

36-303: Sampling, Surveys and Society

Statistics of Surveys II
Brian Junker
132E Baker Hall
brian@stat.cmu.edu

2 February 2010

1

Handouts

- Only the lecture notes!

2 February 2010

2

Outline

- Team Project I.1 Feedback
 - Team Project I.2 Due Thurs Feb 4 in Email by Midnight
 - Please send pdf if possible, otherwise msword
- Urn Models
- A Survey Sampling Experiment
- Elementary Statistics
 - SRS with replacement
- Survey Sampling
 - SRS (and other probability samples) without replacement
- FOR NEXT WEEK Groves Ch 7: Question Design

2 February 2010

3

Team Projects

- It took me longer than I expected!
- I.1: I was somewhat harsh
 - I don't dislike you; I want the projects to be
 - Interesting
 - Doable
 - Please use my comments to improve your proposals
- I.2: Revised full proposals for both projects due Midnight Thur Feb 4 2010 (new due date)
 - Send pdf or msword by email
- I gave everyone 100 pts for I.1. I will assign points more carefully and strictly for I.2.

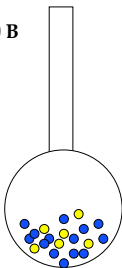
2 February 2010

4

Urn Models

30 Y & 70 B

$N = 100$



- Draw $n=10$ balls from the urn
 - What proportion are yellow?
 - How much variability in the proportion, if I repeat the experiment?
- The properties of the sample depend on how the sample was drawn.

2 February 2010

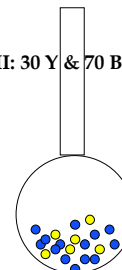
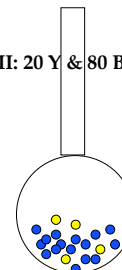
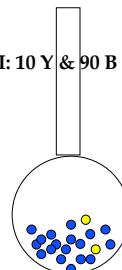
5

A Survey Sampling Experiment

I: 10 Y & 90 B

II: 20 Y & 80 B

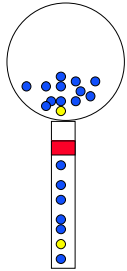
III: 30 Y & 70 B



2 February 2010

6

Sampling From Urns



Urn I

- Take a sample of size $n=10$, by shaking urn and moving 10 balls into neck.
- Repeat process 20 times.
- Write down the number of yellows you got for each time.

Repeat for Urns II and III

2 February 2010

7

Sampling From Urns (cont.)

- Circulate all three urns
- Each student should mix the balls; then draw a sample and record # of yellows out of 10
 - Turn in a piece of paper with your name, and 3 neat columns of 20 results each (20 for each urn!)
- Today: Preliminary look at Urn 3
- Thursday: Compare our results with the actual probability distribution for each urn.

Brian Junker

Urn 1	Urn 2	Urn 3
2	1	3
0	2	5
0	1	2
0	2	5
3	2	4
1	2	2
0	0	4
2	5	2
1	2	1
0	2	3
1	2	1
1	3	1
2	1	3
1	4	3
0	1	4
1	1	3
0	5	2
0	0	3
0	2	0
0	3	3

2 February 2010

8

What do we remember from Elementary Statistics?

- For simple random sampling (SRS) with replacement,

$$E[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

- The Central Limit Theorem then tells us

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

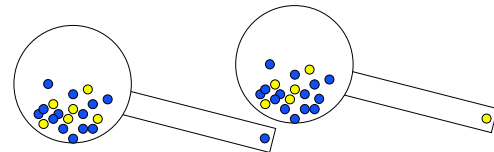
- σ is the SD of X_i ; $\sigma/\sqrt{n}$ is the SE of $\bar{X}$
- But in survey sampling we sample w/o replacement!

2 February 2010

9

SRS With Replacement

- Draw one ball at a time
- Replace ball and re-shake urn for next draw
- Stop when you get n balls
- The composition of the urn never changes



2 February 2010

10

SRS With Replacement

- Let $X_i = 1$ if i^{th} ball in sample is yellow, else $X_i=0$, $i=1, 2, \dots, n$
- $E[X_i] = 1 \cdot P[X_i=1] + 0 \cdot P[X_i=0] = p = 30/100$, so

$$E[\bar{p}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} np = p$$

- Because we always replace the ball, one draw cannot affect the next, and so $\text{Cov}(X_i, X_j)=0$. So

$$\text{Var}(\bar{p}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \left[\sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(X_i, X_j) \right]$$

$$= \frac{1}{n^2} np(1-p) + \sum \sum 0 = p(1-p)/n$$

- So

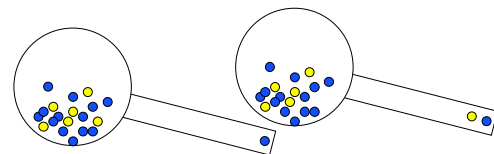
$$SE(\bar{p}) = \sqrt{p(1-p)/n}$$

2 February 2010

11

SRS Without Replacement

- Draw one ball at a time
- Do not replace ball after you draw it
- Stop when you draw n balls
- The composition changes with every draw



2 February 2010

12

SRS Without Replacement

- Let $X_i = 1$ if i^{th} ball in sample is yellow, else $X_i=0$, $i=1, 2, \dots, n$
- $E[X_1] = 30/100 = p$
- What about X_2 ?
$$\begin{aligned} E[X_2] &= E[X_2|X_1=1]P[X_1=1] + E[X_2|X_1=0]P[X_1=0] \\ &= \frac{29}{99} \frac{30}{100} + \frac{30}{99} \frac{70}{100} \\ &= \frac{30}{100} \left(\frac{29}{99} + \frac{70}{99} \right) = \frac{30}{100} = p, \text{ *whew*} \end{aligned}$$
- What about X_3 ?
- What about $E[\hat{p}]$ and $Var(\hat{p})$?

2 February 2010

13

Results of Experiment

2 February 2010

14

Conjectures from the Experiment

2 February 2010

15

Review

- Team Project I.2 Due Thurs in Email by Midnight
 - Please send pdf if possible, otherwise msword
- Elementary Statistics: SRS with replacement
- Survey Sampling: SRS without replacement
- Our Survey Sampling Experiment
 - Will look at results further on Thursday
- Please read Groves Ch 7 for next week
 - Team Assignments I.4 and I.5 are about question design!

2 February 2010

16