

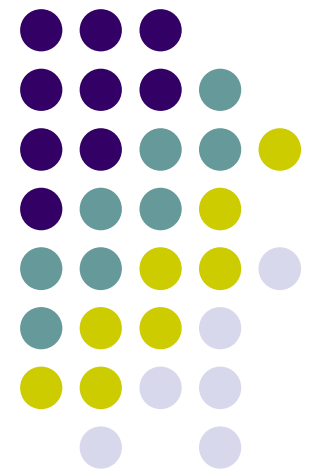
ISSUES RELATING TO METHODS FOR ANALYSIS OF SURVEY DATA

David Binder

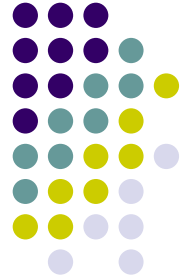
(in collaboration with Georgia Roberts)

Statistical Society Of Ottawa
Société Statistique d'Ottawa
2006 Fall Seminar

November 9, 2006



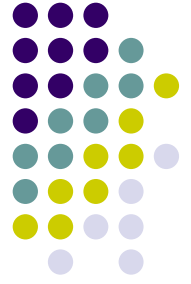
CRITICAL ASPECTS OF DATA ANALYSIS



Appropriateness of methods for analysis depends on:

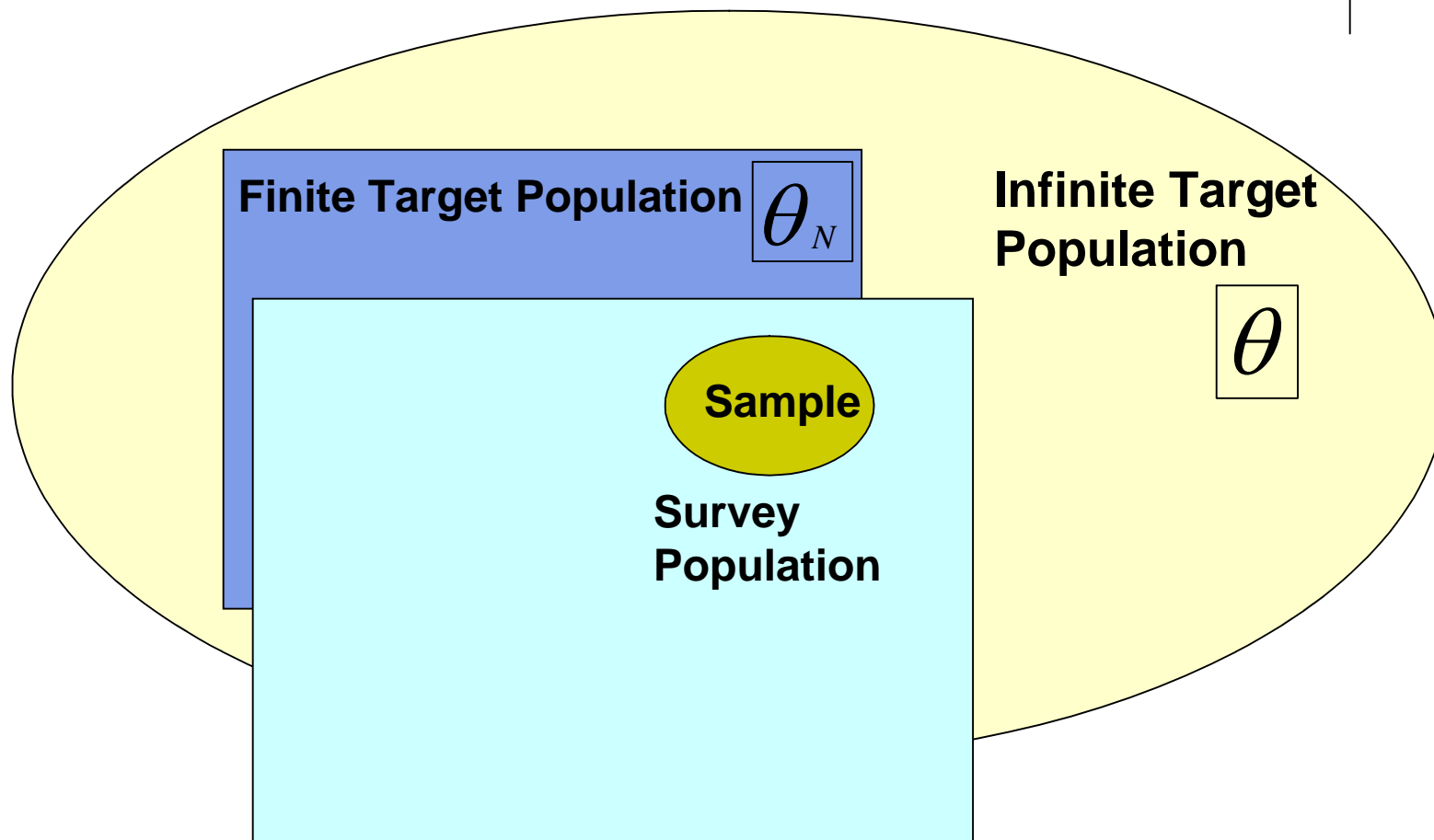
- The target population(s)
- The survey population
- The survey design

TARGET POPULATION OF AN ANALYSIS



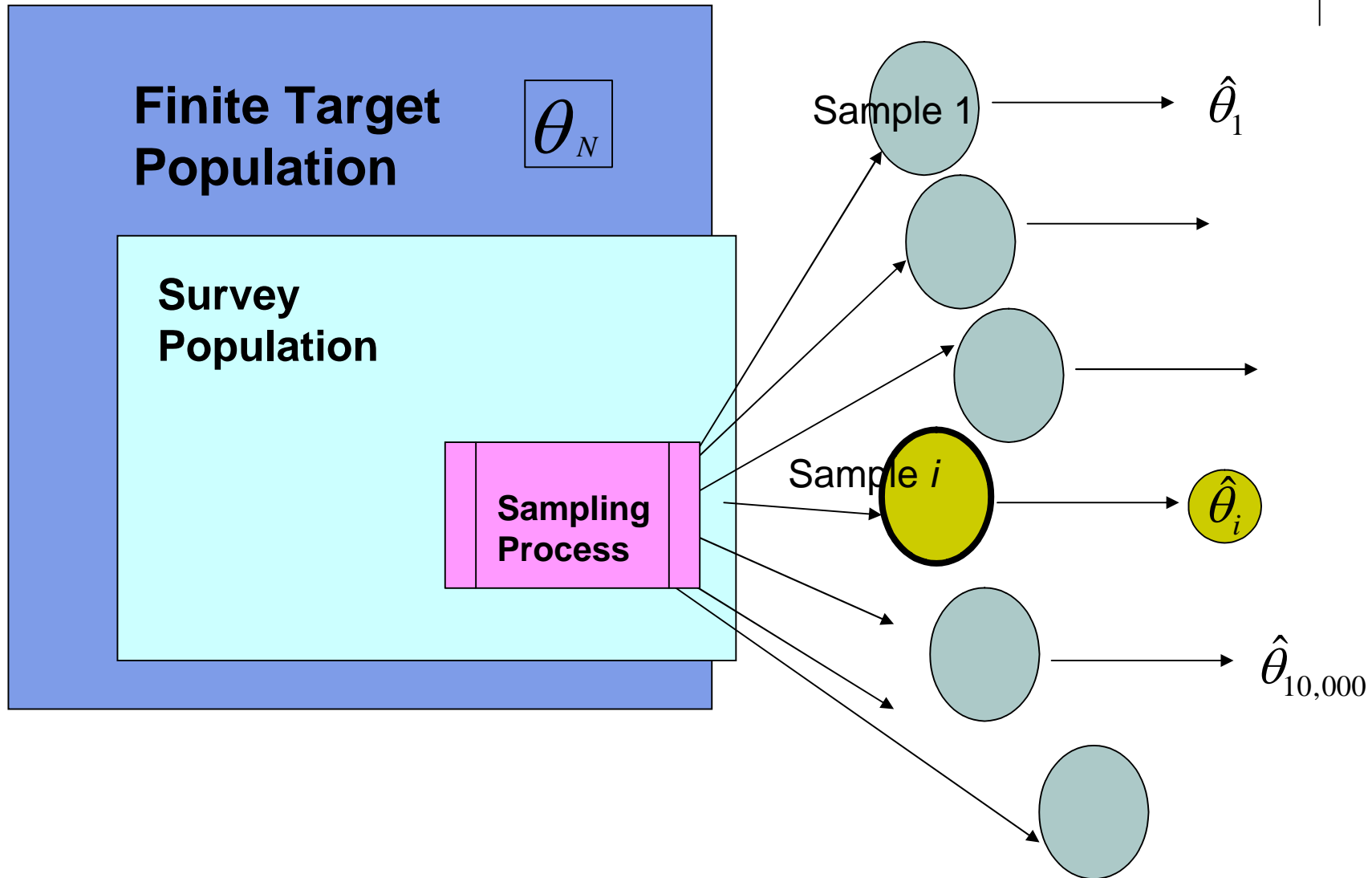
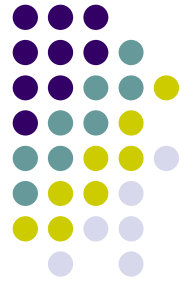
- The **target population** is the population about which the researcher wishes to make conclusions.
- The **survey population** consists of all the units that are eligible for selection through the frame and survey design being used.

TARGET POPULATION OF AN ANALYSIS



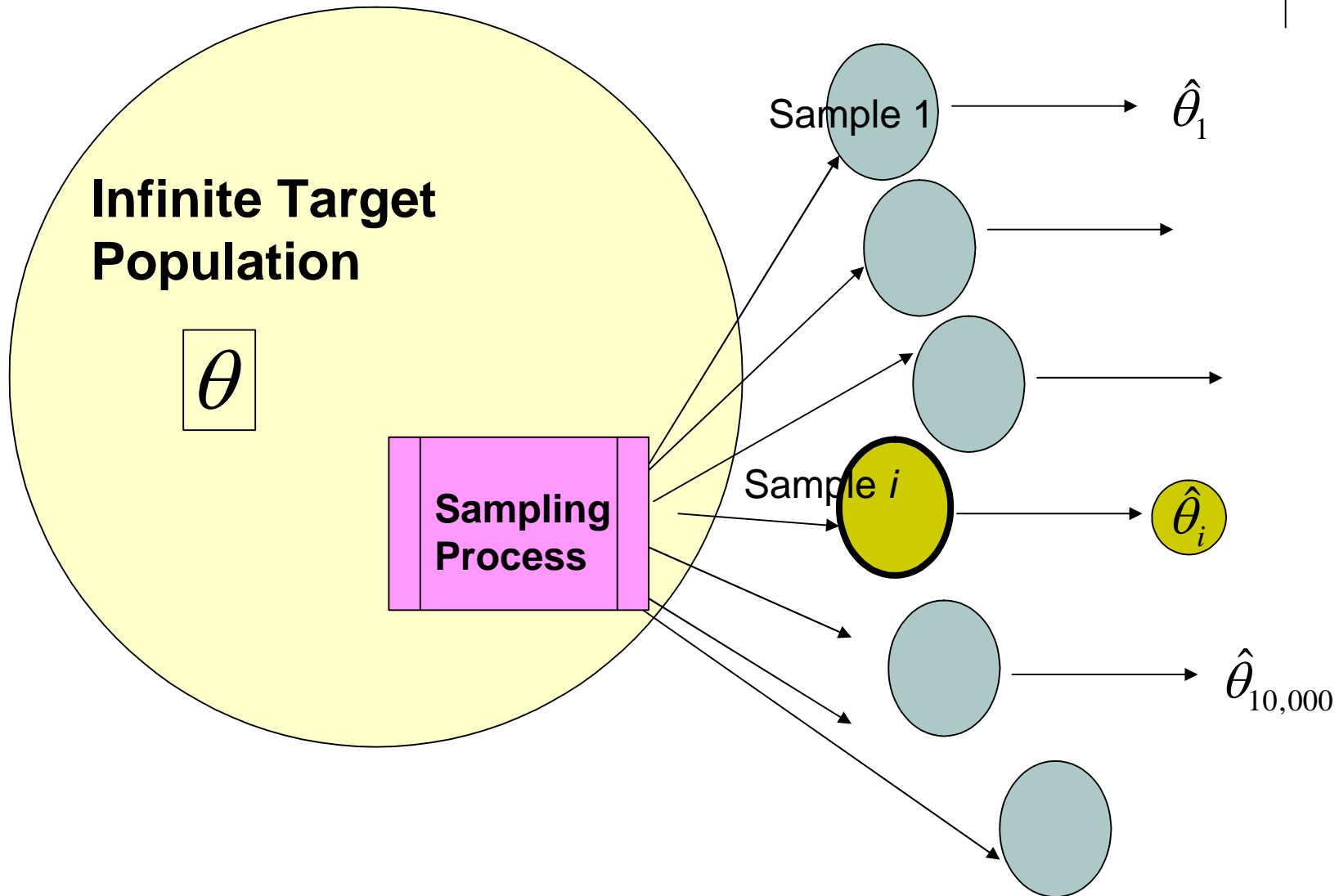
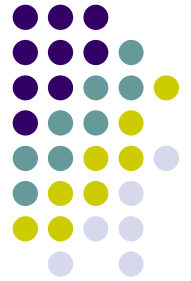
PRINCIPLES FOR MAKING STATISTICAL CONCLUSIONS

Finite target population – **design-based** randomization



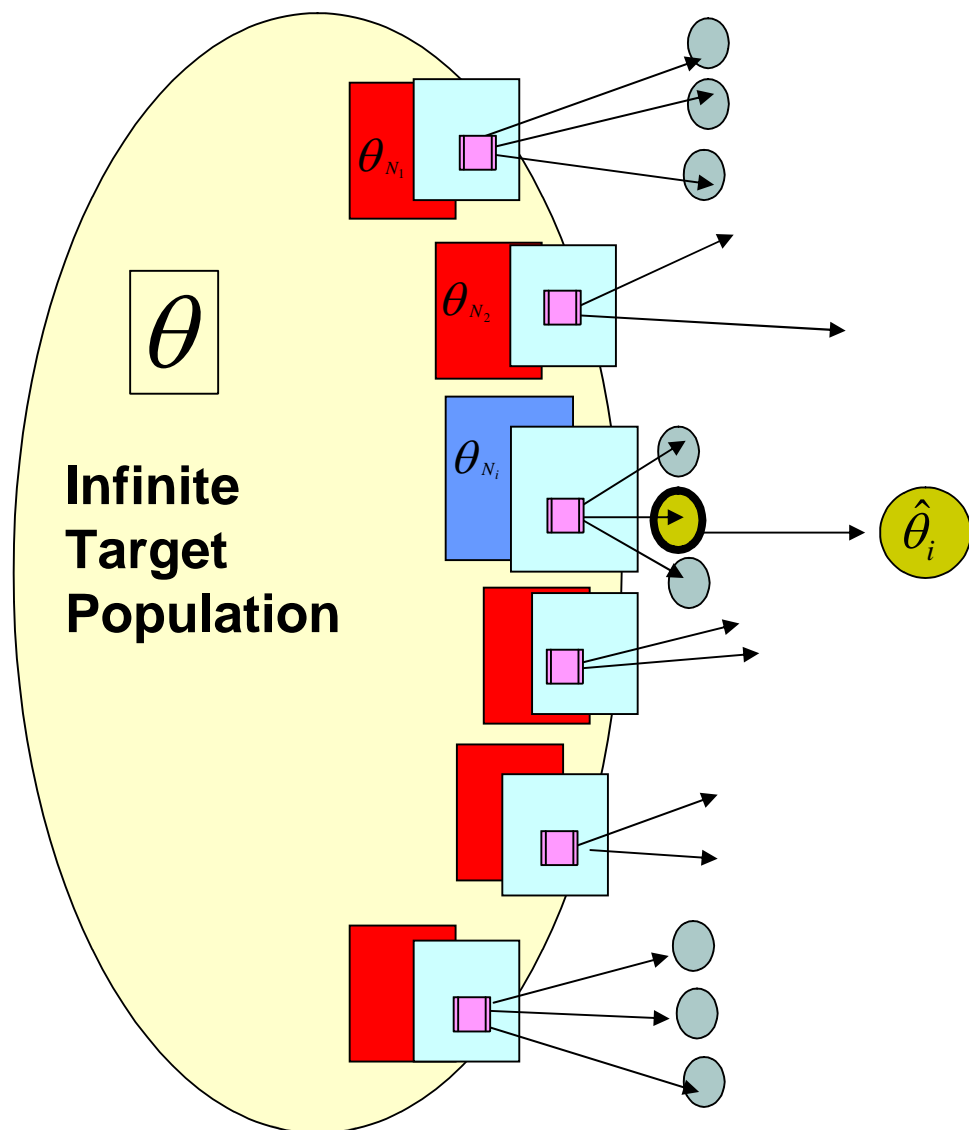
PRINCIPLES FOR MAKING STATISTICAL CONCLUSIONS

Infinite target population – **model-based** randomization



PRINCIPLES FOR MAKING STATISTICAL CONCLUSIONS

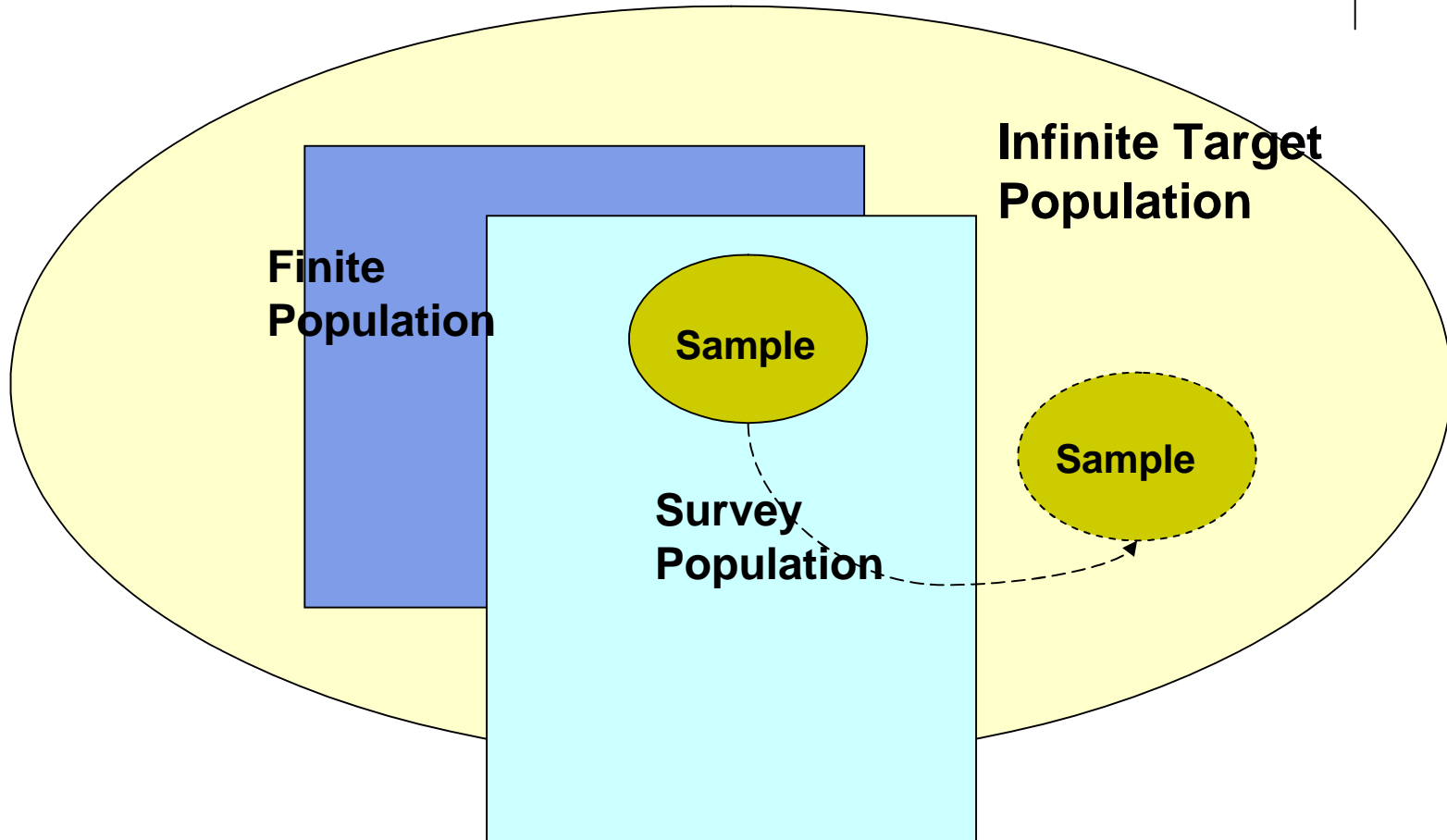
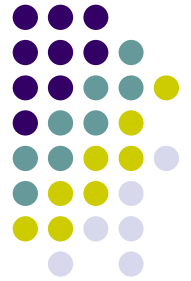
Infinite target population – **model-design-based** randomization



- Finite populations are generated from the infinite population.
- Randomization for estimator is based on both the model and the design.

PRINCIPLES FOR MAKING STATISTICAL CONCLUSIONS

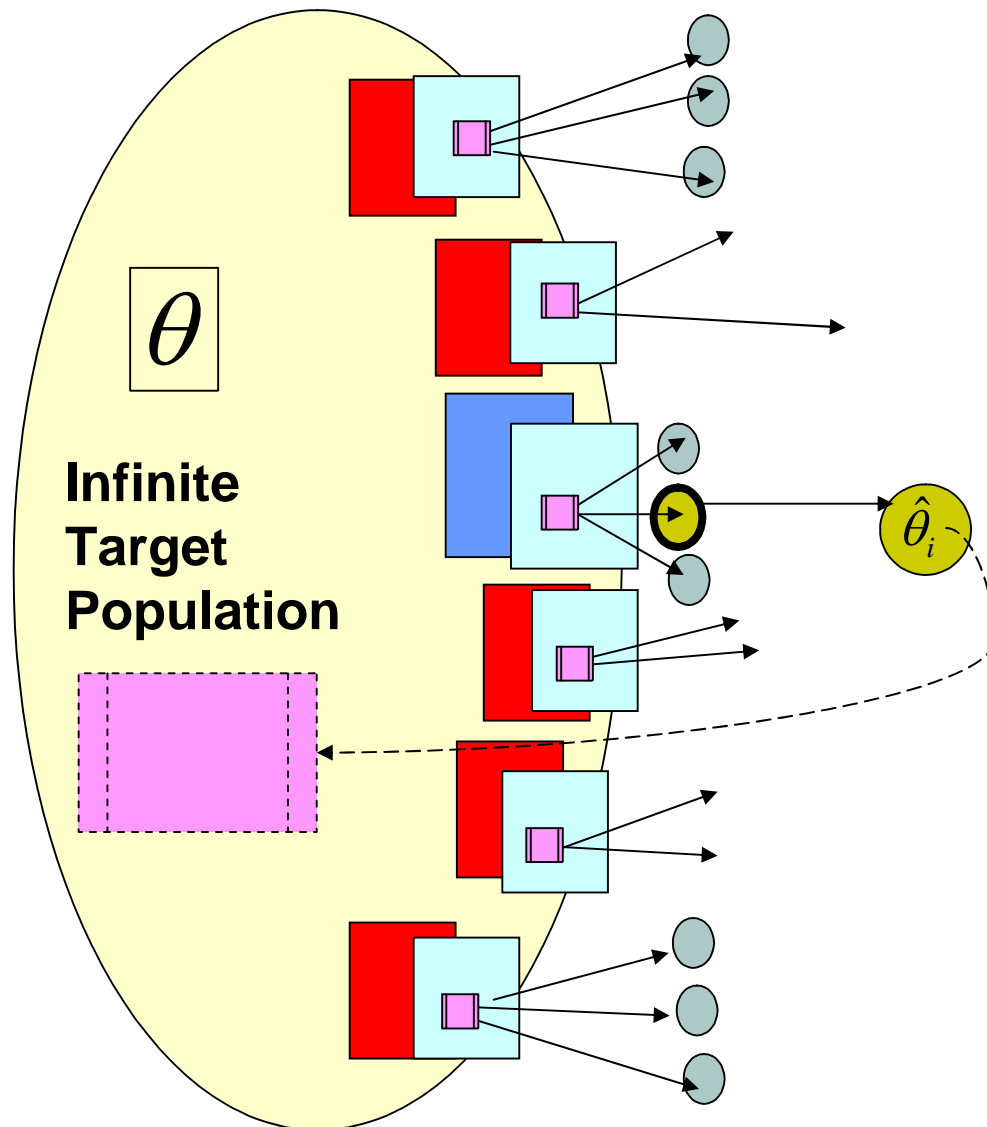
Infinite target population – **informative sampling**



When the sample can be assumed to have been generated from the model, the sampling is not informative. Otherwise, it is informative.

PRINCIPLES FOR MAKING STATISTICAL CONCLUSIONS

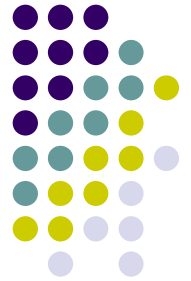
Infinite target population – **ignorability**



- When, for a particular analysis, model-based methods for inference are valid under the model-design randomization, the sampling is **ignorable**. Otherwise, it is non-ignorable for that analysis.

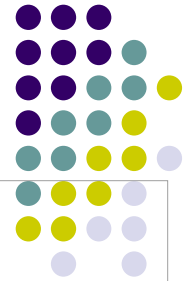
- Diagnostics are available to assess this.

WHAT DESIGN-BASED OPTIONS EXIST FOR VARIANCE ESTIMATION?

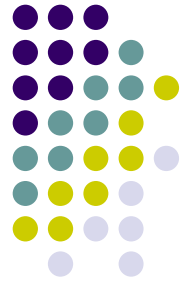


- **Analytic**
(also called linearization or Taylor or delta)
- **Replication**
 - Jackknifing (invalid results for percentiles)
 - Balanced Repeated Replication (BRR)
 - Survey Bootstrapping
 - Other methods

THE DECISION ON WHICH APPROACH TO USE



	Assumed model is valid	Assumed model is not valid
Model -based	<ul style="list-style-type: none"> • Consistent • Efficient • Valid variance estimates • Valid inferences • May be best 	<ul style="list-style-type: none"> • May be inconsistent if sample design is informative or if model means and/or variances are incorrectly specified • Variance estimates may be invalid • Inferences may be invalid
Design -based	<ul style="list-style-type: none"> • Consistent • May be inefficient • Valid variance estimates • Valid inferences 	<ul style="list-style-type: none"> • Valid variance estimates if model means are correctly specified • Consistent for a finite population quantity (FPQ) associated with the model • Valid inferences for the FPQ for large sample sizes

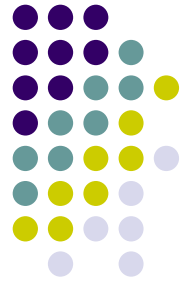


SUMMARY

When a researcher has identified an issue and wishes to examine survey data to answer his questions, there are a number of things to consider, including:

- his target population**
- his quantities of interest**
- survey design(s) and population(s)**
- suitable randomization assumptions**
- possible analytic approaches**
- possible design-based variance estimation methods (depending on nature of available survey design information)**

7. SOME REFERENCES

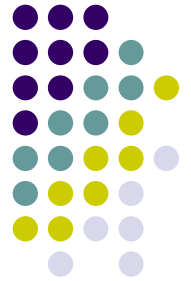


Binder, David A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279- 292.

Binder, David A. and Georgia R. Roberts (2001). Can informative designs be ignorable? *Newsletter of the Survey Research Methods Section, Issue 12*, American Statistical Association.

Binder, David A. and Georgia R. Roberts (2003). Design-based and Model-based Methods for Estimating Model Parameters. In *Analysis of Survey Data*, (eds. R.L. Chambers and Chris Skinner) Wiley, Chichester, 29-48.

7. SOME REFERENCES



- Chambers, R.L. and Chris Skinner, eds. (2003) *Analysis of Survey Data*. Wiley, Chichester.
- Korn, Edward L. and Barry I. Graubard (1999). *Analysis of Health Surveys*. New York: Wiley
- Lehtonen, Risto and Erkki J. Pahkinen (1995). *Practical Methods for Design and Analysis of Complex Surveys*. England: Wiley.
- Lohr, Sharon (1999). *Sampling: Design and Analysis*. Duxbury Press.
- Skinner, C.J., D. Holt, and T.M.F Smith. (1989). *Analysis of Complex Surveys*. New York: Wiley
- Thomas, D. Roland (1993). Inference using complex data from surveys and experiments. *Canadian Psychology* 34, 415-431.