36-303: Sampling, Surveys and Society

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

Class Schedule

- For some reason I'm almost exactly a week behind where the "schedule of topics" says I should be
- I propose to delay the first midterm by one week
  - Old Midterm Date: 2/23
  - New Midterm Date: 3/01

Handouts

- In Class:
  - Lecture Notes
- Online:
  - Team Working Agreements
  - HW04

Upcoming Team Activities

- Team Project Assignment I.3 [Due Thu Feb 16]
  - CHOOSE a single project to do this semester, based on my feedback to I.2
  - TURN IN on Blackboard: a revised version of A-G for the single project you choose, Feb 16.
- Team Working Agreement [Due Thu Feb 23]
  - GET the TWA pdf from the "twa" directory on the class website.
  - TURN IN on blackboard: final TWA Feb 23
- Team Project Assignment II.4 [Due Thu Mar 01]
  - Sampling scheme – questionnaire – sample size

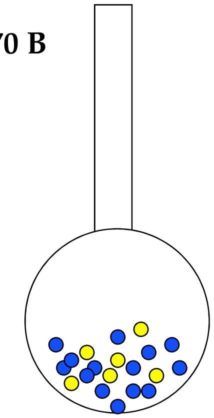
Outline

- 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's 7 & 8: Question Design

Urn Models

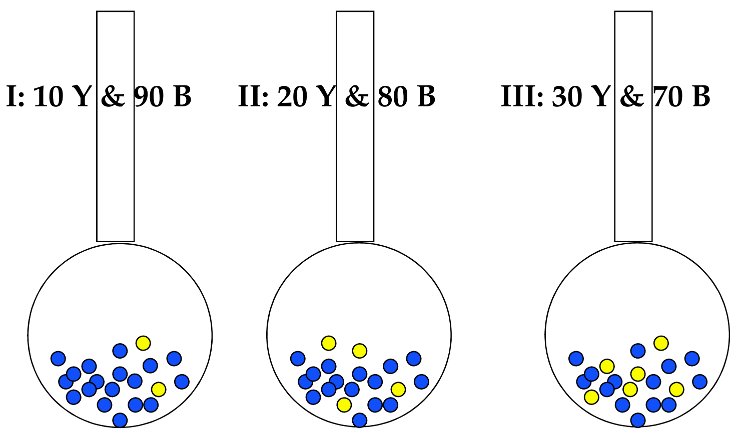
30 Y & 70 B

N = 100

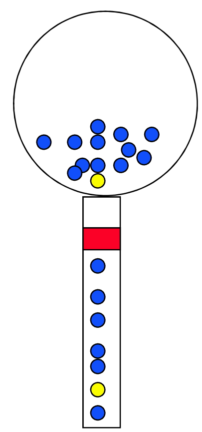


- Draw n=10 balls from the urn
  - What proportion are yellow?
  - How much variability in the proportion, if I repeat the experiment?
- The properties of the sample depend on how the sample was drawn.

A Survey Sampling Experiment



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

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.
- WORK IN PAIRS
  - A shakes, B records
  - B shakes, A records

Brian Junker

10/90 Urn	20/80 Urn	30/70 Urn
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

What do we remember from Elementary Statistics?

- For *simple random sampling (SRS) with replacement,*

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

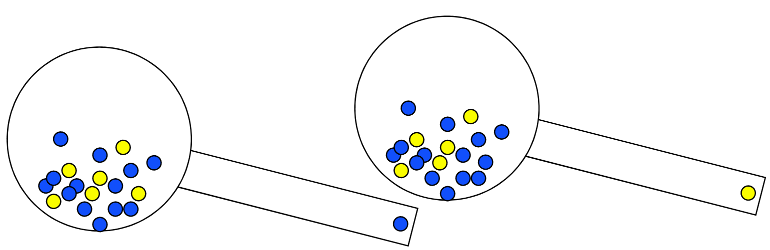
- The Central Limit Theorem then tells us

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

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

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*



SRS With Replacement

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

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

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

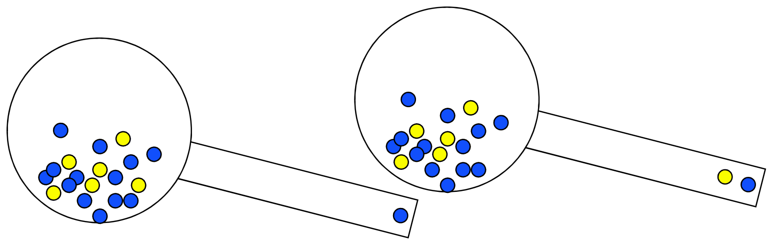
$$\begin{aligned} Var(\hat{p}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} Cov(X_i, X_j)\right] \\ &= \frac{1}{n^2}np(1-p) + \sum\sum 0 = p(1-p)/n \end{aligned}$$

- So

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

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*



SRS Without Replacement

- Let X_i = 1 if i^{th} ball in sample is yellow, else X_i=0, i=1, 2, ..., n
- $E[X_1]$ = 30/100 = p
- What about X_2?

$$\begin{aligned} E[X_2] &= E[X_2|X_1=1]P[X_1=1] + E[X_2|X_1=0]P[X_1=0] \\ &= \frac{29}{99}\frac{30}{100} + \frac{30}{99}\frac{70}{100} \\ &= \frac{30}{100}\left(\frac{29}{99}+\frac{70}{99}\right) = \frac{30}{100} = p, \quad \textit{*whew*} \end{aligned}$$

- *What about X_3?*
- What about $E[\hat{p}]$ and $Var(\hat{p})$?

Results of Experiment

Conjectures from the Experiment

Review

- 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, 8
  - Team Assignments I.4 and I.5 are about question design!
- See HW and team due dates at beginning of lecture