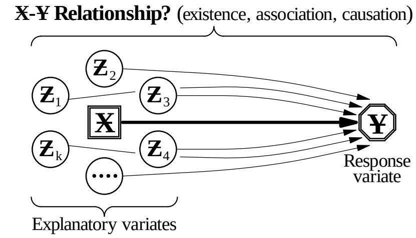
Figure 9.8. OBSERVATIONAL PLANS: Simpson's Paradox

1. Background Review – Lurking Variates, Confounding and Comparison Error [optional reading]



As background to answering Question(s) about an X-Y relationship between a focal variate X and a response variate Y, Z_1, Z_2, ..., Z_k in the schema at the right are called **lurking variates**, a phrase that means lurking *explanatory* variates in that each Z accounts, at least in part, for changes from element to element in the value of the response variate. The importance of lurking variates is that if the distributions of their values *differ* between groups of elements [like (sub)populations or samples] with different values of the focal variate, an Answer about the X-Y relationship may differ from the true state of affairs unless the differences in the values of the relevant Zs are taken into account.

The *same* statistical issue raised by lurking variates is involved, with different terminology, in **confounding**; the difference is that the behaviour of lurking variates (the entity responsible) is *why* confounding (the statistical issue) occurs.
An explanatory variate responsible for confounding is called a **confounder** or **confounding variate**; these two terms are synonyms for a lurking variate whose distribution of values (over a group of elements) differs for different values of the focal variate.
The following definitions summarize the foregoing discussion:

∗ **Lurking variate:** a non-focal explanatory variate whose differing distributions of values (over groups of elements) for different values of the focal variate, if taken into account, would meaningfully change an Answer about an X-Y relationship.

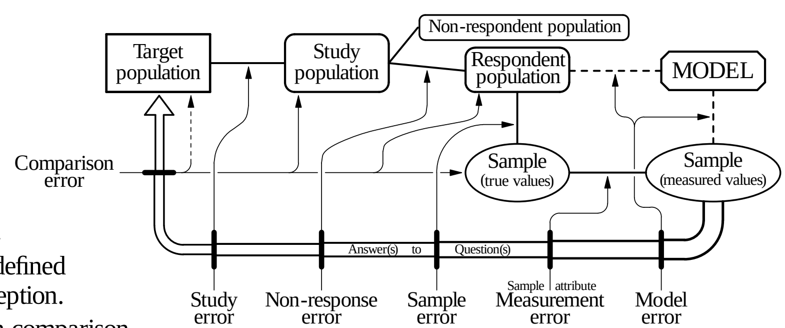
∗ **Confounding:** differing distributions of values of one or more *non*-focal explanatory variate(s) among two (or more) groups of elements [like (sub)populations or samples] with different values of the focal variate.
  – **Confounder (confounding variate):** a non-focal explanatory variate involved in confounding.

'Confounding' and 'confounder' have the convenience of being one-word terminology rather than the multi-word phrases involving 'lurking variates' which convey the same ideas.

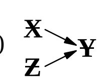
∗ **Comparison error:** for an Answer about an X-Y relationship that is based on comparing attributes of groups of elements with different values of the focal variate, comparison error is the difference from the *intended* (or *true*) state of affairs arising from:
  – differing distributions of lurking variate values between (or among) the groups of elements OR – confounding.
  The alternate wording of the last phrase accommodates the equivalent terminologies of lurking variates and confounding; in a particular context, we use the version of the definition appropriate to that context:
  - 'lurking variates' can more readily accommodate phenomena like Simpson's Paradox discussed in this Figure 9.8;
  - 'confounding' is more common in the context of comparative Plans, as in Section 7 which starts on page 9.12 of Figure 9.2 of these Course Materials, but the variety of usage of 'confounding' can be a source of difficulty (see the following Figure 9.9 on pages 9.65 to 9.68).

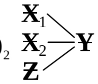
  Comparison error in experimental and observational Plans is discussed in Section 15 on pages 9.26 to 9.28 in Figure 9.2.

The schema at the right (from, for instance, page 9.6 of Figure 9.2) reminds us of several matters.



○ Data-based investigating is concerned initially with four groups of units – the *target* population, the *study* population, the *respondent* population and the *sample*.
○ Associated with each of these groups are (one or more) *attributes* if interest.
○ Answer(s) to Questions(s) are usually given in terms of attributes, often their values.
○ Our six categories of error, of which five are defined in terms of attributes – model error is the exception.
  – In the schema, the four arrows arising from comparison error point to *boxes* representing *groups* of elements or units (a population or a sample) rather than, as for the other five error categories, to *lines joining boxes*; the comparison error arrow at the right is to be taken as pointing to *both* sample ellipses.
    + *Multiple* comparison error arrows are a consequence of its different manifestations in different Question contexts.

(8)


$(A)_2$


In earlier discussion (*e.g.*, in Section 2 on page 9.6 in Figure 9.2), the context for comparison error due to lurking variate(s)/confounding is comparative investigating of a *treatment* effect; the relevant *causal* structure (*e.g.*, from near the middle of page 9.10 of Figure 9.2), is case (8), shown at the upper right, with *focal* variate X, *respone* variate Y and lurking variate/confounder Z. In this Figure 9.8, as summarized in the structure $(A)_2$ at the lower right, we broaden the discussion in two ways:
- we have two (or three) 'focal' variates [not necessarily all of equal interest in the Question context];

(*continued overleaf*)

- we are *un*concerned with *causation* as the reason for the X_i-Y and Z-Y associations, because the nature of the focal variates is such that we can*not* set their levels and this precludes using such focal variate(s) to manipulate the value of Y;
  – this is why the lower structure at the right overleaf on page 9.57 has *lines rather than arrows* between the variate symbols.

The phenomenon known as Simpson's Paradox can arise in a comparative investigation where the attributes are proportions – that is, the response variate Y is *qualitative* [discrete (categorical)] in nature; the dramatic name ('Paradox') is a reflection of how the effect of lurking variate(s) can *reverse* the sign of a relationship. The data in twenty of the first twenty-one Tables 9.8.1 to 9.8.21 used in discussion of Simpson's Paradox in this Figure 9.8 are hypothetical. The discussion is in six sections:

2. Illustrations of Simpson's Paradox.
3. 'Simpson's Paradox' with a quantitative response variate.
4. Reasons for Simpson's Paradox – properties of proportions.
5. Reasons for Simpson's Paradox – population subgroups and weighted averages.
6. Reasons for Simpson's Paradox – probability distributions.
7. A Plan for an investigation to answer the Question of sex discrimination.

The discussion is framed in terms of *populations*, because there are no *inherent* sampling issues in Simpson's Paradox; when the groups being compared are *samples*, there is the additional statistical issue of managing sample error.

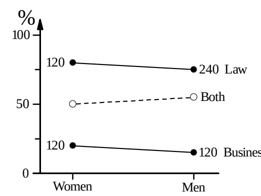
2. Illustrations of Simpson's Paradox.

The data in Table 9.8.1 below come from the discussion of Simpson's Paradox in Program 11 of *Against All Odds: Inside Statistics*; the context is possible sex discrimination in graduate admissions. Overall, the admission *rate* [or *proportion* (an *attribute*)] is *lower* for women (50% *vs.* 55% for men – see the bottom line of the Table) but, when the data are subdivided by school (Law and Business), the female admission rate is *higher* (by 5 percentage points) for *each* school. The (binary) response variate is school admission (Yes, No) and the lurking variate is women-to-men ratio among applicants; its effect is because:

∗ the two schools had appreciably *different* admission rates: 80 and 75% for Law, 20 and 15% for Business;
∗ *half* as many women as men (120 *vs.* 240) applied to Law but *equal* numbers of women and men (120) applied to Business.

Table 9.8.1:

SCHOOL	WOMEN Number of Applicants	WOMEN ADMISSIONS Number	WOMEN ADMISSIONS %	MEN Number of Applicants	MEN ADMISSIONS Number	MEN ADMISSIONS %
Law	120	96	**80**	240	180	**75**
Business	120	24	**20**	120	18	**15**
Both	240	120	**50**	360	198	**55**



The diagram to the right of Table 9.8.1 shows its data in graphical form; Simpson's Paradox is the *positive* slope of the middle dashed line for the data for *both* schools changing to a *negative* slope in the upper and lower lines for the schools *individually.*

In this illustration, the variates in the lower structure $(A)_2$ at the lower right overleaf on page 9.57 are:

X_1 is an applicant's sex (female, male), X_2 is the school applied to (Law, Business),
[In Tables 9.8.5 and 9.8.6 on the facing page 9.59, X_3 is the level of study (Masters, Doctoral)],
Z is the (lurking variate) women-to-men ratio among applicants (discussed further in Sections 3 and 5 on pages 9.59 and 9.60),
Y is the response to an applicant (admitted, not admitted). [On page 9.59, Y is time for degree completion (minimum, longer).]

*Un*like investigating a treatment effect when there is more than one focal variate (*e.g.*, using a factorial treatment structure), the focal variate of primary interest in *this* Question context is X_1, an applicant's sex.

The limitation imposed by lurking variates on an Answer to a Question about an X-Y relationship is illustrated further by the data in Tables 9.8.2 to 9.8.4; as the diagrams to the right of the tables emphasize, it is also possible to have:

∗ the *same* overall admission rate for women and men but a *higher* rate for women in the two schools individually (Table 9.8.2);
∗ a *lower* overall admission rate for women but the *same* rate for women and men in the two schools individually (Table 9.8.3);
∗ a *higher* rate overall *and* in the two schools individually for women (Table 9.8.4).

Table 9.8.2:

SCHOOL	WOMEN Number of Applicants	WOMEN ADMISSIONS Number	WOMEN ADMISSIONS %	MEN Number of Applicants	MEN ADMISSIONS Number	MEN ADMISSIONS %
Law	120	96	**80**	168	126	**75**
Business	120	24	**20**	120	18	**15**
Both	240	120	**50**	288	144	**50**

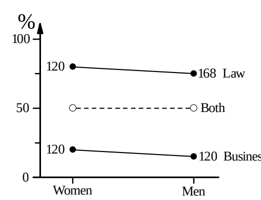
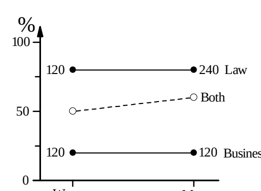


Table 9.8.3:

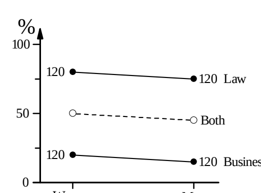
SCHOOL	WOMEN Number of Applicants	WOMEN ADMISSIONS Number	WOMEN ADMISSIONS %	MEN Number of Applicants	MEN ADMISSIONS Number	MEN ADMISSIONS %
Law	120	96	**80**	240	192	**80**
Business	120	24	**20**	120	24	**20**
Both	240	120	**50**	360	216	**60**



The effect of lurking variates on an X-Y relationship at a *second* level of subdivision is illustrated in Tables 9.8.5 and 9.8.6 at the upper right of the facing page 9.59; a context for

Table 9.8.4:

SCHOOL	WOMEN Number of Applicants	WOMEN ADMISSIONS Number	WOMEN ADMISSIONS %	MEN Number of Applicants	MEN ADMISSIONS Number	MEN ADMISSIONS %
Law	120	96	**80**	120	90	**75**
Business	120	24	**20**	120	18	**15**
Both	240	120	**50**	240	108	**45**



(*continued*)

Figure 9.8. OBSERVATIONAL PLANS: Simpson's Paradox (continued 1)

these data is the proportion of graduate students who complete their degree in the minimum time. In Table 9.8.5, the proportion for women is *lower* overall, *higher* when subdivided by subject area (Law or Business) but again *lower* when subect area is subdivided by level (Masters or Doctoral). Similar effects are seen in Table 9.8.6, except the proportions for women become *equal* when subdivided by subject area and *higher* when further subdivided by level.

Probabilistically, subdividing is *conditioning* so that Tables 9.8.1 to 9.8.6, in illustrating Simpson's Paradox, show the limitation on an Answer which involves comparing *conditional* probabilities for a response variate with *different* conditionings; that is, comparing probabilities for $\mathbf{Y}$ given $\mathbf{X}_1$ and $\mathbf{X}_2$ with $\mathbf{Y}$ given only $\mathbf{X}_1$ (in Tables 9.8.1 to 9.8.4) or for $\mathbf{Y}$ given $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ with $\mathbf{Y}$ given $\mathbf{X}_1$ and $\mathbf{X}_2$ or $\mathbf{Y}$ given only $\mathbf{X}_1$ (in Tables 9.8.5 and 9.8.6) – see Section 6 overleaf on page 9.60. Four other illustrations of Simpson's Paradox are given in Note 2 on pages 9.61 and 9.62 and three more illustrative tables (like Table 9.8.9 overleaf on page 9.60) are discussed on pages 9.62 and 9.63 in the Appendix of this Figure 9.8.

Table 9.8.5:	............**WOMEN**............			................**MEN**..............		
SCHOOL	Number of Students	COMPLETIONS Number	%	Number of Students	COMPLETIONS Number	%
Law: Masters	60	51	**85**	60	54	**90**
Doctoral	60	33	**55**	300	180	**60**
Bus.: Masters	60	27	**45**	20	10	**50**
Doctoral	60	9	**15**	100	20	**20**
Law	120	84	**70**	360	234	**65**
Business	120	36	**30**	120	30	**25**
Both	240	120	**50**	480	264	**55**

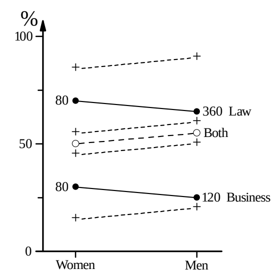
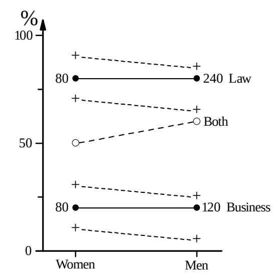


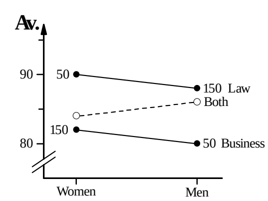
Table 9.8.6:	............**WOMEN**............			................**MEN**..............		
SCHOOL	Number of Students	COMPLETIONS Number	%	Number of Students	COMPLETIONS Number	%
Law: Masters	60	54	**90**	240	204	**85**
Doctoral	60	42	**70**	80	52	**65**
Bus.: Masters	60	18	**30**	120	30	**25**
Doctoral	60	6	**10**	40	2	**5**
Law	120	96	**80**	320	256	**80**
Business	120	24	**20**	160	32	**20**
Both	240	120	**50**	480	288	**60**



3. 'Simpson's Paradox' with a Quantitative Response Variate

Simpson's Paradox is usually presented in the context of comparing *proportions* but the same phenomenon can occur with a *continuous* response variate. As illustrated by the data in Table 9.8.7 and the diagram to its right, whose context is graduate studies admission averages, the average is *lower* overall for women than men (84% *vs.* 86%) but, when the data are subdivided by school, both averages are *higher* for women. The response variate here is an applicant's average, the attribute is the *average* of these averages (*e.g.*, 90 and 88 for Law, 82 and 80 for Business) and the lurking variate is women-to-men ratio among applicants (1:3 for Law, 3:1 for Business). With 1:1 ratios, there is *no* 'paradox'.

Table 9.8.7:	........**WOMEN**........		...........**MEN**...........	
SCHOOL	Number of Applicants	Applicants' Average (%)	Number of Applicants	Applicants' Average (%)
Law	50	**90**	150	**88**
Business	150	**82**	50	**80**
Both	200	**84**	200	**86**

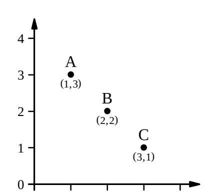


The illustration in Table 9.8.7 shows that Simpson's Paradox is *not* solely a phenomenon which may arise when comparing *proportions*. Its origin lies in the *relative* 'natural' group sizes arising from the process of subdividing (or its inverse of combining) used to manage comparison error in observational Plans. Such a lurking variate (called the women-to-men ratio in the discussion of Table 9.8.1 on page 9.58 and Table 9.8.7 above) is *different* in nature to $\mathbf{Z}$ in the upper causal structure of case (8) at the lower right of page 9.57, which we think of as being able to *cause* an element to change the value of its response variate. Thus, we now recognize *two* ways a change in a lurking variate can affect attribute value(s):

- by *causing* elements' response variate (and, hence, their attribute) values to change, AND:
- by distorting attribute *calculation* when subdividing is used to manage comparison error in an observational Plan.

4. Reasons for Simpson's Paradox – properties of proportions

Quantities (like variate and attribute values) which are *single* numbers are relatively straightforward to compare: 4 is greater than 2 is greater than −6, although the latter has a larger *magnitude* than the first two. However, when quantities (like proportions or fractions and the coordinates of points on a scatter diagram) involve *two* numbers, comparisons may raise complications. For example, in the diagram at the right, points A and C with *different* coordinates are the *same* distance from the origin and point B is *closer* to the origin than A and C despite its coordinates being *larger* than one of those of A and C. The (surprising) result for fractions (or proportions), exhibited



(*continued overleaf*)

as Simpson's Paradox, is that for eight (positive) integers a, b, ..., h, it is possible to have (as in Table 9.8.1 on page 9.58):

$$\frac{a}{b} > \frac{c}{d} \text{ and } \frac{e}{f} > \frac{g}{h} \text{ but at the } \textit{same time} \text{ to have: } \frac{a+e}{b+f} < \frac{c+g}{d+h}; \qquad \textit{e.g.}, \frac{32}{40} > \frac{90}{120} \text{ and } \frac{12}{60} > \frac{9}{60} \text{ but } \frac{44}{100} < \frac{99}{180}.$$

This property also applies to *more* than two pairs of fractions (as in Table 9.8.15 on page 9.61) and for other combinations of inequality and equality (as in Tables 9.8.2, 9.8.3, 9.8.5 and 9.8.6 on page 9.58 and overleaf on page 9.59)

When the fraction $\frac{90}{120}$ is instead $\frac{30}{40}$, we see at the right that there is *no* 'paradox' (as in Table 9.8.4), reminding us that it is the group sizes (in the denominators) under subdividing that may engage the property of proportions which generates the 'paradox' – recall Section 3 on page 9.59.

$$\frac{32}{40} > \frac{30}{40} \text{ and } \frac{12}{60} > \frac{9}{60} \textit{ and } \frac{44}{100} > \frac{39}{100}$$

5. Reasons for Simpson's Paradox – population subgroups and weighted averages

The distorted calculation of the values of (population) attributes (like proportions and averages), which generates the 'paradox' illustrated in Sections 2 and 3, is an instance of *weighted* combinations of the corresponding attributes of population *subgroups*. As shown in Table 9.8.8 at the right, the attribute values in the last line of each of Tables 9.8.1 to 9.8.4 are weighted combinations of the attributes in the two table lines above them; what produces the changes in attribute values *relative* to each other is a change in *weights*. Each weight is determined by the (natural) *size* of a population subgroup; this size is the *lurking* variate whose change is responsible for the change in (the sign of) the **X**-**Y** relationship. The same idea applies to *each* of the *two* levels of subdivision in Tables 9.8.5 and 9.8.6 and to the averages in Table 9.8.7. When the weights are *equal* (as in Table 9.8.4), there is *no* 'paradox'.

Table 9.8.8: Weighted percentage

	Weighted percentage	Weights	
Table 9.8.1:	$\frac{120}{240}\times\mathbf{80} + \frac{120}{240}\times\mathbf{20} = \mathbf{50}$	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{240}{360}\times\mathbf{75} + \frac{120}{360}\times\mathbf{15} = \mathbf{55}$	$\frac{2}{3}$	$\frac{1}{3}$
Table 9.8.2:	$\frac{120}{240}\times\mathbf{80} + \frac{120}{240}\times\mathbf{20} = \mathbf{50}$	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{168}{288}\times\mathbf{75} + \frac{120}{288}\times\mathbf{15} = \mathbf{50}$	$\frac{7}{12}$	$\frac{5}{12}$
Table 9.8.3:	$\frac{120}{240}\times\mathbf{80} + \frac{120}{240}\times\mathbf{20} = \mathbf{50}$	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{240}{360}\times\mathbf{80} + \frac{120}{360}\times\mathbf{20} = \mathbf{60}$	$\frac{2}{3}$	$\frac{1}{3}$
Table 9.8.4:	$\frac{120}{240}\times\mathbf{80} + \frac{120}{240}\times\mathbf{20} = \mathbf{50}$	$\frac{1}{2}$	$\frac{1}{2}$
	$\frac{120}{240}\times\mathbf{75} + \frac{120}{240}\times\mathbf{15} = \mathbf{45}$	$\frac{1}{2}$	$\frac{1}{2}$

6. Reasons for Simpson's Paradox – probability distributions

Table 9.8.1 on page 9.58 provides data from which the probability function of a discrete trivariate distribution can be estimated. To obtain this model, we first extend Table 9.8.1 as in Table 9.8.9 below to include three extra columns for 'Both sexes'. We then define five events and use estimates for ten probabilities – the vertical line means 'given that' in the eight *conditional* probabilities and ∩ denotes an *intersection* of events.

Table 9.8.9:

	WOMEN			MEN			BOTH SEXES		
SCHOOL	Number of Applicants	ADMISSIONS Number	%	Number of Applicants	ADMISSIONS Number	%	Number of Applicants	ADMISSIONS Number	%
Law	120	96	**80**	240	180	**75**	360	276	**76.$\dot{6}$**
Business	120	24	**20**	120	18	**15**	240	42	**17.5**
Both schools	240	120	**50**	360	198	**55**	600	318	**53**

The (joint) trivariate model is shown in Table 9.8.10 at the right below; summing its probabilities for one variate, we obtain the three (marginal) *bi*variate models in Tables 9.8.11 to 9.8.13. The smaller **bold** annotations in Tables 9.8.10 to 9.8.12 show how eight of the nine percentages in Table 9.8.9 arise; for example, the 80% of women admitted to Law is $\frac{0.16}{0.2}$.

Event A:	Applicant is admitted (**Y** = yes; the complement $\bar{\text{A}}$ is **Y** = no)		
Event F:	Applicant is female (X_1 = female)	Pr(F) = 0.4	Pr(A|F) = 0.5	Pr(A|F∩L) = 0.8
Event M:	Applicant is male (X_1 = male)	Pr(M) = 0.6	Pr(A|M) = 0.55	Pr(A|F∩B) = 0.2
Event L:	Applicant applies to Law (X_2 = Law)		Pr(A|L) = 0.7$\dot{6}$	Pr(A|M∩L) = 0.75
Event B:	Applicant applies to Business (X_2 = Business)		Pr(A|B) = 0.175	Pr(A|M∩B) = 0.15

Table 9.8.10: Trivariate model for Y, X_1 and X_2

	F: L	F: B	M: L	M: B	
A	0.16	0.04	0.3	0.03	0.53
$\bar{\text{A}}$	**0.8** 0.04	**0.2** 0.16	**0.75** 0.1	**0.15** 0.17	0.47
	0.2	0.2	0.4	0.2	

Table 9.8.11: Bivariate model for Y and X_1

	F	M	
A	0.2	0.33	0.53
$\bar{\text{A}}$	**0.5** 0.2	**0.55** 0.27	0.47
	0.4	0.6	

We see that Table 9.8.1 on page 9.58 involves *parts* of the two multivariate distributions in Tables 9.8.10 and 9.8.11; it is therefore *un*surprising if comparisons among these parts, taken in isolation, yield seeming 'paradoxes'. It can be confusing that Table 9.8.1 and those like it do not show *explicitly* percentages involving *complements* [like applicants '*not* admitted' (event $\bar{\text{A}}$)].

Table 9.8.12: Bivariate model for Y and X_2

	L	B	
A	0.46	0.07	0.53
$\bar{\text{A}}$	**0.7$\dot{6}$** 0.14	**0.175** 0.33	0.47
	0.6	0.4	

Table 9.8.13: Bivariate model for X_1 and X_2

	L	B	
F	0.2	0.2	0.4
M	0.4	0.2	0.6
	0.6	0.4	

7. A Plan for an Investigation to Answer the Question of Sex Discrimination

Comparing proportions of women and men admitted among applicants to graduate studies (as in the context of Table 9.8.1 on page 9.58) is *not* an adequate Plan to answer the Question of possible sex discrimination, for two reasons:

- there is the possibility of Simpson's Paradox and no clear way to define the level of subdivision at which to make comparisons;
- applicants' *qualifications* are not taken into account.

Both matters are addressed by a Plan which involves taking *pairs* of applicants, one female and one male, with the *same* qualifications for admission and then comparing the proportions of women and men who are admitted across a number of such

(*continued*)

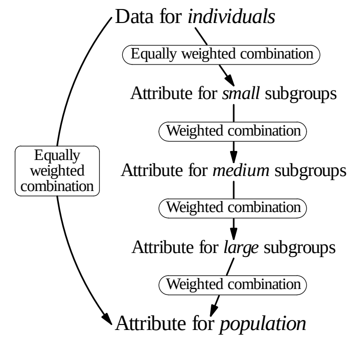
Figure 9.8. OBSERVATIONAL PLANS: Simpson's Paradox (continued 2)

pairs that is adequate, in the investigation context, to manage all relevant categories of error.

∗ For comparison error, pairing manages the *group sizes* (and hence, the weights in the attribute calculations) in a way that precludes Simpson's Paradox; matching manages equality of qualifications for the groups of women and men being compared.

However, as with *any* observational Plan (that gathers data from a population in its *natural* state), there is still the limitation on Answer(s) imposed by comparison error due to other (unrecognized) lurking variates.

∗ When investigating the much-discussed issue of comparable worth (whether women are paid the same as men for the same work), relevant explanatory variates to manage include qualifications, experience and hours worked per month or per year.



The schema at the right is a pictorial reminder of the lurking variate of group (population or sample) sizes when developing an observational Plan to answer a Question with a *causative* aspect, which (usually) involves comparing attribute values (calculated or obtained from a scatter diagram in the *Analysis* stage of the FDEAC cycle) for broad subpopulations (like women and men). By contrast, when answering a Question with a *descriptive* aspect (*e.g*., a Question about *both* sexes), differing attribute values at different levels of subdivision are more obvious and so lurking variate(s) are usually less troublesome. These matters are illustrated, using information from Table 9.8.5 (near the top of page 9.59), in Table 9.8.14 at the right below.

Table 9.8.14: (Based on Table 9.8.5 data)

Group		**Women** Group size	**Women** %	**Men** Group size	**Men** %	**Both sexes** Group size	**Both sexes** %
Individuals		1	--	1	--	1	--
Smaller subgroups	Law: Masters	60	85	60	90	120	88
	Doctoral	60	55	300	60	360	59
	Bus.: Masters	60	45	20	50	80	46
	Docroral	60	15	100	20	160	18
Larger subgroups	Law	120	70	360	65	480	66
	Business	120	30	120	25	240	27
Population		240	50	480	55	720	53

NOTES: 1. Simpson's Paradox is so surprising, particularly when first encountered, that it is easy to lose sight of key *statistical* issues.

- The proportions *are* correctly calculated – Simpson's Paradox is *not* the result of mistakes in arithmetic.
- Simpson's Paradox is *not* confined to attributes that are proportions (as discussed in Section 3 on page 9.59).
- Simpson's Paradox occurs when subdividing (or combining) data for categories and only in *some* circumstances.

Lessons for data-based investigating are:

∗ recognize and manage the (surprising) property of proportions discussed in Section 4 on pages 9.59 and 9.60;

∗ manage relevant non-focal explanatory variates – this includes the possibility of sometimes being able to identify an appropriate level of subdivision at which to make comparisons (as in Table 9.2.6 on page 9.15 of Figure 9.2).

There is then no 'paradox' for a *clear* Question investigated with an adequate Plan, suggesting that the name Simpson's *Paradox* can be misleading;

2. Four more illustrations of Simpson's Paradox are:

Table 9.8.15: The context is the same as that of Table 9.8.1 on page 9.58 but there are now *six* programs (A, ..., F) instead of two schools (Law, Business). Like Table 9.8.1, there is a *lower* percentage of women admitted overall but a *higher* percentage for *each* of the six programs.

Table 9.8.15:

PROGRAM	WOMEN Number of Applicants	WOMEN ADMISSIONS Number	WOMEN ADMISSIONS %	MEN Number of Applicants	MEN ADMISSIONS Number	MEN ADMISSIONS %
Archeology	108	89	**82**	825	512	**62**
Biology	25	17	**68**	560	353	**63**
Chemistry	593	219	**37**	325	114	**35**
Drama	375	131	**35**	417	138	**33**
English	393	106	**27**	191	48	**25**
French	341	27	**8**	373	22	**6**
All	1,825	589	**32**	2,691	1,187	**44**

Table 9.8.16: Baseball batting averages – the batter with the *lower* average for the whole season has a *higher* average in both *half* seasons. Recalling Section 7 and Note 1 above, it is of interest to develop a Plan to answer the Question of which batter to take if only one can be chosen.

Table 9.8.16:

Time Period	BATTER #1 Hits	BATTER #1 At bats	BATTER #1 **Average**	BATTER #2 Hits	BATTER #2 At bats	BATTER #2 **Average**
First half	15	70	**.214**	25	130	**.192**
Second half	15	50	**.300**	80	280	**.286**
Whole season	30	120	**.250**	105	410	**.256**

Table 9.8.17: Death rates (per 1,000 lives) in two regions of the U.S. for smokers and non-smokers.

Table 9.8.17:

LOCATION	SMOKERS Deaths	SMOKERS Policies	SMOKERS **Rate**	NON-SMOKERS Deaths	NON-SMOKERS Policies	NON-SMOKERS **Rate**
Nashville	6	900	**6.67**	7	1,100	**6.36**
Los Angeles	5	1,100	**4.55**	3	700	**4.29**
Either	11	2,000	**5.50**	10	1,800	**5.56**

[These data were gathered by a life insurance company which was issuing whole life policies countrywide on a non-medical issue basis; in 1986, 3,800 policies were issued to males aged 40-45. The company's files were kept in two locations – Nashville for policies issued east of the Mississippi and Los Angeles for policies issued west of the Mississippi. Nashville issued 2,000 policies and processed 13 deaths, Los Angeles issued 1,800 policies and processed 8 deaths.]

REFERENCE: Dolins, J.G.: Actuaries ... be careful! *The Actuary*, March, 1989, page 11.

(*continued overleaf*)

NOTES: 2. **Table 9.8.18:** Effect of jury challenges on conviction rates in trials in the U.K.
(cont.)

Table 9.8.18: DEFENDENT STATUS	...NO CHALLENGE... Number of Trials	CONVICTIONS Number	%	.......CHALLENGE...... Number of Trials	CONVICTIONS Number	%
Guilty	20	16	**80**	70	42	**60**
Innocent	10	0	**0**	0	0	**0**
Either	30	16	**53**	70	42	**60**

[In early 1987, an article by Bernard Levin in *The Times* raised the question of whether jury challenges assist those who are guilty in avoiding conviction. Mr. Levin concluded this was *not* the case, on the basis of data showing a conviction rate of 53% in trials with no challenges, *lower* than the conviction rate of 60% in trials *with* challenges. However, this answer does *not* necessarily follow from these conviction rates; in the *hypothetical* data in Table 9.8.18 (at the right above), the conviction rate for *guilty* defendents is substantially *higher* in trials with *no* challenges. Unfortunately, this counter-argument is speculative because the number of defendents *actually* guilty and innocent, and the rates of challenge and of conviction in both these groups, are not readily accessible. Nevertheless, an article in a major newspaper which uses flawed reasoning from data to answer a Question on a substantive issue is a serious matter.]

REFERENCE: Hill, I.D.: Rebutting the media. *The Royal Statistical Society NEWS & NOTES* **16** (#1), September, 1989, page 4.

There is discussion and further illustrations of Simpson's Paradox in Wagner, C.H.: Simpson's Paradox in Real Life. *American Statistician* **36** (#1, February): 46-48 (1982).

1. Referring to the data in Table 9.8.17 overleaf at the lower right of page 9.61, suggest a plausible explanation for the *lower* death rates for both smokers and non-smokers whose files were kept in Los Angeles, compared with those kept in Nashville.

2. Referring to the matters raised by the numbers in Table 9.8.18 at the right above, outline how you would try to *reduce* the uncertainties which are present and so obtain an Answer with fewer limitations about the effect(s) of jury challenges on conviction rates for the guilty in the U.K.
   - What effect(s) of jury challenges on the conviction of *innocent* defendents is indicated by the numbers in Table 9.8.18? Explain briefly.

8. Appendix: Simpson's Paradox and Interaction

For extending the discussion of Simpson's Paradox on the first six sides (pages 9.57 to 9.62) of this Figure 9.8, for convenience in this Appendix (including labelling the three diagrams to the right of Tables 9.8.19 to 9.8.21 below and on the facing page 9.63) we use the notation defined near the middle of page 9.58:

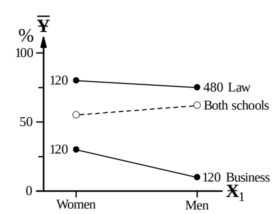
$\mathbf{X}_1$ is an applicant's sex (female, male), $\mathbf{X}_2$ is the school applied to (Law, Business), $\mathbf{X}_3$ is the level of study [Masters, Doctoral],
$\mathbf{Y}$ is the response to an applicant (admitted, not admitted) or time for degree completion (minimum, longer),
$\overline{\mathbf{Y}}$ [the *average* of ($\mathbf{Y}$)] is the *percentage* of applicants admitted or who complete their degree in the minimum time.

The diagrams illustrating Simpson's Paradox to the right of Tables 9.8.1 to 9.8.6 (on pages 9.58 and 9.59) are reminiscent of a diagram showing interaction (*e.g.*, in Note 28 on page 9.21 in Figure 9.2); however, there are *differences*:

- the Simpson's Paradox diagrams have an additional (dashed) line for the overall $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship;
- the instances of Simpson's Paradox in Tables 9.8.1 to 9.8.6 have only *parallel* (solid) lines for the $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationships for different values of $\mathbf{X}_2$ – that is, there is *no* interaction of $\mathbf{X}_1$ and $\mathbf{X}_2$ in their effects on $\mathbf{Y}$.

This restriction is *removed* in (another) reworking of Table 9.8.1 and its diagram in Table 9.8.19 below, where there *is* interaction of $\mathbf{X}_1$ and $\mathbf{X}_2$ in their effects on $\mathbf{Y}$ because the two solid lines in the diagram to the right of the Table are *not* parallel.

Table 9.8.19: SCHOOL	..........WOMEN........... Number of Applicants	ADMISSIONS Number	%	..............MEN............. Number of Applicants	ADMISSIONS Number	%	.......BOTH SEXES...... Number of Applicants	ADMISSIONS Number	%
Law	120	96	**80**	480	360	**75**	600	456	**76**
Business	120	36	**30**	120	12	**10**	240	48	**20**
Both schools	240	132	**55**	600	372	**62**			



Thus, interaction *may* be involved in Simpson's Paradox but is not *required* for it to occur.

Earlier discussion at the upper left of page 9.59 and on page 9.60 in Section 6, and in this Appendix, reminds us that Simpson's Paradox and interaction *both* involve (estimated) values of *conditional* probabilities for $\mathbf{Y}$, BUT:

- Simpson's Paradox involves comparing these probabilities conditioned on two (or three) of the $\mathbf{X}$s with probabilities conditioned on one *fewer* (one or two) $\mathbf{X}$s; WHEREAS:
- interaction is absent or present depending on the values of probabilities with the *same* conditioning on the $\mathbf{X}$s – these values determine whether the corresponding lines are or are not parallel.

NOTES: 3. Illustration of Simpson's Paradox from comparing *across* Tables 9.8.1 to 9.8.6 can overshadow comparisons *down* such tables. For example, in Table 9.8.1 (reworked as Table 9.8.9 on page 9.60), the six **bold** percentages for $\mathbf{X}_2$ (80 and 20, 75 and 15, 76.$\dot{6}$ and 17.5) address a Question *different* from possible sex discrimination:

(*continued*)

Figure 9.8. OBSERVATIONAL PLANS: Simpson's Paradox (continued 3)

NOTES: (cont.) 3. ● *How do the admission standards of the Law and Business schools compare?*
The (hypothetical) data in Table 9.8.9 on page 9.60 (and Tables 9.8.19 to 9.8.21 on the facing page 9.62 and below on this page) indicate an appreciably *higher* admission standard for Business than for Law, unless the abilities of the two applicant pools are remarkably different.

4. A second reworking of Table 9.8.1 and its diagram is given below in Table 9.8.20; as in Table 9.8.19, there *is* interaction of $\mathbf{X}_1$ and $\mathbf{X}_2$ in their effects on $\mathbf{Y}$ but, for both schools combined, there is *no* sex difference in proportions, due to cancellation of effects in *opposite* directions for the schools individually.

Table 9.8.20:

SCHOOL	WOMEN: Number of Applicants	WOMEN: ADMISSIONS Number	WOMEN: ADMISSIONS %	MEN: Number of Applicants	MEN: ADMISSIONS Number	MEN: ADMISSIONS %	BOTH SEXES: Number of Applicants	BOTH SEXES: ADMISSIONS Number	BOTH SEXES: ADMISSIONS %
Law	120	96	**80**	180	153	**85**	300	249	**83**
Business	120	24	**20**	180	27	**15**	300	51	**17**
Both schools	240	120	**50**	360	180	**50**			

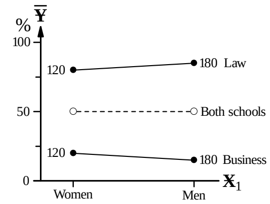
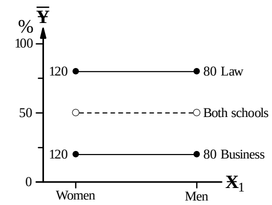


Table 9.8.21 below and its diagram show, like Table 9.8.20, *no* sex difference for both schools combined but this is now a consequence of the *individual schools* also showing this same behaviour – there is *no* interaction.

Table 9.8.21:

SCHOOL	WOMEN: Number of Applicants	WOMEN: ADMISSIONS Number	WOMEN: ADMISSIONS %	MEN: Number of Applicants	MEN: ADMISSIONS Number	MEN: ADMISSIONS %	BOTH SEXES: Number of Applicants	BOTH SEXES: ADMISSIONS Number	BOTH SEXES: ADMISSIONS %
Law	120	96	**80**	80	64	**80**	200	160	**80**
Business	120	24	**20**	80	16	**20**	200	40	**20**
Both schools	240	120	**50**	160	80	**50**			



5. Across Tables 9.8.1 to 9.8.6 on pages 9.58 and 9.59 and Tables 9.8.19 to 9.8.21 on the facing page 9.62 and above, different weights in the proportion calculations (like those in Table 9.8.8 on page 9.60) yield a noteworthy *variety* in the percentages for women compared to those for men. This is summarized in Table 9.8.22 at the right below; three categories are distinguished.

- ○ In four tables, there *is* an $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship, there is no interaction of $\mathbf{X}_1$ and $\mathbf{X}_2$ in their effects on $\mathbf{Y}$ and, in two of the tables, the $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship is *un*exceptional in light of the effect of subdivision by $\mathbf{X}_2$; by contrast, in Table 9.8.3 and Table 9.8.6 between the first and second levels of subdivision by $\mathbf{X}_2$, the exceptional behaviour is the $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship *disappearing* when the data are subdivided by $\mathbf{X}_2$.
  - – Notation like (1,3) [or (1,2)] on Table 9.8.5 (or Table 9.8.6) in Table 9.8.22 refers to the first and third (or first and second) levels of subdivision by $\mathbf{X}_2$.
- ○ In four tables, there is *no* $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship but three of these are '*false* negative' Answers – when the data are subdivided by $\mathbf{X}_2$, there *is* an $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship and so they are designated 'exceptional' in the fourth column.
  - – In Table 9.8.20, interaction is the *reason* for the exceptional behaviour but interaction is absent in the other three tables.
- ○ In five tables, there *is* (again) an $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship, interaction is absent or incidental, and each is a case of the exceptional behaviour known as Simpson's Paradox.

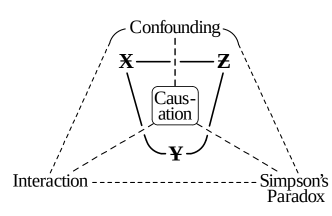
Table 9.8.22: $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ **RELATIONSHIPS IN NINE TABLES**
(SP in the fourth column denotes 'Simpson's Paradox')

Table	Relationship	Interaction	Exceptional behaviour
9.8.3	Yes	No	Yes
9.8.4	Yes	No	No
9.8.5 (1,3)	Yes	No	No
9.8.6 (1,2)	Yes	No	Yes
9.8.2	No	No	Yes
9.8.6 (2,3)	No	No	Yes
9.8.20	No	Essential	Yes
9.8.21	No	No	No
9.8.1	Yes	No	Yes: SP
9.8.5 (1,2)	Yes	No	Yes: SP
9.8.5 (2,3)	Yes	No	Yes: SP
9.8.6 (1,3)	Yes	No	Yes: SP
9.8.19	Yes	Incidental	Yes: SP

Apart from understanding the properties of proportions and weighted averages and using an adequate Plan (discussed in Sections 4, 5 and 7 and Note 1 on pages 9.59 to 9.61), the summary in Table 9.8.22 above reminds us that:

- ⊙ Simpson's Paradox is merely the *most* exceptional case (*change of direction* of an $\mathbf{X}_1$-$\overline{\mathbf{Y}}$ relationship) in a context (involving *discrete* variates) that can give rise to less exceptional or even *un*exceptional behaviour;
- ⊙ interaction is rarely the reason for the exceptional behaviour (only in Table 9.8.20 above).

6. In meeting the obligation to deal with *relationships* in an introductory statistics course, the lengthy discussion (*e.g.*, in Figure 9.2, Sections 1 to 15 and Appendix 1 on pages 9.5 to 9.29 and Figures 9.8 to 9.13 on pages 9.57 to 9.80) shows the (unexpected) complexities, for only *three* variates, arising from issues of causation, confounding, interaction and Simpson's Paradox. The schema at the right reminds us there are common themes *and* differences among these four matters – see also Appendix 2 on pages 9.67 and 9.68 in the following Figure 9.9.

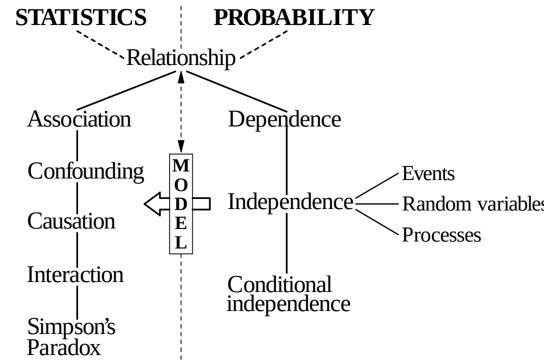


(*continued overleaf*)

NOTES: 7. As summarized on the left of the schema at the right, a relationship in *statistics* is often considered in terms of one or more of association, confounding, causation, interaction and Simpson's Paradox.

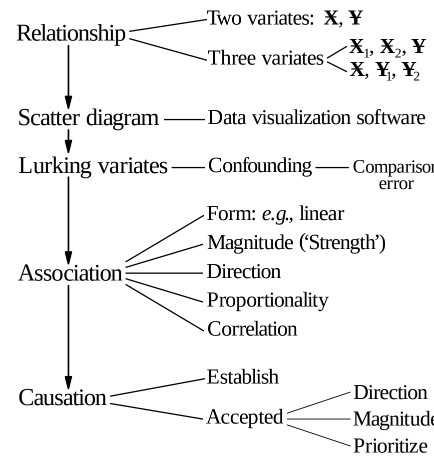
In *probability* (on the right of the schema), a relationship is considered in terms of *dependence*, which comes in great variety and is often difficult to mathematize; as a consequence, introductory courses emphasize ***in***dependence, as it applies to events, random variables and processes. Even the first two of these three involve an appreciable set of ideas and may be all a course has time to discuss.

Connection between statistical and probabilistic considerations of a relationship arises in the probability *models* statistics uses in the Analysis stage of the FDEAC cycle.

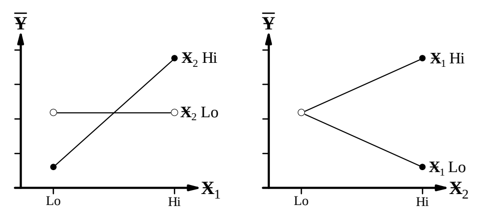


- Emphasis on *in*dependence in introductory courses can obscure the fact that independence is a mathematical *idealization*. In the real world, *de*pendence is the norm – it may be that the behaviour of *every* particle in the universe depends on (*i.e.*, is affected by) *every other* particle, no matter how minute the degree of dependence.
  - This may be why lurking variates are usually so *numerous* when answering Questions with a causative aspect.

The schema on page 9.5 in Figure 9.2, which provides a framework for our discussion of data-based investigating of statistical relationships, is shown again at the right; *un*surprisingly, it is a more detailed version of the left-hand ('statistics') side of the schema above it.



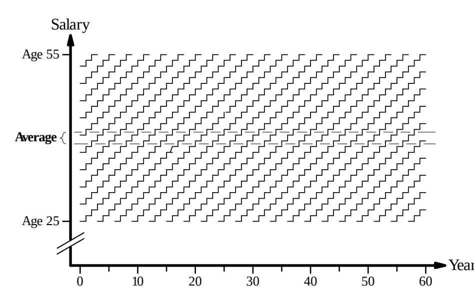
8. The (equivalent) diagrams at the right show the effects of two (binary) focal variates X_1 and X_2 on (the average of) Y; one focal variate is on the horizontal axis of a diagram, the other distinguishes the two lines by its level.



- The *non*parallel lines show there is an X_1-X_2 *interaction*.
- The left-hand diagram shows that X_1 and Y are *conditionally independent* when X_2 is Lo (the relevant line has *zero* slope) but *not* when X_2 is Hi.
- The right-hand diagram shows there is *no* conditional independence of X_2 and Y – *neither* line has zero slope [see also diagram (4) on the lower half of page 9.75 and its discussion below this diagram in Figure 9.12].

  Thus, equivalences between statistical and probabilistic views of relationships are not always straight forward.

9. Table 9.8.7 on page 9.59 illustrates, for *averages* of a *continuous* response variate, a phenomenon analogous to Simpson's Paradox. A similar 'paradox' occurs for an organization of (say) 10 employees who each work for the organization for (say) 30 years or until age 55, whichever comes first. In constant dollars (*i.e.*, with the effect of inflation removed), each employee is hired at a salary determined by their experience, taken as proportional to their age which is at least 25; each receives the *same* salary *in*crease each year. At start-up, the organization hires people with a variety of levels of experience/age but a steady state is then maintained by hiring a new employee aged 25 to replace a person leaving the organization at age 55. This situation is portrayed graphically at the right above for the first 60 years of the organization's existence – each stepped line represents one employee's salary over time. Despite every employee's *in*creasing salary, the organization's employee experience/age structure (a 'lurking variate') and, hence, the *average* salary for the organization, remain nearly *constant* over time.



Although the organization described is an idealization, it is reasonable model for a college or university with a constant faculty complement. The 'paradox' could result in a specious claim, at the time of salary negotiations, that faculty salaries are being unfairly constrained because the *average* salary is not increasing over time.