

Classical Experiments

- Most clear-cut example of how data are collected and used to test hypotheses
- Closest thing social scientists have to the labs of physical scientists
- Best for small-group research

Classical Experiments: Why important?

- 1) better understand the logic of all research designs
- 2) a strong research design for *inferring causality*

Classic experimental design

<u>Classic experimental design</u>		Time \longrightarrow			
		<u>Pretest</u>		<u>Posttest</u>	
Experimental group	R	O ₁	X	O ₂	O ₂ -O ₁ =d _e
Control group	R	O ₃		O ₄	O ₄ -O ₃ =d _c

Definitions

- R = random assignment
- O = observation
- X = experimental stimulus
(= independent var)
- **Randomization**: divides systematic biases between two groups

Decisions

- If $d_e > d_c$, then + relationship
- If $d_e < d_c$, then – relationship
- If $d_e = d_c$, then no relationship

Example: food aversion therapy (hypothetical)

- ✓ Smokers, food therapy
- ✓ *Hypothesis*: those who undergo food therapy eat less

Example: food aversion therapy

<u>Classic experimental design</u>		Time \longrightarrow			
		<u>Pretest</u>		<u>Posttest</u>	
Experimental group	R	$O_1=0$	X	$O_2=10$	$O_2-O_1=10$
Control group	R	$O_3=0$		$O_4=30$	$O_4-O_3=30$

Example: Manhattan Bail Project

- Bernard Botein, 1965, "The Manhattan Bail Project: Its Impact in Criminology and the Criminal Law Process," *Texas Law Review* 43:319-331.
- Initiated by Vera Institute, NYC
- Can people be safely released from jail prior to trial *without* bail
- Would default rate decline?

Manhattan Bail Project: methods

- Restricted to people accused of felonies and misdemeanors (nothing more serious)
- NYU law students and Vera staff members evaluated defendants' records re employment, family, residence, references, current charges, and previous records
- Randomly split *those recommended* into experimental and control groups
- Experimental group: recommended for pretrial release *without* bail
- Control group: released only *with* bail

Manhattan Bail Project: methods

X=?

Y=?

Hypo: $X \implies Y?$

Manhattan Bail Project: findings

- ✓ Between 1961 and 1964: less than 1% of the experimental group defaulted
- ✓ Considerably less than the control group
- ✓ $d_e < d_c$ (negative relationship)
- ✓ Pretrial release without bail reduced default rates

Pygmalion in the Classroom

- Robert Rosenthal and Lenore Jacobson, 1968, *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*, New York: Holt, Rinehart and Winston.
- Central idea: others' expectations may change behavior; self fulfilling prophecy?

Pygmalion in the Classroom: methods

- Sampled students in public elementary schools in lower class community (Oak School)
- Administered standard tests of intelligence
- Results to teachers: 20% identified “*intellectual bloomers*”
- **Reality:** 20% were randomly assigned to experimental group; rest to control group
- End of year: standard tests of intelligence again

Pygmalion in the Classroom: methods

- “Harvard Test of Inflected Acquisition”:
- Alias for standardized, test of intelligence: Flanagan’s Tests of General Ability (TOGA)
- Measure verbal ability and reasoning
- Called them “IQ tests”

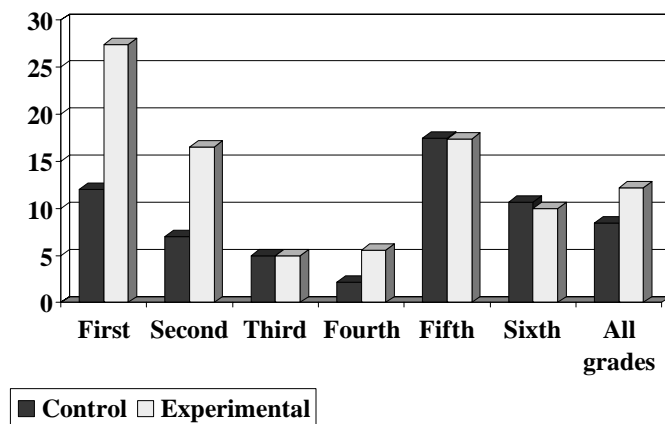
Pygmalion in the Classroom: methods

$X=?$

$Y=?$

Hypo: $X \implies Y?$

Pygmalion in the Classroom: Gains in IQ points, by grade



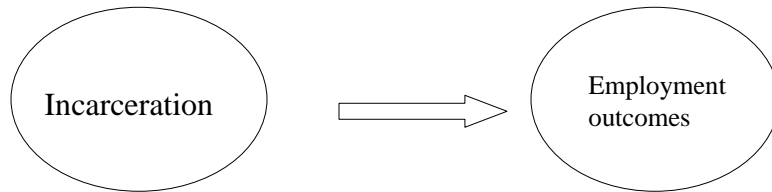
Experimental design:
audit technique

Devah Pager. 2003. "The Mark
of a Criminal Record."
American Journal of Sociology
108:937-75.

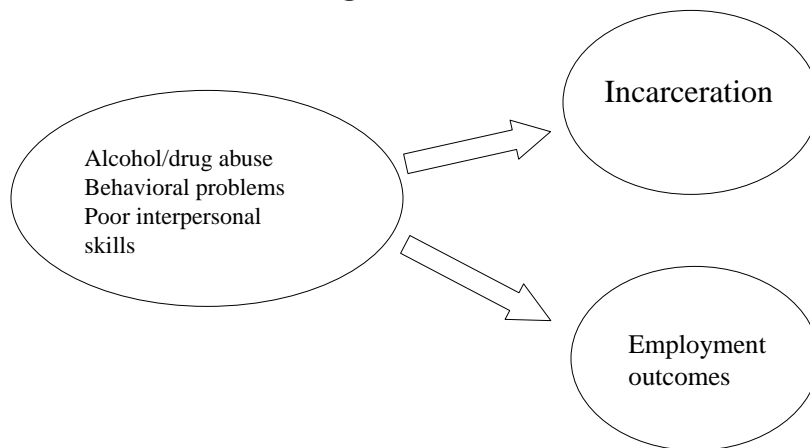
Pager, 2003: Theory

- Large and growing number of men with criminal careers (up 600% in three decades)
- U.S. has highest incarceration rate in the world
- What's effect on employment outcomes?

Pager, 2003



Pager, 2003



Pager, 2003: Audit Methods

- Pairs: two blacks and two whites
- Testers: 23-year old Milwaukee college students
- “All other things equal”: education, work experience, self presentation
- One of each race pair given criminal record
- Answered 15 job openings per week (white pair did 150 “employer audits,” black pair 200)
- Question: how often did they get callbacks

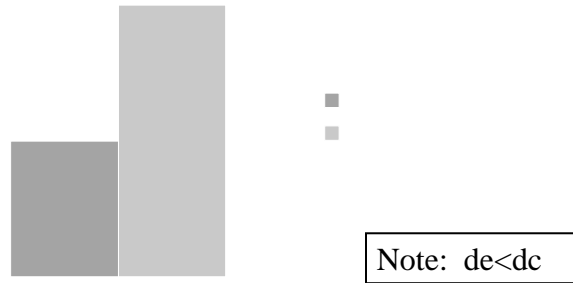
Pager, 2003

$X=?$

$Y=?$

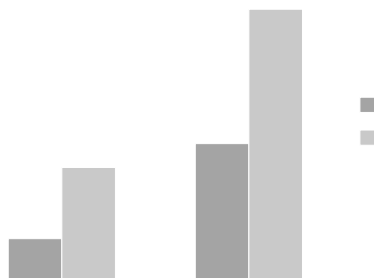
Pager, 2003: findings

Percentage Called Back, for Whites



Pager, 2003: findings

Percentage Called Back, by Race



Use of control group

- Why use control group: “R O₃ O₄”
- Separate effects of independent variable (experimental stimulus) from effect of test itself (e.g., teachers’ expectations vs. IQ test)
- Hawthorne test: effect of test itself

Hawthorne Effect

Fritz Roethlisberger and William Dickson,
1939, *Management and the Worker*,
Cambridge, MA: Harvard University Press.

“Hawthorne Effect”:
why do you need a control group?

- Analysis of working conditions and worker satisfaction in Western Electric Works, Hawthorne, Illinois
- Improved working conditions → improvement in satisfaction and productivity
- Improved working conditions again → yet another increase in satisfaction and productivity
- Is there a relationship?

“Hawthorne Effect”:
why do you need a control group?

- To substantiate their claims, they dimmed the lights
- Whoops! → satisfaction and productivity improved again!
- What’s going on?

“Hawthorne Effect”:
why do you need a control group?

- Workers responding to study, not improved working conditions
- Attention to them → increased satisfaction and productivity

Internal validity:
problems in experiments

Did the independent
variable really *cause* the
dependent variable?

Internal validity:

12 problems leading to internal invalidity

- | | |
|---------------------------|---|
| 1) history | 8) causal time order |
| 2) maturation | 9) diffusion or imitation of treatments |
| 3) testing and retesting | 10) compensation |
| 4) instrumentation | 11) compensatory rivalry |
| 5) statistical regression | 12) demoralization |
| 6) selection biases | |
| 7) experimental mortality | |

Internal validity

- Classic experimental design: guards against history, maturation, testing, instrumentation, statistical regression, selection bias, experimental mortality
- Rest (8-12) handled through careful administration of design (keep 2 groups separate)

External validity: problems in experiments

- ***External validity***: can the results be generalized beyond experiment?
- Question: how representative is sample?
- Question: does the artificial nature of the experiment affect generalizability?

Solomon 4-Group Design

- Classic experimental design is weak on external validity (generalizability)
- Handle via Solomon 4-Group Design
- Described as useful for:
 - 1) Addressing generalizability
 - 2) Addressing external validity
 - 3) Addressing the effects of the pretest

Solomon 4-group design

<u>Solomon 4-group design</u>		Time \longrightarrow		
		<u>Pretest</u>		<u>Posttest</u>
Experimental group I	R	O ₁	X	O ₂
Control group I	R	O ₃		O ₄
Experimental group II	R		X	O ₅
Control group II	R			O ₆

Solomon 4-group design

- Does the pretest have an independent effect?
- Does pretest sensitize people so that posttest gives different response, over and above the effect of experimental stimulus?

Decisions:

To judge effect of pretesting, compare:

- ✓ O_2-O_5 (experimental group with and without pretest)
- ✓ O_4-O_6 (control group with and without pretest)

Solomon 4-group design: Example (hypothetical)

X = horror movie

O = observation on fear
stimulation scale (stress test)

Solomon 4-group design (hypothetical)

<u>Solomon 4-group design</u>		Time —————>		
		<u>Pretest</u>		<u>Posttest</u>
Experimental group I	R	O ₁ =3	X	O ₂ =10
Control group I	R	O ₃ =3		O ₄ =5
Experimental group II	R		X	O ₅ =8
Control group II	R			O ₆ =3

Solomon 4-group design: Example

Michael Robinson, 1976, "Public Affairs Television and the Growth of Political Malaise: The Case of the Selling of the Pentagon." *American Political Science Review* 70:409-432.

Michael Robinson: “The Selling of the Pentagon”

- Up to 1960’s, political scientists believed television and mass media had little effect on attitudes and behaviors
- Then, shift in attitude: TV could have major impact on beliefs
- **Hypothesis:** TV fosters cynicism and feelings of helplessness

“The Selling of the Pentagon”: Methods

- Applied the Solomon 4-group design
- Tested impact of a CBS documentary, “The Selling of the Pentagon”
- Opinions about the behavior and credibility of social and public institutions, public officials, private citizens, and news organizations via public opinion questionnaires
- Pretest to two experimental groups: November, 1971
- Posttest to all four groups: December, 1971

Solomon 4-group design

$X = ?$

$Y = ?$

“The Selling of the Pentagon” Conclusions

- ✓ CBS documentary changed people’s beliefs about the behavior of the American military
- ✓ Program reduced support for the military
- ✓ Program increased belief in governmental misconduct

Pre-experimental designs: natural settings

- 1) One-shot case study: X O₁
- 2) Pretest-posttest design: O₁ X O₂
[Before-and after design]
- 2) Posttest-comparison group
design [ex post facto control group design]:

$$\begin{array}{c} X \quad O_e \\ \text{-----} \\ O_c \end{array}$$

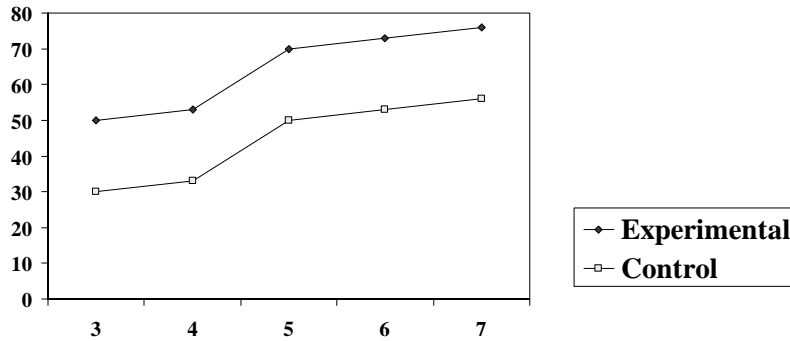
Quasi-experimental designs (n=3)

- No lab setting possible
- No possibility of randomizing into experimental and control groups
- *Purpose:* to enhance causal inference by strengthening internal and external validity

Contrasted groups design

(Chambliss & Schutt: nonequivalent control groups)

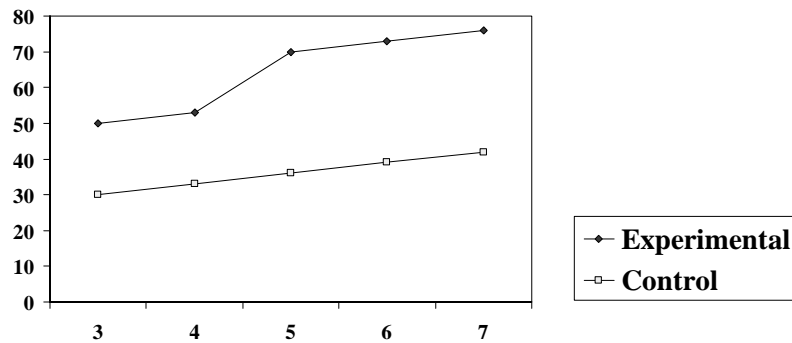
Figure A: Reading scores by grade



Contrasted groups design

(Chambliss & Schutt: nonequivalent control groups)

Figure B: Reading scores by grade



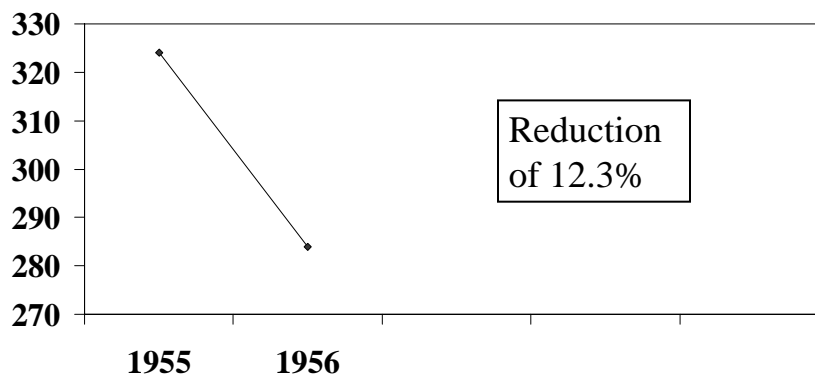
Time-series design

O1 O2 O3 X O4 O5 O6

- Multiple observations over time
- Example: Connecticut crackdown on speeding (1955)

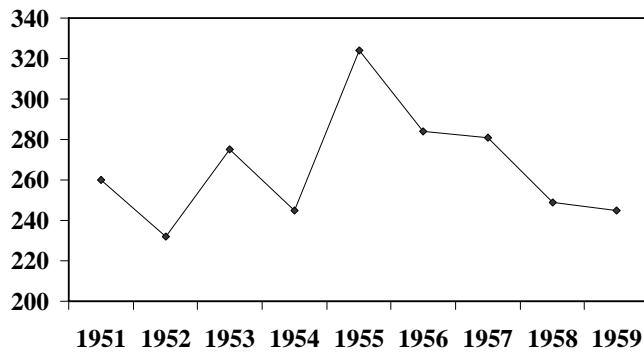
Time-series design

Figure A: # of fatalities, CT., 1955-56



Time-series design

Figure B: Number of fatalities, CT., 1951-59



Control-series design

Fatality Rates, 1951-59

