

# Statistical Studies

## Observational Study

- No treatment imposed
- No cause and effect relationship can be concluded
- Includes surveys

## Experiment

- Treatment imposed
- Cause and effect relationship can be inferred

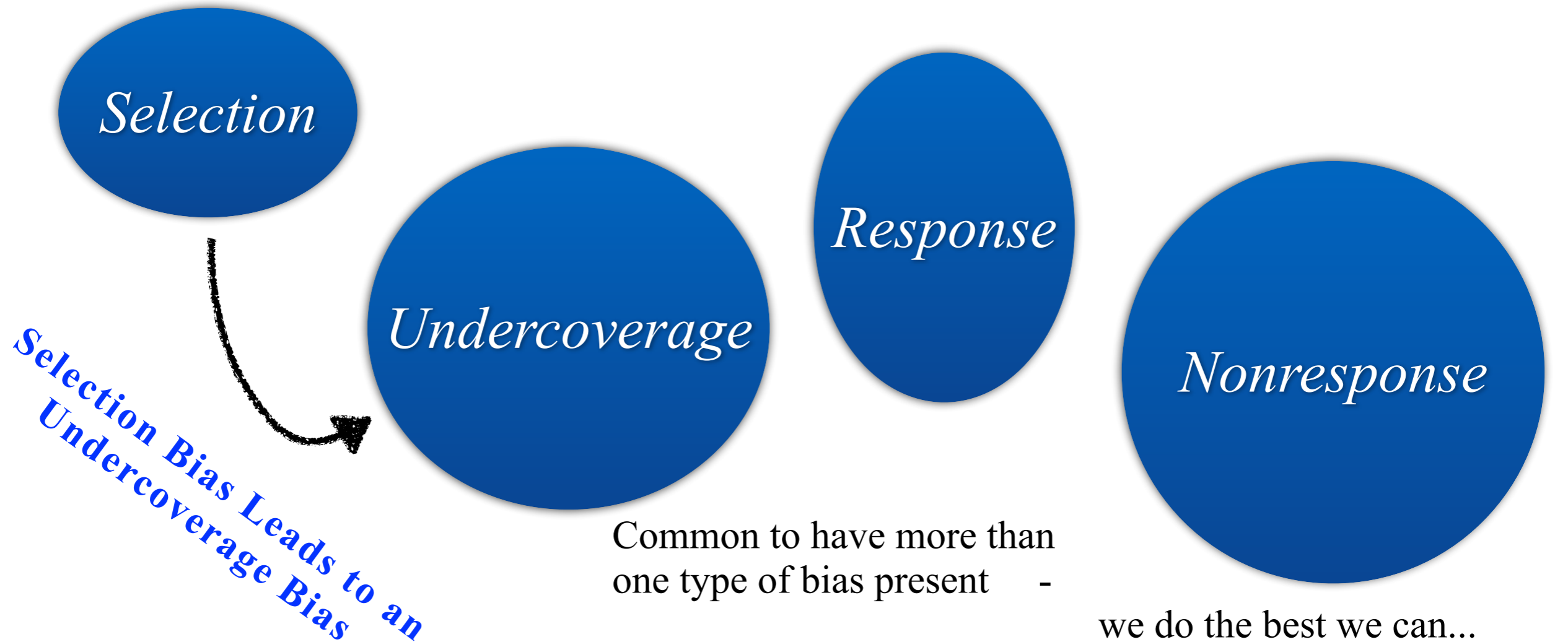
When designing a study or experiment, we always need to look for sources of *bias*

Bias in statistics is not necessarily personal. It can be anything that would adversely affect or *confound* the results of the study or experiment.

*Bias* occurs when certain responses are systematically favored over others

A sample is *biased* if it systematically over-represents or under-represents a segment of our population of interest

# Types of Bias



Common to have more than one type of bias present -

we do the best we can...  
but once bias is present,  
adding more observations to  
the sample will not erase  
the bias

# Types of Bias



*Selection*

**Non-random  
sampling  
methods:**

Samples chosen for convenience, using voluntary response, or which miss out on a segment of a population



*Undercoverage*

**Undercoverage Bias:** A sampling scheme that fails to sample from some part of the population or that gives part of the population less representation than it has in the population.

Problems in the data gathering instrument or process can lead to *Response Bias*

# Types of Bias



*Selection*

*Undercoverage*

*Response*

## **Voluntary Response Bias:**

When a sample is comprised solely of volunteers (people who chose to participate), the sample will typically not be representative of the population.

**Nonresponse Bias:** This occurs in a sample design when individuals selected from the sample fail to respond, cannot be contacted, or decline to participate.

**Wording Bias:** When questions are misleading or confusing, or when there is a component of “social desirability” in the question (where the respondent feels social pressure to respond a certain way).

## **Self-Reported Responses:**

When people self-assess and report the results such as under or over estimating how much junk food one consumes, there is often significant bias.

# Types of Bias

*Selection*

*Undercoverage*

*Response*

*Nonresponse*

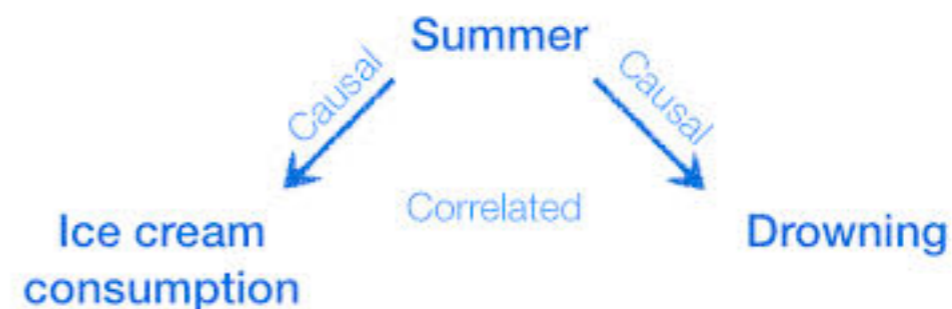
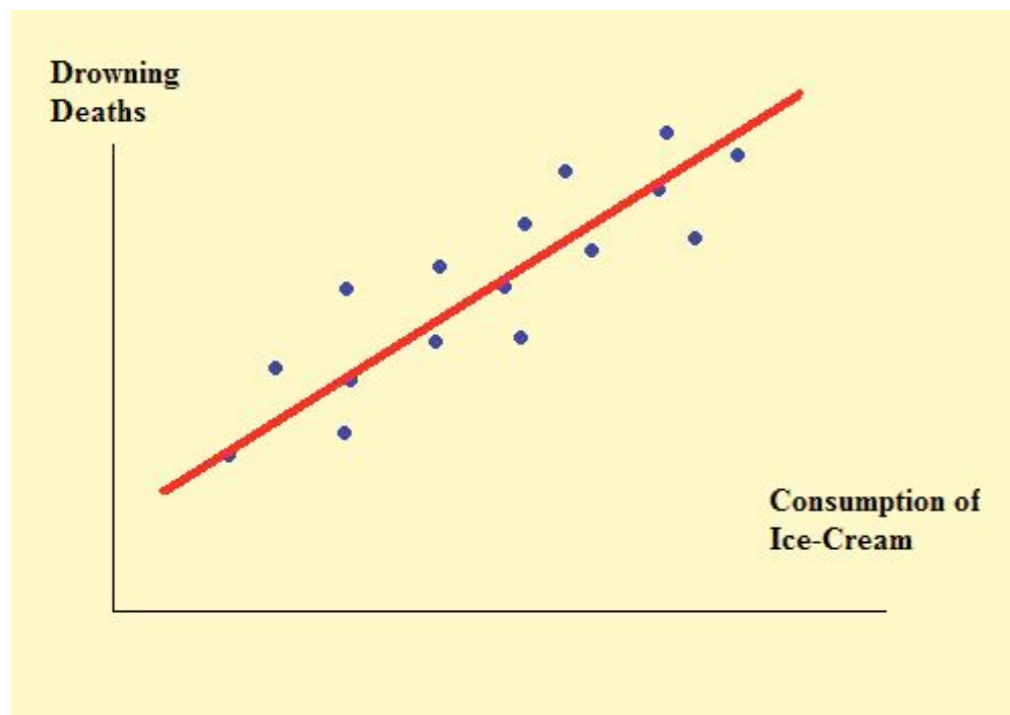
**Non-response bias:** Subjects do not or can not respond such that the results of the study are significantly affected

## Vocab ...Yup, more vocab

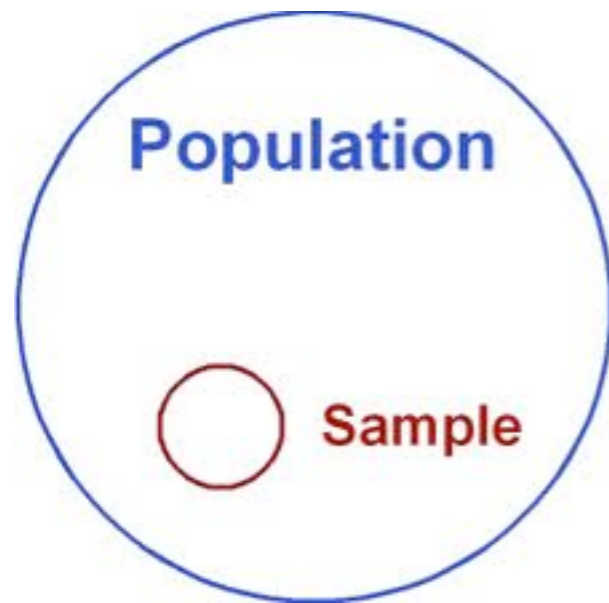
**Confounding variable** - also **confounding factor**, **hidden variable**, **lurking variable**, a **confound**, or **confounder** - is an **extraneous variable** in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable.

### Example - Ice Cream Consumption and Number of Drownings

These two variables are positively related, i.e. as ice cream consumption rises, so does the number of drownings. Does ice cream consumption cause drownings?

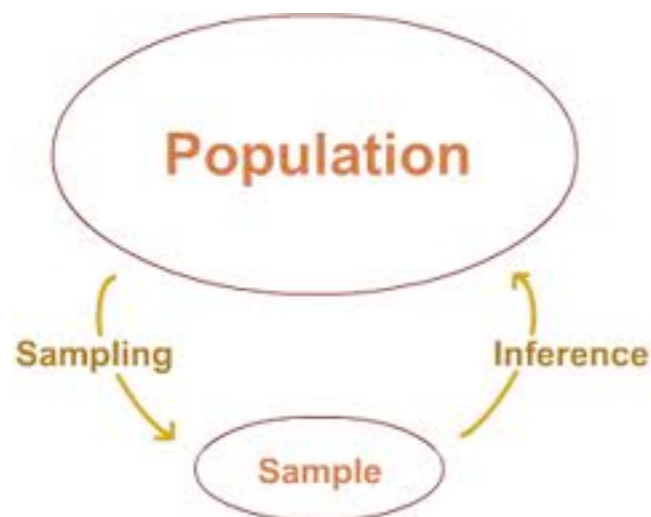


Remember...



A subset of the population.

We randomly select our sample so that we can generalize our findings to the population.



But ... we randomly assign treatments to subjects to reduce *bias* and variation.

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

```
randInt(1,5,6)
00000
00000
00000
04000
24000
40000
00004
00002
00004
00004
00002
00002
00001
00004
00004
00002
00001
00004
00004
00001
00004
00001
00004
00004
00004
00004
```

