University of Waterloo W. H. Cherry

RELATIONSHIPS IN STATISTICS: Terminology for Comparative Plans
The Protocol for Choosing Groups

1. Background I – Causation in Statistics: a Definition [optional reading]

To define *formally* in statistics what it means to say **X** *causes* **Y** in a (target) population, we state three criteria:

(1) **LURKING VARIATES:** Ensure *all other* explanatory variates $Z_1, Z_2, \ldots, Z_k$ hold their (same) values for *every* population element when $X = 0$ and $X = 1$ (sometimes phrased as: *Hold all the* Z_i ***fixed*** *for* .....).

(2) **FOCAL VARIATE:** Observe the population **Y**-values, and calculate an appropriate attribute value, under *two* conditions:
- ⊙ with *every* element having $X = 0$;
- ⊙ with *every* element having $X = 1$.

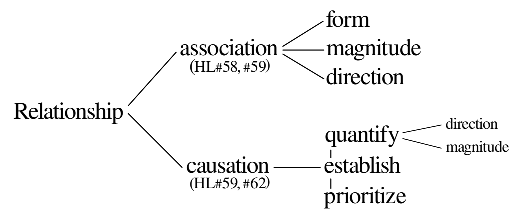
(3) **ATTRIBUTE:** Attribute(Y, perhaps some of $Z_1, Z_2, \ldots, Z_k \mid X = 0$) $\neq$ Attribute(Y, perhaps some of $Z_1, Z_2, \ldots, Z_k \mid X = 1$); those of $Z_1, Z_2, \ldots, Z_k$ *included* in the attribute will have the *same* values when $X = 0$ and $X = 1$ under (1).

For example, the z values must be the *same* when using least squares estimates [as given in equation (HL63.1) at the right] to *compare* simple linear regression *slopes* when $X = 0$ and $X = 1$.

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{n} y_j (z_j - \bar{z})}{\sum_{j=1}^{n} (z_j - \bar{z})^2} \quad \text{-----(HL63.1)}$$

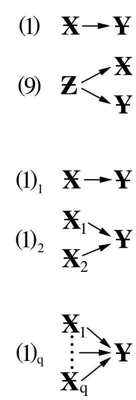
2. Background II – Investigating Statistical Relationships: Three Types of Causal Questions [optional reading]

Relationships investigated in statistics, which *we* describe in terms of variates, are often encountered as *associations*; investigating associations includes identifying their characteristics and/or the reasons (causal or otherwise) for them (see also Statistical Highlights #58 and #59). This Section 2 is concerned with comparative Plans for investigating relationships where causation is to be established or *is* involved; the focus on the **X**-**Y** relationship being *causal* means that a *change* can (potentially) be induced in **Y** by *changing* **X**. These matters are summarized in the schema at the right, which reminds us that:



∗ association is usually characterized by its *form*, *magnitude* or *direction*;
  – correlation (see Statistical Highlight #66) is one measure of magnitude ('strength') for a straight-line association; form can also be *non*-linear;

∗ it is useful to distinguish three types of Questions with a causative aspect:
  – **Establishing** whether **X** *is* a cause of **Y**, usually with a view to manipulating **X** to produce a (desired) change in **Y** – the quintessential example is whether cigarette smoking is a cause of lung cancer (and other life-threatening diseases), the topic of tens of thousands of data-based investigations over several decades starting in the 1940s. Establishing that an observed association of **X** and **Y** is causation of **Y** by **X** is answering the Question whether the relevant causal structure (shown again at the right from page HL59.1 in Statistical Highlight #59) is case (1) or case (9) [= case (12)].
  – **Quantifying** the relationship between **X** (or, more commonly, $X_1, X_2, \ldots, X_q$) and **Y**; this arises in the statistical area of *Design of Experiments* (DOE) – for example, the effect of temperature, humidity, light, fertilizer and insecticide levels on the growth of seedlings in a greenhouse. Quantifying a causal relationship is, in essence, investigating the case (1) causal structure – the subscripts on the case number now remind us that the Plan needs to reflect the number of focal variates involved.
  – **Prioritizing** causes by the size of their effect is the domain of (data-based) process improvement – trying to identify the *most important* cause (usually of excessive variation in the process output, **Y**) from among many causes $X_1, X_2, \ldots, X_q$.



Questions which involve *establishing* and *quantifying* causal relationships are typically part of the *same* investigation. For example, in the Physicians' Health Study (described in Figure 10.2 of the STAT 231 Course Materials) of the effect of taking aspirin on heart disease, *two* Questions, in the context of an appropriate target population, are:

- does taking aspirin reduce heart-attack risk?
- is the reduction in heart-attack risk due to taking aspirin large enough to be practically important?

The Physicians' Health Study had to answer *both* Questions; in *in*formal discussion, it is easy to consider only *one* of the Questions and overlook the other.

Similarly, when *prioritizing* causes in process improvement investigations, investigators should:

- verify that the suspected (most important) cause *is* a cause of the (variation in the) response variate(s);
- validate that the proposed Answer *does* address the Question – that the proposed 'solution' *does* solve the 'problem'.

NOTE: 1. In STAT 231, *establishing* causation is discussed in Chapter 11 of the 2004 Course Notes but the emphasis is on *quantifying* the relationship between *one* focal variate and a response variate – for example, see Chapters 7, 10 and 15; extension to more than one focal variate is taken up in STAT 332. *Prioritizing* causes is pursued in STAT 435.

3. Terminology for Comparative Plans – The Protocol for Choosing Groups [The title matter of this Highlight #63.]

The three criteria defining what *we* mean by causation, reviewed in Section 1 at the top of this page, involve observing a *population* under two values of the focal variate: with *all* the elements having $X = 0$ and with *all* the elements having $X = 1$. We try to approach this ideal in a *sampling* context by having *two* samples, one with its units having $X = 0$ and the other with its units

 (*continued overleaf*)

having $\mathbf{X} = 1$; each sample 'represents' the population under one of the two conditions, in the usual statistical sense of sample attributes being *estimates* of respondent population attributes. When the two samples are *compared* to quantify the change in (the average of) $\mathbf{Y}$ corresponding to a change in $\mathbf{X}$, each *non*-focal explanatory variate must have the *same* value in both samples; otherwise, there is (likely to be) comparison error. For comparative Plans for quantifying relationships, we distinguish:

- ∗ an **experimental** Plan – a comparative Plan in which the *investigator(s)* (*actively*) assign the value of the focal (explanatory) variate to each unit in the sample (or in each block);
- ∗ an **observational** Plan – a comparative Plan in which, for each unit selected for the sample, the focal (explanatory) variate (*passively*) takes on its 'natural' value ***un****influenced* by the investigator(s).

This distinction reflects two types of populations encountered in data-based investigating of relationships.

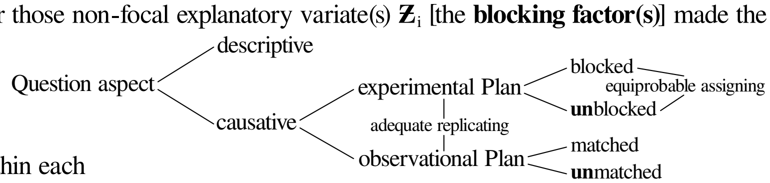
- ● A population in which all (or most) elements have *one* value of a focal variate of interest, whose value it *is* feasible to change.
  - – An example is a new drug to treat a serious disease – no one would already be taking the drug but it could be given to some participants ($\mathbf{X} = 1$) and withheld from others ($\mathbf{X} = 0$) in a clinical trial (an *experimental* Plan – see Note 7 on page HL63.4).
- ● A population in which each element has one of *two (or more)* values ($\mathbf{X} = 0, 1, ...$) of a focal variate of interest, whose value it is *not* feasible to change for any element – see Note 9 on pages HL63.5 and HL63.6.
  - – Instances of such focal variates are age, sex, marital status and income – their investigation necessarily involves an *observational* Plan; changes in people's dietary or exercise habits can be imposed but compliance is difficult to achieve.

It is investigators' *in*ability to assign elements' (or units') focal variate values that restricts choice of Plan type and so weakens ability to manage comparison error; this matter is pursued in Statistical Highlights #10 and #9.

For comparative Plans to answer a Question with a causative aspect, the **protocol for choosing groups** specifies whether the units of the sample will be selected so they form groups that can be used to reduce the limitation imposed on an Answer(s) by comparison error – relevant Plan components are shown in the schema at the right below.

∗ **Blocking** in an *experimental* Plan: forming groups of units (the **blocks**) with the *same* values of one or more non-focal explanatory variates; units within a block are then assigned *different* values of the *focal* variate. THUS:

Blocking meets 'lurking variates' criterion (1) for those non-focal explanatory variate(s) $\mathbf{Z}_i$ [the **blocking factor(s)**] made the same within each block. SO THAT:



Whether the Question involves establishing causation or quantifying a treatment effect, blocking *prevents confounding* of the focal variate with the $\mathbf{Z}_i$ made the same within each block, reducing the limitation imposed on Answer(s) by *comparison* error.

- ● By making their values the *same* for one or more $\mathbf{Z}$s (*i.e.*, holding their values 'fixed') within blocks in an experimental Plan, blocking reduces variation in $\mathbf{Y}$ and so has the additional benefit of decreasing *comparing* imprecision.
  - – This additional benefit of blocking is analogous to that of *stratifying* in reducing *sampling* imprecision, as indicated in last lines of the two branches of the schema at the centre right of page HL63.6 in Note 10. [This analogy is sometimes interpreted as showing that stratifying in survey sampling is merely an instance of blocking, but this interpretation (unhelpfully) downplays the different contexts and intents of blocking and stratifying.]

∗ **Equiprobable assigning (EPA) [random assigning** or **randomization]:** using a probabilistic mechanism (described in the protocol for choosing groups) in an *experimental* Plan to assign the values of the focal variate with *equal* probability:

\+ across the units of each block in a blocked Plan; \+ to each unit in the sample in an *un*blocked Plan.

Equiprobable assigning provides a basis for theory which relates comparing imprecision to level of replicating; thus, EPA, *in conjunction with EPS and adequate replicating*, provides for quantifying comparing imprecision arising from unblocked, unknown and unmeasured non-focal explanatory variates and so allows a particular investigation to set group sizes which are likely to yield an Answer(s) with limitation imposed by comparison error that is acceptable in the Question context.

∗ **Matching** in an *observational* Plan: forming groups of units with the *same* values of one or more non-focal explanatory variates but *different* values of the *focal* variate. THUS:

Matching meets 'lurking variates' criterion (1) [overleaf at the top of page HL63.1] for those non-focal explanatory variate(s) $\mathbf{Z}_i$ made the same within each group. SO THAT:

Whether the Question involves establishing causation or quantifying a treatment effect, matching *prevents confounding* of the focal variate with the $\mathbf{Z}_i$ made the same within each group, thus decreasing comparing imprecision and so reducing the limitation imposed on Answer(s) by *comparison* error.

- – **Subdividing:** a form of *matching* used in an *observational* Plan in which the each value of the focal variate for the units of the sample is *subdivided* on the basis of the values of one or more *non*-focal explanatory variates that may be *confounded* with the focal variate under the Plan – see Table HL63.3 and its discussion on pages HL63.4 and HL63.5.

  We can think of *subdividing* as *matching* at an *aggregate* (rather than an *individual*) level; subdividing therefore has the *same* statistical benefit as matching for the non-focal explanatory variate(s) that are the basis for the subdividing.
  - ○ If subdividing is going to manage *only one* non-focal explanatory variate that is a (potential) source of comparison error, it *may* not be cost effective to devote the resources needed to obtain the relevant additional data.

(*continued*)

University of Waterloo

W. H. Cherry

RELATIONSHIPS IN STATISTICS: Terminology for Comparative Plans (continued 1)

NOTES: 2. Where the definitions given on the facing page HL63.2 of blocking and matching refer to values of non-focal explanatory variates being the *same*, in practice the values may only be *similar.*

3. The groups of elements (or units) are called *blocks* in an experimental Plan but there is no such general term in an observational Plan; however, when the groups contain *two* elements (or units), they may be referred to as *matched pairs* – see Table HL63.1 at the right – but a *block* of two elements (or units) may also be referred to as a 'pair.'

Table HL63.1
Terminology for Comparative Plans

Plan	**Process**	**Group**
Experimental	Blocking	Block
Observational	Matching	(Matched pair)

- A comparative Plan involving pairing is usually our first encounter with the concepts of blocking or matching, to illustrate their role in managing comparison error.

4. In DOE, non-focal explanatory variate(s) made the same within blocks are called **blocking factor(s)**; in data-based investigating to improve industrial processes, typical blocking factors are days, shifts, batches of raw material, machine spindles or filler heads, moulding machines, moulds, or cavities within moulds.
   - The values of a blocking factor among blocks should be chosen to make its sample attribute (*e.g.*, its average or distribution) similar to its respondent (or study) population attribute.
   - An entity that is the same both within *and* among blocks (like the measuring process) is *not* a blocking factor but is part of what *defines the study population/process* – for example, data for an investigation collected on *one* day and *one* production shift. *If* such factors as day or shift have an appreciable effect on the response, the limitation imposed on the Answer by *study* error is more severe (comparison error is traded for study error).

5. Just as equiprobable *selecting*, in conjunction with *adequate replicating*, provides a theoretical basis for quantifying the likely size of sample error when estimating a (respondent) population average, so equiprobable *assigning*, in conjunction with *both* EPS *and* adequate replicating, provides the *same* benefit when estimating an average difference in (two) populations in an experimental Plan. This and other parallels between EPS and EPA are discussed in Note 10 on page HL63.6.

6. *We* use *different* terms for two processes which are similar but are used to manage different categories of error.
   - *Subdividing* (of a sample) on an *explanatory* variate to manage comparison error due to confounding by this variate, usually in an observational Plan used to answer a Question with a causative aspect.
   - *Stratifying* (of a population) on a *response* variate (or, in practice, on an explanatory variate that *stands in* for it) to make an Answer(s) more useful and/or to manage sample error – see Statistical Highlight #85.

   Elsewhere, *both* processes may be called 'stratifying.' There is further discussion of subdividing in Section 5 on pages HL63.4 and HL63.5 and of stratifying on the lower half of page HL63.5 later in Note 8.

4. Experimental Plans – Sample selecting and Blocking

The statistical ideal for sample selecting in *any* Plan is to have a *known* inclusion probability for each element of the respondent population; an example is *equi*probable selecting. For a Question with a *descriptive* aspect, if this ideal is not met, severe limitation is imposed on an Answer by *sample* error. However, experimental Plans to answer Questions with a *causative* aspect commonly do *not* use probability selecting because it is not feasible to implement it.

∗ For example, in data-based investigating to improve a manufacturing process (*e.g.*, by identifying and removing causes of excessive variation in the process output), the items manufactured by the process are often shipped away from the manufacturing plant as they are made and investigators are then forced (quickly) to use recent production, or a subset of it, as the sample – a *sample of convenience.* Three factors alleviate this *statistically* unsatisfactory state of affairs:

– With *stable* processes [where the distribution(s) of the output response variate values remain (essentially) the same from one time period to another], a 'snapshot' of the process in time (like recent production) may often have attribute values that are *close* to those of the process in the long-term.

– Answers are derived from *differences* in sample attributes; such Answers may have less severe limitation imposed by sample error than Answers based on sample attribute values which do *not* involve taking a difference.

  + An illustration is the Physicians' Health Study (of the effect of aspirin on heart disease), which used about 22,000 male doctors as the sample – half the doctors took aspirin and half took a placebo. It is likely that the incidence of heart attacks among doctors differs appreciably from that for the target population of all males, but the *difference* in incidence of heart attacks caused by taking aspirin may be much more similar among doctors and all males. (Two newspaper reports of this investigation are reprinted in Figure 10.2 of the STAT 231 Course Materials.)

– Investigators may have a level of (extra-statistical) process knowledge that enables them to assess how close relevant attributes of recent production are likely to be to the corresponding long-term process attributes – informed human judgement seems to be better at sample selecting in such situations than it does when answering a Question with a *descriptive* aspect, but it is still far from the statistical ideal. [This matter is pursued in Statistical Highlight #83.]

(*continued overleaf*)

The use of judgement selecting in the sampling protocols of comparative Plans illustrates the divergence between the statistical ideal and statistical practice under real-world constraints. Limitations imposed by the use of judgement selecting are:

- ⁎ we can no longer say increased replicating reduces *sampling* imprecision; that is, we can no longer say increased sample size reduces the *likely* magnitude of sample error – see Appendix 3 on pages HL74.5 to HL74.8 in Statistical Highlight #74;
- ⁎ more generally, the theoretical basis is gone for interpreting formal methods of data analysis like confidence intervals and tests of statistical significance. [Regrettably, both limitations are commonly ignored in practice.]

When answering a Question with a *causative* aspect, statistical best practice to manage comparison error (to reduce the limitation it imposes on the Answer) is to:

- ● block (to the extent that is feasible in the Question context) on known and measured lurking variates,
- ● use EPA to manage unblocked, unmeasured and unknown lurking variates,

[summarized in the precept: *Use blocking to manage what is known, probability assigning to manage what is* un*known*].

Unfortunately, there may be practical or ethical constraints on investigators' freedom to implement best practice; for instance:

- ⁎ A block which is an individual participant in an investigation may not practically be able to be assigned both values of the focal variate. For example, in the Physicians' Health Study (see Figure 10.2 of the STAT 231 Course Materials) of the effect of aspirin on heart disease in males, the investigation would have gone on for too long if each participant had been required to take aspirin for several years and *not* to take aspirin for another period of the same length. For this (and other) reasons, the experimental Plan for the Physicians' Health Study was *un*blocked [see also the last bullet (●) of Note 9 on page HL63.6].
- ⁎ It would be unethical to assign human participants to the smoking group when investigating health effects of cigarette smoking;
  - = in addition to ethical considerations, it is *un*likely that many non-smokers would be able to take up smoking for the investigation or that most smokers would be prepared to quit if assigned to the non-smoking group.

  Ethical issues *can* be managed but considerable resources may be needed to achieve compliance among participants when the focal variate in medical investigations with an experimental Plan involves exercise levels or dietary practices.

NOTE: 7. A special class of comparative experimental investigation is a **clinical trial**, used in medical research to assess the efficacy of new forms of treatment (*e.g.*, drugs, surgery); because the elements are *humans*, a technique called **blinding** is used (where feasible) because of its statistical benefits. [To be *blind* means not to know, for any element, whether it is in the *treatment* group or the *control* group (which usually receives a dummy treatment known as a **placebo**)]. As shown in Table HL63.2 at the right, blinding is used to manage comparison error and/or measuring inaccuracy, depending on the degree to which it is (or can be) implemented – for instance, blinding of participants is often *not* feasible when the focal variate involves exercise level or diet.

Table HL63.2

Blinding of....	Short name	Statistical benefit
Participants	Single blind	Reduced risk of *comparison error*
Treatment administrators	Double blind	Reduced risk of *comparison error*
Treatment assessors	Triple blind	Reduced *measuring inaccuracy*

5. Observational Plans – Sample selecting, Matching and Subdividing

The comments in Section 4 (overleaf on page HL63.3 and above) about the use of *judgement selecting* in experimental Plans are also generally applicable to observational Plans; similarly, *matching* reduces the limitation due to comparison error on Answer(s) from an observaltional Plan but, like blocking, matching may not be feasible in a particular Question context.

Subdividing samples from the respondent subpopulations with different values of the focal variate in an observational Plan, on the basis of a possible confounder Z_i, is illustrated in Table HL63.3 at the right for the case of *two* subpopulations. These hypothetical data for two samples (selected from subpopulations of non-smokers and smokers) of 10,000 people involve a response variate Y which is lung cancer status, a focal variate X which is smoking status, and Z_i is whether a unit has a family history of lung cancer, as a possible indicator of genetic predisposition to the disease; for simplicity, X, Y and Z_i are *binary* variates in this illustration. Each of the six sets of three table entries is the sample size ('Number') and the lung cancer 'Cases' as a number and a percentage of the sample size.

Table HL63.3	**Non-smokers ($X = 0$)**			**Smokers ($X = 1$)**		
	Number	**Cases**	**%**	**Number**	**Cases**	**%**
No family history ($Z_i = 0$)	9,000	63	0.7	8,900	712	8
Family history ($Z_i = 1$)	1,000	7	0.7	1,100	88	8
Both	10,000	70	0.7	10,000	800	8

The bottom line of Table HL63.3 shows a substantially higher proportion of lung cancer cases among the smokers; because this pattern *persists* in the upper two lines of the table when the data are subdivided by Z_i value, the association between smoking status and lung cancer status appears *not* to be due to (our type 2b) confounding by a genetic factor which determines *both* a unit's smoking status *and* its lung cancer status, at least in so far as family history is a measure of such a factor.

Unfortunately, such subdividing of sample data to manage the limitation imposed by comparison error on an Answer about an X-Y relationship from an observational Plan encounters three potential difficulties.

- ● Investigators have no control over the sample sizes after subdividing; if one or more of the X-Z_i combinations is rare, the resulting small sample size(s) *in*crease comparing imprecision and so increase(s) the limitation imposed by comparison error on an Answer about an X-Y relationship (in even the 'best case' situation of probability selecting of the samples).

RELATIONSHIPS IN STATISTICS: Terminology for Comparative Plans (continued 2)

- Obtaining the $\mathbf{Z}_i$ value for each unit in the samples may be difficult (and, hence, expensive) and such resource-intensive data manage only *one* possible confounder.
  - If data for two (or more) $\mathbf{Z}_i$ are collected, the ensuing subdividing into more numerous subsamples is likely to increase the limitation imposed [under probability selecting] by small sample size(s).
- Subdividing data in the manner of Table HL63.3 raises the possibility (*not* realized here) of the phenomenon known as Simpson's Paradox (and its accompanying limitation imposed on an Answer) – see Statistical Highlight #51.

NOTES: 8. In an (observational) **Case-Control** Plan (used in medical research, for example), units with a response of interest (say, lung cancer) [the 'Cases'] are matched on relevant explanatory variates (like, sex, age, region of residence) with units with*out* the response of interest (the 'Controls'). The two groups are then compared on the basis of the value of a focal variate of interest (cigarette smoking, say); appreciably higher levels of smoking among the *cases* would show *association* of smoking and lung cancer, indicating smoking may be a *cause* of the disease.

- A Case-Control Plan is used commonly:
  - when an experimental Plan would require resources beyond those available, OR:
  - as a cheaper forerunner to a possible experimental Plan to assess a promising but unconfirmed treatment effect.
- A Case-Control Plan makes the response and focal variates *appear* to be interchanged.
  - An illustration is in the 1993 newspaper article EM9359 *Fats raise risk of lung cancer in non-smokers*, which describes an investigation that compared the diets of 429 non-smoking women who had lung cancer with the diets of 1,021 non-smoking women who did *not* have lung cancer. The women all lived in Missouri, were of about the same age and represented "a typical American female population." The women filled out forms that asked about their dietary habits and they were divided into five groups based on the amount of fat and other nutrients they said they consumed. The investigation found that those with diets with the lowest amount of saturated fat and the highest amount of fruits, vegetables, beans and peas were the least likely to develop lung cancer. At the other end of the scale, 20 per cent of the women with the highest consumption of fat and diets lowest in fruits, vegetables, beans and peas had about six times more lung cancer.

    The *actual* response variate (lung cancer) and focal variate (level of dietary fat) *appear* to be interchanged solely as an artifact of the Case-Control Plan.
- Probability selecting is commonly *not* used for the cases and/or the controls, which has consequences for the limitation imposed by comparison error on Answer(s).
  - Cases are often a **sample of convenience** – units with a response of interest conveniently *available* to the investigator(s), like people with a particular disease in a hospital or clinic nearby to the investigator(s).
    - + Consequences of non-probability selecting to answer Question(s) with a descriptive or a causative aspect are discussed in Statistical Highlight #83 – recall also the discussion on the lower half of page HL63.3.
  - Controls are often selected *non*-probabilistically to meet the matching criteria; this *in*creases the limitation imposed by comparison error due to the selecting method, to be set against the *de*creased limitation imposed by comparison error due to the confounding which is managed by the matching.
    - + A way of selecting controls probabilistically is to form *strata* (or groups) of controls where the units in one stratum match one case; controls for the investigation are then selected probabilistically from these strata.
      - ⊙ While decreasing the limitation imposed by *comparison* error, such stratifying *in*creases the limitation imposed by *study* error, because the matching criteria which define the strata *restrict* the elements (or units) which can make up the study (and respondent) population of controls.

        A Plan should carefuly consider whether error from one source should be managed in a way that *in*creases error from another source – that is, whether there *is* a net gain in reducing the limitation on an Answer by managing one category of error in a way that *in*creases limitation due to *another* category – recall the discussion of comparison error and study error in Note 4 on the upper half of page HL63.3.
    - + When controls are selected *non*-probabilistically, there is no theoretical basis for an inverse relationship between sampling imprecision and (the square root of) of the sample size – see Statistical Highlight #21 – so there is no *statistical* reason why a larger sample size for controls will decrease comparing imprecision.
    - + The blocks in a blocked experimental Plan are also often selected *non*-probabilistically but, as discussed in Statistical Highlight #83, judgement selecting *may* still allow an experimental Plan to have *acceptable* limitation imposed by comparison error on an Answer to a Question with a causative aspect.

9. For 'focal variate' criterion (2), there are focal variates (like age and sex) whose values can*not* be *assigned* to units by the investigator(s) in an experimental Plan. For such variates, we avoid the stronger language of saying *increasing age* ***causes*** *loss of visual acuity* in favour of *increasing age is* ***associated with*** *loss of visual acuity.*

- Such associations are important in contexts like discrimination by sex or race where, for example, we compare the relevant population proportion with the proportion of women or a racial group in an employment or

(*continued overleaf*)

NOTES: (cont.) 9. ● other category. *Causation* (in the sense of our three criteria given in Section 1 on page HL63.1) by sex or race is not the issue with such associations, because there is no intention to change the value of the focal variate.

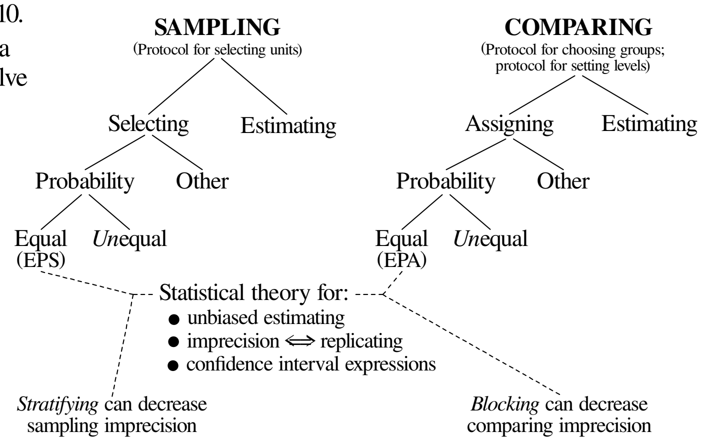
– We may also speak of the *reason* (rather than the *cause of*) why a population subgroup is under- or over-represented – for example, in an employment context we may consider relevant *qualifications*.

● Some focal variates (like cigarette smoking) cannot *ethically* be assigned to human units, which imposes limitations that arise from using animal units in an experimental Plan or human units in an observational Plan.

These matters are pursued in a discussion of Simpson's Paradox in Statistical Highlight #51.

● The ideal of criterion (2) ignores any *time* difference between the realization of the two conditions $\mathbf{X} = 0$ and $\mathbf{X} = 1$. In actual investigations, the two groups (usually samples) with units having $\mathbf{X} = 0$ and $\mathbf{X} = 1$ are observed concurrently but, in a cross-over Plan (like the oat bran investigation described in Note 3 on page HL9.6 in Statistical Highlight #9), there *is* a time difference between $\mathbf{X} = 0$ and $\mathbf{X} = 1$ for both half samples; any changes in units' *other* explanatory variates values over time may then be a source of comparison error.

10. Equiprobable *selecting* and equiprobable *assigning* are key components of the processes of sampling and (experimental) comparing, whose *similarities* are illustrated in the two tree diagrams in the schema at the right below – it is taken from page HL10.6 in Statistical Highlight #10.



● Investigations involving *comparing* (to answer a Question with a *causative* aspect) usually involve *sampling*; investigations involving *sampling* to answer a Question with a *descriptive* aspect need *not* involve *comparing*.

● **Probability selecting** means having *known* unit inclusion probabilities in the selecting process; introductory statistics courses emphasize *equi*probable selecting (EPS) as the basis of statistical theory for the behaviour of *sample* error under repetition.

– Here, *we* coin the term **probability assigning** (EPA) for having known *assigning* probabilities; *we* encounter mainly the special case of *equi*probable assigning, and we *hope* for a fortunate outcome of (roughly) equal numbers of units in the groups (*e.g.*, control and treatment) being compared.

+ Analogous to EPS, EPA is the basis of statistical theory for the behaviour of *comparison* error under repetition – see the discussion on pages HL10.3 and HL10.4 of Table HL10.4 in Statistical Highlight #10.

+ Surprisingly, 'probability assigning' is not currently used elsewhere, perhaps reflecting separate development of the two large statistical areas of survey sampling and design of experiments.

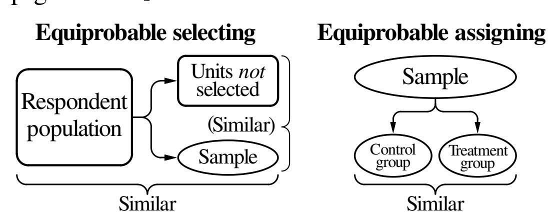
Our *equiprobable selecting* is usually *simple random selecting* or *random selecting* elsewhere;
our *equiprobable assigning* is **random assigning** or **randomization** elsewhere.

– Statistical theory is *used* in the estimating branches of the two tree diagrams in the schema at the right above; these branches are part of the Analysis stage of the FDEAC cycle.

+ Selecting/assigning probabilities as the basis of the theory used for estimating is noteworthy.

– The schema at the right above reminds us of the analogous roles of stratifying and blocking in sampling and comparing [but recall the comment (–) near the middle of page HL63.2].

● As shown pictorially at the right, a common theme of EPS and EPA is dividing a group of elements into *sub*groups that are likely to be *similar* enough *under adequate replicating* for the respective limitations imposed on Answer(s) by sample error and comparison error to be acceptable in the investigation context.



– When *selecting* the sample, the group of elements is the respondent population, the subgroups are the units *not* selected and the sample.

11. Investigators may have the opportunity to trade study error and sample error; a Plan involving *smaller* study error and (likely) *larger* sample error is usually preferred because:

– study error requires *extra*-statistical knowledge to assess it and its behaviour can seldom be quantified; BUT:

+ under EPS, sampling theory describes the behaviour of sample (and, perhaps, measurement) error under repetition of selecting and estimating [see the second bullet (●) above Note 1 on page HL21.3 in Statistical Highlight #21].

[Notes 9 to 11 on pages HL63.5 and HL63.6, given here for more convenient cross reference, are (respectively) Notes 3, 11 and 6 on pages HL62.2, 10.5 and 9.8 in Statistical Highlights #62, #10 and #9.]