Chapter 5

Experiments, Good and Bad

Point of both observational studies and designed experiments is to identify variable or set of variables, called *explanatory variables*, which are thought to predict outcome or *response variable*. *Confounding* between explanatory variables occurs when two or more explanatory variables are not separated and so it is not clear how much each explanatory variable contributes in prediction of response variable. *Lurking* variable is explanatory variable not considered in study but confounded with one or more explanatory variables in study.

Confounding with lurking variables effectively reduced in *randomized comparative experiments* where subjects are assigned to treatments at random. Confounding with *a (only one at a time)* lurking variable reduced in observational studies by *controlling* for it by *comparing matched groups*. Consequently, experiments much more effective than observed studies at detecting which explanatory variables *cause* differences in response. In both cases, *statistically significant* observed differences in average responses implies differences are "real", did not occur by chance alone.

Exercise 5.1 (Experiments, Good and Bad)

1. *Randomized comparative experiment: effect of temperature on mice rate of oxygen consumption.* For example, mice rate of oxygen consumption 10.3 mL/sec when subjected to 10^o F.

temperature (F^o)	0	10	20	30
ROC (mL/sec)	9.7	10.3	11.2	14.0

   (a) Explanatory variable considered in study is (choose one)

      i. temperature
      ii. rate of oxygen consumption
      iii. mice
      iv. mouse weight

(b) Response is (choose one)

i. temperature
ii. rate of oxygen consumption
iii. mice
iv. room temperature

(c) Possible explanatory variable *not* considered in study (choose *two*!)

i. temperature
ii. rate of oxygen consumption
iii. noise level
iv. mouse weight

(d) Mouse weight is lurking variable if *confounded* with temperature in, for example, following way.

temperature (F^o)	0^o	10^o	20^o	30^o
mouse weight (oz)	10	14	18	20
ROC (mL/sec)	9.7	10.3	11.2	14.0

Weight is a confounding (and so lurking) variable with temperature in determining ROC because (circle one)

i. heavier mice had higher ROC than lighter mice.
ii. higher ROC for higher temperatures than lower temperatures.
iii. it was not clear at end of study whether change in ROC was a consequence of change in temperature or change in weight.

(e) **True** / **False**. One way to reduce confounding effect of weight with temperature on ROC is to assign mice to temperatures at random. Randomization reduces not only weight lurking variable but many other possible lurking variables such as mouse health, mouse gender and so on.

(f) Number mice 1, 2, 3 and 4. Assign four mice at random to four temperatures. Use random digits table, line 106.

random assignment ↗	Group 1 (1 mouse) →	Treatment 1 (0^o)	↘	observe ROC
random assignment ↗	Group 2 (1 mouse) →	Treatment 2 (10^o)	↘	observe ROC
random assignment ↘	Group 3 (1 mouse) →	Treatment 3 (20^o)	↗	observe ROC
random assignment ↘	Group 4 (1 mouse) →	Treatment 4 (30^o)	↗	observe ROC

Mice (choose one)

i. 6, 8, 4, 1
ii. 4, 1, 3, 2
iii. 2, 4, 1, 3

assigned to temperatures 0^o, 10^o, 20^o and 30^o respectively.

2. *Randomized comparative experiment: effect of temperature on mice ROC again.*

temperature →	70^o F	-10^o F
	10.3	9.7
	14.0	11.2
	15.2	10.3

(a) Explanatory variable or, equivalently, factor is (choose one)

i. temperature
ii. mice
iii. room temperature, 70^o F
iv. 70^o F and -10^o F

(b) Two treatments or, equivalently, two levels of factor are (choose one)

i. temperature
ii. mice
iii. room temperature, 70^o F
iv. 70^o F and -10^o F

(c) Experimental units are (choose one)

i. temperature
ii. mice
iii. room temperature, 70^o F
iv. 70^o F and -10^o F

(d) Control, or "do-nothing" treatment is (choose one)

i. temperature
ii. mice
iii. room temperature, 70^o F
iv. 70^o F and -10^o F

Although it makes sense to designate room temperature as control, it is possible to designate cold temperature as control. Control, then, can actually be either treatment. Also, if only two treatments and one treatment is control, the other is, confusingly, referred to as "treatment". So, if "room temperature" is control, "treatment" is "cold temperature".

(e) Number of replications (mice per treatment) is **2** / **3** / **4**.

(f) To reduce lurking (confounding) variables, assign six mice at random to two temperatures. Number mice 1, 2, 3, 4, 5 and 6. Use random digits table, line 117.

↗ Group 1 (3 mice) → Treatment 1 (70^o) ↘
random assignment observe ROC
↘ Group 2 (3 mice) → Treatment 2 (-10^o) ↗

Mice (choose one)

i. 3, 8, 1
ii. 3, 1, 2
iii. 3, 1, 6

assigned to temperature 70^o and so mice (choose one)

i. 4, 5, 2
ii. 2, 4, 5
iii. 5, 2, 4

assigned to temperature -10^o.

(g) The average ROC of mice subjected to 70^o is
$\frac{10.3+14.0+15.2}{3} \approx$ **10.4 / 11.3 / 13.2**,
whereas the average ROC of mice subjected to -10^o is
$\frac{9.7+11.2+10.3}{3} \approx$ **10.4 / 11.3 / 13.2**.
Since the difference between these two averages, $13.2 - 10.4 = 2.8$, **did / did not** occur by chance alone, it is "real", the difference is **statistically significant / statistically insignificant**.

3. *Randomized comparative experiment: effect of drug on patient response.*

drug →	A	B	C
	120	97	134
	140	112	142
	125	100	129
	133	95	137

(a) Explanatory variable or, equivalently, factor is (choose one)

i. drug
ii. patients
iii. drug A
iv. drug A, B and C

(b) Three treatments or, equivalently, three levels of factor are

i. drug
ii. patients
iii. drug A
iv. drug A, B and C

(c) Experimental units or, better, subjects are (choose one)

i. drug

ii. patients
iii. drug A
iv. drug A, B and C

(d) Number of replications (patients per treatment) is **2** / **3** / **4**.

(e) If drug A is a *placebo*, a drug with no medical properties, control is

i. drug
ii. patients
iii. drug A
iv. drug A, B and C

(f) To reduce lurking (confounding) variables, assign twelve patients at random to three drugs. Which diagram best describes experimental design? Choose one.

i. Diagram A:

	↗	Group 1 (3 patients) →	Treatment 1 (drug A)	↘	
random assignment					drug response
	↘	Group 2 (3 patients) →	Treatment 2 (drug B)	↗	
	↘	Group 3 (3 patients) →	Treatment 3 (drug C)	↗	

ii. Diagram B:

	↗	Group 1 (4 patients) →	Treatment 1 (drug A)	↘	
random assignment					drug response
	↘	Group 2 (4 patients) →	Treatment 2 (drug B)	↗	
	↘	Group 3 (4 patients) →	Treatment 3 (drug C)	↗	

iii. Diagram C:

	↗	Group 1 (4 patients) →	Treatment 1 (drug A)	↘	
doctor's choice					drug response
	↘	Group 2 (4 patients) →	Treatment 2 (drug B)	↗	
	↘	Group 3 (4 patients) →	Treatment 3 (drug C)	↗	

(g) Number patients 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11 and 12. Use random digits table, line 101. Patients (choose one)

i. 05, 09, 07, 02
ii. 03, 11, 02
iii. 19, 22, 39, 50

assigned to drug A and patients (choose one)

i. 01, 08, 11, 06
ii. 34, 05, 75, 62
iii. 05, 12, 04, 03

assigned to drug B and so patients (choose one)

i. 12, 04, 03, 10
ii. 33, 11, 22, 44

iii. 87, 13, 96, 40

are assigned to drug C.

(h) The average patient response subjected to drug A is
$\frac{120+140+125+133}{4} \approx$ **101.0 / 129.5 / 135.5**,
whereas the average patient response subjected to drug B is
$\frac{97+112+100+95}{4} \approx$ **101.0 / 129.5 / 135.5**,
and the average patient response subjected to drug C is
$\frac{134+142+129+137}{4} \approx$ **101.0 / 129.5 / 135.5**.
Since the differences between these three averages, $129.5, 101.0, 135.5$, **did / did not** occur by chance alone, they are "real", the differences are **statistically significant / statistically insignificant**.

4. *Matched group observed study: effect of drinking on traffic accidents.*
Indiana police records from 1999–2001 are analyzed to determine if there is an association between drinking and traffic accidents. One heavy drinker had 3 accidents for example.

drinking	heavy drinker	3	6	2
	light drinker	1	2	1

(a) Match columns.

Terminology	Example
(a) explanatory variable	**(A)** driver's age
(b) response	**(B)** amount of drinking
(c) lurking variable	**(C)** number of traffic accidents

Terminology	(a)	(b)	(c)
Example			

(b) Suppose age influences number of traffic accidents. Age is a confounding (and so lurking) variable with drinking in number of traffic accidents of Indiana drivers if (circle one)

i. young drivers had more traffic accidents than older drivers.
ii. intoxicated drivers had more traffic accidents than sober drivers.
iii. it was not clear at end of study whether number of traffic accidents was a consequence of being intoxicated or not, or whether it was a consequence of age.

(c) One way to eliminate confounding effect of age with drinking on traffic accidents (to control for age) in this observed study would be to (choose one)

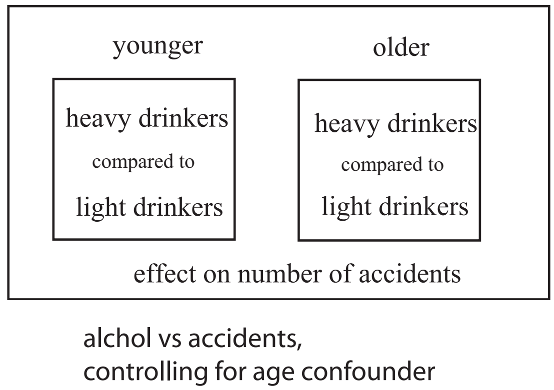

Figure 5.1 (Drinking and number of traffic accidents, controlling for age)

i. assign drivers to be either drunk or sober at random (Is this possible, since the data was collected from police records?)
ii. compare number of traffic accidents of drunk drivers with sober driver who both have similar ages, to *match* ages.
iii. compare number of traffic accidents of drunk drivers with sober driver who both have different ages

5. *Matched group observed study: effect of teaching method on academic achievement.* A recent study compares academic achievement (measured by final examination scores) of online students with classroom students.

(a) Suppose average GPA influences academic achievement. Average student GPA is a confounding (lurking) variable with teaching method on academic achievement of students if (circle one)

i. students with high average GPAs had better final examination scores than students with low average GPAs.
ii. Online students had better final examination scores than classroom students.
iii. it is not clear at end of study whether students' academic achievement is a consequence of being either online students or classroom students, or is a consequence of average GPA.

(b) One way to eliminate confounding effect of average GPA with teaching method on academic achievement of students (to control for average GPA) in this observed study would be to (choose one)

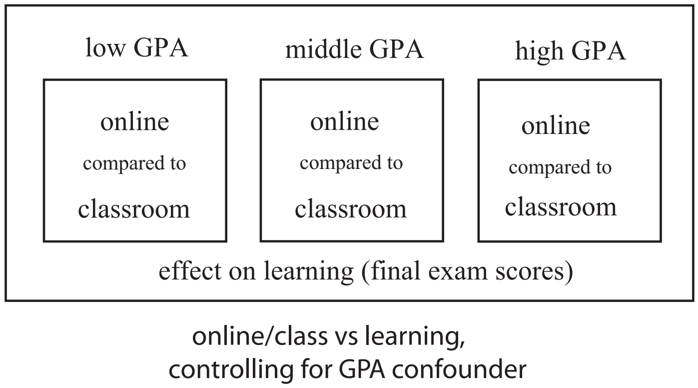

Figure 5.2 (Academic achievement and teaching method, controlling for GPA)

i. assign students to be either "classroom" or "online" students at random (Is this possible, since these students choose between these two options themselves?)

ii. compare academic achievement of classroom students with online students who both have similar average GPAs, to *match* GPAs.

iii. compare the academic achievement of classroom students with online students who both have different average GPAs

(c) **True** / **False** Matching students for confounder average GPA in this study does not match students for any other confounder. Each confounder (lurking variable) must be matched for separately from every other confounder in an observational study. Or careful measurement and adjustment for other confounders must be undertaken.

This is unlike in a randomized designed experiment, where randomization takes care of all confounders all at once.

6. *Principals of a good experiment (observational study)?* Choose one or more.

(a) Randomize: assign subjects to treatments at random to reduce bias.

(b) Control: match groups to reduce effects of lurking variables.

(c) Replicate: use many subjects to decrease variability in results.

Chapter 6

Experiments in the Real World

Conducting experiments in real world face practical problems. *Placebo effect*, positive or negative influence on patient (or experimenter) not due to drug, is offset by *double-blinding*, if both patients and experimenters do not know which drug is assigned to which patient, until after experiment is over. Experiments suffer from *refusals*, subjects who refuse to participate; *nonadherers*, subjects who do not follow prescribed treatment; and *dropouts*, subjects who do not complete experiment.

Convincing experiments are well designed. Three types of designs are described. *Completely randomized design* investigates how one or more explanatory variables (each with two or more levels or treatments) influences a response variable. *Block design* investigates how one explanatory variable influences a response variable, but also, to reduce variability of statistical inference, groups experimental units into homogeneous blocks. An important special case of block design is *matched-pair design* where there are only two experimental units per block.

Exercise 6.1 (Experiments in the Real World)

1. *Blinding to reduce placebo effect: effect of drug on flu.*
   Patients are assigned to receive drug or placebo at random. For example, flu symptoms of patient who receives drug improves 2.5 units.

drug	placebo
+2.5	-8
+1.5	-6
+2	-5
+0.5	-2

   (a) Match columns.

Column I	Column II
(a) treatment	**(A)** placebo
(b) control	**(B)** drug
(c) response	**(C)** patients
(d) subject(s)	**(D)** flu symptoms

Column I	(a)	(b)	(c)	(d)
Column II				

(b) *Placebo effect.* Patients who receive drug may feel better simply because they received drug and *not* because of any actual medical relief drug might provide. A bias **in favor of** / **against** drug occurs.

(c) Doctors might tend to promote positive aspects and downplay negative aspects of drug. A bias **in favor of** / **against** drug occurs.

(d) To offset possible bias, double-blind. This means *blinding* both patient and doctor to which patient is receiving which drug. **Neither** / **both** / **either** patients nor/and/or doctors know who receives which drug.

2. *Appropriate type of blinding.*
Identify type of blinding appropriate in following situations.

(a) **blind researcher** / **blind subjects** / **blind both**
Company research investigates effectiveness of new suntan cream on a number of subjects.

(b) **blind researcher** / **blind experimental units** / **blind both**
Agricultural researcher investigates yield of crop subjected to different fertilizers.

(c) **blind researcher** / **blind subjects** / **blind both**
Educator investigates impact of different teaching methods on student academic achievement.

3. *Experimental designs: completely randomized, block, matched-pair.*
Consider experiment to determine effect of temperature on mice ROC.

(a) *Completely Randomized Design.*

temperature →	70^o F	-10^o F
	10.3	9.7
	14.0	11.2
	15.2	10.3

This is a completely randomized design with (choose *one or more*!)

i. one factor (temperature) with two (but could be more) treatments,

ii. three pair of mice matched by age,
iii. mice assigned to temperatures at random.

(b) *Block design.*

age ↓ temperature →	70^o F	-10^o F
10 days	10.3	9.7
20 days	14.0	11.2
30 days	15.2	10.3

This is a block design with (choose *one or more*!)

i. one factor (temperature) with two (but could be more) treatments,
ii. one block (age) with three levels (where age is *not* "assigned" to mice)
iii. mice assigned to temperatures within each age block at random.

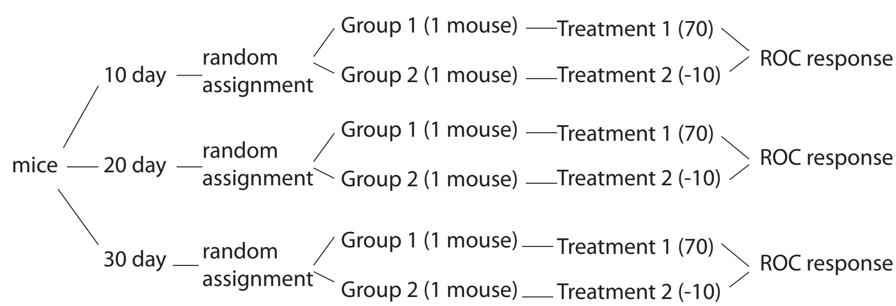

Figure 6.1 (Temperature (factor), Age (block)

(c) *Matched-pair design.*

age ↓ temperature →	70^o F	-10^o F
mouse 1 or two "similar" mice	10.3	9.7
mouse 2 or two "similar" mice	14.0	11.2
mouse 3 or two "similar" mice	15.2	10.3

This is a matched-pair design with (choose *one or more*!)

i. one factor (temperature) with *exactly* two treatments,
ii. three mice or three pair of "similar" mice,
iii. mice assigned to temperatures within each age at random.

A pair of mice might be "similar" due to age only; that is, 10 days, 20 days and 30 days, but, typically, are paired with greater number of similar characteristics such as similar age, gender, general health and so on. Using the *same* mouse is also possible; matched-pair design is special case of blocked design.

4. *Experimental designs: completely randomized, block, matched-pair?*
Consider experiment to determine effect of drug on patient response.

(a) *Study A.*

drug →	A	B	C
	120	97	134
	140	112	142
	125	100	129
	133	95	137

Since this experiment consists of one factor with three levels, no blocks and where patients are assigned to drugs at random, this is a (choose one)

i. completely randomized design
ii. block design
iii. matched-pair design

(b) *Study B.*

age↓ drug →	A	B	C
20-25 years	120	97	134
25-30 years	140	112	142
30-35 years	125	100	129
35-40 years	133	95	137

Since this experiment consists of one factor (drug) with three levels, one block (age) with four levels and where patients are assigned to drugs within common ages at random, this is a (choose one)

i. completely randomized design
ii. block design
iii. matched-pair design

Although a block design, it is not also a matched-pair design because there are three, not two, levels of the drug factor.

5. *More completely randomized design: effect of music (artist, loudness) on heart rate.* Effect of volume and different musical artists on heart rate of students, chosen at random from PNW, is investigated. For instance, heart rates of two students listening to Natalie Merchant playing softly are 7.2 and 8.1 units.

artist ↓ volume →	soft	medium	loud
Natalie Merchant	7.2, 8.1	8.4, 8.2	8.9, 9.2
Matchbox Twenty	9.1, 8.7	9.2, 9.5	10.2, 12.7
Handel	3.2, 4.1	4.3, 4.1	4.7, 4.5

(a) Match columns.

Column I	Column II
(a) explanatory variables	**(A)** artist, volume
(b) subjects	**(B)** Handel, soft
(c) response	**(C)** heart rates
(d) control	**(D)** PNW students

Column I	(a)	(b)	(c)	(d)
Column II				

(b) Number of treatments is 3 artists × 3 volumes = **8** / **9** / **10**.

(c) Number of replications (students per treatment) is **2** / **9** / **18**.

(d) Number of students required for experiment is **2** / **9** / **18**.

(e) *Block or explanatory variable?* Choose one.

i. Both artist and volume are explanatory variables
ii. Artist is block; volume is explanatory variable
iii. Artist is explanatory variable; volume is block
iv. both artist and volume are blocks

because both artist and volume are *assigned* to students.

6. *More completely randomized design: effect of clay type, oven on proportion cracked.* Clay tiles are fired in a kiln where, unfortunately, some of the tiles crack during the firing process. Batches of clay are molded into clay tiles using one of three ovens. A manufacturer of clay roofing would like to investigate the effect of clay type and oven on proportion of cracked tiles. Two different types of clay are to be considered. Both clay types are identical, aside from addition of ingredient A for clay type A and addition of ingredient B for clay type B. Plates of one hundred tiles can be placed in kiln at any one time. Higher firing temperature increases cracking. There are slight variations in firing temperature at different locations in kiln. Where tiles are placed in kiln is *not* considered (used) in design of experiment.

(a) Which design is correct? Choose one table.

i. Table A.

temperature ↓ oven →	1	2	3
warm	100 tiles	100 tiles	100 tiles
hot	100 tiles	100 tiles	100 tiles

ii. Table B.

clay type ↓ oven →	1	2	3
ingredient A	100 tiles	100 tiles	100 tiles
ingredient B	100 tiles	100 tiles	100 tiles
ingredient C	100 tiles	100 tiles	100 tiles

iii. Table C.

clay type ↓ oven →	1	2	3
ingredient A	100 tiles	100 tiles	100 tiles
ingredient B	100 tiles	100 tiles	100 tiles

(b) Match columns.

Column I	Column II
(a) explanatory variable	**(A)** clay ingredient A, oven 1
(b) experiment unit(s)	**(B)** proportion of cracked clay tiles
(c) response	**(C)** clay tiles
(d) control	**(D)** clay type, oven
(e) confounder	**(E)** location (heat) of tiles in kiln

Column I	(a)	(b)	(c)	(d)	(e)
Column II					

(c) Number of treatments is **6** / **8** / **10**.

(d) Number of replications (tiles per treatment) is **50** / **100** / **200**.

(e) Number of tiles required for experiment is **600** / **800** / **1000**.

(f) *Block or explanatory variable?* Choose one.

i. Both clay type and oven are explanatory variables
ii. Clay type is block; oven is explanatory variable
iii. Clay type is explanatory variable; oven is block
iv. Clay type and oven are blocks

because both clay type and oven are *assigned* to tiles.

(g) Identify possible confounders (lurking variables). Choose one or more.

i. **confounder** / **not confounder** gender of kiln operator
ii. **confounder** / **not confounder** humidity in kiln
iii. **confounder** / **not confounder** temperature in kiln
iv. **confounder** / **not confounder** air pressure in kiln