

Chapter 3: Producing Data

3.2 Samples

- Population is to parameter as sample is to statistic.
- *e.g.* μ in a population corresponds to \bar{x} in a sample
- Types of samples: convenience, voluntary, random

Non-Random Samples

- (a) web survey on music of the 1970s on itunes (selection bias)
- (b) mutual fund returns (survivor bias)

Historical Example

- During World War II, the statistician Abraham Wald was asked the following question.
- Of returning bombers 75% had holes in the wings and 25% had holes in the tails.
- So shall we reinforce the wings?

- Wald's answer:
- Where do you think the holes were located in the bombers that did not return?

Random Samples

- With random sampling, all members of the population have an equal chance to be selected.
- How can you randomly sample from a small group?
- From a large group use Excel:
`=round(rand()*100,0)+1` gives integers between 1 and 100.

- Cautions: non-response bias, response bias, wording effects, undercoverage.
- *Stratified* random sampling is reweighting *e.g.* using census data

3.3 Designing Experiments

- What is a *controlled* experiment?
- The main idea is to defeat lurking/confounding variables to isolate the cause and effect.
- The cause sometimes is called the *treatment*.

- *e.g.* the alleged Mozart effect
- The treatment group listens to music composed by Mozart for 20 minutes. The control group does nothing.
- They then write a test and the treatment group does better.
- Is this a valid control group? Does it isolate the Mozart effect?

- Please define these terms:
 - (i) random assignment
 - (ii) placebo effect
 - (iii) double blind
 - (iv) matched pairs

- There is enormous growth in RCTs (randomized control trials) in education, development economics, and other fields.
- It is important to know how they work.
- A key question is scalability: Can they be expanded to a population?