# DAY 2 – OCTOBER 7
SELECTION BIAS

Selection bias

- An example:

- *The "supported work program" provides individuals with severe employment problems with work experience of a year or so, under conditions of gradually increasing demands, close supervision, and work in association with a crew of peers.*

- *The guiding principle is that by participating in the program, a significant number of people who are severely handicapped for employment may join the labor force, cease socially-destructive or dependent behavior, and become self-supporting.*
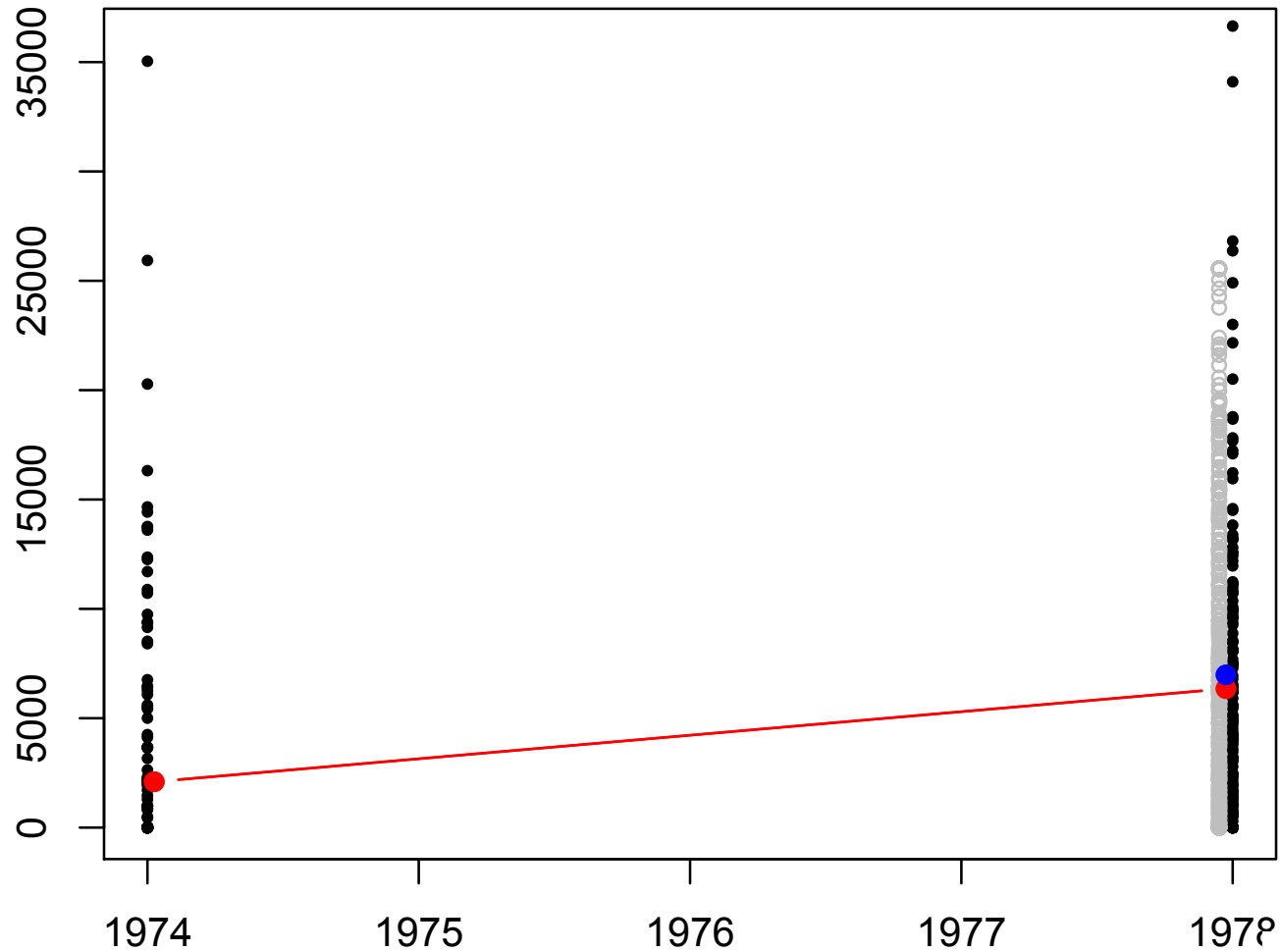
Annual Income Before and After Program
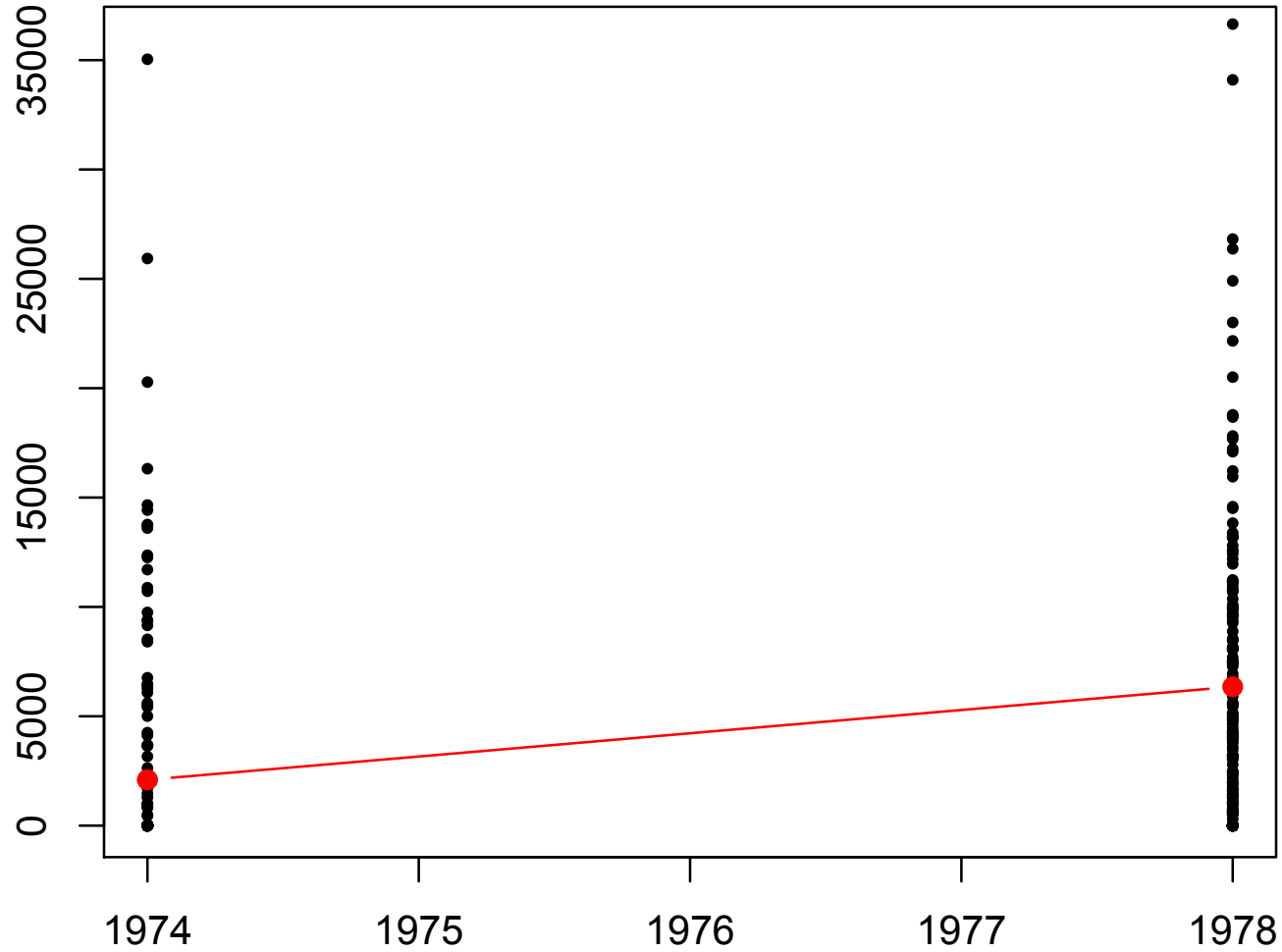
Selection bias

- Can we claim success?

- Looking at this in isolation may be misleading.

- We have a survey of incomes in the population.

- What if we used this a basis for comparison?
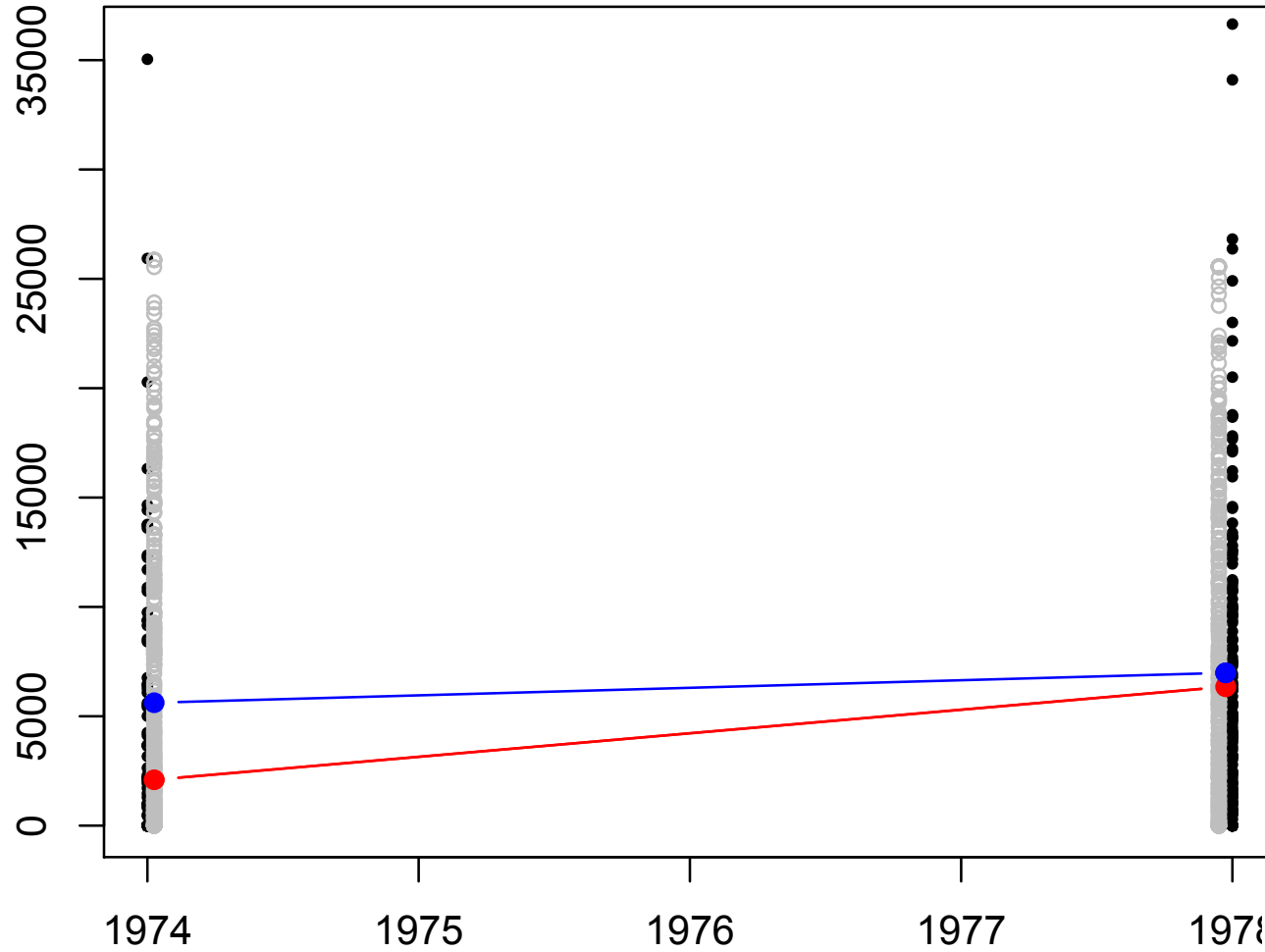
- For example, compare post-program incomes…

Annual Income Before and After Program

Selection bias

- So post-program incomes were about equal to the overall average.

- Can we claim success yet?

- May be misleading to ignore trends…

Annual Income Before and After Program

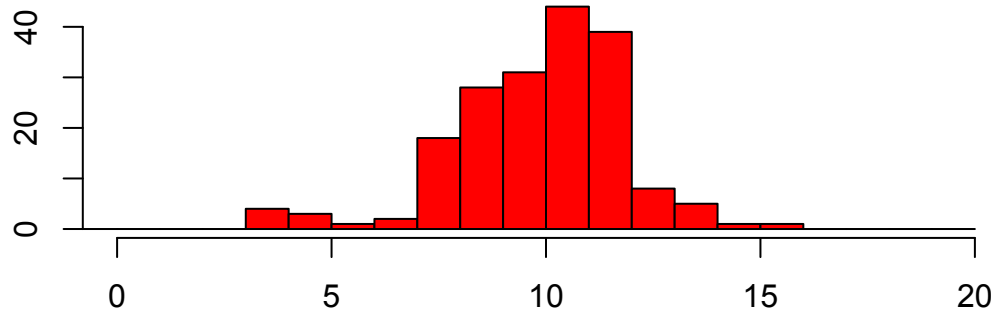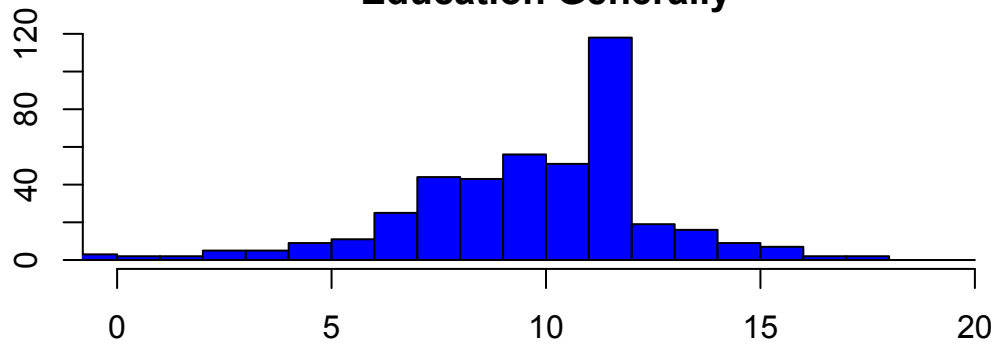# Selection bias

- Looks encouraging, but…

- …what should we make of baseline differences?

- …should we assume comparable trends?

- …do the trends form the survey provide a valid counterfactual for our beneficiaries?

- …how comparable are these groups, really?

Beneficiaries' Age
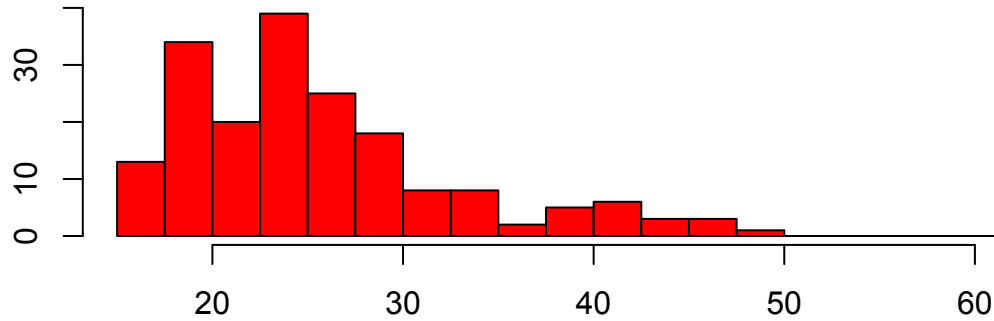
Age Generally
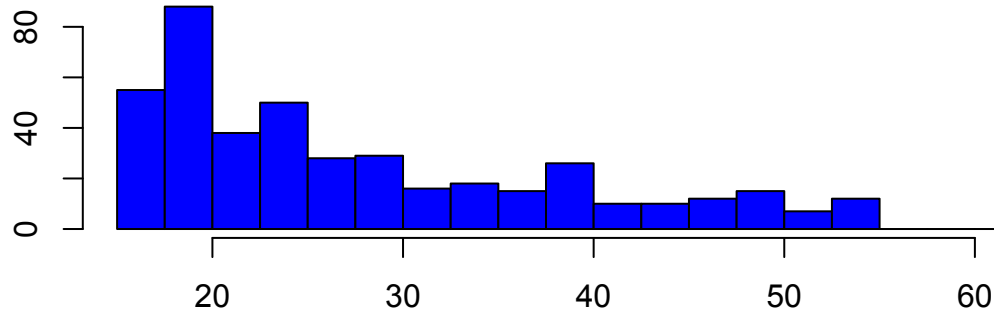
Selection bias

- Substantial differences in age, education, and 1974 income

- Who knows what else might differ???

- Differences threaten validity of the comparison.

- Not clear that we have constructed a valid counterfactual for our beneficiary group.

- Thus, there is the potential for selection bias.

Selection bias

- "Selection on observables":
  - Key differences between beneficiaries & comparison group based on characteristics that are *measured*.
  - Suppose two people have same measured characteristics but differ in their program status.
  - Selection on observables means that this is due only to "random" or "unimportant" reasons.
  - This is a key litmus test.

Selection bias

- "Selection on unobservables":

    – Some key things that may determine program status have not been measured.

    – Suppose two people have same measured characteristics but differ in their program status.

    – If the reason for this difference is not innocuous, then we have selection on unobservables.

    – Then, the comparison is haunted by "lurking variables" that "confound" estimation of impact.

Selection bias

- Supported Work example:
  - We have seen differences in measured characteristics (age, education, 1974 income).
  - These likely affect 1978 income *as well as* potential to benefit from program.
  - We could "control" for these factors.
  - But for people with same age, education, & 1974 income, why might some be in program & others not?
  - We need to know how beneficiaries were selected!

Selection bias

- Supported Work beneficiary selection:
  - Suppose beneficiaries selected by advertising project & taking people who showed up at registration event.
    - How might they differ from general population?
    - How might this bias the comparison?
    - Can we reliably measure all of these things?
  - Suppose beneficiaries selected by asking community leaders to choose "at risk youth" in their communities.
    - How might they differ from general population?
    - How might this bias the comparison?
    - Can we reliably measure all of these things?

Selection bias

- These are examples of…
  - "Self-selection": whether someone takes up program depends on anticipated benefit.
  - "Program placement selection": admission to program is based on anticipated benefit.

- Clearly these result in *selection bias* for naïve comparison to general population.

- Whether biases can be removed by accounting for the measured differences is unclear.

Dealing with selection bias

- Need to use experimental or quasi-experimental methods to cope with this; this is what has been meant by rigorous impact evaluation

- Experimental:
  - Randomized controlled trials.

- Quasi-experimental:
  - Control strategies (matching, regression, differences-in-differences).
  - "Natural experiments" and regression discontinuity designs that exploit haphazard events or arbitrariness in selection processes.

Internal validity

- **Remember, randomization ensures that**
- **– On average, those in the treatment group are the same as those in the control group**
- **– This means that the impact will be the same if either one of the groups were treated**

A successful randomized evaluation requires:

1. Those in the treatment group receive the treatment, while those in the control group do not

2. Those in the control group are not affected by the treatment in any way

3. Outcomes for members of both treatment and control are measured at the end of the program

Internal validity

- **Remember, randomization ensures that**
- **On average, those in the treatment group are the same as those in the control group**
- **This means that the impact will be the same if either one of the groups were treated**

A successful randomized evaluation requires:

1. Those in the treatment group receive the treatment, while those in the control group do not
2. Those in the control group are not affected by the treatment in any way
3. Outcomes for members of both treatment and control are measured at the end of the program

Internal validity

When these requirements aren᾽t fully met:

1) *Those in the treatment group receive the treatment, while those in the control group do not*

→ PARTIAL COMPLIANCE

Individuals assigned to the treatment group may not receive the program or those in the control group do receive it

2) *Those in the control groups are not affected by the treatment in any way*

→ EXTERNALITIES

Individuals or communities which did not receive the treatment may nonetheless be affected by it, either positively or negatively

3) *Outcomes for members of both treatment and control are measured at the end of the program*

→ ATTRITION

Members of the original sample drop out between the beginning of the program and the endline survey. The subsample may differ systematically from the remaining sample

Internal validity

When these requirements aren't fully met:

1) *Those in the treatment group receive the treatment, while those in the control group do not*

→ PARTIAL COMPLIANCE

Individuals assigned to the treatment group may not receive the program or those in the control group do receive it

2) *Those in the control groups are not affected by the treatment in any way*

→ "SPILL OVER" EFFECTS

Individuals or communities which did not receive the treatment may nonetheless be affected by it, either positively or negatively

3) *Outcomes for members of both treatment and control are measured at the end of the program*

→ ATTRITION

Members of the original sample drop out between the beginning of the program and the endline survey. The subsample may differ systematically from the remaining sample

External validity

- The ability to use the results of the experiment to know what will happen when we carry out the program more broadly in the future, in other communities, etc.

- Can we generalize from the people and communities who participated in the experiment to the people and communities we will be targeting in the future?

Internal versus External validity

- The focus thus far has been "internal validity":
  - Have you constructed a proper counterfactual for *your beneficiaries*?

- Another issue that arises is "external validity":
  - Assuming internal validity, are your beneficiaries representative of *the population that you plan to target for future programming*?
  - Example: vaccine studies often have to exclude children and elderly. But for whom are vaccines potentially most important?!

Exercise

- How are beneficiaries typically selected for your case study? List two to three examples.

- For each of these beneficiary selection processes, list possible sources of selection bias (threats to internal validity) if one were to try to measure program impacts by comparing to the broader population.

- Are there external validity problems for applying results of an evaluation of the program in your case study to Afghanistan (context, beneficiaries, etc.)?

- What kind of "counterfactual comparison" would you like to make to evaluate the impact of your program?