Note: using s_h^2 not S_h^2

Ex 4.12 , N_h s_h^2 we're using ACRES92 from text $\rightarrow$ R $H=4$

optimal = $\underline{N_h s_h}$ 192 389 63 etc.

$n=300$ $\frac{N_h s_h}{\sum N_h s_h} \times 300$

7	N	21
69	NC	103
122	S	135
101	W	41

'optimal' for ACRES92 — agstrat.dat

In R

```
> tapply(agpop$ACRES92,
    agpop$REGION, length)
```

N_h's

class variable

1. Notation for sample quantities — Ch 5 cluster sampling, equal cl. sizes

a) $\hat{t} = \frac{N}{n}\sum_{i=1}^{n} t_i$ (*) — N # psu's in pop, n # psu's in sample — SRS

$SE(\hat{t}) = N\sqrt{\left(1-\frac{n}{N}\right)\frac{s_t^2}{n}}$ — same as SRS

$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(t_i - \frac{\hat{t}}{N}\right)$ — usual s^2 for cluster totals

$\hat{\bar{y}} = \frac{\hat{t}}{NM}$ — estimate of $\bar{y}_u$, mean over all NM individuals

$SE(\hat{\bar{y}}) = \frac{SE(\hat{t})}{NM} = \frac{1}{M}\sqrt{\left(1-\frac{n}{N}\right)\frac{s_t^2}{n}}$ — Note (*) 1-stage clustering

See Table 5.1 for another way to summarize population quantities.

Analysis of Variance (Anova) Table [for the sample]

Source	df	SS	MS
→ between c.	$n-1$	$\sum_i\sum_j(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2$	$SS/(n-1) = \hat{MSB}$
within c.	$n(m-1)$	$\sum_i\sum_j(y_{ij}-\bar{y}_{i\cdot})^2$	$SS/n(m-1) = \hat{MSW}$
	$nm-1$	$\sum_i\sum_j(y_{ij}-\bar{y})^2$ (SSTOT)	

s_t^2 is computed as between SS — n.t.b.c. exactly how

§5.2.3 unequal cluster sizes $M_i \neq M$ not all same i

- formula for $\hat{t}$ is unchanged
- trickier to estimate $\bar{y}_u = \sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} \Big/ \sum_{i=1}^{N} M_i$

bec. we don't know $M_1, \ldots, M_N$ ← Wed.

Exercise 5.9.1 p. 169

$N = 3000$ $n = 100$ residential phone numbers

total # eligible voters 157

bias 23 no answer

134 responses 112 opposed

$\hat{p} = \frac{112}{134} = 0.84$

$\hat{V}(\hat{p}) = \frac{112}{134}\left(1-\frac{112}{134}\right)\cdot\frac{1}{134}$ $\quad \frac{\hat{p}(1-\hat{p})}{n_v}$ } assumes all responses independent

ok?

$y_{ij} = \begin{cases}1\\0\end{cases}$ $\quad j=1,\ldots,M_i$ # of voters in household i

$i = 1,\ldots,100 = n$

t_i = # voters opposed

$\hat{t} = 112$ $\quad s_t^2 = \sum(t_i - \hat{t}/N)^2/(n-1)$ ← can't get this

Exercise 5.9.2 — HW

Exercise 5.9.3 compliance audit in accounting

$N = 828$ claims $n = 85$ every claim has 215 entries M

claims

1 . . . 85

entries

$$t_i \xrightarrow{215} 4 \quad 3 \quad 2 \quad 2 \quad 2 \quad 2 \quad \underbrace{1 \cdots 1}_{22} \quad \underbrace{0 \cdots 0}_{57}$$

$t_i \quad i = 1, \ldots, 85$

$\hat{t} = \frac{N}{n} \sum t_i = \frac{828}{85} 37 = \boxed{360.4}$ $\qquad \frac{\hat{t}}{N} = 0.4353$

$s_t^2 = \frac{1}{n-1} \sum (t_i - \frac{\hat{t}}{N})^2 = 0.5582633$

$SE(\hat{t}) = N \sqrt{(1 - \frac{n}{N}) \frac{s_t^2}{n}} = \boxed{63.56}$ $\leftarrow$ (b)

error rate $\hat{\bar{y}} = \frac{\hat{t}}{NM} = 0.0020 = \frac{\hat{t}/N}{M}$ $\qquad .2\%$

$SE(\hat{\bar{y}}) =$ ~~0.0046~~ 0.00036 $\qquad .036\%$

(c) $SE(\hat{p})$ under SRS 0.00031 $\begin{cases} n & 18275 \\ N & 178020 \end{cases}$