

# Stat 572 Sampling Theory & Practice

## Homework 1

Erik Erhardt

February 1, 2006

### Assignment 1.1

**target population** All books.

**sampling frame** Shelf locations.

**sampling unit** Selected shelf location.

**observation unit** Book at selected shelf location.

**selection bias/inaccuracy** Need to enumerate Lib. of Cong. call numbers in catalog to numbers, then select random numbers covering ranges.

### Assignment 1.2

**target population** HMO consumers.

**sampling frame** Readers of magazine, May '94 – April '95.

**sampling unit** Self-selected respondents to survey.

**observation unit** Same, provided respondent is involved with HMO.

**selection bias/inaccuracy** Self-selection, misreporting, sampling frame does not cover target population.

**valuable information** Results are limited to CR population, may be misleading since self-selection may be more likely to include people with certain characteristics, such as discontent.

**Assignment 1.3**

**target population** All people with BRCA1 gene.

**sampling frame** Families in N.A. and W.E., with a history of breast or ovarian cancers.

**sampling unit** Families with at least 4 persons diagnosed with breast and ovarian cancer before age 60.

**observation unit** Unsure: Women in these families who consented.

**selection bias/inaccuracy** Resonant self-selection by way of consent, chose families with diagnosed cancers, sampling frame does not cover target population.

**An estimate** No, results do not generalize to all women, but only to those women who could have been sampled using their sampling design.

**Assignment 1.5**1. *mean of population*

$$\begin{aligned}\bar{y}_s &= \frac{1}{6}(98 + 154 + 190 + 102 + 133 + 175) = \mathbf{142} \\ \sigma_y &= \mathbf{34.53}\end{aligned}$$

2. *sampling plans*

$$(1a) \quad E(\bar{y}_s) = \sum_{i=1}^6 \Pr(S_i)\bar{y}_i = \frac{1}{6}(140.3 + 154.3 + 111 + 125 + 159 + 173) = 862.6/6 = \mathbf{143.8}$$

$$(1b) \quad \text{Var}(\bar{y}_s) = \sum_{i=1}^6 \Pr(S_i)(\bar{y}_i - E(\bar{y}_s))^2 = \mathbf{439.2}$$

$$(1c) \quad \text{Bias}(\bar{y}_s) = E(\bar{y}_s) - \bar{y}_s = 143.8 - 142 = \mathbf{1.8}$$

$$(1d) \quad \text{MSE}(\bar{y}_s) = \text{Var}(\bar{y}_s) + (\text{Bias}(\bar{y}_s))^2 = 439.2 + 1.8^2 = \mathbf{442.4}$$

$$(2a) \quad E(\bar{y}_s) = \frac{1}{4}(125) + \frac{1}{2}(173) + \frac{1}{4}(140.3) = \mathbf{152.825}$$

$$(2b) \quad \text{Var}(\bar{y}_s) = \mathbf{436.3}$$

$$(2c) \quad \text{Bias}(\bar{y}_s) = 152.825 - 142 = \mathbf{10.8}$$

$$(2d) \quad \text{MSE}(\bar{y}_s) = 436.3 + 10.8^2 = \mathbf{552.9}$$

3. *best sampling plan*

While sampling plan 1 has a slightly larger variance, it has a much smaller bias and MSE, thus I prefer 1 over 2.

**Assignment 2.1** *Why not a SRS?*

Districts have different numbers of houses. While districts have equal probability of being chosen, and houses within districts have equal probability of being chosen, groups of houses in districts 51–75 do not have equal probability. Thus, this is not a SRS.

**Assignment 2.2** *sample 10 houses*

From p. 416 Figure A.2, using the cumulative house count, houses in district 51–75 are house numbers 12326–31989. Using Table E.1 on p. 457, starting at the first row going across, and continuing until I get 10 numbers in the range of house numbers, I draw the sample shown in the table below. Returning to Figure A.2, the house numbers are identified as addresses by finding the closest cumulative house count greater than the sampled house number to decide the district, then taking the difference with the previous cumulative house count for the house within the district.

Sample	District	House
26428	68	803
23607	65	236
20708	62	513
13557	52	707
17870	58	988
17371	58	489
29089	72	173
26149	68	524
21223	63	133
31842	75	655

**Assignment 2.3** *Use sample*

(a) *average number of TVs in Lockhart City (Q3)*

Mean (SE) of TVs:  $1.5 \left( \frac{1.2042}{\sqrt{10}} = 0.3808 \right)$

(b) *average price willing to pay for cable TV (Q4)*

Mean (SE) of price:  $8.0 \left( \frac{5.5678}{\sqrt{10}} = 1.7607 \right)$

ADDRESS	VALUE	1	2	3	4	5	6	7	8	9
68 803	75746	1	0	4	10	7	0	0	0	7
65 236	72794	2	7	2	15	141	14	42	25	26
62 513	73956	3	0	0	0	0	0	0	0	0
52 707	59455	3	1	2	10	77	7	30	3	15
58 988	68743	2	3	2	10	66	6	31	8	15
58 489	48082	1	0	0	0	0	0	0	0	0
72 173	79626	3	0	2	10	14	1	9	0	4
68 524	76573	2	1	2	15	57	4	16	4	8
63 133	56233	1	0	0	0	0	0	0	0	0
75 655	73116	2	0	1	10	56	5	14	0	11