Experiments vs. Observational Studies

In an **experiment** investigators apply treatments to experimental units (people, animals, plots of land, etc.) and then proceed to observe the effect of the treatments on the experimental units.

In a **randomized experiment** investigators control the assignment of treatments to experimental units using a chance mechanism (like the flip of a coin or a computer's random number generator).

1

Experiments vs. Observational Studies (cont.)

In an **observational study** investigators observe subjects and measure variables of interest without assigning treatments to the subjects. The treatment that each subject receives is determined beyond the control of the investigator.

For example, suppose we want to study the effect of smoking on lung capacity in women.

2

Experiment

- Find 100 women age 20 who do not currently smoke.
- Randomly assign 50 of the 100 women to the smoking treatment and the other 50 to the no smoking treatment.
- Those in the smoking group smoke a pack a day for 10 years while those in the control group remain smoke free for 10 years.
- Measure lung capacity for each of the 100 women.
- Analyze, interpret, and draw conclusions from data.

3

Observational Study

- Find 100 women age 30 of which 50 have been smoking a pack a day for 10 years while the other 50 have been smoke free for 10 years.

- Measure lung capacity for each of the 100 women.

- Analyze, interpret, and draw conclusions from data.

4

Fisher's Hypothesis

- Suppose there is a gene that causes smoking to appear to be a very pleasurable experience.

- Suppose the same gene also causes emphysema, lung cancer, throat cancer, etc.

- People who have the gene will be more likely to smoke than people who do not have the gene.

- People who have the gene will be more likely to get emphysema, lung cancer, throat cancer, etc.

5

Fisher's Hypothesis (cont.)

- So is it really smoking that causes health problems? Maybe it is just the gene?

- A **confounding** variable is related both to group membership and to the outcome of interest. Its presence makes it hard to establish the outcome as being a direct consequence of group membership.

6

Always Randomize if Possible

Consider a field experiment intended to compare the yield of two corn varieties (A and B).

Suppose the field is divided into 20 plots that run from one end of the field to the other.

Is there anything wrong with the following assignment of varieties to field plots?

A B A B A B A B A B A B A B A B A B A B

7