Week 12-13--IES 612.doc

IES 612 Winter/Spring 2009

Cluster Sampling

Defn: Cluster Sample is any probability sample in which a sample of clusters is chosen and every element in the cluster is sampled.

VOCABULARY

Cluster = arrangement or collection of population elements that are grouped together under some common theme (often based on natural delineations – e.g. herd of animals, block of houses)

Think of clusters as being "mini-populations" which are "easy" to sample.

Primary Sampling Unit (PSU) = any of the defined clusters

Secondary Sampling Unit (SSU) = the actual element within a cluster whose observation is recorded for analysis

NOTES/COMMENTS

1. Cluster sampling is most commonly done when construction of a complete frame of population elements is difficult, but construction of a frame of clusters is straightforward and easy.

   An example is that the US Census Bureau has very detailed information on city blocks, but very poor information on individuals in a city.

2. Cluster sampling is also more "cost" efficient compared to SRS if the cost of obtaining sample information increases as the distance between sample elements increases.

PROCEDURE

1. Obtain population frame, which is just a list of the clusters and identify every cluster in the frame with a unique identifier.

2. Obtain a SRS of n distinct clusters from the frame and sample every element in the cluster.

NOTATION

Population

N = number of clusters or Primary Sampling Units in the population

$\mathbf{M_i}$ = number of elements in the i^{th} cluster, i = 1, …, n

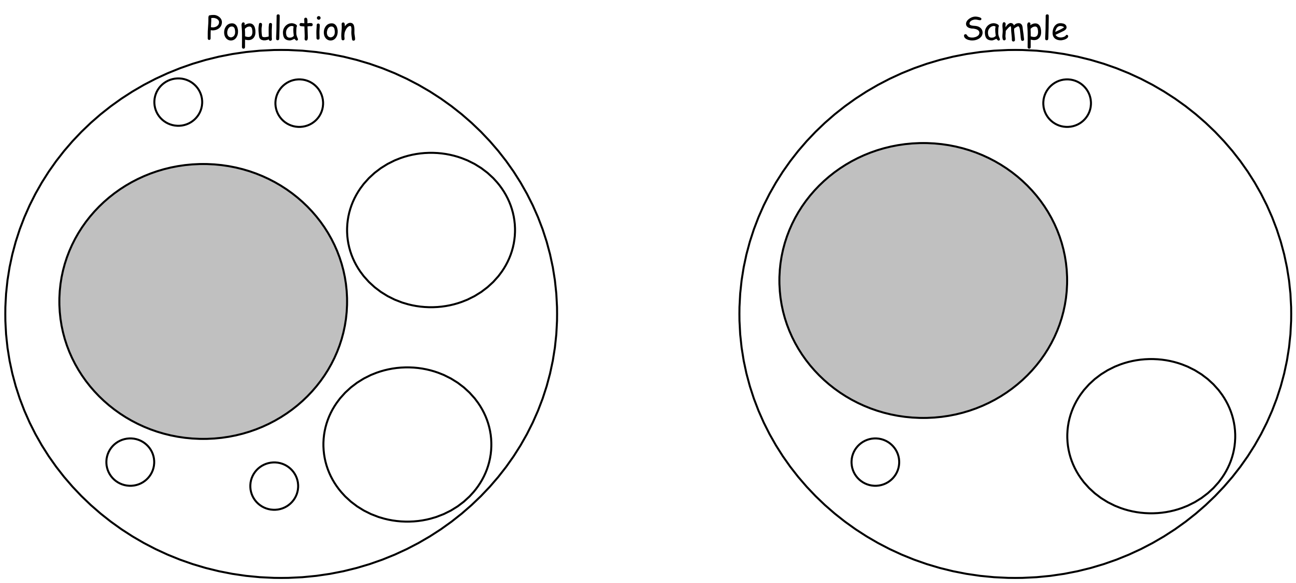
M = number of elements in the population = $\sum_{i=1}^{N} K_i$

$\bar{\mathbf{M}}$ = average cluster size (avg number of observations per cluster) = $\frac{1}{N}\sum_{i=1}^{N} M_i = \frac{M}{N}$

$\mathbf{y_{ij}}$ = j^{th} element's observed outcome from i^{th} cluster, j = 1, …, M_i

τ_i = total of measurements in the i^{th} cluster, i = 1, …, N; $\left(\tau_i = \sum_{j=1}^{M_i} \mathbf{y_{ij}} = \mathbf{y_{i\bullet}}\right)$

$\mu_i = \frac{\tau_i}{\mathbf{M_i}} = \frac{\mathbf{y_{i\cdot}}}{\mathbf{M_i}} = \bar{\mathbf{y}}_{i\cdot}$ = mean of measurements in the i^{th} cluster, i = 1, …, N



Sample

n = number of clusters sampled

$\bar{\mathbf{m}}$ = average cluster size in the sample = $\frac{1}{n}\sum_{i=1}^{n} M_i$

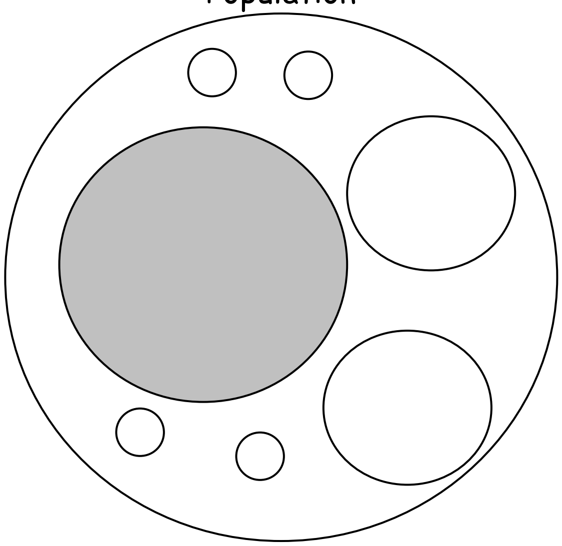
Population Parameter

$$\mu = \text{populatioin mean} = \frac{\text{sum of all observations in popln}}{\text{total number of observations}} = \frac{\sum_{i=1}^{N}\tau_i}{\sum_{i=1}^{N}M_i} = \frac{\frac{1}{N}\sum_{i=1}^{N}\tau_i}{\frac{1}{N}\sum_{i=1}^{N}M_i} = \frac{\frac{1}{N}\sum_{i=1}^{N}\tau_i}{\bar{M}} = \frac{\bar{\tau}}{\bar{M}}$$

$$= \frac{\text{mean cluster total}}{\text{mean cluster size}} = \text{ratio of two unknowns}$$

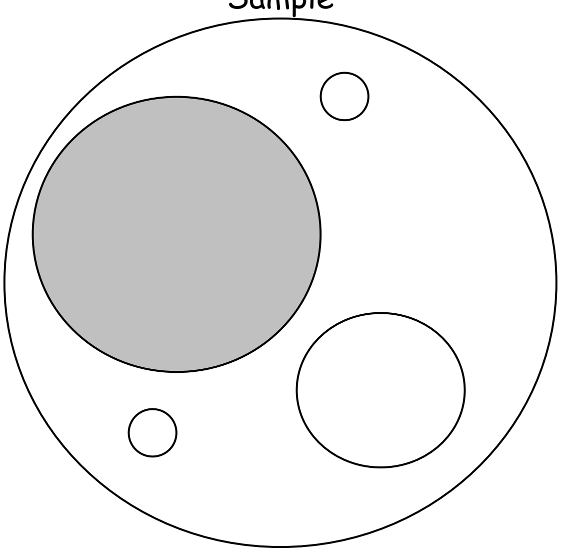
$$\tau = \text{population total} = \text{sum of all observations in popln} = \sum_{i=1}^{N}\tau_i$$

Population



How would we summarize the sample? What would be our SAMPLE STATISTICS?

Sample



$y_{i\bullet}$ = total of measurements in the i^{th} sampled cluster, i = 1, …, n; $\left(y_{i\bullet} = \sum_{j=1}^{M_i} y_{ij} \right)$

ESTIMATING PARAMETERS IN A CLUSTER SAMPLE

1. ESTIMATING THE POPULATION MEAN (μ) IN A CLUSTER SAMPLE

a. Point Estimate of μ is $\bar{y} = \dfrac{\text{Sum of observations in the \textbf{sampled} clusters}}{\text{Sum of the cluster sizes in the \textbf{sampled} clusters}} = \dfrac{\sum_{i=1}^{n} y_{i\cdot}}{\sum_{i=1}^{n} M_i}$

b. Standard Error of PE = $se(\bar{y}) = \sqrt{\dfrac{s_r^2}{n}\left(\dfrac{N-n}{N}\right)\dfrac{1}{\bar{M}^2}}$

where $s_r^2 = \dfrac{\sum_{i=1}^{n}(y_{i\cdot} - \bar{y}m_i)^2}{n-1}$ and if $\bar{M}$ unknown, use $\bar{m}$ as an estimate of $\bar{M}$.

c. CI for μ: approximate 1 – α confidence interval for μ is

$$\bar{y} \pm z_{\frac{\alpha}{2}} se(\bar{y}) \qquad \text{or} \qquad \bar{y} \pm t_{(\frac{\alpha}{2}, n-1)} se(\bar{y}).$$

2. ESTIMATING THE POPULATION TOTAL ($\tau = M\mu$) IN A CLUSTER SAMPLE

a. Point Estimate of τ is $\hat{\tau} = M\bar{y} = M\dfrac{\sum_{i=1}^{n} y_{i\cdot}}{\sum_{i=1}^{n} M_i}$.

b. Standard Error of PE = $se(\hat{\tau}) = se(M\bar{y}) = Mse(\bar{y}) = M\sqrt{\dfrac{s_r^2}{n}\left(\dfrac{N-n}{N}\right)\dfrac{1}{\bar{M}^2}}$.

c. CI for μ: approximate 1 – α design-based confidence interval for τ is

$$M\bar{y} \pm z_{\frac{\alpha}{2}} Mse(\bar{y}).$$

Example: SMO Example 8.2

Interviews are conducted in each of 25 blocks sampled from a city. The n = 25 blocks represent of cluster sample of blocks from the city of N = 415 blocks. The data on incomes are presented below.

Cluster (i)	Number of Residents in cluster (M_i)	Total Income of Residents in cluster ($y_{i\bullet}$)
1	8	$96,000
2	12	121,000
3	4	42,000
4	5	65,000
5	6	52,000
6	6	40,000
7	7	75,000
8	5	65,000
9	8	45,000
10	3	50,000
11	2	85,000
12	6	43,000
13	5	54,000
14	10	49,000
15	9	53,000
16	3	50,000
17	6	32,000
18	5	22,000
19	5	45,000
20	4	37,000
21	6	51,000
22	8	30,000
23	7	39,000
24	3	47,000
25	8	41,000
	$\Sigma M_i = 151$	$\Sigma y_{i\bullet}$ = $1,329,000

Some further summary, via calculator, Excel, or (heaven forbid) R,

Cluster (i)	Number of Residents in cluster (Mi)	Total Income of Residents in cluster (yi·)	(yi·)-ybar*Mi
1	8	96,000.00	25,589.40
2	12	121,000.00	15,384.11
3	4	42,000.00	6,794.70
4	5	65,000.00	20,993.38
5	6	52,000.00	-807.95
6	6	40,000.00	-12,807.95
7	7	75,000.00	13,390.73
8	5	65,000.00	20,993.38
9	8	45,000.00	-25,410.60
10	3	50,000.00	23,596.03
11	2	85,000.00	67,397.35
12	6	43,000.00	-9,807.95
13	5	54,000.00	9,993.38
14	10	49,000.00	-39,013.25
15	9	53,000.00	-26,211.92
16	3	50,000.00	23,596.03
17	6	32,000.00	-20,807.95
18	5	22,000.00	-22,006.62
19	5	45,000.00	993.38
20	4	37,000.00	1,794.70
21	6	51,000.00	-1,807.95
22	8	30,000.00	-40,410.60
23	7	39,000.00	-22,609.27
24	3	47,000.00	20,596.03
25	8	41,000.00	-29,410.60
Sum	151	1,329,000.00	0.00
Average	6.04	53,160.00	0.00
StDev	2.371356855	21,784.32	25,189.31
y-bar =	8801.324503		

So

	n	Avg	Std Dev
Number of Residents in cluster (M_i)	25	6.040	2.371
Total Income of Residents in cluster ($y_{i\bullet}$)	25	53,160	21,784
($y_{i\bullet}$) – 8,801M_i	25	0	25,189

```
> Cluster
   Cluster ClusterSize ClusterTotal
1        1           8        96000
2        2          12       121000
3        3           4        42000
4        4           5        65000
5        5           6        52000
6        6           6        40000
7        7           7        75000
8        8           5        65000
9        9           8        45000
10      10           3        50000
11      11           2        85000
12      12           6        43000
13      13           5        54000
14      14          10        49000
15      15           9        53000
16      16           3        50000
17      17           6        32000
18      18           5        22000
19      19           5        45000
20      20           4        37000
21      21           6        51000
22      22           8        30000
23      23           7        39000
24      24           3        47000
25      25           8        41000
> sum(ClusterSize)
[1] 151
> sum(ClusterTotal)
[1] 1329000
> ybar = sum(ClusterTotal)/sum(ClusterSize)
> ybar
[1] 8801.325
> sd(ClusterTotal-ybar*ClusterSize)
[1] 25189.31
```

1. ESTIMATING THE POPULATION MEAN INCOME (μ) IN THE CITY

a. Point Estimate of μ is $\bar{y} = \frac{\sum_{i=1}^{n} y_{i\cdot}}{\sum_{i=1}^{n} M_i} =$

b. Standard Error of PE = $se(\bar{y}) = \sqrt{\frac{s_r^2}{n}\left(\frac{N-n}{N}\right)\frac{1}{\bar{M}^2}} =$

c. CI for μ: approximate 1 – α confidence interval for μ is $\bar{y} \pm z_{\frac{\alpha}{2}} se(\bar{y})$

2. ESTIMATING THE CITY'S TOTAL INCOME (τ = Mμ)

Assuming there are M = 2,500 residents in the city:

a. Point Estimate of τ is $\hat{\tau} = M\bar{y} =$

b. Standard Error of PE = $se(\hat{\tau}) = se(M\bar{y}) = Mse(\bar{y}) =$

c. CI for μ: approximate 1 – α confidence interval for τ is

$$M\bar{y} \pm z_{\frac{\alpha}{2}} Mse(\bar{y}).$$

Randomized Response

PROBLEM or ISSUE

At times you need to obtain answers to questions that are sensitive and which respondents are unlikely or unwilling to provide honest/correct answers. For example, "Have you used illegal drugs in the last six months?" will likely result in respondents in lying if they have used drugs.

Our goal will be to obtain an estimate of p = proportion of the population that have used illegal drugs in the past 6 months.

PROCEDURE

1. Start with a deck of cards that are all identical except that a fraction, θ, have "Yes, I have used illegal drugs in the past 6 months" on them and the remaining fraction, $1 - \theta$, have "No, I have not used illegal drugs in the pat 6 months."

2. Take a SRS of n people.

3. Each person randomly selects a card at AND without the interviewer seeing which card they have selected, they respond YES or NO as to whether they are a member of the group on their selected card.

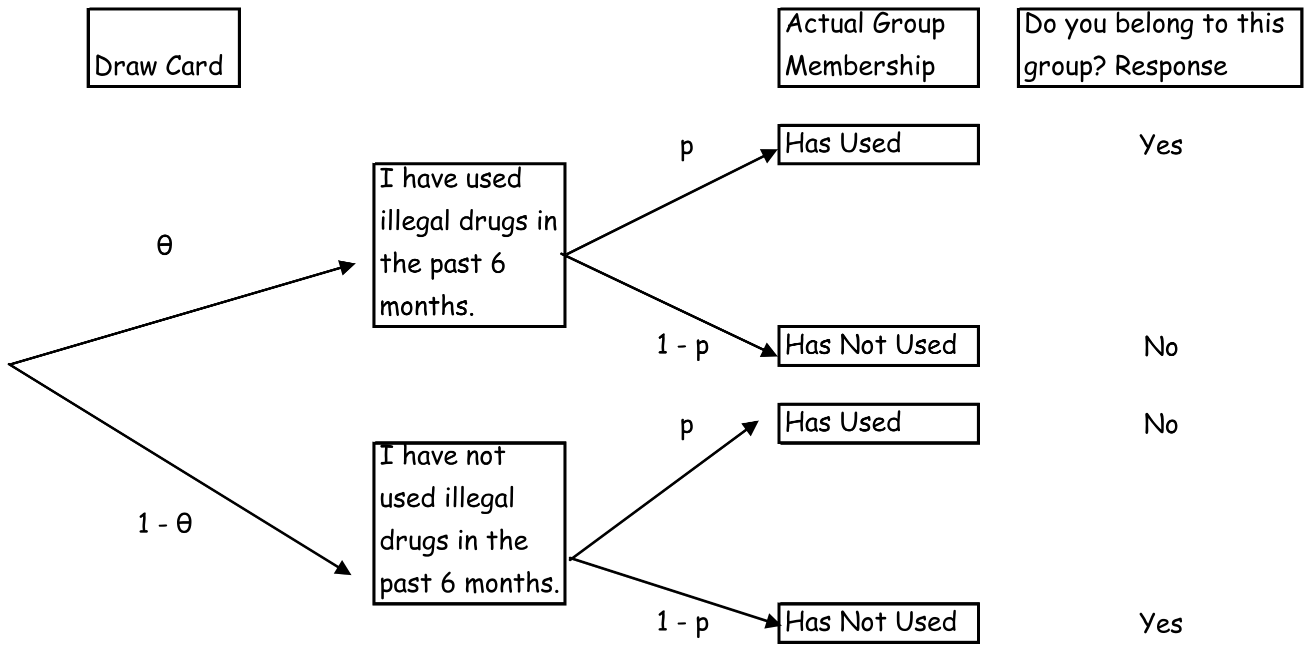
4. The interviewer simply records a YES or NO.

5. This is done for each person in the sample, WHERE the card selected is replaced and the cards shuffled.

6. At the end, the interviewer will have recorded n_1 Yes's out of the total of n responses AND knows what fraction of the cards, θ, had the statement "Yes, I have used illegal drugs in the past six months."

To see how we would estimate p, the true popln proportion, using only these three pieces of information, we will use probability ideas and a tree diagram.

The resulting tree diagram would look like this:



With our results, we have an estimate of the Pr(Yes), namely n_1/n. Using simple conditional probability results we know that Pr(Yes) = $\theta p + (1 - \theta)(1 - p)$. Which can be solved for p as $p = \frac{1}{(2\theta - 1)}\Pr(Yes) - \left(\frac{1-\theta}{(2\theta-1)}\right)$.

Hence our estimate of p is: $\hat{p} = \frac{1}{(2\theta - 1)}\left(\frac{n_1}{n}\right) - \left(\frac{1-\theta}{(2\theta-1)}\right)$.

An estimate of the variance of $\hat{p}$ is: $\hat{Var}(\hat{p}) = \frac{1}{(2\theta-1)^2}\hat{Var}\left(\frac{n_1}{n}\right) = \frac{1}{(2\theta-1)^2}\frac{1}{n}\left(\frac{n_1}{n}\right)\left(1-\frac{n_1}{n}\right)$.

Our Bound on the Error of Estimation is: $2\sqrt{\hat{Var}(\hat{p})}$ from which we could obtain an approximate 95% confidence interval for p.

Example: SMO Example 11.5

A SRS of 400 people were used to estimate the proportion of people who falsified information on their income tax return. Using a deck of cards with $\frac{3}{4}$ of the cards indicating info had been falsified and $\frac{1}{4}$ of the cards with a statement tax return info had been correct, resulted in 120 Yes's.

The resulting estimate of p is $\hat{p}$ = 0.10 with an estimated variance of 0.0021 and a bound of 0.09. We would conclude, with ~ 95% confidence that between 1 to 19% have falsified info.

Specialized Techniques for Environmental Sampling

CAPTURE-RECAPTURE SAMPLING

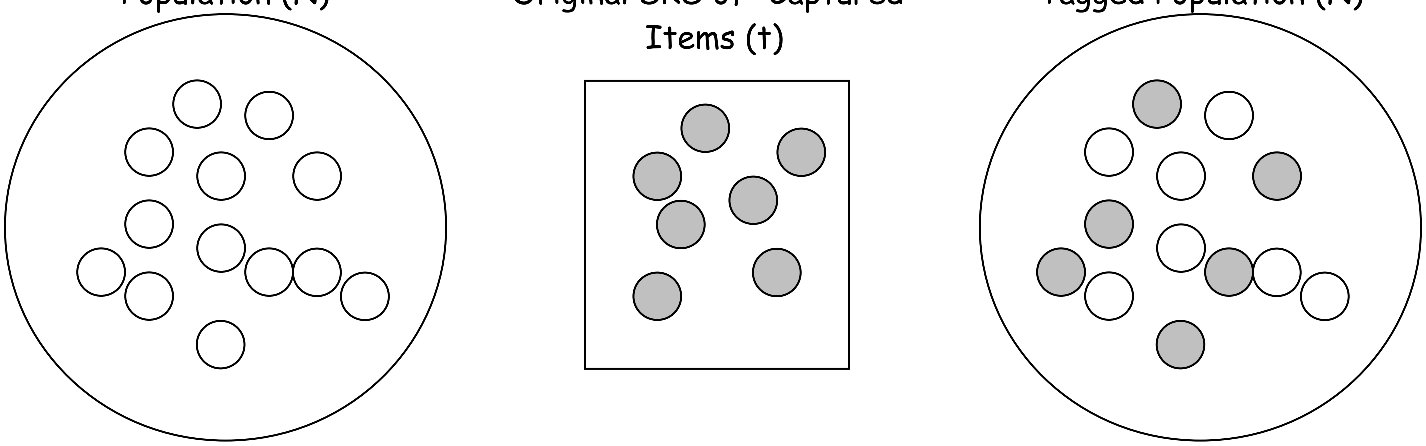
GOAL: Estimate the size of a single population, N (assume N > 0)

PROCEDURE

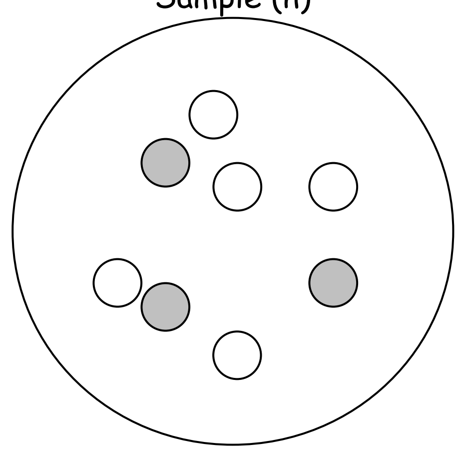
1. Take a SRS of t > 0 elements from the population and tag or mark them in some unambiguous fashion

2. A short time later, take another SRS of size n (where each population element has an equal probability of capture in either sampling sweep) and count the number of elements, M, that exhibit the previous tag

Population (N) | Original SRS of "Captured" Items (t) | Tagged Population (N)



Sample (n)



The capture-recapture method is based on estimating the proportion captured and tagged in the population, ie, $\pi = \frac{t}{N}$. Our sample estimate of π is the sample proportion of tagged elements, namely $\hat{\pi} = \frac{M}{n}$. Under Simple Random Sampling we know that $\hat{\pi} \approx \pi$ or $\frac{M}{n} \approx \frac{t}{N}$. Hence we estimate N by solving $\frac{M}{n} = \frac{t}{\hat{N}}$ for $\hat{N}$ to obtain: $\hat{N} = \frac{tn}{M}$.

NOTES/COMMENTS

1. Capture-Recapture sampling is also known as tag-recapture and mark-recapture sampling.

ESTIMATING PARAMETERS IN A CAPTURE-RECAPTURE SAMPLE

1. Estimating the POPULATION SIZE (N) in CAPTURE-RECAPTURE–DIRECT Sampling

a. Point Estimate of N is:

i. Lincoln-Petersen estimator (goes back to 1896 and 1930): $\hat{N}_{LP} = \frac{nt}{M}$ (biased)

ii. Chapman estimator (goes back to 1948): $\hat{N}_C = \frac{(n+1)(t+1)}{M+1} - 1$

b. Standard Error of PE = $se(\hat{N}_C) = \sqrt{\frac{(t+1)(n+1)(t-M)(n-M)}{(M+1)^2(M+2)}}$

c. CI for N: approximate 1 – α confidence interval for N or ψ is

$\hat{N}_C \pm z_{\frac{\alpha}{2}} se(\hat{N}_C)$ (best for large samples).

2. Estimating the POPULATION SIZE (N) in CAPTURE-RECAPTURE–INVERSE Sampling (Fix "t" and let "n" be random)

a. Point Estimate of N is: $\hat{N} = \frac{nt}{s}$

b. Standard Error: $se(\hat{N}) = \sqrt{\frac{t^2 n(n-s)}{s^2(s+1)}}$

c. CI for N: $\hat{N} \pm z_{\alpha/2} se(\hat{N})$

(often use B=2*se($\hat{N}$) here)

Example (Bobwhite Quail Abundance)

GOAL: Estimate the population of bobwhite quail (*Colinus virginianus*) in a Florida field research station.

Initially, t = 148 quail were trapped and banded (the 'capture' sweep) over a 20-day period. After release, n = 82 birds were recaptured at a later time of which a total of M = 39 had been banded.

Estimates of N are:

$$\hat{N}_{LP} = \frac{(82)(148)}{39} = 311.18\text{, or } 312 \text{ birds}$$

$$\hat{N}_{C} = \frac{(83)(149)}{40} - 1 = 308.175\text{, or } 309 \text{ birds}$$

An approximate 95% CI for N is

$$\hat{N}_C \pm z_{\frac{\alpha}{2}}\, se(\hat{N}) = 308.175 \pm (1.96)(29.7254) = 308.175 \pm 58.2618$$

We would estimate, with 95% confidence that there are between 249 and 367 quail.

NOTES/COMMENTS

3. Assumptions:

   Closed population – no migration (or birth or death)—WHY?

   Probability of capture is _____________ —WHY?

4. Alternative Method Uses the Hypergeometric Distribution:

   Applying the hypergeometric-based confidence limits, iterative calculations show that at a 95% confidence interval for N is [259, 394] birds.

QUADRAT SAMPLING

GOAL: Estimate the density of the population and the size of a single population, M (M > 0)

DEFINITIONS AND PARAMETERS OF INTEREST

Population Density = $\frac{\text{number in population}}{\text{area encompassed by population}} = \frac{M}{A} = \lambda$ = density per unit area

Population Size = M = λA

BACKGROUND

Assume

1. population encompasses an overall area of size A

2. population area can be partitioned into N equally size quadrats of size "a" (A = Na)

PROCEDURE

1. Take a SRS of size n of the quadrats

2. Let m_i be the count in the i^{th} sample quadrat, i = 1, 2, ..., n

SAMPLE RESULTS

1. Let $\bar{m}$ be the average number of elements per quadrat, $\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n}$

2. Let $s_m^2 = \frac{\sum_{i=1}^{n}(m_i - \bar{m})^2}{n-1}$ be the sample variance of the sampled quadrat counts.

3. Let $\hat{Var}(\bar{m})$ be the estimated variance of $\bar{m}$, $\hat{Var}(\bar{m}) = \frac{s_m^2}{n}$

ESTIMATING PARAMETERS IN A QUADRAT SAMPLING

1. ESTIMATING THE POPULATION DENSITY (λ = density per unit area)

a. Point Estimate of λ is $\hat{\lambda} = \frac{\bar{m}}{a}$

b. Standard Error of PE = $se(\hat{\lambda}) = se(\frac{\bar{m}}{a}) = \frac{1}{a} se(\bar{m}) = \frac{1}{a}\sqrt{\frac{s_m^2}{n}}$

c. CI for μ: approximate 1 – α confidence interval for λ is

$$\hat{\lambda} \pm z_{\frac{\alpha}{2}} se(\hat{\lambda})$$

2. ESTIMATING THE POPULATION SIZE (M = λA)

a. Point Estimate of M is $\hat{M} = \hat{\lambda} A = \frac{\bar{m}}{a} A = \frac{A}{a}\bar{m}$

b. Standard Error of PE = $se(\hat{M}) = se(\hat{\lambda} A) = A se(\hat{\lambda}) = A \frac{1}{a}\sqrt{\frac{s_m^2}{n}} = \frac{A}{a}\frac{s_m}{\sqrt{n}}$

c. CI for μ: approximate 1 – α design-based confidence interval for M is

$$A\hat{\lambda} \pm z_{\frac{\alpha}{2}} A se(\hat{\lambda})$$

COMMENT:
If a variable has a Poisson distribution, then its variance = mean. Thus, if the number of events follows a Poisson distribution, then an estimate of the MEAN number of events also would be an estimate of the VARIANCE.

Example: (Fire Ant Hill density – SMO)

GOAL: Estimate the number of fire ant hills per unit area and total number of fire ant hills in St. John's County, FL. (NOTE: St. John's county Florida = 1,574,712,771 m^2 = A.)

A SRS of fifty quarats, each 16 m^2, (NOTE that a = 16 m^2) yielded the following:

Number of Hills (m_i)	0	1	2	3	4	5	Total
Frequency	13	8	12	10	5	2	50

Ie, { 0, 0, ..., 0, 1, 1, ..., 1, 2, 2, ..., 2, 3, 3, ..., 3, 4, 4, 4, 4, 4, 5, 5 }

Sample results

Total number of hills found = 0(13) + 1(8) + 2(12) + ... + 5(2) = 92

Average number of hills found $\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} = \frac{0+0+...+5}{50} = \frac{92}{50} = 1.84$

Sample variance of the m_i's is $s_m^2 = \frac{\sum_{i=1}^{n}(m_i - 1.84)^2}{50-1} = 2.1779592 = 2.1780$

1. ESTIMATING THE FIRE ANT HILL DENSITY PER SQUARE METER (λ)

a. Point Estimate of λ is $\hat{\lambda} = \frac{\bar{m}}{a} =$

b. Standard Error of PE = $se(\hat{\lambda}) = se(\frac{\bar{m}}{a}) = \frac{1}{a} se(\bar{m}) = \frac{1}{a}\frac{s_m}{\sqrt{n}} =$

c. CI for μ: approximate 1 – α confidence interval for λ is $\hat{\lambda} \pm z_{\frac{\alpha}{2}} se(\hat{\lambda}) =$

2. ESTIMATE TOTAL # OF FIRE ANT HILLS IN THE COUNTY (M = λA)

a. Point Estimate of M is $\hat{M} = \hat{\lambda} A = \frac{\bar{m}}{a} A =$

b. Standard Error of PE = $se(\hat{M}) = se(\hat{\lambda} A) = Ase(\hat{\lambda}) =$

c. Approximate 1 – α confidence interval for M is $A\hat{\lambda} \pm z_{\frac{\alpha}{2}} Ase(\hat{\lambda})$

LINE-INTERCEPT SAMPLING

GOAL: Estimate the population total or population density in a region with area of size A of some measurement of interest (say number of downed logs in a given forest, length of downed logs in a given forest, or volume of downed logs in the forest)

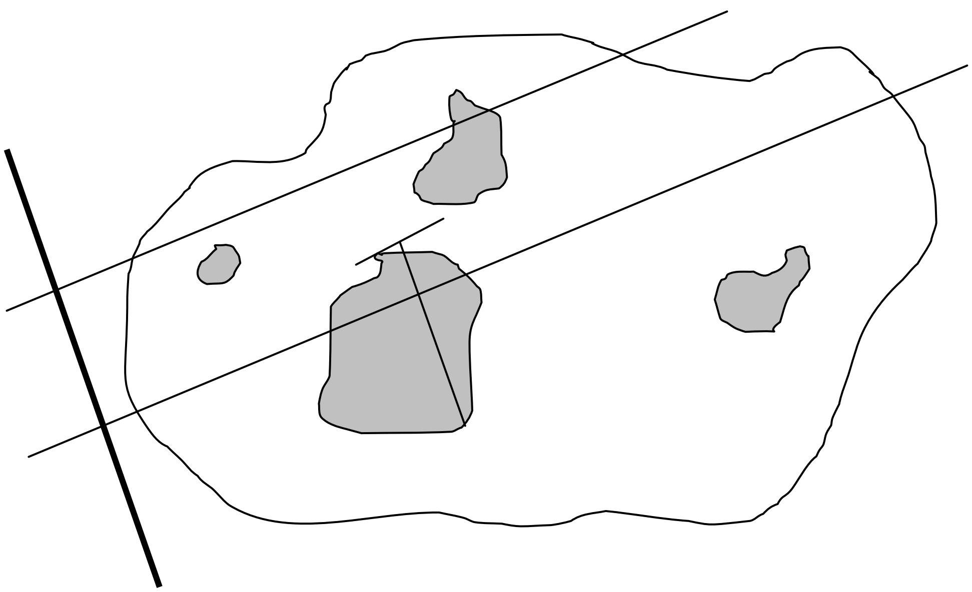
DEFINITIONS AND PARAMETERS OF INTEREST

Measurement of Interest = x
Number of Elements in Population = N
Population Total of Measurement = $\tau = \Sigma$ x in population
Population Density = $\lambda = \tau/A$

PROCEDURE

1. Establish a base line, of length B, across the region of interest

2. Take a sample (via SRS or Systematic Sampling) of n starting points along the baseline and set a perpendicular transect from each of these points



3. Each time a population element is found to cross the transect ('intercepted'), th measurement of interest, x_j, is recorded

4. Let w_j be the width of this element.

5. Define p_j to be w_j/B to be the detection probability of the j^{th} element

6. Repeat 1 – 5 for a total of n different transects, indexed by i = 1, 2, ..., n

SAMPLE RESULTS

1. Let φ_i be the number of elements encountered along the i^{th} transect, i = 1, 2, ..., n

2. Let $y_i = \sum_{j=1}^{\phi_i} \frac{x_j}{p_j} = B \sum_{j=1}^{\phi_i} \frac{x_j}{w_j}$ in the i^{th} transect, i = 1, 2, ..., n

3. Define $\overline{y}$ as usual, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

4. And $s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}$

ESTIMATING PARAMETERS IN A LINE INTERSECT SAMPLING

1. ESTIMATING THE POPULATION TOTAL (τ)

a. Point Estimate of τ is $\hat{\tau} = \overline{y}$

b. Standard Error of PE = $se(\hat{\tau}) = se(\overline{y}) = \sqrt{\frac{s_y^2}{n}} = \frac{s_y}{\sqrt{n}}$

c. CI for μ: approximate 1 – α confidence interval for τ is $\hat{\tau} \pm z_{\frac{\alpha}{2}}\, se(\hat{\tau})$

2. ESTIMATING THE POPULATION DENSITY ($\lambda = \tau/A$)

a. Point Estimate of λ is $\hat{\lambda} = \frac{\hat{\tau}}{A}$

b. Standard Error of PE = $se(\hat{\lambda}) = se\left(\frac{\hat{\tau}}{A}\right) = \frac{1}{A} se(\hat{\tau}) = \frac{1}{A}\sqrt{\frac{s_y^2}{n}} = \frac{1}{A}\frac{s_y}{\sqrt{n}}$

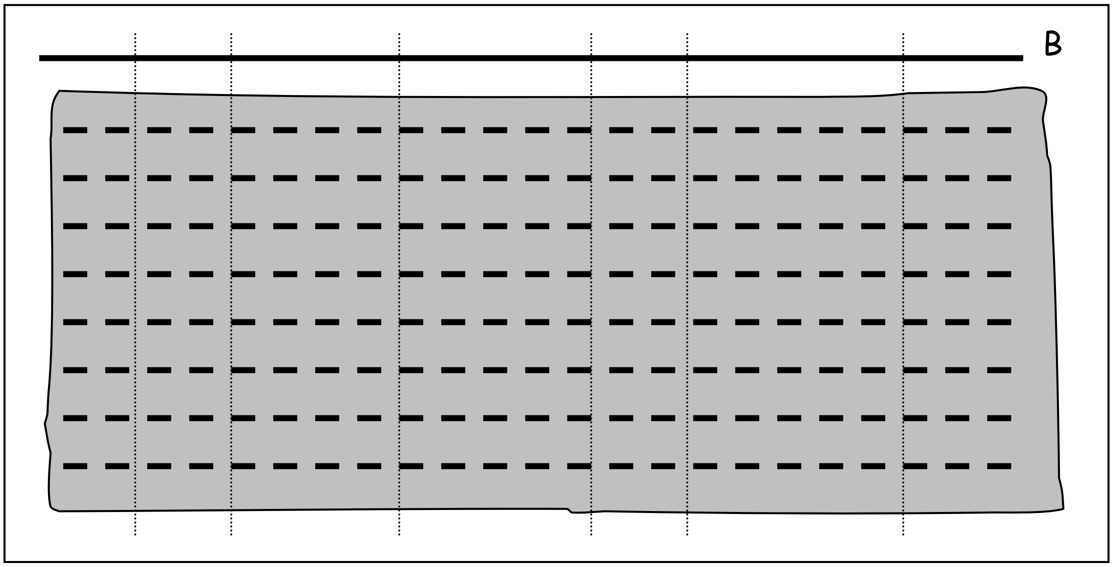
c. CI for μ: approximate 1 – α confidence interval for λ is $\frac{1}{A}\hat{\tau} \pm z_{\frac{\alpha}{2}} \frac{1}{A} se(\hat{\tau})$

Example: Abundance of tarnished plant bugs (PB Exercise 8.21)

GOAL: Estimate the total number of insects in a cotton field with 8 rows of cotton

Sample Method

1. A baseline of 450 ft is established and 6 transects randomly selected. Since the transects cross the rows of cotton on the perpendicular, each transect will have φ value of 8.



2. At each row of cotton a 2.5 x 2.5 ft drop cloth is placed under each intersection point and the number of insects is counted. This is the x_j value. Note that the w_j value is fixed at 2.5 ft.

Below is the data collected and represents the number of bugs (x_j) collected at each site.

Transect	1	2	3	4	5	6
Row 1	2	0	0	0	0	0
Row 2	2	2	1	0	0	0
Row 3	2	0	0	0	0	0
Row 4	0	1	0	0	0	0
Row 5	2	0	0	0	0	0
Row 6	0	0	0	0	0	2
Row 7	2	0	0	1	1	0
Row 8	3	0	0	0	0	0

Sample results

Note that since B = 450 and w_j = 2.5 ft in every case, the y_i are simply:

$$y_i = \sum_{j=1}^{\phi_i} \frac{x_j}{p_j} = B\sum_{j=1}^{\phi_i} \frac{x_j}{w_j} = \frac{B}{w}\sum_{j=1}^{8} x_j = \frac{450}{2.5}\sum_{j=1}^{8} x_j = 180\sum_{j=1}^{8} x_j$$

Transect	1	2	3	4	5	6	
	2	0	0	0	0	0	
	2	2	1	0	0	0	
	2	0	0	0	0	0	
	0	1	0	0	0	0	
	2	0	0	0	0	0	
	0	0	0	0	0	2	
	2	0	0	1	1	0	
	3	0	0	0	0	0	
$y_i = 180\sum_{j=1}^{8} x_j$	2,340	540	180	180	180	360	$\bar{y} = 630$
							$s_y = 850.01$

1. ESTIMATING THE POPULATION TOTAL (τ)

a. Point Estimate of τ is $\hat{\tau} = \bar{y} =$

b. Standard Error of PE = $se(\hat{\tau}) = se(\bar{y}) = \sqrt{\frac{s_y^2}{n}} = \frac{s_y}{\sqrt{n}} =$

c. CI for μ: approximate $1 - \alpha$ confidence interval for τ is $\hat{\tau} \pm z_{\frac{\alpha}{2}}\, se(\hat{\tau})$