Design based inference

Week 2, day 3 (2.2)

Stat 260, St. Clair

Content:

- Small example
- Comparing estimators
- Notation

Small example

Suppose we have $N = 3$ lakes in a town. We are asked to estimate the total number residents who live around the lakes.

Sampling Design: We are going to compare three different designs where

- Population = three lakes
- Sampling unit = Observation unit = lake
- Measurement = y_i is the number of people who reside around lake i
- **Parameter** of interest = $t = \sum_{i=1}^{3} y_i$ total number of residents around all three lakes

Small example

We want to estimate t using a sample of size $n = 2$ lakes.

Why does it make sense to use

$$\hat{t} = 3\bar{y}$$

as an estimator of t where $\bar{y}$ is our average response in our sample?

Design-based idea

What is the **sampling distribution** of $\hat{t}$ **under a particular sampling design?**

1. List out all possible samples of n sampling units from a population of size N.
2. For each set of sampled units $\mathcal{S}$:
    - Compute the probability of each sample **based on your chosen sampling design**
    - Compute the estimator value $\hat{t}$

Design-based idea

Step (2) generates the **sampling distribution** for $\hat{t}$.

Properties:

- **Expected value**: $E(\hat{t})$
- **Bias**: $bias(\hat{t})$
- **SD**: $SD(\hat{t})$
- **Mean Square Error**: $MSE(\hat{t})$

Different sampling designs will yield different sampling distributions *and* different estimator properties.

Small example

Design 1: Roll a 6-sided die twice.

- A 1 or 2 samples unit 1
- a 3 or 4 samples unit 2
- a 5 or 6 samples unit 3

Use this same scheme to sample two lakes (with replacement so repeats are possible!)

What samples are possible?

Small example

Design 1: Roll a 6-sided die twice.

- A 1 or 2 samples unit 1
- a 3 or 4 samples unit 2
- a 5 or 6 samples unit 3

What is the probability of each sample?

Small example

No.	Sample	D1: $P(\mathcal{S}_i)$	Data	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$
1	$\mathcal{S}_1 = \{1, 1\}$	1/9	?	?	?
2	$\mathcal{S}_2 = \{1, 2\}$	2/9	?	?	?
3	$\mathcal{S}_3 = \{1, 3\}$	2/9	?	?	?
4	$\mathcal{S}_4 = \{2, 2\}$	1/9	?	?	?
5	$\mathcal{S}_5 = \{2, 3\}$	2/9	?	?	?
6	$\mathcal{S}_6 = \{3, 3\}$	1/9	?	?	?

We can't fill in the table (with numbers) unless we know what the lake responses y_i are.

Small example

Suppose our population looks like:

lake i	**1**	**2**	**3**
y_i	10	8	12

then:

No.	Sample	D1: $P(\mathcal{S}_i)$	Data	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$
1	$\mathcal{S}_1 = \{1, 1\}$	1/9	10,10	10	30
2	$\mathcal{S}_2 = \{1, 2\}$	2/9	10,8	9	27
3	$\mathcal{S}_3 = \{1, 3\}$	2/9	10,12	11	33
4	$\mathcal{S}_4 = \{2, 2\}$	1/9	8,8	8	24
5	$\mathcal{S}_5 = \{2, 3\}$	2/9	8,12	10	30
6	$\mathcal{S}_6 = \{3, 3\}$	1/9	12,12	12	36

Small example

Under **design 1**, what is the sampling distribution of $\hat{t}$? (hint: it will be a **discrete probability mass function**)

Small example

Under **design 1**, what is the (estimator) bias of $\hat{t}$?

Small example

Under **design 1**, what is the SD of $\hat{t}$?

Small example

Under **design 1**, what is the MSE of $\hat{t}$?

Small example

Design 2: Toss two darts at a map of the town where

- lakes 1 and 2 are the same area and
- lake 3 is three times the area of 1 and 2.

Use this same scheme to sample two lakes (with replacement so repeats are possible!)

Fill in the ? probabilities

No.	Sample	Data	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$	D1: $P(\mathcal{S}_i)$	D2: $P(\mathcal{S}_i)$
1	$\mathcal{S}_1 = \{1, 1\}$	10,10	10	30	1/9	?
2	$\mathcal{S}_2 = \{1, 2\}$	10,8	9	27	2/9	?
3	$\mathcal{S}_3 = \{1, 3\}$	10,12	11	33	2/9	?
4	$\mathcal{S}_4 = \{2, 2\}$	8,8	8	24	1/9	?
5	$\mathcal{S}_5 = \{2, 3\}$	8,12	10	30	2/9	?
6	$\mathcal{S}_6 = \{3, 3\}$	12,12	12	36	1/9	?

Small example

Design 2: Toss two darts at a map of the town where

- lakes 1 and 2 are the same area and
- lake 3 is three times the area of 1 and 2.

No.	Sample	Data	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$	D1: $P(\mathcal{S}_i)$	D2: $P(\mathcal{S}_i)$
1	$\mathcal{S}_1 = \{1, 1\}$	10,10	10	30	1/9	1/25
2	$\mathcal{S}_2 = \{1, 2\}$	10,8	9	27	2/9	2/25
3	$\mathcal{S}_3 = \{1, 3\}$	10,12	11	33	2/9	6/25
4	$\mathcal{S}_4 = \{2, 2\}$	8,8	8	24	1/9	1/25
5	$\mathcal{S}_5 = \{2, 3\}$	8,12	10	30	2/9	6/25
6	$\mathcal{S}_6 = \{3, 3\}$	12,12	12	36	1/9	9/25

Write down the sampling distribution of $\hat{t}$ under design 2 and computed the expected value, bias, SD and MSE.

Small example

Design 2: Toss two darts at a map of the town where

- lakes 1 and 2 are the same area and
- lake 3 is three times the area of 1 and 2.

Sampling distribution of $\hat{t}$ under **design 2**

$\hat{t}$	**24**	**27**	**30**	**33**	**36**
Probability	1/25	2/25	7/25	6/25	9/25

- Expected value: $E(\hat{t}) = 32.4$
- Bias: $Bias(\hat{t}) = 2.4$
- SD: $SD(\hat{t}) \approx 3.39$
- MSE: $MSE(\hat{t}) \approx 17.28$

Small example

Design 3: Put three pieces of paper in a hat labeled 1-3, draw 2 pieces at random, without replacement.

Fill in the ? probabilities

No.	**Sample**	**Data**	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$	**D1:** $P(\mathcal{S}_i)$	**D2:** $P(\mathcal{S}_i)$	**D3:** $P(\mathcal{S}_i)$
1	$\mathcal{S}_1 = \{1,1\}$	10,10	10	30	1/9	1/25	?
2	$\mathcal{S}_2 = \{1,2\}$	10,8	9	27	2/9	2/25	?
3	$\mathcal{S}_3 = \{1,3\}$	10,12	11	33	2/9	6/25	?
4	$\mathcal{S}_4 = \{2,2\}$	8,8	8	24	1/9	1/25	?
5	$\mathcal{S}_5 = \{2,3\}$	8,12	10	30	2/9	6/25	?
6	$\mathcal{S}_6 = \{3,3\}$	12,12	12	36	1/9	9/25	?

Small example

Design 3: Put three pieces of paper in a hat labeled 1-3, draw 2 pieces at random, without replacement (Simple Random Sample)

No.	**Sample**	**Data**	$\bar{y}_s$	$\hat{t} = 3\bar{y}_s$	**D1:** $P(\mathcal{S}_i)$	**D2:** $P(\mathcal{S}_i)$	**D3:** $P(\mathcal{S}_i)$
1	$\mathcal{S}_1 = \{1,1\}$	10,10	10	30	1/9	1/25	0
2	$\mathcal{S}_2 = \{1,2\}$	10,8	9	27	2/9	2/25	1/3
3	$\mathcal{S}_3 = \{1,3\}$	10,12	11	33	2/9	6/25	1/3
4	$\mathcal{S}_4 = \{2,2\}$	8,8	8	24	1/9	1/25	0
5	$\mathcal{S}_5 = \{2,3\}$	8,12	10	30	2/9	6/25	1/3
6	$\mathcal{S}_6 = \{3,3\}$	12,12	12	36	1/9	9/25	0

Write down the sampling distribution of $\hat{t}$ under design 3 and computed the expected value, bias, SD and MSE.

Small example

Design 3: Put three pieces of paper in a hat labeled 1-3, draw 2 pieces at random, without replacement.

Sampling distribution of $\hat{t}$ under **design 3**

$\hat{t}$	**24**	**27**	**30**	**33**	**36**
Probability	0	1/3	1/3	1/3	0

- Expected value: $E(\hat{t}) = 30$
- Bias: $Bias(\hat{t}) = 0$
- SD: $SD(\hat{t}) \approx 2.45$
- MSE: $MSE(\hat{t}) = 6$

Small example

Which design is "better"?

design	Bias	SD	MSE
1 (die)	0	3.5	12
2 (darts)	2.4	3.39	17.28
3 (srs)	0	2.45	6