SAMPLING

SYRACUSE UNIVERSITY
School of Information Studies

SAMPLING

Why sampling?

Sampling when obtaining or analyzing the entire set of data of interest is too expensive or time consuming,

And the sample is representative, meaning, it has approximately the same property (of interest) as the original set of data.

Therefore, the analysis results on the sample data may be reliably generalized to the entire data set.

SAMPLING METHODS

A sampling task: To sample 300 college students in Syracuse University to study their social media use patterns

Convenience sampling:

Sample iSchool students to represent SU students.

Random sampling:

Randomly sample students around campus.

Stratified sampling:

Sample equal number of students from each school.

Systematic sampling:

E.g., sort students' SUID numbers in increasing order; pick the 1st, 11th, 21st, 31st, … , students until 300 students are sampled.

SAMPLING WITH AND WITHOUT REPLACEMENT

Sampling without replacement:

One item would occur in the sample at most once.

Sampling with replacement:

One item may occur in the sample multiple times.

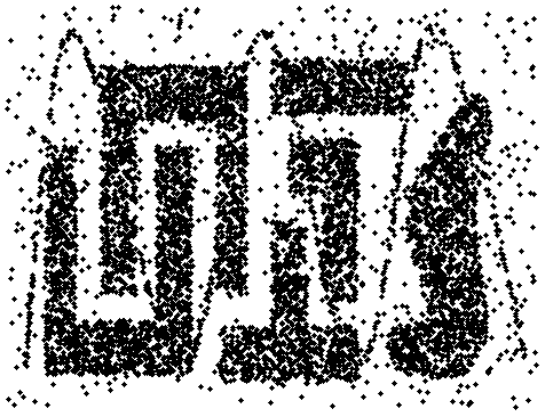
RANDOM SAMPLING

```
> sample <- titanic[sample(1:nrow(titanic), 100, replace=FALSE),
]
> table(titanic$Survived)

  0   1
549 342
> table(sample$Survived)

 0  1
62 38
```

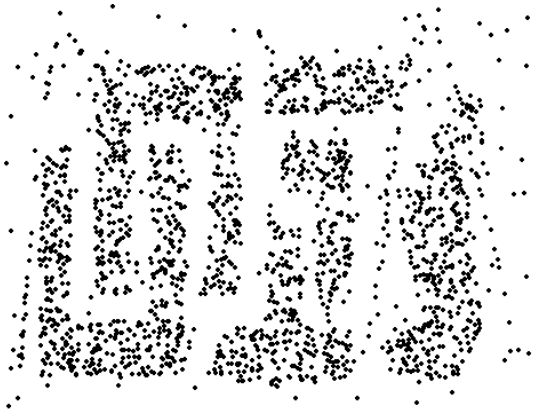
SYSTEMATIC SAMPLING

```
> #sample lines #1, #11, #21, ...
> ss=titanic[seq(1, nrow(titanic),10),]
> nrow(ss)
[1] 90
```

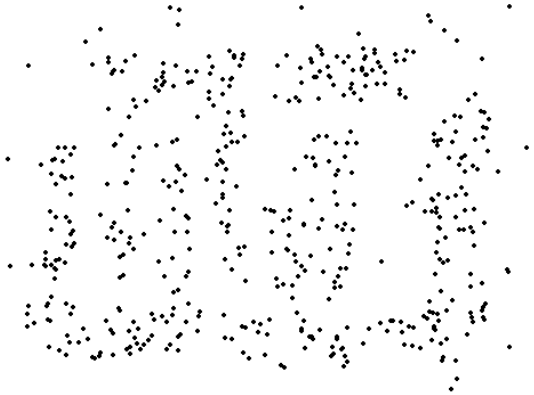
SAMPLE SIZE



8,000 points



2,000 Points



500 Points

A REVIEW OF SAMPLING METHODS

Convenience sampling

Random sampling

Stratified sampling

Systematic sampling

Sample representativeness