Soc351
Introduction to
Survey Methods for Social Research

Lab

WK02
Sampling Distributions
Sampling Frames

Questions about previous assignments?

Week 1 Assignment

- Make sure you are familiar with the different types of language used to describe survey error (especially those in the lecture figures from Groves et al.)

- Systematic versus variable errors are very important and both are almost always possible. Be sure you are able to identify when each might be at play and how they are different.

Week 2 Reading Response

- What are the main arguments about why cell phone samples are good?

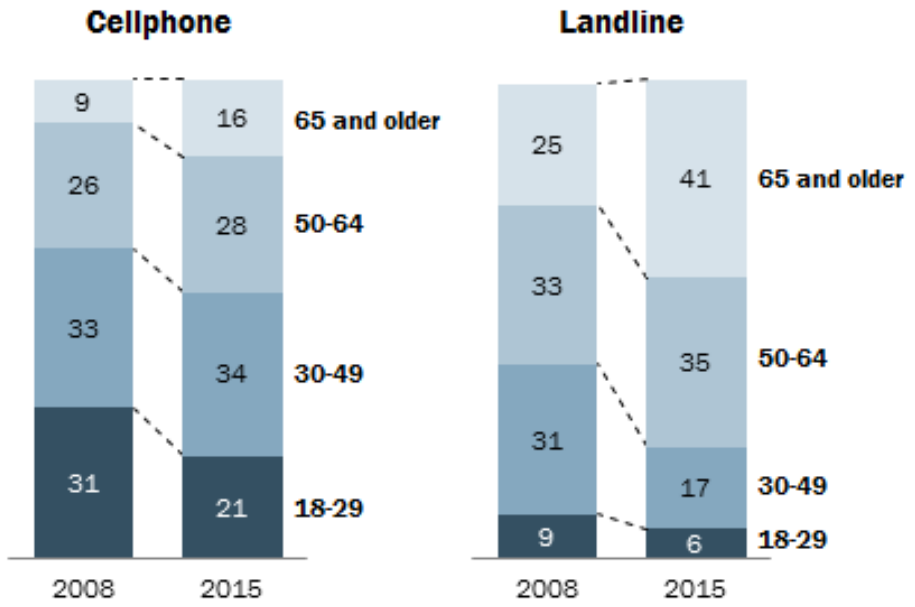
- Is the argument the authors present convincing? Why?

Pew Research Center:
Twilight of Landline Interviewing

Pew: Twilight of Landline Interviewing

Age profile of cellphone samples improves as landline samples continue to gray

Unweighted % of interviews from each sample



Note: Landline respondents include adults who do not have cellphones and those who have cellphones but were reached to complete the survey on their landlines. This second group may or may not have been willing to comply if they had been reached on their cellphones. "Don't know" and refused respondents are not shown.

Source: Surveys conducted Sept. 9-14, 2008 and Sept. 22-27, 2015.
"The Twilight of Landline Interviewing"

PEW RESEARCH CENTER

Pew: Twilight of Landline Interviewing

Cellphone sample more closely matches population distribution for race, age

Demographic profiles of adult population, cellphone respondents and landline respondents

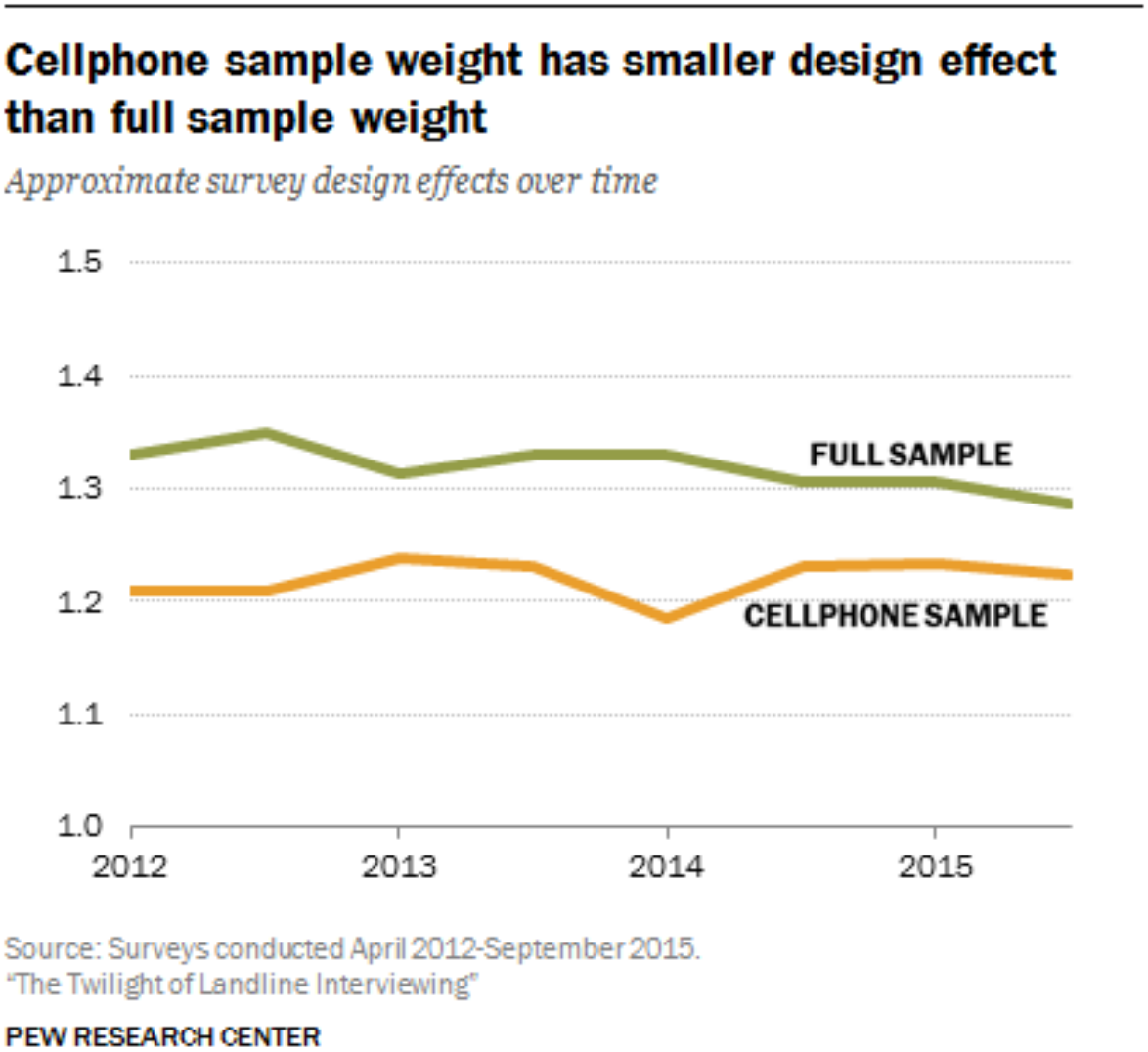
	U.S. population benchmark	Cellphone sample	Landline sample
	Weighted	*Unweighted*	*Unweighted*
	%	%	%
White, non-Hispanic	65	64	**82**
Hispanic	15	**14**	6
Black, non-Hispanic	12	**11**	6
Other, non-Hispanic	8	**9**	3
18-29	22	**21**	6
30-49	34	**34**	17
50-64	26	28	**35**
65+	19	16	**41**
Male	48	**56**	45
Female	52	44	**55**
High school grad or less	41	30	30
Some college/associate degree	31	28	26
Bachelor's degree or more	28	41	44
Unweighted n	2,403,157	977	525

Note: Landline respondents include adults who do not have cellphones and those who have cellphones but were reached to complete the survey on their landlines. This second group may or may not have been willing to comply if they had been reached on their cellphones. "Don't know" and refused respondents are not shown.

Source: Survey conducted Sept. 22-27, 2015; for non-institutionalized adult population, 2014 American Community Survey (IPUMS). Statistically significant differences between cellphone and landline sample in **bold**.
"The Twilight of Landline Interviewing"

PEW RESEARCH CENTER

Pew: Twilight of Landline Interviewing (Design Effects)



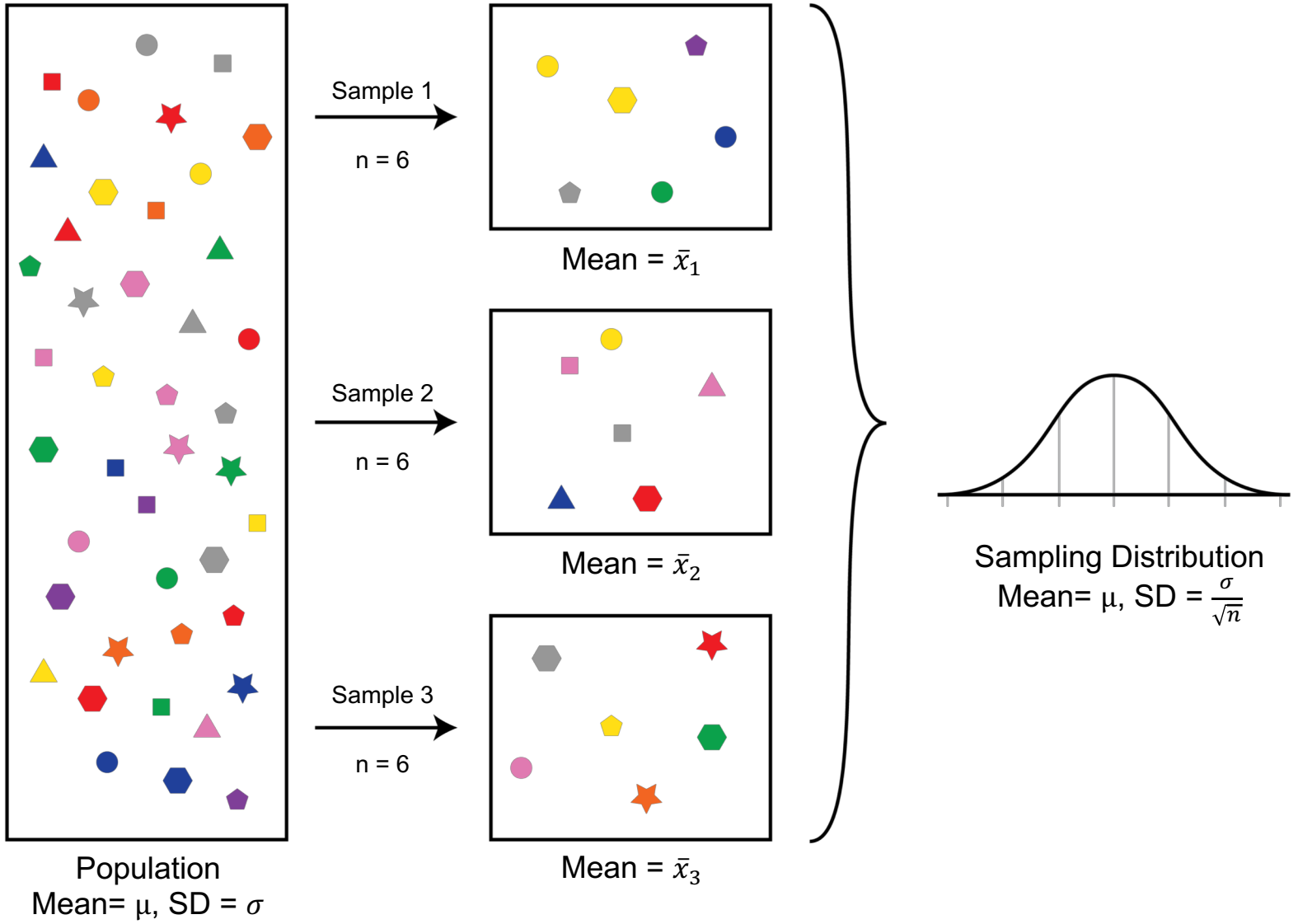
Sampling Distributions

Population Parameters vs. Sample Statistics

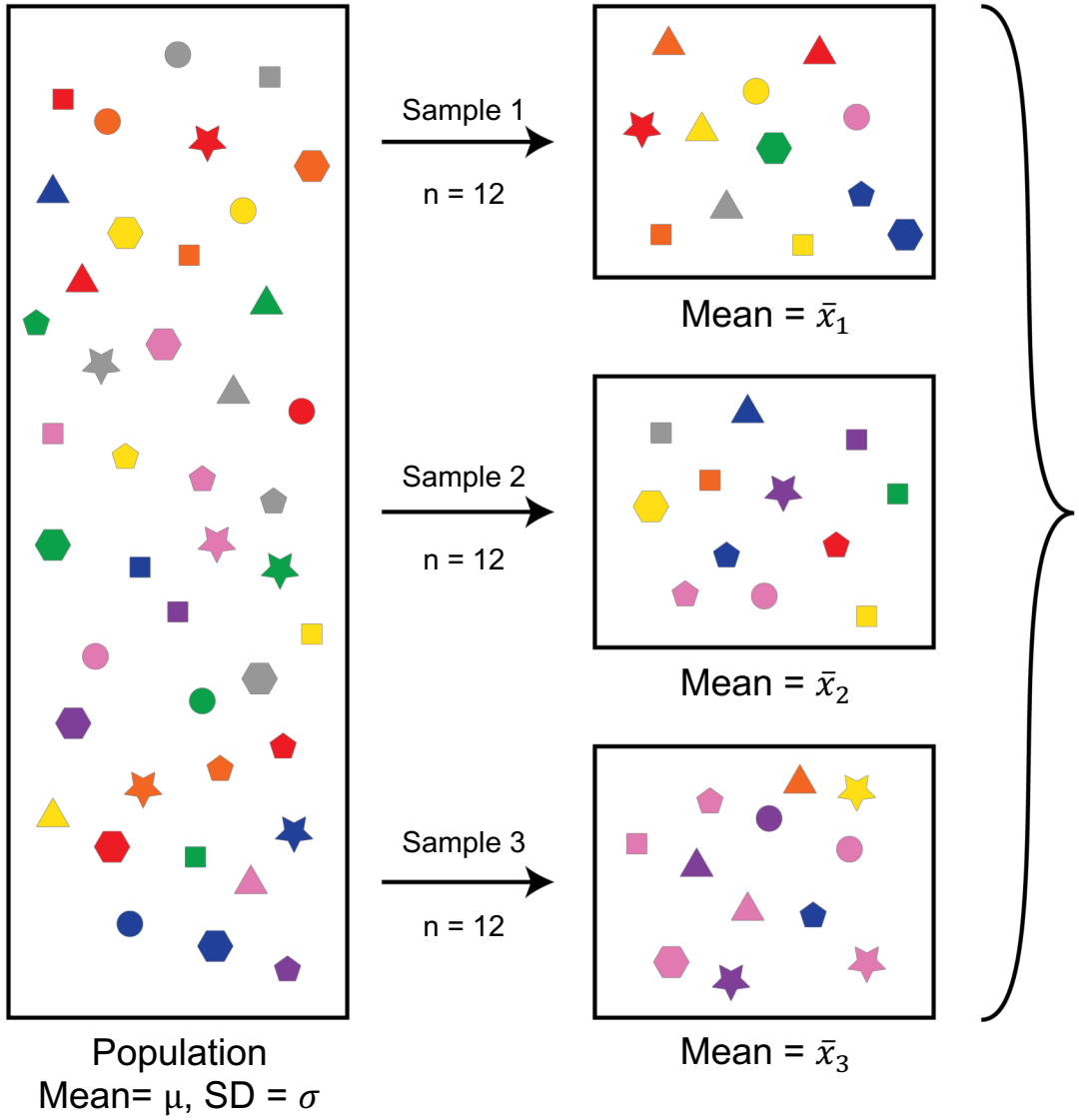
- Population parameters
  - Usually not feasible to calculate
  - What we are trying to estimate with a sample
  - Exact

- Sample statistics
  - Much more feasible to calculate
  - Meant to be an estimate of the population parameter
  - Error induced because of sampling

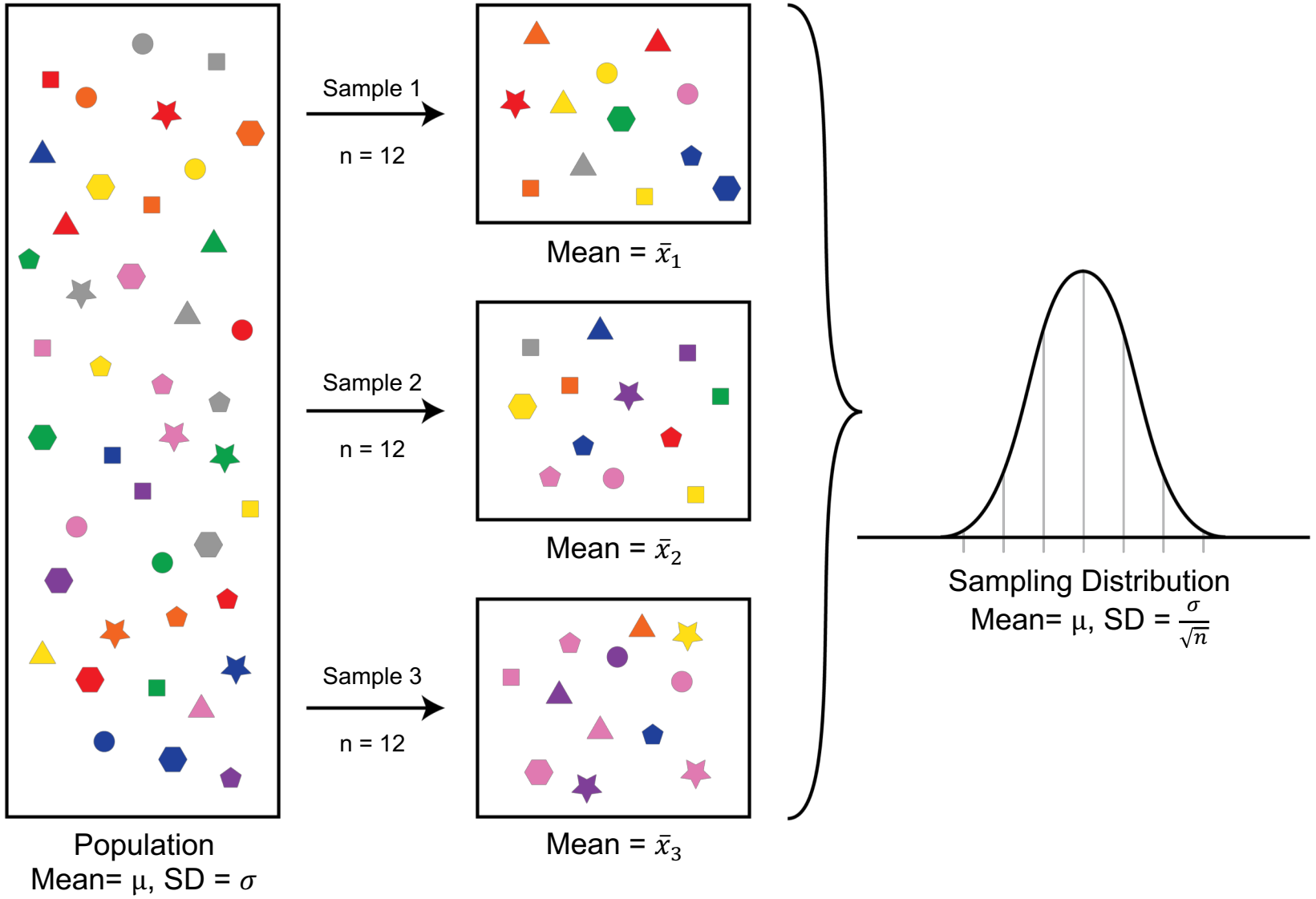
Sampling Distributions (n=6)



Sampling Distributions (n=12)



Sampling Distributions (n=12)



Sampling Distributions – Exercise

- In your own words, describe what a sampling distribution is

- What two components define the sampling distribution? How do they relate to the population parameters?

- Compare your answers to a partner's and come up with a unified answer together

Sampling Distributions - Exercise

- Single department from hypothetical university
  - 7 faculty members
- Calculate population mean and variance for `eval`
  - $\mu_{eval} =$
  - $\sigma_{eval} =$

- $\mu = \frac{1}{N}\sum X_i$
- $\sigma^2 = \frac{1}{N}\sum (X_i - \mu)^2$

id	dept	years	eval
870520	Women's, gender, and sexuality studies	29	100
870382	Women's, gender, and sexuality studies	7	99
870034	Women's, gender, and sexuality studies	9	81
870081	Women's, gender, and sexuality studies	11	71
870597	Women's, gender, and sexuality studies	7	77
870090	Women's, gender, and sexuality studies	23	63
870528	Women's, gender, and sexuality studies	5	74

Sampling Distributions - Exercise

- Single department from hypothetical university
  - 7 faculty members
- Identify all possible samples of n=2, record the `eval` values for each of those samples (you will need this for your assignment this week)

- 1 and 2: 100, 99
- 1 and 3: 100, 81
- ...
- 6 and 7: 63, 74

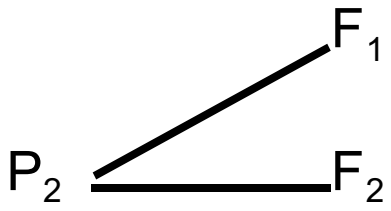
id	dept	years	eval
870520	Women's, gender, and sexuality studies	29	100
870382	Women's, gender, and sexuality studies	7	99
870034	Women's, gender, and sexuality studies	9	81
870081	Women's, gender, and sexuality studies	11	71
870597	Women's, gender, and sexuality studies	7	77
870090	Women's, gender, and sexuality studies	23	63
870528	Women's, gender, and sexuality studies	5	74

Sample Frame Evaluation

Sample Frames

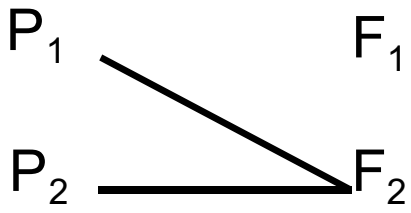
- What is a sample frame?
- How does it relate to a target population?
- What are the most common issues we face when evaluating the quality of a sample frame?

Coverage: Types of problems

Sampling Frame	Label	Example	Solution
	Undercoverage P(selection)=0	Household population, telephone frame	Additional frames, coverage improvement, evaluate impact
	Duplication P(selection) = unequal	Household population, multiple telephone numbers in one unit	Remove duplicates from frame, randomly omit all but one, estimate frequency and weight by inverse

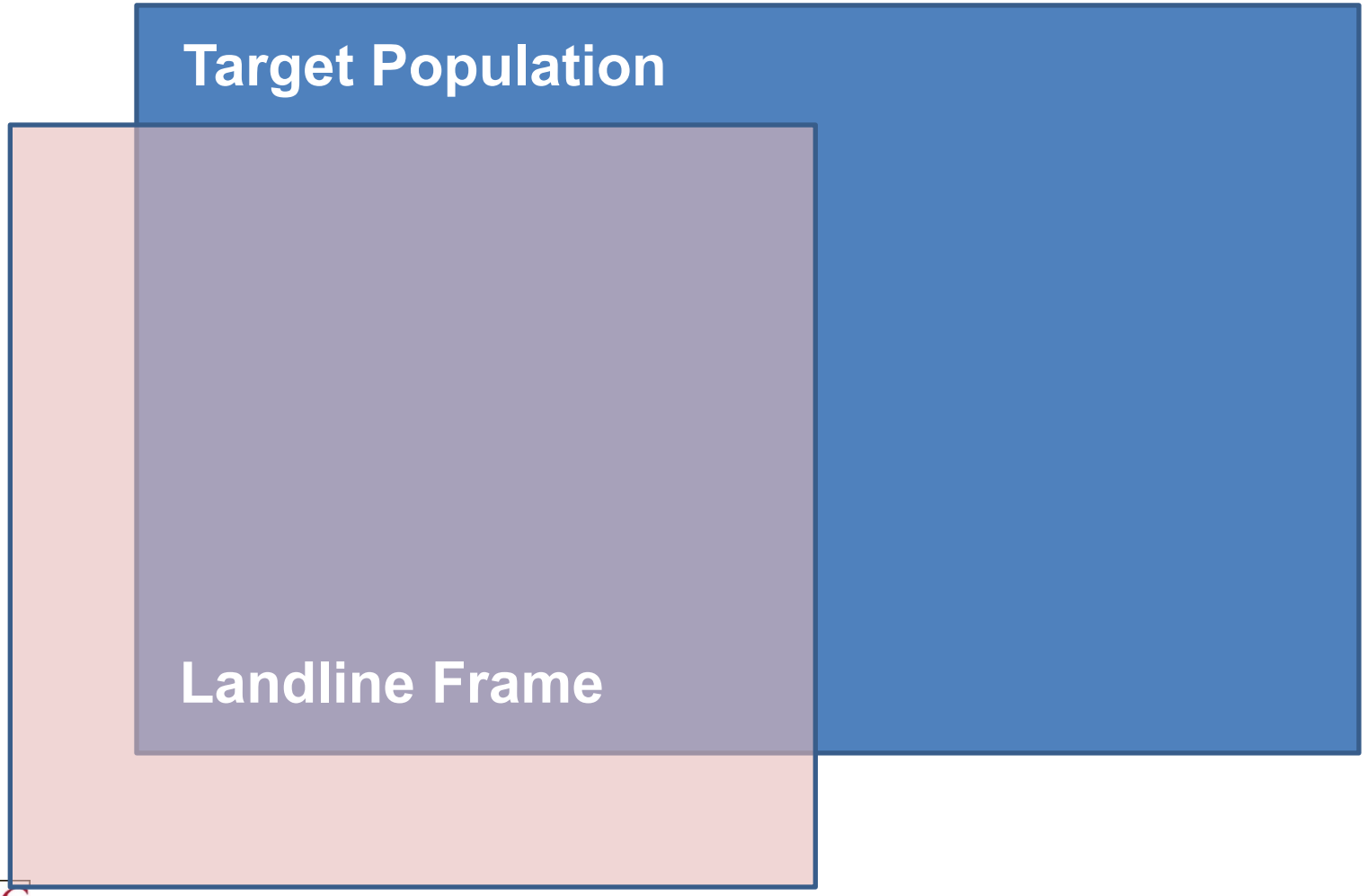
P = Element in the population
F = Unit in the frame

Coverage: Types of problems (continued)

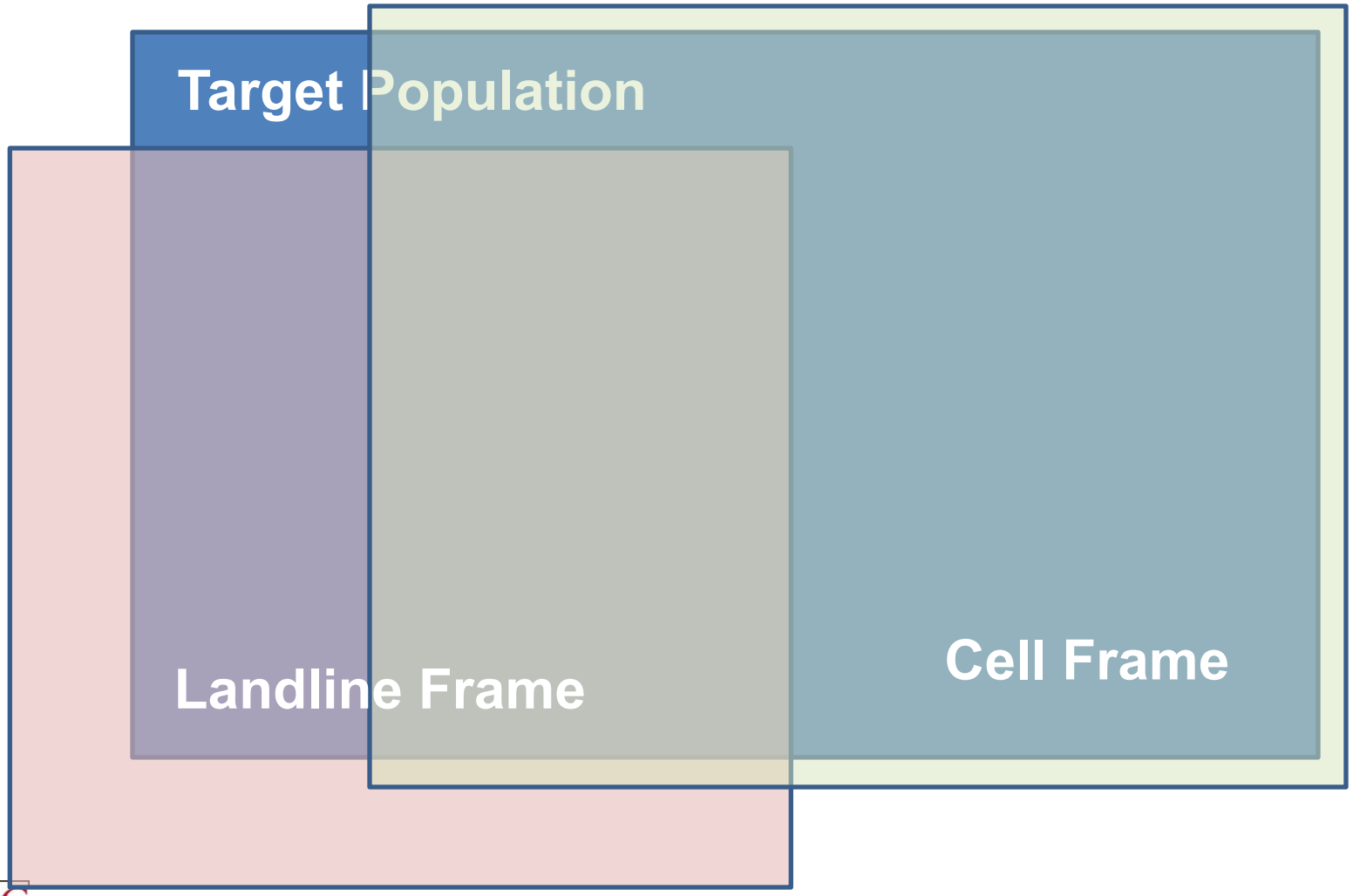
Survey Population	Sampling Frame	Label	Example	Solution
		Clustering	Population of individuals, address frame	Weight
		Foreign elements or blanks	Household population, address frame with vacant units	Discard blanks Increase sampling fraction

Sampling Frames and Coverage

Sampling Frames and Coverage



Sampling Frames and Coverage



Our Population – Overview

- Faculty members at a hypothetical university
- We are interested in estimating information about the permanent (tenure and tenure-track) faculty at the university
- We asked the university for a list of faculty members with details about their department and contact information

Our Population

id	college	dept	rank	tenure	ten_tr~k	phone	email
870003	Business administration	Accounting and finance	5_emeritus	.	0_no	555-1085	hz468038
870005	Education and human sciences	Health and human performance	4_full	1_tenured	1_yes	555-1544	iq221345
870009	Forestry and conservation	Forest management	4_full	1_tenured	1_yes	555-1821	mt627980
870012	Health professions and biomedical sciences	Pharmacy practice	2_assistant	0_untenured	1_yes	555-1287	ra146414
870013	Humanities and sciences	Biological sciences	3_associate	1_tenured	1_yes		gy702988
870004	Humanities and sciences	Chemistry and biochemistry	4_full	1_tenured	1_yes	555-1557	bk495626
870010	Humanities and sciences	Economics	3_associate	1_tenured	1_yes	555-1760	op765675
870014	Humanities and sciences	Economics	2_assistant	0_untenured	1_yes	555-1048	zz731018
870001	Humanities and sciences	Geography	3_associate	1_tenured	1_yes		fx967666
870011	Humanities and sciences	Geosciences	4_full	1_tenured	1_yes		lo571783
870007	Humanities and sciences	Modern and classical languages and literatures	4_full	1_tenured	1_yes		wj698230
870002	Law	Law	4_full	1_tenured	1_yes	555-1097	kz215349
870008	Law	Law	4_full	1_tenured	1_yes	555-1144	
870006	Visual and performing arts	Theatre and dance	1_adjunct	.	0_no	555-1258	ds537851

Useful Stata Commands

- Browse
  - **browse** varname
- Tabulate
  - **tab** varname
- Duplicates
  - **duplicates report** varname
- Codebook
  - **codebook** varname