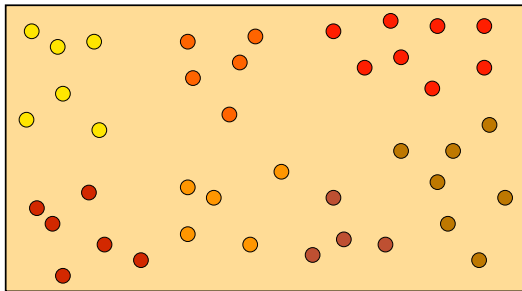
Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment



The stratified sample

# Example

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Example (Stratified random sample)

- Let the population consist of males Bill, Danny, Fred, Henri, Joaquin, Larry, Nicholas, and Peter and females Ana, Claudette, Erika, Grace, Ida, Kate, Mindy, and Odette.
- Choose a stratified sample of size  $n = 8$ , where the strata are the two sexes.
- Is the sample representative with regard to sex?

# Estimating Parameters

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Suppose that
  - The population is 60% male and 40% female.
  - Our sample is 50% male and 50% female.
  - Our variable has an average value of 10 for the males and 15 for the females.
- What is our best estimate for the variable's average for the population?

# Estimating Parameters

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Or we compute proportions and get 20% for the males and 35% for the females.
- What is our best estimate for the proportion for the population?

# Estimating Parameters

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- We need to compute a *weighted average*.

$$\begin{aligned}\text{average} &= (0.60)(10) + (0.40)(15) \\ &= 6 + 6 \\ &= 12\end{aligned}$$

# Cluster Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Definition (Heterogeneous)

A group is **heterogeneous** if its members vary in regard to the variables of interest in the same way that the population varies.

## Definition (Cluster)

A **cluster** is a heterogeneous subset of the population.

## Definition (Cluster random sampling)

**Cluster random sampling** is a sampling method in which the population is first divided into clusters. Then a simple random sample of clusters is taken. All the members of the selected clusters together constitute the sample.

# Cluster Sampling

## Stratified and Cluster Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- Note that it is the clusters that are selected at random, not the individuals.
- It is hoped that each cluster by itself is representative of the population, i.e., each cluster is heterogeneous.

# Cluster Random Sampling

## Stratified and Cluster Sampling

Robb T.  
Koether

Introduction

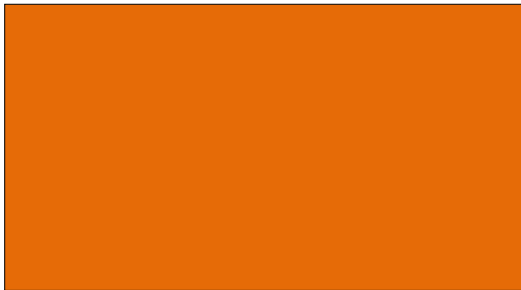
Stratified  
Random  
Samples

Estimating  
Parameters

**Cluster  
Samples**

Stratified vs.  
Cluster

Assignment



The population



# Cluster Random Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

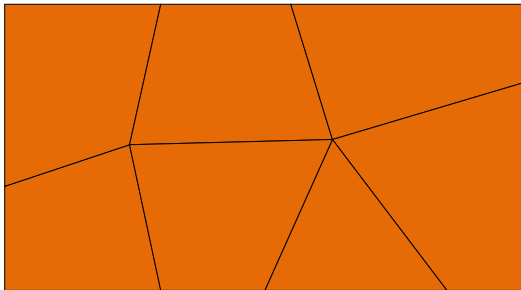
Stratified  
Random  
Samples

Estimating  
Parameters

**Cluster  
Samples**

Stratified vs.  
Cluster

Assignment



The clusters

# Cluster Random Sampling

## Stratified and Cluster Sampling

Robb T. Koether

Introduction

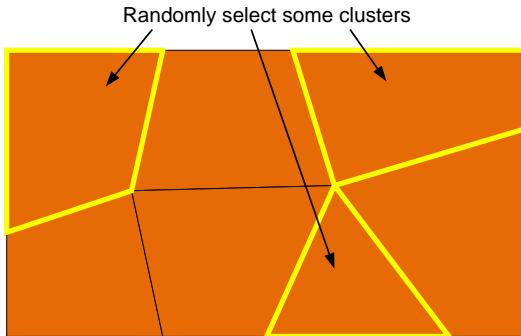
Stratified Random Samples

Estimating Parameters

Cluster Samples

Stratified vs. Cluster

Assignment



The clusters

# Cluster Random Sampling

## Stratified and Cluster Sampling

Robb T. Koether

Introduction

Stratified Random Samples

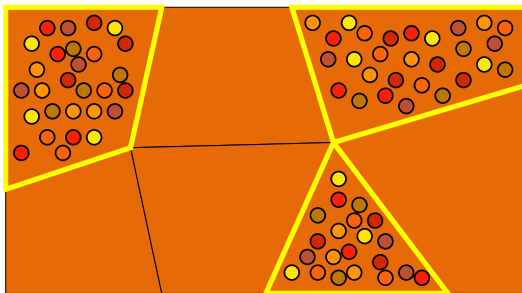
Estimating Parameters

Cluster Samples

Stratified vs. Cluster

Assignment

Selected all member in the clusters



The clusters

# Cluster Random Sampling

## Stratified and Cluster Sampling

Robb T. Koether

Introduction

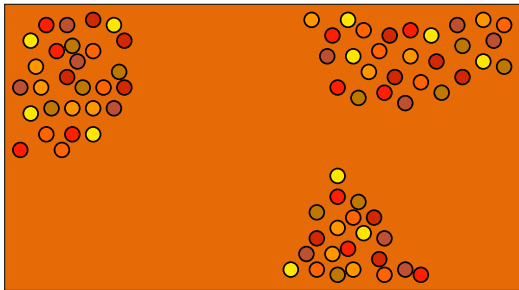
Stratified Random Samples

Estimating Parameters

**Cluster Samples**

Stratified vs. Cluster

Assignment



The cluster sample

# Example

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Example (Cluster sample)

- Now suppose that
  - Ana, Bill, Claudette, and Danny live in Fredericksburg.
  - Erika, Fred, Grace, and Henri live in Richmond.
  - Ida, Joaquin, Kate, and Larry live in Charlottesville.
  - Mindy, Nicholas, Odette, and Peter live in Roanoke.

# Example

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Example (Cluster sample)

- Use cluster sampling to choose a sample of size  $n = 8$ , where the clusters are the cities.
- Is the sample representative with regard to sex?

# Stratified Sampling vs. Cluster Sampling

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

- In stratified sampling
  - From **all of the strata** we take **randomly selected individuals**.
- In cluster sampling
  - From **randomly selected clusters** we take **all of the individuals**.

# Assignment

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Homework

- Read Sections 2.6, 2.8, pages 108 - 115, 122 - 126.
- Let's Do It! 2.6, 2.8.
- Page 115, exercises 19 - 23, 25.
- Page 126, exercises 35 - 38.



# Answers to Even-numbered Problems

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Page 115, Problems 20, 22

- 2.20 (a) Good. Textbook prices tend to vary dramatically across different majors. A stratified sample will guarantee that all majors are represented.
- (b) Not so good. Textbook prices do not vary by the gender of the student, except that females are slightly less likely to be taking science courses, where the textbooks are very expensive. But that takes us back to part (a).
- (c) Not so good. Textbook prices may vary somewhat by class rank, but not much.
- 2.22 (a)  $\frac{100}{200} \cdot$
- (b)  $\frac{100}{1000} \cdot$
- (c) 0.10. No. Not all samples are equally likely. Only samples containing 100 males and 20 females are possible.

# Answers to Even-numbered Problems

Stratified and  
Cluster  
Sampling

Robb T.  
Koether

Introduction

Stratified  
Random  
Samples

Estimating  
Parameters

Cluster  
Samples

Stratified vs.  
Cluster

Assignment

## Page 126, Problems 36, 38

2.36 (a) Physics and Mathematics.

(b) 60.

2.38 (a) Cluster sampling.

(b) No. The more classes a student takes, the more likely he is to be in the sample.

(c) Poor design together with bad luck. The procedure is not biased, except for the small effect described in part (b). However, the individual classes, especially the education classes, are not heterogeneous, so cluster sampling should not be used.