You should know...

- stratum, strata, stratified sampling
- why/when we prefer stratified sampling to SRS
- inference: use SRS theory in each stratum h
- $\hat{t}_{str} = \sum_{h=1}^{H} \hat{t}_h, \quad \widehat{V}(\hat{t}_{str}) = \sum_{h=1}^{H} \underbrace{(1 - n_h/N_h) N_h^2 s_h^2/n_h}_{\text{SRS for } h}$ (4.4)
- $\bar{y}_{str}, \quad \widehat{V}(\bar{y}_{str})$
- sampling weights:
  $\bar{y}_{str} = \sum_{h=1}^{H} \sum_{j \in \mathcal{S}_h} w_{hj} y_{hj} / \sum_{h=1}^{H} \sum_{j \in \mathcal{S}_h} w_{hj}$
- proportional allocation: $w_{hj} = N/n$; self-weighting sample
- optimal allocation: $n_h \propto (N_h S_h)/\sqrt{c_h}$
- how to define strata; when is stratified sampling better (= smaller variance)
- **Final Exam: December 15, 9 - 11 a.m., EX 200 (255 McCaul Street)**
- $$: Samuel Beatty Scholarship November 13

  `http://www.physics.utoronto.ca/undergraduate/scholarship.2009`

... you should know...

- HW: Exercises 3.5, 3.13a, 3.15, 3.24; Examples 4.2, 4.3; Exercises 4.2, 4.12, 4.10; Ex. 4.15 – new
- 3.5, 3.13a done on Friday; 3.15 posted online; 3.24 – you do it
- Example 4.2 done in class; Example 4.3 see text
- Exercise 4.2: posted online and see R code from Oct 23
- Exercise 4.12: optimal allocation for the agriculture data set

HW for me

On p.97, we see have estimates of S_h^2 for each of the 4 strata (for ACRES92), and we know the population sizes N_h:

agsrs. dat

p. 97

N_h	Stratum	Sample Size	s_h^2	optimal sample size
220	Northeast	21	7,647,472,708	69
1054	North Central	103	29,618,183,543	7
1382	South	135	53,587,487,856	122
422	West	41	396,185,950,266	101

agpop.dat 3087 300

77

avail from agpop

- optimal: $n_h \propto N_h S_h^2/\sqrt{c_h}$; if c_h's all equal then $n_h \propto N_h S_h^2$; we use s_h^2 as estimates
- Exer. 4.10: It is WRONG on Rcode from October 23 – Friday

$n_h \propto N_h S_h^2 / \Sigma(N_h S_h^2) \times 300$

Exercise 4.10 (see R code from today)

	N_h	n_h	t_h	s_h^2
Bio	102	7	2+3+10+7	$\dfrac{2+9+50+49-22^2/7}{6}$
Phy	310	19	2+3+8+5+6+16	
Soc	217	13	2+8+6	
Hum	178	11	2+3	
	807	50		

1^2+1^2

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

$2+9+\cdots$

Cluster sampling Ch.5

stratified	cluster
variance within	variance between
small	small
strata $1, \ldots H$	psu's $1, \ldots, N$ – sampling unit
population $N_1, \ldots, N_h$	ssu's $M_1, \ldots M_n$ – observation unit
observation y_{hj}	observation y_{ij}

See Figure 5.1

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...) 500 such psu's
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...) 500
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...) 500 psu's
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...) 400 total s.s.
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with ≥ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with ≥ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with ≥ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ...
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling: Examples

- Example, p.131: 10,000 households; divide into blocks of 20 households (= ...)
- psu: sample 20 of the 500 blocks
- ssu: sample all 20 households on the block (total sample size = ...)
- cheaper, easier to implement
- values on a single block more similar than 20 values taken at random from all 10,000 households
- so less information than in an SRS of size 400
- Example 5.2: 400 students in a dorm, in suites of size 4
- sample 5 suites at random
- interview all 4 students
- Example 5.6: clutches (= nests) with $\geq$ 2 eggs each ...
- 2 eggs in each nest chosen at random ... ← n.b. 2 stage c.s.
- Example: nearly all household surveys: `http://www.statcan.gc.ca/concepts/index-eng.htm`

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

... cluster sampling

- one stage cluster sampling: sample psu's by SRS, sample all ssu's
- two stage cluster sampling: sample psu's by SRS, sample ssu's using some probability method
- why?
- may not have a sampling frame of observation units (individuals in a city, customers of a store)
- population may be widely distributed
- population may occur in natural clusters
- example: nursing home residents
- usually, cost
- stratified sampling is more efficient (...)
- cluster sampling is less efficient (...)

Mortality in Iraq

- "New study estimating number of dead in Iraq hotly contested" (Globe & Mail)
- "The human cost of the war in Iraq" (Economist)
- "A statistical study claims that many more Iraqis have died than was thought"
- "Mortality after the 2003 invasion of Iraq: a cross-sectional study" (The Lancet, 2006)
- "Iraqi death estimates called too high: methods faulted" (Science)

Mortality in Iraq

- "New study estimating number of dead in Iraq hotly contested" (Globe & Mail)
- "The human cost of the war in Iraq" (Economist)
- "A statistical study claims that many more Iraqis have died than was thought"
- "Mortality after the 2003 invasion of Iraq: a cross-sectional study" (The Lancet, 2006)
- "Iraqi death estimates called too high: methods faulted" (Science)

Mortality in Iraq

- “New study estimating number of dead in Iraq hotly contested” (Globe & Mail)
- “The human cost of the war in Iraq” (Economist)
- “A statistical study claims that many more Iraqis have died than was thought”
- “Mortality after the 2003 invasion of Iraq: a cross-sectional study” (The Lancet, 2006)
- “Iraqi death estimates called too high: methods faulted” (Science)

Mortality in Iraq

- "New study estimating number of dead in Iraq hotly contested" (Globe & Mail)
- "The human cost of the war in Iraq" (Economist)
- "A statistical study claims that many more Iraqis have died than was thought"
- "Mortality after the 2003 invasion of Iraq: a cross-sectional study" (The Lancet, 2006)
- "Iraqi death estimates called too high: methods faulted" (Science)

Mortality in Iraq

- "New study estimating number of dead in Iraq hotly contested" (Globe & Mail)
- "The human cost of the war in Iraq" (Economist)
- "A statistical study claims that many more Iraqis have died than was thought"
- "Mortality after the 2003 invasion of Iraq: a cross-sectional study" (The Lancet, 2006)
- "Iraqi death estimates called too high: methods faulted" (Science)

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  http://www.emro.who.int/iraq/ifhs_faq.htm
- Journal of Peace Research, 2008: “Bias in epidemiological studies of conflict mortality”
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: "Bias in epidemiological studies of conflict mortality"
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  http://www.emro.who.int/iraq/ifhs_faq.htm
- Journal of Peace Research, 2008: "Bias in epidemiological studies of conflict mortality"
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: “Bias in epidemiological studies of conflict mortality”
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: "Bias in epidemiological studies of conflict mortality"
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: “Bias in epidemiological studies of conflict mortality”
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: "Bias in epidemiological studies of conflict mortality"
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  http://www.emro.who.int/iraq/ifhs_faq.htm
- Journal of Peace Research, 2008: “Bias in epidemiological studies of conflict mortality”
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

... mortality in Iraq

- Iraq body count: 48, 693
- Burnham et al. : 601,027 (427,000 – 739,700)
- NEJM, Jan 2008 151,000 (104,000 – 223,000)
- based on IFHS study
  `http://www.emro.who.int/iraq/ifhs_faq.htm`
- Journal of Peace Research, 2008: “Bias in epidemiological studies of conflict mortality”
- select a random main street
- choose a random cross street to the main street
- select a random household on the cross street to start the process
- interview that house and proceed to adjacent house until 40 houses have been surveyed

~~Formulas~~ Notation

For cluster sampling

population quantities :

N # psu's ← sampling units ✓

M_i # ssu ← observation units ✓

$K = \sum_{i=1}^{N} M_i$ total pop. size ✓

$t_i = \sum_{j=1}^{M_i} y_{ij}$ $t = \sum_{i=1}^{N} t_i$ X

pop$^{=}$ quantities

$\bar{y}_u = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij} / K$ $S^2 = \sum_i \sum_j (y_{ij} - \bar{y}_u)^2 / (K-1)$

$\bar{y}_{iu} = \sum_{j=1}^{M_i} y_{ij} / M_i$ $S_i^2 = \sum_{j=1}^{M_i} (y_{ij} - \bar{y}_{iu})^2 / (M_i - 1)$

unknown, we want to estimate it

sample n psu's m_i obs$^{=}$ in each psu

$\bar{y}_i$ sample mean in ith psu ← data

$\hat{t}_i = M_i \bar{y}_i$ $\quad \hat{t} = \sum_{i=1}^{n} \hat{t}_i \cdot \frac{N}{n}$ ← est. of t $= \hat{t}_{unb}$

$$s_t^2 = \sum_{i=1}^{n} \left(\hat{t}_i - \frac{\hat{t}}{N}\right)^2 / (n-1)$$

↑ this is for Ch 5 & 6 ; but in Ch 5 $M_i \equiv M$

Two results: 1) $\hat{t} = \frac{N}{n} \sum \hat{t}_i$

$$SE(\hat{t}) = \sqrt{N^2\left(1 - \frac{n}{N}\right) \frac{s_t^2}{n}} \qquad (5.3)$$

2) $\hat{\bar{y}} = \frac{\hat{t}}{NM}$ $\qquad SE(\hat{\bar{y}}) = \frac{1}{M} \sqrt{1 - \frac{n}{N}} \, \frac{s_t^2}{n} \qquad (5.6)$

$$var\left(\frac{\hat{t}}{NM}\right) = \frac{1}{(NM)^2} \cdot var(\hat{t})$$

Example 5.2

$N = 100$ dorm ~~rooms~~ suites psu's

$M = 4$ rooms in each suite

$n = 5$ psu's ← sample SRS

$m = 4$

Cluster = suite

p 137

	1	2	3	4	5
1	3.08	2.36			
2	2.60	.		.	.
3	3.44	.			
4	3.04	.			
	12.16	11.36	8.96	12.96	11.08

← $\hat{t}_i$'s

$$\hat{t} = \frac{N}{n}\sum \hat{t}_i = \frac{100}{5}\left(12.16 + \cdots + 11.08\right) = 1130.4$$

$$\hat{\bar{y}} = \frac{\hat{t}}{NM} = \frac{1130.4}{400} = 2.\cancel{2}\cancel{5}6 \quad 2.826 \quad \text{OK ??}$$

← est. ~~of~~ of $\bar{y}_u$

2826

$\frac{1130.4}{4}$

$$s_t^2 = \frac{1}{4}\left\{(12.16 - 11.304)^2 + (11.36 - 11.304)^2 + \cdots + (11.08 - 11.304)^2\right\}$$

↖ 1130.4 / N = ? 2.256 ?

$$SE(\hat{\bar{y}}) = \sqrt{\left(1 - \frac{5}{100}\right) \frac{s_t^2}{n}} \cdot \frac{1}{M}$$

$$= 0.164 \quad \text{ntbc.}$$

$$2.826 \pm 1.96 \times 0.164$$

(Central limit theorem)

What's going on?

pop: psu $i = 1, \ldots, N$

ssu $j = 1, \ldots, M$ ← ch. 5

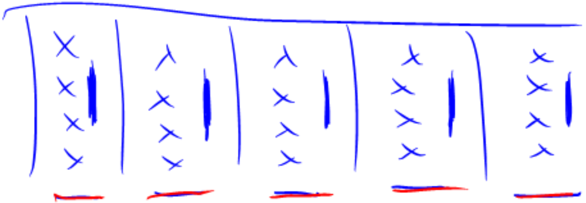
pop values: $y_{ij}\ ;\ i = 1, \ldots, N$
$j = 1, \ldots, M$

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (y_{ij} - \bar{y}_u)^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{iu} + \bar{y}_{iu} - \bar{y}_u)^2$$

$$\overset{\shortparallel}{(K-1) S^2} \qquad = \sum_i \sum_j (y_{ij} - \bar{y}_{iu})^2 + \sum_i \sum_j (\bar{y}_{iu} - \bar{y}_u)^2 + 2 \times 0$$

total SS in pop $= \sum_i \sum_j (y_{ij} - \bar{y}_{iu})^2 + M \sum_i (\bar{y}_{iu} - \bar{y}_u)^2$

within psu SS — between psu SS



E_x

$- \frac{\hat{t}}{N})^2$

SSTO = SSW + SSB — cluster sampling

SSTO ↑ SRS

SSW ↑ stratified

Def: ICC = "intraclass correlation coefficient" or intracluster

$$= 1 - \frac{M}{M-1} \frac{SS_{within}}{SS_{total}}$$

It can be shown

$$\frac{V(\hat{t}_{clu})}{V(\hat{t}_{SRS})} = \frac{MSB}{S^2} = \frac{SS_{betw}/(M-1)}{S^2} \cong 1 + (M-1)\,ICC$$