You should know...

- stratum, strata, stratified sampling
- why/when we prefer stratified sampling to SRS
- inference: use SRS theory in each stratum h
- $\hat{t}_{str} = \sum_{h=1}^{H} \hat{t}_h, \quad \widehat{V}(\hat{t}_{str}) = \sum_{h=1}^{H} (1 - n_h/N_h) N_h^2 s_h^2/n_h$ (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

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:

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

- 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

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 (= ...)
- 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 

- 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

- 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