Design Effect

$$Deff = \frac{S^2_{cluster} / a}{S^2_{SRS} / n} = [1 + roh(b - 1)]$$

SRS

Cluster

a # of sampled clusters from A number of population clusters
n sample size b # of sampled elements / cluster

Rate of Homogeneity and Design Effect

Roh=0 — *Roh*=1

Deff=1 — *Deff*=b

Using the Design Effect (Deff) to Estimate Sample Size

Let: $n = 1150$ $A = 87$ (# of clusters in the population)
$a = 46$* $b \approx n/a = 25$ (sample size per strata)
*minimum of 2 clusters per strata (23 strata)

- If design effect is known, use it to calculate sample size for clustered design.

- To calculate the cluster sample size for this example, multiply the stratified random sample size by the design effect (3).

  $n_{cluster} = n_{STR} * deff$

  For example: $n_{cluster} = 1150*3 = 3450$

4.02 Multiple Answer Poll

Which of the following statements are true? (Choose all that apply.)

a. The design effect (*deff*) for a survey item is related to the intra-class correlation
b. Design effects can be calculated prospectively for a survey item of interest.
c. When calculating sample size for a clustered design, you find the sample size for an SRS (or stratified design) and then use the *deff* as a multiplier.
d. Large design effects are desirable for survey items.

4.02 Multiple Answer Poll – Correct Answers

Which of the following statements are true? (Choose all that apply.)

a. The design effect (*deff*) for a survey item is related to the intra-class correlation

b. Design effects can be calculated prospectively for a survey item of interest.

c. When calculating sample size for a clustered design, you find the sample size for an SRS (or stratified design) and then use the *deff* as a multiplier.

d. Large design effects are desirable for survey items.

Probability Proportional to Size (PPS) Sampling

π_k

π_k

π_k

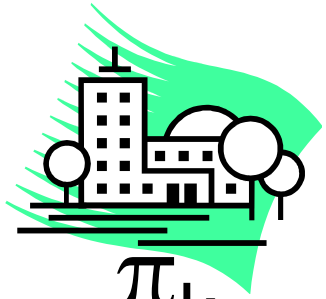
π_k

π_k

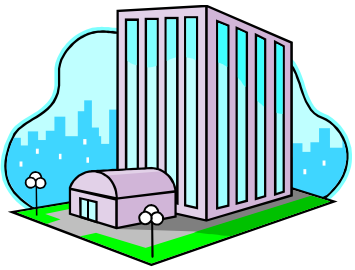
π_k

π_k

π_k



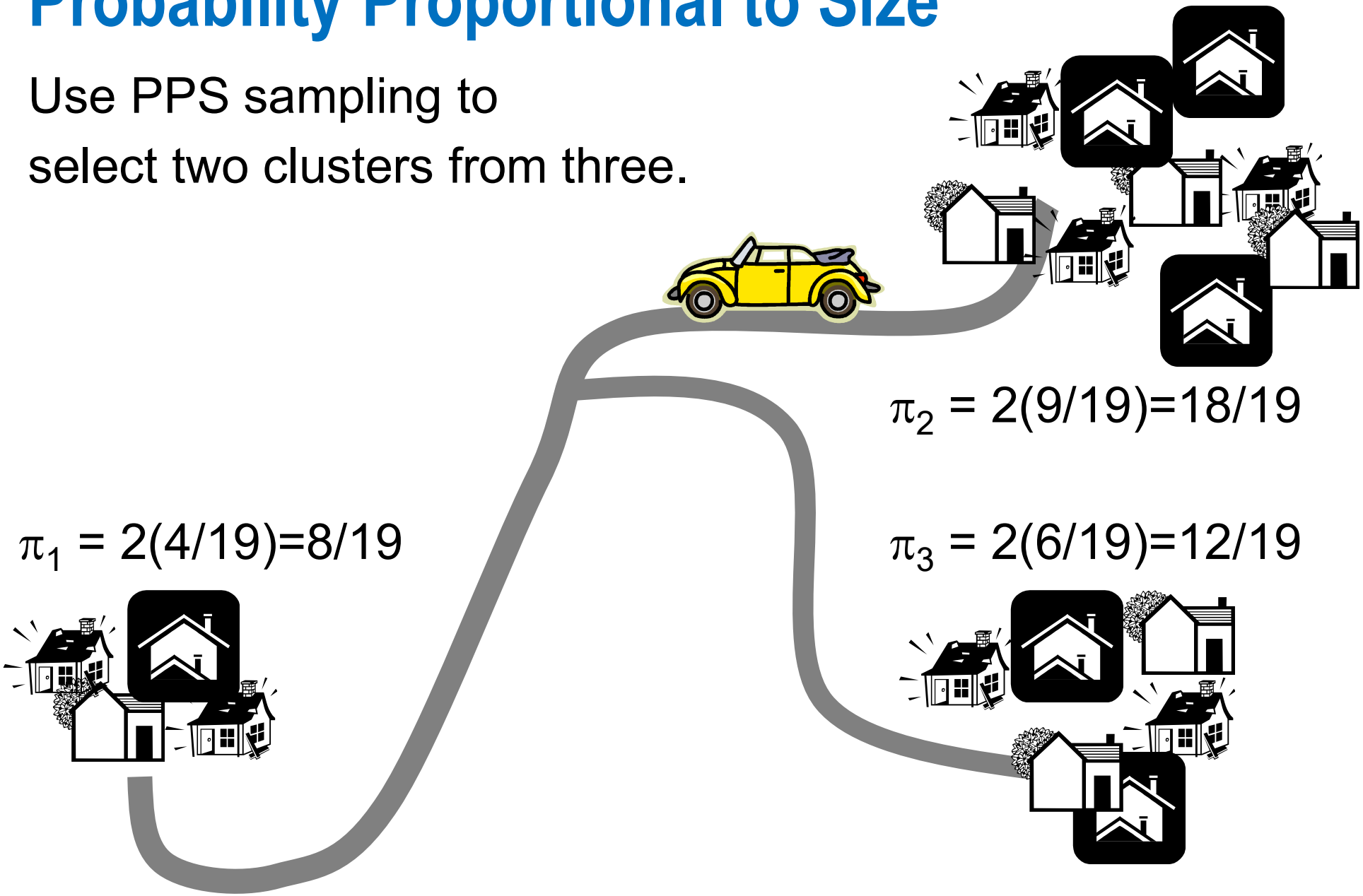
π_k



π_k

Selecting Clusters of Households with Probability Proportional to Size

Use PPS sampling to select two clusters from three.



$\pi_1 = 2(4/19)=8/19$

$\pi_2 = 2(9/19)=18/19$

$\pi_3 = 2(6/19)=12/19$

Selecting Equal Number of Households from the Selected Clusters

Use SRS to select three households from the selected clusters.

$\pi_1 = 8/19$

$\pi_{j\in 1} = 3/4$

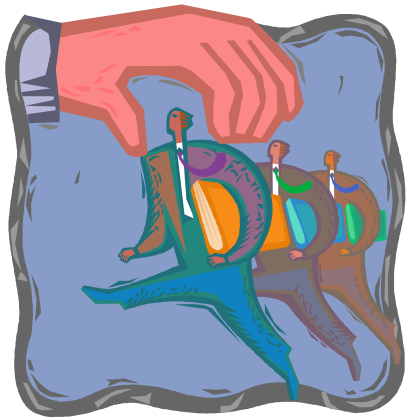
$\pi_{j1} = 3/4*8/19 = 6/19$

$\pi_2 = 18/19$

$\pi_{j\in 2} = 3/9$

$\pi_{j2} = 3/9*18/19 = 6/19$

Sequential Sampling



- In *sequential sampling*, each unit's selection is determined individually, and in order.
- Each unit is *visited* and its sample membership is determined independently from the other units.
- Two sequential methods are available.
  - Sequential with equal probabilities
  - Sequential with probability proportional to size (also known as Chromy's method)

Sequential Sampling

Sequential sampling is appropriate in the following situations:

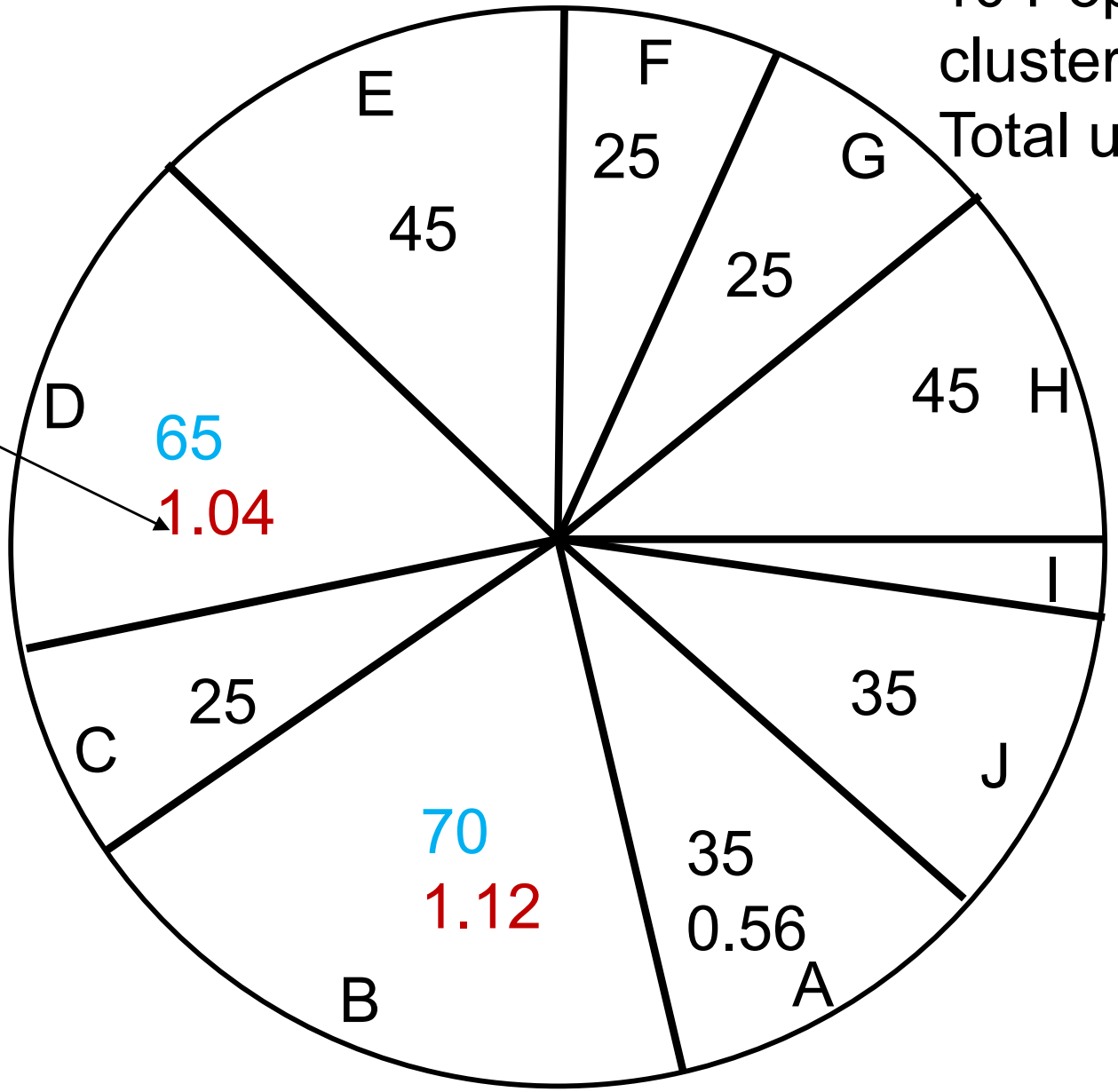
- when the sampling frame is extremely large
- when implicit stratification is desired
- when PPS with minimal replacement is desired

Chromy's Circular Zone Selection.

10 Population clusters.
Total units=375

Sampling Units 6

| A |



Expected number of hits= (6 * cluster size) / total

Chromy's Circular Zone Selection

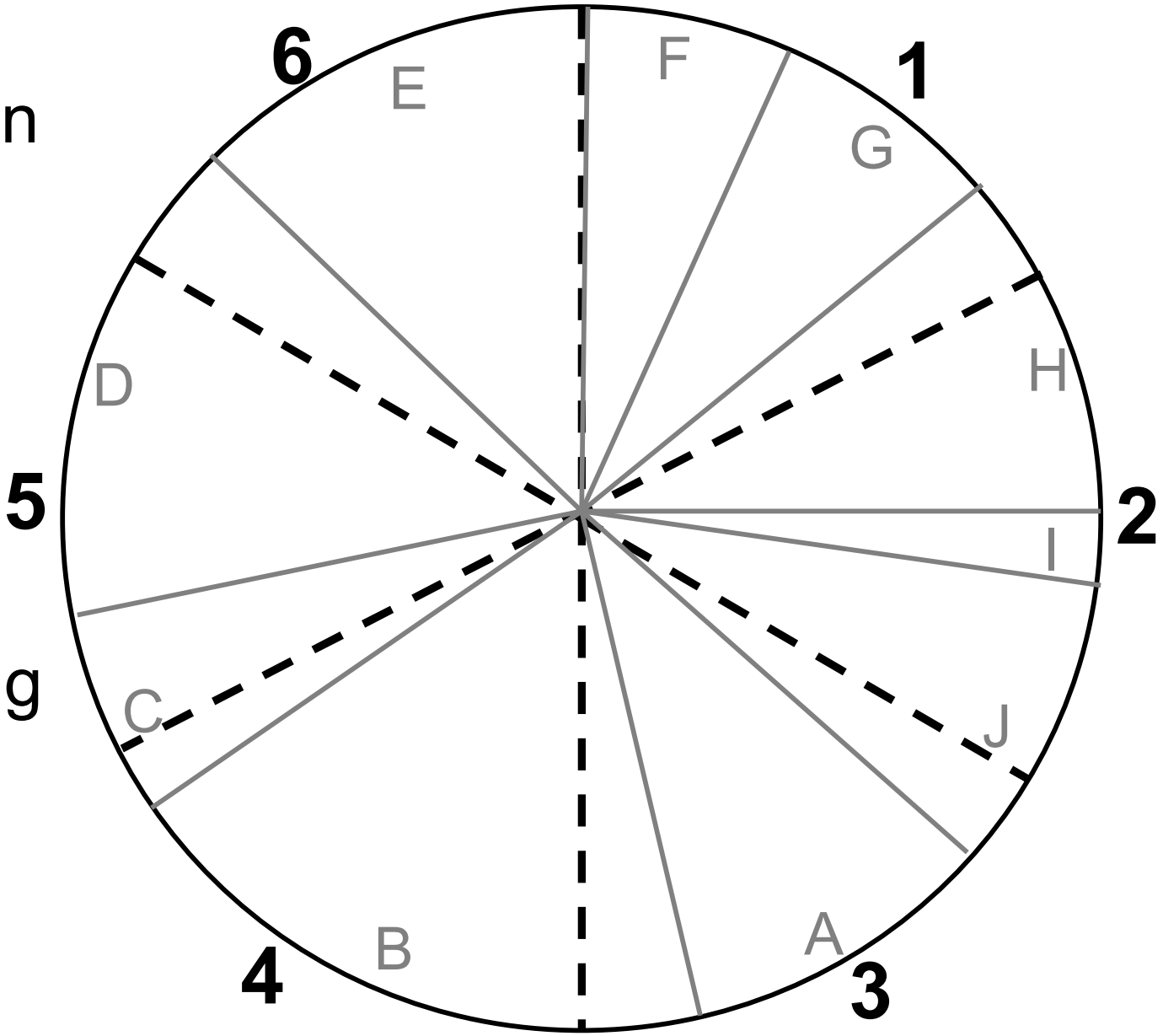
Selection Zones

1

n = 6

Sampling Units

A



Possible Effect of Cluster Sampling Designs on Estimation

4.03 Multiple Choice Poll

Sampling methods that result in equal sampling weights include which of the following?

a. SRS
b. STR with a constant sampling rate
c. Cluster sampling with PPS sampling at the first stage followed by an SRS of constant size at the second stage
d. a and b only
e. a, b, and c

4.03 Multiple Choice Poll – Correct Answer

Sampling methods that result in equal sampling weights include which of the following?

a. SRS

b. STR with a constant sampling rate

c. Cluster sampling with PPS sampling at the first stage followed by an SRS of constant size at the second stage

d. a and b only

e. a, b, and c

Two-Stage Cluster Household Survey

This demonstration illustrates the concepts discussed previously.

4.04 Poll

Only the Taylor linearization variance estimation method accounts for the sample design when calculating variances.

- ❍ True
- ❍ False

4.04 Poll – Correct Answer

Only the Taylor linearization variance estimation method accounts for the sample design when calculating variances.

- ○ True
- ○ False

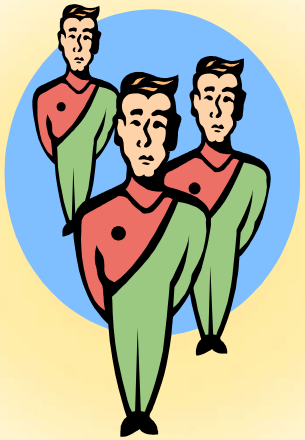
All three methods (Taylor linearization, BRR, and delete-1 jackknife) account for the sample design when calculating variances.

Variance Estimation for Complex Surveys

Three variance estimation methods are available in the SAS SURVEY procedures:

Taylor Series Linearization

$$g(y) = g(\theta) + \sum_{j=1}^{k} g'(\theta)(y_j - \theta_j) + \text{Remainder}$$



Balanced Repeated Replication

Jackknife