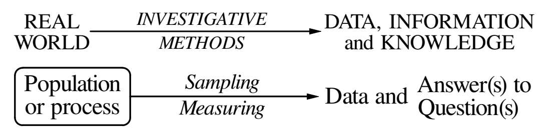
University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Statistics – Distinctions and Notation

(about 400 entries, about 315 definitions and descriptions)

1. Background I – What is Statistics About? [optional reading]

As summarized in the two schemas at the right, statistics is concerned with *data-based investigating* (or empirical problem solving) of the real world, which means investigating some population or process on the basis of *data* to *answer* one or more *questions*. For this introductory discussion, only the investigative methods of **sampling** and **measuring** are shown in the lower schema.



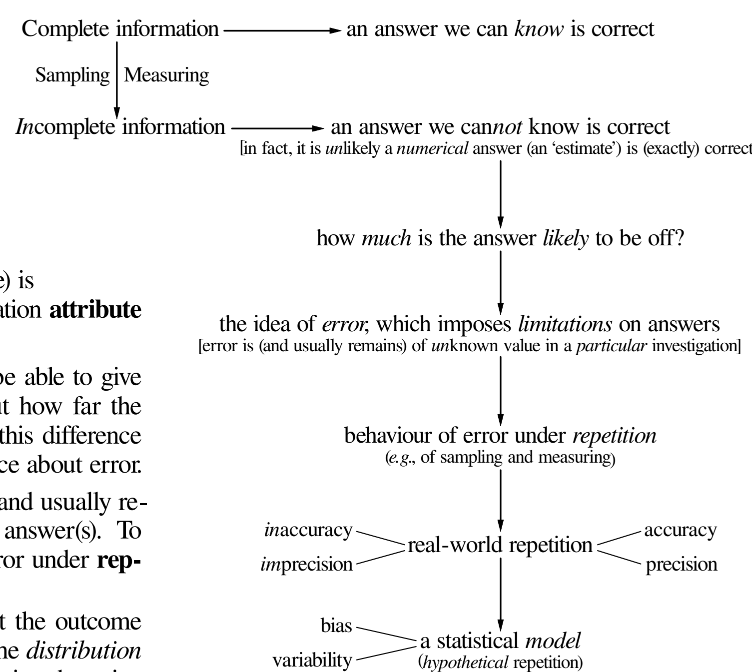
This introduction presupposes initially that data-based investigating is concerned with answer(s) to question(s) about a collection of **elements** which comprise a **population**. For example:

- a computer manufacturer may want to assess warranty claims for one of its products – an *element* would be one item of the product and the *population* is all such items sold over some specified period;
- a government department may want to know the proportion of Canadian adults who have concerns about the level of funding of the health care system – an *element* would be a Canadian adult and the *population* is all such Canadians;
- a financial institution may wish to find people whose profile makes them more likely to accept an unsolicited offer of a credit card – an *element* would again be a person but the *population* is harder to specify; it could be people whose profile puts them at or above an acceptance level deemed likely to make the credit card offering profitable for the financial institution.

The question(s) to be answered may *sometimes* be concerned with an *individual* to be selected in future from the population.

2. Background II – Key Ideas: Uncertainty, Error, Repetition, Accuracy and Precision [optional reading]

A central issue in data-based investigating is that external constraints and the type of information we require *impose* sampling and measuring processes on most investigating; we then need to assess the likely 'correctness' of the answer from the investigating in light of the **uncertainty** introduced by these two processes. In essence, this is the problem of **induction**. An overview of this matter is given in the schema at the right below; the main ideas are summarized at the left.



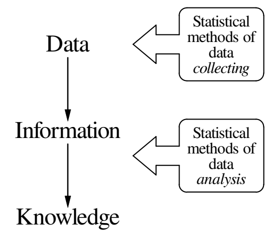
- ∗ If we have *complete* information, we can obtain a *certain* answer; that is, an answer we can *know* is correct.
- ∗ If we have *in*complete information, we can*not* know an answer is correct (an *un*certain answer) – in fact, it is *un*likely a *numerical* answer (an 'estimate') is (exactly) correct;
  - – sampling and measuring yield data (and, hence, information) that are *inherently* incomplete.
- ∗ An answer which is a *number* (like a sample average) is called a **point estimate** of the corresponding population **attribute** (the population average).
- ∗ To make such an answer more useful, we want to be able to give an **interval estimate**, a quantitative statement about how far the point estimate is *likely* to be from the true value – this difference is the **error** of the estimate. **Uncertainty** is ignorance about error.
- ∗ In a *particular* investigation, the size of the error is (and usually remains) *un*known; this is reflected in **limitations** on answer(s). To quantify uncertainty, we turn to the behaviour of error under **repetition** of the sampling and measuring processes;
  - – an analogy is tossing a coin – we cannot predict the outcome of a *particular* toss but *probability* can describe the *distribution* of outcomes under *repetition* of the process of tossing the coin.
- ∗ In the classroom, we can do *actual* repetition (repeating over and over) to demonstrate, for example, the statistical behaviour of the values of sample attributes (*e.g.*, averages) and of the values which arise from measuring processes;
  - – the two characteristics of *sign* and *magnitude* of (numerical) error lead, under repetition, to what we call **inaccuracy** and **imprecision**; the *inverses* of these two ideas – **accuracy** and **precision** – provide more familiar terminology, but we must realize that statistical methods (try to) manage the (undesirable) *former* to achieve needed levels of their (desirable) inverses.
- ∗ *Out*side the classroom, we recognize that *actual* repetition is usually not a viable option – we use instead *hypothetical* repetition based on an appropriate statistical *model* (for the selecting and measuring processes, for example);
  - – *we* help maintain the distinction between the real world and the model by using bias and variability as the *model* quantities which represent inaccuracy and imprecision in the real world.

(*continued overleaf*)

∗ A statistical **model**, including its *assumptions*, allows us to achieve our aim of quantifying (some of) the uncertainty in our estimate(s); part of identifying the limitations on an Answer is to assess model error arising from the difference between ('idealized') **modelling assumptions** and the real-world situation.

∗ We emphasize **sample error** and **measurement error** in this introductory overview; other categories of error, which need more background to understand, are (besides **model error**) **study error**, **non-response error** and **comparison error**.
– There can also be *non*-statistical error, although extra-statistical knowledge is generally required to assess it.
Error in our six categories is discussed in detail in the seventeen Statistical Highlights #6 to #22; Highlight #6 provides an introductory overview.

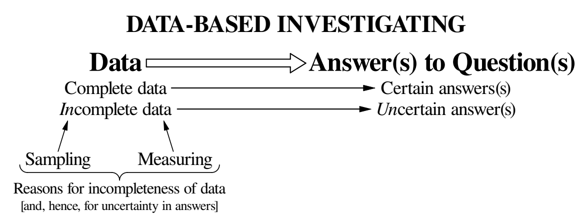
NOTES: 1. Another diagrammatic summary of what statistics is about is given at the right; it shows a broad division of the focus of statistical methods into data *collecting* and data *analysis*.



Instances of important questions to be answered (*i.e.*, matters needing data-based investigating) are given by P. Calamai: *Why journalists can't add* [Chapter 3 (pages 15-24) in *Statistics, Science, and Public Policy*, A.M. Hertzberg and I. Krupta (editors), Queen's University, Kingston, Ontario K7L 3N6, 1998]:

> What are the big things? They are issues such as silicone breast implants, the downsizing of the middle class, gender pay equity, the consumer price index, violence against women, the poverty line, drinking and pregnant women, race and crime, international debt comparisons, unemployment rates, the disappearance of the Atlantic cod, the tainted blood scandal and, of course, BSE and CJD, two sets of initials known only to a few researchers until Mad Cow Disease hit the headlines in 1996.

2. The following are illustrations of the precept that *certainty* requires *complete information* and, conversely, *in*complete information inescapably yields *un*certainty [ignorance (usually of the size or magnitude) of error]:
   - In the proof of a theorem in mathematics, we have a set of axioms and the rules of logic; within this system of complete information, there can be *certainty* the theorem is true.
   - Games like chess and bridge involve small populations of elements (32 pieces and 52 cards, respectively) and a set of rules; within these systems of complete information, there can be *certainty* which player has won a game.
   - In data-based investigating, information boundaries are seldom as clearly defined as in mathematics, chess and bridge, and practical considerations impose sources of *in*completeness and, hence, of *un*certainty, which include:
     – sampling: although we want Answer(s) to apply to a *population*, we seldom have the resources to investigate more than a *sub*set of the population (a *sample*) – the idea of sample error reminds us a sample seldom agrees *exactly* with the population;
     – measuring: even apparently straight-forward quantities like length and weight have *inherent* uncertainty in the values we obtain (the idea of measurement error), while measuring quantities like the number of cod in the North Atlantic fishing grounds or people's *attitudes* or *opinions* (*e.g.*, on mandatory registration of firearms, capital punishment, abortion) may introduce much *greater* measuring uncertainty.



A diagrammatic summary of these matters is given at the right above.

3. An illustration of *sample* error is polling to estimate a proportion of interest, typically from a national sample of about 1,000 to 1,500 adults. The *sample* proportion is a point estimate of the *population* proportion but is unlikely to be *exactly* equal to it; however, in a properly designed and executed poll, it is likely close enough (*i.e.*, to have small enough sample error) to be an Answer with *acceptable* limitations, provided that *measurement* error (*e.g.*, arising from the questionnaire, the interviewer and the interview process) has also been adequately managed.

3. Background III – Why Do Terminology and Notation Matter in Statistics? [optional reading]

Literary discipline is usually needed for clear and succinct presentation of ideas in any respectable subject area, but introductory statistics teaching materials are notable for one or more of their obscurity, carelessness and muddled thinking. This unhappy state of affairs may be because:
- statistical ideas tend to be more than usually troublesome to explain,
- statistical writers tend to be more than usually undisciplined and careless.

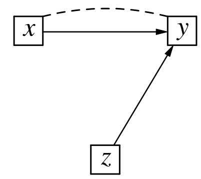
Whatever the reasons for inadequate teaching materials, the consequences are that:
– it is difficult for students trying to learn from such materials to maintain their intellectual integrity,
– students' intellectual challenges arising from inadequacy of presentation compromise their capacity to engage with the (important and useful) *statistical* ideas being presented.

(*continued*)

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 1)

Three excerpts (from among many like them) illustrate the foregoing bleak assessment – should the reader laugh or cry?

> Two explanatory variables or lurking variables are **confounded** when their effects on a response variable are mixed together. When many variables interact with each other, confounding of several variables can prevent us from drawing conclusions about causation. In the diagram at the right which illustrates confounding, both the explanatory variable *x* and the lurking variable *z* may influence the response variable *y* – the dashed line shows an association, the arrows show a cause-and-effect link. Because *x* is confounded with *z*, we cannot distinguish the influence of *x* from the influence of *z*. We cannot say how strong the direct effect of *x* on *y* is. In fact, it is hard to say if *x* influences *y* at all.



> A Google search in May, 2022, yields a variety of results for **the law of large numbers**; for example:
>
> ○ *As a sample size grows, its mean gets closer to the average of the whole population.* (Investopedia)
>
> ○ *A theorem that describes the result of performing the same experiment a large number of times.* (Wikiedia)
>
> ○ *If you repeat an expriment a large number of times, what you obtain should be close to the expected value.* (Probability Course)
>
> ○ *As the number of identically distributed randomly generated variables increases, their sample mean (average) approaches their theoretical mean*. (Encyclopedia Britannica)
>
> ○ *A theorem that describes the result of repeating the same experiment a large number of times.* (Corporate Finance)
>
> ○ *If we take a sample of n observations of a random variable, the sample mean approaches the expected value of the random varable as n approaches infinity.* (Khan Academy)
>
> ○ *The frequencies of events with the same likelihood of occurrence even out, given enough trials or instances. As the number of experiments increases, the actual ratio of outcomes will converge on the expected, or theoretical, ratio of outcomes.* (Whatis.com)
>
> ○ *As a sample size becomes larger, the sample mean gets closer to the expected value.* (Statology)
>
> ○ *When we do an experiment a large number of times, the average result will be very close to the expected result. In other words, in the long run, random events tend to average out at the expected value.* (Math is Fun)
>
> ○ *The relative frequency of an event will converge on the probability of the event, as the number of trials increases.* (Statistics Dictionary)

> The statistical terminology of **degrees of freedom** is often first encountered in introductory statistics courses when the divisor, $\mathrm{n}-1$, in the calculation of s^2 is introduced. The sample $\overline{\mathrm{y}}$ is used as the location from which to measure deviations when computing s^2, so it has been arranged arbitrarily to make the sum of the deviations $(\mathrm{y}-\overline{\mathrm{y}})$ add to zero, which it would not ordinarily do if we used the location value μ. This constraint on the deviations is called loss of a degree of freedom. Its effect is most severe when the sample size is 1. Then y is the only observation, $\overline{\mathrm{y}} = \mathrm{y}$, and the deviation is $\mathrm{y}-\overline{\mathrm{y}} = 0$. However, when we square this and try to divide by $\mathrm{n}-1$, we find ourselves illegally trying to divide by zero. With only one observation, we thus have no way of using only sample data to compute s^2 as an estimate of σ^2. When we have only 1 observation and want to estimate σ^2, we lose the only degree of freedom we have by estimating μ by y, and we say we have no degrees of freedom left for estimating σ^2.
>
> When we want to estimate σ^2, we have to 'pay' a degree of freedom for having to estimate μ, and this leaves us with $\mathrm{n}-1$ degrees of freedom, or essentially $\mathrm{n}-1$ observations worth of information, for estimating σ^2. A 2×2 contingency table loses 3 degrees of freedom because of constraints imposed by marginal totals. The idea of degrees of freedom keeps coming up in more complicted statistical methods, and so we need to become comfortable with it. Such statistical methods may lose additional degrees of freedom by estimating more things.

Introductory statistics teaching materials are a vast body but, rather than quantity, we need materials that are correct, clear and useful. Achieving any *one* of these three characteristics is a demanding task; together, the three are a daunting undertaking. If the current dismal situation is to be rectified, teaching materials need to be built around three distinctions between:

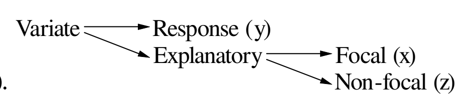
∗ the population and the sample, ∗ the real world and the model, ∗ the individual case and behaviour under repetition.

Such teaching materials can be achieved by appropriate terminology and notation presented and maintained with unremitting discipline; it is also useful to discard some (*un*helpful) current terminology. Rigorously maintaining these distinctions fosters a mind-set which routinely recognizes what statistical methods can, and can*not*, accomplish. [These three distinctions are a seminal insight of the writer's colleague, Prof. R. Jock MacKay.]

4. Background IV – Terminology, Notation and Distinctions [optional reading]

From the start, we avoid unthinking adoption from mathematics of matters *un*suited to statistics.

∗ Rather than (dependent and independent) 'variables', we have 'variates' which (as summarized in the schema at the right) are designated as response (letter y) or explanatory, with the latter being focal (letter x) or non-focal (letter z).



∗ Instead of only using *italics* as is common for typesetting mathematics, we exploit more broadly letter case and face/style:
- upper case **bold** letters are for **population** quantities,
- lower case Roman letters are for (sample) data,
- upper case *italic* letters are for ***random variables***,
- lower case *italic* letters are for *values* of random variables;
- also, lower case Greek letters [*e.g.*, μ (mu), σ (sigma), π (pi)] are used for (probability and response) **model parameters**.

(*continued overleaf*)

Table HL91.1

Real world	Model
¥, y	Y, y
$\overline{\textbf{¥}}$, $\overline{\mathrm{y}}$	$\overline{Y}$, $\overline{y}$, μ
Ᵽ, p	P, p, π

As indicated for a response variate in the first line of Table HL91.1 at the right, these notation conventions enable maintaining the population-sample and real world-model distinctions.

- The line through the **bold** letter (we *say* 'y cross') is to distinguish **¥** from Y in handwritten symbols.
- From the beginning, *we* use letter y as the *generic* variate, which is a *response* for much of the discussion in introductory statistics; this helps avoid confusion from (the surprisingly common practice in introductory texts of) starting with x as the generic variate (a carry-over from mathematics?) and then having to switch to y when (say) linear regression is discussed.

$$Y_j = \mu + R_j, \quad j = 1, 2, ..., \mathrm{n}, \quad R_j \sim N(0, \sigma), \quad \text{prob. independent, EPS} \qquad \text{-----(HL91.1)}$$

The second line of Table HL91.1 is for an **attribute** [of a *group* of elements (or units) like a population (or a sample)], specifically an *average* in the context of our simplest response model of equation (HL91.1). We say (or write): *the* ***population*** *average* $\overline{\textbf{¥}}$ *is represented by the* ***model*** *mean* μ, *for which the* ***estimator*** *is the random* variable $\overline{Y}$ *and the* ***estimate*** *is the sample average* $\overline{\mathrm{y}}$ (= $\overline{y}$).

- Implicit (and easily overlooked) in this statement is the *investigator's* responsibility to ensure that the Plan for the investigation makes it reasonable to treat $\overline{\mathrm{y}}$ (the sample average – a *number* – calculated from data) as $\overline{y}$ (the value of the random variable representing the sample average in the *model*).
- The *length* of the statement makes it tempting to try to shorten it but, as words are omitted, the distinctions it makes become obscured or are lost altogether (*e.g.*, 'μ is estimated by $\overline{\mathrm{y}}$') – see also Table HL91.2 below.
- Symbols are often subscripted; *e.g.*, Y_j and R_j in model (HL91.1) and **Ƶ**$_1$, **Ƶ**$_2$, ..., **Ƶ**$_k$ for multiple non-focal explanatory variates.

The third line of Table HL91.1 shows another (useful) attribute, a *proportion* (letter p); reserving letter 'p' for proportions entails *our* using 'Pr( )' (Roman letters) for *probability*, in contrast to the (random variable look-alike) 'P' of most introductory statistics texts.

A feature of these Materials is their emphasis on **standard deviation** rather than **variance** – the latter, which (unhelpfully) is the focus of traditional statistics teaching, is rarely mentioned here and, when it is, with reluctance for reasons given on page HL91.22 [see also the comment (●) above Note 9 on page 2.138 in Figure 2.17 in the STAT 332 Course Materials (1995 curriculum)].

– Standard deviations entail the routine typesetting of square roots, an *onerous* task for achieving aesthetic outcomes.

For standard deviations (*and* averages), maintaining the real-world/model distinction is aided by terminology given in Table HL91.2 at the right below; we exploit the availability of two words in English for measures of location but, with *one* word for **variation** and **variability**, these Materials suggest, particularly for beginning students, (mentally) adding the adjectives 'data' and 'probabilistic'.

Table HL91.2

Attribute	Real World	Model
Location	Average	Mean
Variation/variability	(Data) standard deviation	(Probabilistic) standard deviation

+ The **average-mean** distinction is absent from current statistical discourse – there also seems to be almost wilful (unnecessary) use of 'mean', perhaps to sound erudite, because we learn about averages in grade school where it would be inappropriate to raise the (nuanced) matter of a divisor *other than* the number of observations (widely denoted n, Roman 'n' in these Materials).
+ The (unprofitable) **n−1** *vs.* **n saga** is about the 'correct' divisor for calculating (under constraints) the average deviation from the average for a (data) standard deviation – see Appendix 3 on pages HL100.7 to HL100.9 in Statistical Highlight #100.
+ In contrast to (nuanced) averaging of **data**, at least in introductory statistics there is *no* disagreement about how to calculate the mean and standard deviation of a random variable – see, for example, Figures 5.9 and 7.11 in the STAT 220 Course Materials.

The foregoing matters can be seen *in use* in Appendix 1 on pages HL91.22 to HL91.24 and, in a regression context, in Figure 16.1 in the STAT 231 Course Materials; relevant Glossary entries on pages HL91.7 to HL91.22 also provide more detail.

We distinguish *four* populations by evocative adjectives:

∗ the **target** population; ∗ the **study** population; ∗ the **respondent** population; ∗ the **non-respondent** population.

These populations are central to developing the first two categories of **error** in Table HL91.3 at the right below; we see them, for instance, as part of the right-hand schema on the lower half of page HL91.6 and of the diagram at the upper right of page HL91.11. [Some contexts involve a **process** rather than a population – for example, see Statistical Highlight #94].

These Materials maintain throughout that (real-world) populations in statistics are *finite*; infinite 'populations' are a (sometimes useful) *model* – see, for example, Appendix 4 on page HL77.10 in Statistical Highlight #77.

The individual case-repetition distinction is already *implicit* in our use of:

- *error*, which arises in an *individual* investigation, AND:
- a response *model* like the one in equation (HL91.1) above, because such models only describe behaviour under *repetition*.

Table HL91.3

Individual case	...........Repetition...........	
Error	**Inaccuracy**	**Imprecision**
Study	Studying	(Studying)
Non-response	Non-responding	(Non-responding)
Sample	Sampling	Sampling
Measurement	Measuring	Measuring
Comparison	Comparing	Comparing
Model	Modelling	-----

The distinction is *explicit* in the terminology for our six categories of error; as shown in Table HL91.3 at the right, (numerical) error under repetition can become either inaccuracy or imprecision, depending on whether its sign or magnitude is involved. *Repetition* is evoked by adding '-ing' to the eleven adjectives for inaccuracy and imprecision.

- Studying and non-responding imprecision are de-emphasized in the right-hand column of Table HL91.3 because these Materials assume a **deterministic** (*not* a **stochastic**) process for specifying the study population and, for people as elements or

University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 2)

units, deciding whether to respond (under given incentives). [Elsewhere, there *are* stochastic models for the response process.]

- Measurement error arises for an individual measurement but **overall** error involves ***attribute* measurement error**.
  - Usually, attribute measurement error error is managed in an investigation by managing measurement error.
  - The *effect* of measurement error may differ between variates and an attribute (like the slope of a regression line).
  - Our usual concern is with *sample* attribute measurement error because a census is rare in practice.
- Model error (in the last line of Table HL91.3) is the only category of error *not* defined in terms of attributes; its nature means it is seldom appropriate to consider modelling imprecision, as indicated by the final dash (-----) in Table HL91.3.

Our six error categories are useful because, contingent on proper Question formulation in terms of the target population/process, they help us identify *sources* of error; in a particular investigation, we then incorporate Plan components which we expect will manage inaccuracy and manage imprecision (by managing variation) in ways that will reduce, to a level acceptable in the Question context, the limitations imposed on Answer(s) by (the likely size or chance of) each category of error. Plan components to manage the six categories of error are summarized in Table HL6.1 on pages HL6.4 and HL6.5 in Statistical Highlight #6.

NOTE: 4. It is unfortunate that, for two essential concepts, statistics deals *directly* with ***in***accuracy and ***im***precision, whereas common useage involves the (more familiar and, seemingly, more straight forward) inverses, namely accuracy and precision.

5. Background V – The FDEAC Cycle [optional reading]

A characteristic of statistics courses is their concern (some might say obsession) with *data analysis.* For later specialized courses, this concern is usually appropriate but, in introductory courses where students typically encounter the subject at a serious level for the first time, it is *imperative* they learn that there is much *more* to statistics than data analysis. They need to recognize the hard truth that unless many *difficult* tasks of error management are undertaken successfully *before* data analysis, its answer(s) are likely to be fatally compromised by severity of limitations. The tabular summary of the components of the FDEAC cycle in Table HL91.4 below (an overview of a *process* for data-based investigating) provides a basis for students to start to assimilate these vital ideas; Table HL91.4 is elaborated in Statistical Highlights #88 and #89.

Another lesson to learn is that *in*adequate management of any *one* category of error can result in severe limitation on answer(s), despite all *other* error categories being adequately managed.

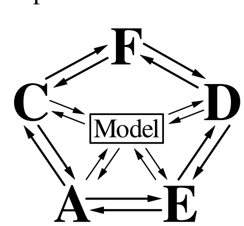
Table HL91.4: The FDEAC cycle: a *structured process* for data-based investigating

Stage	Formulation stage	Design stage	Execution stage	Analysis stage	Conclusion stage
Input	***Question(s)***	***clear Question(s)***	***a Plan***	***Data***	***Information***
Components	Target element Target population/process Variates: ● Response ● Explanatory Attributes Fishbone diagram Aspect: ● Descriptive ● Causative	Study element/unit Study population/process Respondent population/process Refine response variate(s) Deal with explanatory variates Protocol for: ● Selecting units ● Choosing groups ● Setting levels Measuring process(es) Plan for the: ● Execution stage ● Analysis stage	Execute the Plan Monitor the data Examine the data Store the data	Informal analysis: ● Numerical attributes ● Graphical attributes ● Other informal methods Assess modelling assumptions Formal analysis: ● Confidence intervals ○ Prediction intervals ● Significance tests ● Other formal methods	In the language of the Question context: Answer(s) Limitations Recommendations ['*Evidence*-based decisions, improvements, ...' means using Answers from *data*-based investigating with an *adequate* Plan.]
Output	***clear Question(s)***	***a Plan***	***Data***	***Information***	***Knowledge***

The kudos from being party to arcane mathematical and computational techniques, and the 'excitement' of reaching more quickly the goal of answers, combine to make it widespread in statistics teaching and practice to devote *in*adequate resources to error management in the relative drudgery of Question formulation, developing a proper Plan and its careful execution; the consequence of answers that likely embody **falsehoods** can easily be (and routinely is) ignored.

The diagram at the right makes the following points about the FDEAC cycle in relation to the *model*:

- ○ the *peripheral* arrows remind us each stage has implications for the stages before and after it; for example:
  - the Formulation stage may need to consider the type of Plan (*e.g*., experimental or observational) and the Plan must have appropriate components to address the Question(s);
  - the Plan must specify the data which are required and the Execution stage must be sure the Plan *as developed* is carried out to generate such data;
  - the Execution stage generates the data for the Analysis stage and the Analysis stage must use modelling assumptions that can be justified in light of the Execution stage;
  - the Analysis stage must obtain from the data the information needed in the Conclusion stage to



(*continued overleaf*)

answer the Question(s) and the Conclusion stage must give Answer(s) that can be justified from (proper) data analysis;

– the Conclusion stage must give Answer(s) which address the Question(s) with limitations acceptable in the Question context and the Formulation stage must have posed Question(s) for which such Answer(s) *can* be provided;

○ the *radial* arrows remind us the (response) *model* has implications for the last *four* stages of the FDEAC cycle; specifically:

– once there are clear Question(s), part of the Plan to answer these Question(s) is an appropriate model (*e.g.*, for sampling and measuring processes, including calculating sample size) [discussed in Figure 2.13 of STAT 332];

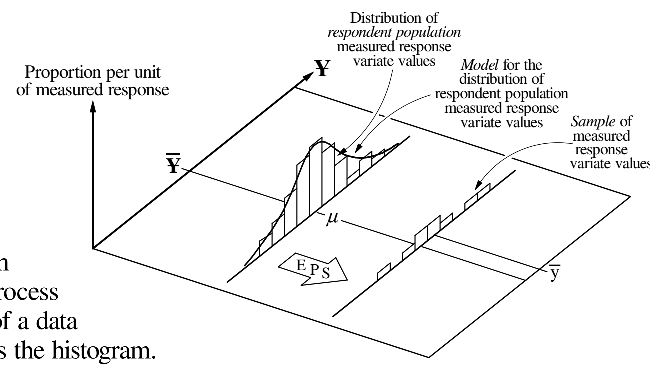
– the Execution stage generates data used to estimate model parameters representing study population attributes, and error in these estimates depends on how the data are generated (*e.g.*, imprecision and inaccuracy of measuring processes);

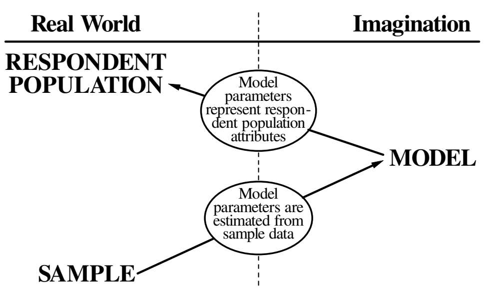
– the Analysis stage may use model-based formal methods of data analysis and then the model must be appropriate for the method(s) of analysis that will obtain from the data the information needed to answer the Question(s);

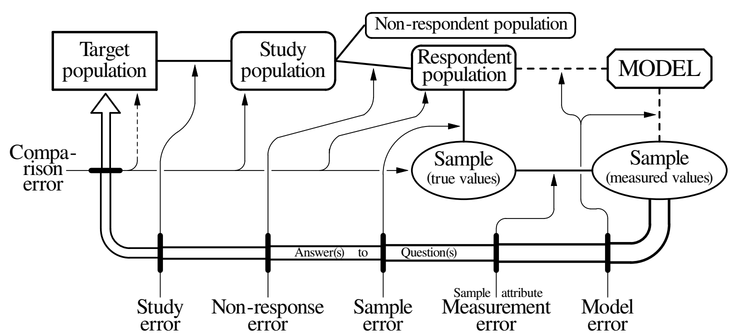
– the Conclusion stage must consider the model and its assumptions in assessing limitations on Answer(s) and the model must be chosen with Answer(s) and their acceptable limitations in the Question context in mind.

The three diagrams below are pictorial reminders of, respectively, the respondent population-model-sample sequence, our view of the model as a link in this sequence, and of how our four populations, the sample and the six categories of error comprise a process for data-based investigating.

Model-based methods of analysis in statistics use data from a sample to *estimate* values of model parameters which then represent plausible values (in light of the data) for respondent population attributes and, hence, for Answer(s) to Question(s); we distinguish a *point* estimate from an *interval* estimate (defined on page HL91.12). When the normal model is appropriate for the distribution of the response variate values, the model mean μ is estimated by the sample average $\overline{\mathrm{y}}$ (= $\overline{y}$) and σ is estimated by the sample standard deviation s (= s) – both *point* estimates. As illustrated at the right, we can think of the process of estimating μ by $\overline{y}$ and σ by s as approximating the histogram of a data set by the normal p.d.f. with the same ‘centre’ and same ‘width’ as the histogram.







6. Background VI – Terminology to Avoid or Use With Care [optional reading]

As well as defining appropriate and evocative terminology, clarity is aided by *not* duplicating existing terms with (equivalent) ‘feel-good’ words (with possibly ambiguous statistical connotations) – in statistics, we eschew ‘elegant variation’ in wording.

∗ **Applicability** (of an Answer) refers to study error and/or sample error.

∗ **Generality** (of an Answer) usually refers to sample error; in DOE, **generality** (or a **wider inductive basis**) may refer to whether the Plan involves a factorial treatment structure so that interaction effect(s) can be estimated.

– **Generalizability** refers to *study* error and **generalization** to *sample* error in the social sciences.

∗ **Reliability** [usually] refers to adequate precision (attained by managing *im*precision) [sometimes to adequate accuracy].

∗ **Sensitivity** (ability to detect an effect) refers to adequate precision (attained by managing *im*precision).

∗ **Strength** (of an Answer) means precision so **weakness** means *im*precision.

∗ **Trustworthiness** (of an Answer) means accuracy so **untrustworthiness** means *in*accuracy.

∗ **Validity** (of an Answer) means accuracy so **invalidity** means *in*accuracy.

As well as the clarity that results from using the ideas only of ‘accuracy,’ ‘precision’ and ‘error,’ the Glossary in Section 7 starting on the facing page HL91.7 suggests restricted use or avoidance as statistical terminology for the following words or phrases:

degrees of freedom, error (in regression models), experiment, hypothesis testing/test of hypothesis, independent, pivotal/pivotal quantity, population parameter, probable error, random, randomization, relative frequency, representative sample, sample statistic, sampling bias, significance level, simple random sampling, statistic, survey, test of significance, test statistic, universe, variance.

7. Glossary

[numbers in brackets () are page references to Parts 4, 5 and 13 of the STAT 231 Course Materials.]

Absolute value signs: a *shorthand* for what can be written in a longer form that may be easier to deal with. For example:
- $\Pr(|Y| > 2)$ means $\Pr(Y < -2)$ or $\Pr(Y > 2)$;
- $\Pr(|Y| < 2)$ means $\Pr(-2 < Y < 2)$;
- $f(y) = \mathrm{e}^{-|y|}$ $(-\infty < y < \infty)$ means $f(y) = \mathrm{e}^{-y}$ $(-\infty < y < 0)$ and $f(y) = \mathrm{e}^{-y}$ $(0 < y < \infty)$.

Accessibility selecting: see **Non-probability selecting**.

Accuracy: the inverse of *in*accuracy. (5.21)
In statistics, an **accurate** answer means an answer that is (or is adequately close to) correct (the "truth"), taking account of its limitations.
Statistical methods deal routinely with **induction** (reasoning from *in*complete information), so accuracy is distinguished from **precision**.

Acronym: the initial letters or syllables of the words in a phrase used as a short form or another word.
Thirteen acronyms in this Glossary are ANOVA, CI, df, DOE, EPA, EPS, EPSWIR, EPSWOR, FDEAC, IQR, mle, rms, SRS.
Abbreviations are p.f. for probability function, p.d.f. for probability density function, r.v. for random variable, s.d. for standard deviation.

Adequate replicating: to conserve resources. See **Replicating**.

Ali-Baba Paradox: see Statistical Highlight #46.

α-level: to be avoided – duplicate terminology for **critical value**.

Alternative hypothesis: see **Hypothesis testing**.

ANOVA: the acronym for **ANalysis Of VAriance**.
Despite its mellifluous appeal, 'ANOVA' is *un*helpful terminology:
- 'analysis' stretches the normal connotations of this word;
- 'variance' does not have its usual statistical meaning and, anyway, 'variance' is unhelpful in practical statistics;
- tabular presentation of ANOVA refers to averages as *means*;
- the divisors which produce these averages are unhelpfully called **degrees of freedom**.

ANOVA actually involves *partitioning sums of squared differences*; the evocative acronym 'PASSDI' has no chance of replacing 'ANOVA' but is useful to keep in mind when *we* encounter ANOVA.

Applicability: a word to be *avoided* as ambiguous duplication of existing statistical terminology; applicability (of an Answer) means (but does not distinguish) study error and/or sample error.

Approximate, Approximation: a value or process that (it is hoped) is roughly correct, typically obtained or used more simply or cheaply than an 'exact' value or process.
We distinguish *approximating* from **estimating** – in English, the two words are often used interchangeably (*e.g.*, estimating a crowd size).

Argument by contradiction: the three steps are:
- an assumption (a *hypothesis*);
- deductive reasoning (a *probability* calculation);
- a contradiction (*strength of evidence*).

The phrases in brackets ( ) are the corresponding *statistical* terminology in a **test of *(statistical)* significance.**
An argument by contradiction is sometimes confused with a paradox – see Statistical Highlight #46.

Aspect: a binary categorization of the primary concern of a Question, identified in the Formulation stage of the FDEAC cycle.
- **Descriptive:** a Question whose Answer will involve primarily values for *population/process* **attributes** (past, present, future).
- **Causative:** a Question whose Answer will involve primarily whether and/or how the focal explanatory variate is **causally** related to the response variate in a population/process. (5.72)

Assigning: in an **experimental** Plan, the process by which the value of the focal variate is set for each unit:
- within each block in a blocked Plan;
- in the sample in an *un*blocked Plan. (5.37, 5.48)

See also **Equiprobable assigning** and **Probability assigning**.

Association: if a scatter diagram shows, say, a clustering of its points about a line with positive slope (*i.e.*, we see that, as **X** increases, **Y** also tends to increase), we say **X** and **Y** show a (positive) *association*.
Characteristics of an association of statistical interest include its:
- **Form:** for example, can the trend be modelled by a *straight line*, indicating *linear* association?
- **Magnitude:** for linear association, what is the magnitude of the *slope*?
- **Direction:** for linear association, is the slope *positive* or *negative*?

Introductory statistics courses often emphasize the distinction between *association* and **causation**.
See also **Proportionality**. (5.30 to 5.31)

Attribute: a quantity defined as a function of the response (and, perhaps, explanatory) variate(s) over a *group* of elements/units, typically:
- the target population/process,
- the study population/process,
- the respondent population,
- the non-respondent population,
- the sample. (5.20)

Our useage of *population attributes* and *sample attributes* (where the latter can yield estimates of the former) is to be contrasted with the unhelpful **population parameter** and **sample statistic** of some introductory texts.

Attribute measurement error: see **Error**.

Average: a measure of location (commonly, for data), calculated as the sum of a set of entities (commonly, numbers), divided by the number of the entities that are 'independent' of each other. (5.28)
A (real world) *average* is to be distinguished from a (model) **mean**.
See also **n – 1 *vs.* n saga**.

Bar graph: see **Histogram**.

Base rate: *un*evocative terminology coined for investigating how people process what can be modelled as *conditional* probabilities.
See discussion of the silver cab/grey cab 'paradox' in Statistical Highlight #50, where the 'base rates' are the population proportions of cabs with the two colours in equations (1) and (2) on page HL50.1.

Bias: the **model** quantity representing **inaccuracy**.
(5.21, 5.46, 5.50, 5.63) See also Statistical Highlight #7.
Bias involves behaviour under **repetition** and is easily confused with error which involves an individual case (an individual investigation).
Our concern is usually with *estimating* bias and/or *measuring* bias.
See pages HL77.12 and HL77.13 in Statistical Highlight #77.

Binary (response) variates take only *two* values (often denoted 0 and 1), such as *Yes* or *No*, *Female* or *Male*, *Success* or *Failure*. (5.62)

Blind, Blinding: to withhold, for any unit, knowledge of whether it is in the **treatment group** or the **control group** (whose units usually receive a dummy treatment known as a **placebo**).
Blinding is typically used in a **clinical trial**, a special class of comparative experimental investigation used in medical research to assess the efficacy of new forms of treatment (*e.g.*, drugs, surgery); up to three levels of blinding may the used – blinding of:
- the participants,
- the treatment administrators,
- the treatment assessors,

depending on feasibility in the investigation context. (5.39, 5.52, 5.61)

Block: see **Blocking**.

Blocking in an **experimental** Plan: forming groups of units (the **blocks**) with the *same* (or similar) values of one or more non-focal explanatory variates; the units within a block are then assigned *different* values of the **focal variate**. [See also **Matching**] THUS:
Blocking *prevents* **confounding** of the focal variate with the non-focal explanatory variate(s) made the same within each block, thereby decreasing the *likely* magnitude of comparison error. SO THAT:
By holding one or more **Z**s fixed within blocks in an experimental Plan, blocking reduces variation in **Y** and so has the additional benefit of decreasing **comparing imprecision**, thus reducing the **limitation**

(*continued overleaf*)

imposed on Answer(s) by **comparison error**. (5.36 to 5.39, 5.52)

Blocking factor: a non-focal explanatory variate used as a basis for forming blocks in a blocked Plan. (5.37)

Blue car-green car: see Statistical Highlight #50.

Box: George E. P. Box, a respected U. S. statistician, coined the maxim: *All models are wrong, some are useful.*

Box plot, **Box and whisker plot:** see **Quantiles**.

Calibrating: using *known* value(s) [**standard(s)**] to quantify **measuring inaccuracy**. (5.20, 5.60)

Case-Control Plan: a (retrospective) observational Plan involving comparing *cases* (with **Y** = 1) and matched **controls** (with **Y** = 0).
Matching the controls to the cases (usually) precludes selecting them *probabilistically*, thus forfeiting its statistical advantages. (5.40)

Categorical: a categorical (or **qualitative**) variate has values which are *categories*; for example, sex or marital status.
Quantitative variate values can become categorical; *e.g.*, ages can be classified into age *groups*. (5.61, 5.62)

Causal: the adjective from *cause* (easily misread as *casual*).

Causal chain: a (usually long) sequence of explanatory variates intermediate between the focal variate **X** and the response variate **Y**.
Preoccupation in statistics with *whether* **X** *causes* **Y** makes it easy to forget that causation *always* involves a causal chain, potentially extending backwards and/or forwards indefinitely.
(5.32, 5.43, 5.76, 5.77)
In outline, Geroge Lakoff's two world views involve 'conservatives' who favour *direct* causation to deal with a problem by direct action, and 'progressives' who recognize causal *chains* or 'systemic causation', which Lakoff divides into direct causes, interacting (or chains of) direct causes, feedback loops and probabilistic causes.
Lakoff claims that direct causation appears to be represented in the grammars of all languages, but systemic causation is not represented in grammar – it has to be learned.

Causal relationship: there is a **causal** relationship between a response variate and a specific explanatory variate (usually the **focal variate**) if the value of an appropriate **attribute** of the response (and, perhaps, explanatory) variate(s) *changes* when, for *every* element of the **target population**:

- the specific explanatory variate value is *changed*, AND:
- *all other* explanatory variates *hold* their (same) values.

(5.32 to 5.35, 5.47)

Causation: ***informally***, the idea that deliberate change of (solely) explanatory variate **X** *brings about* a change in response variate **Y**.
Formally, we state three criteria to define what *we* mean when we say (a change in) **X** *causes* (a change in) **Y** in a **target population**:

(1) **LURKING VARIATES:** Ensure *all other* explanatory variates $\mathbf{Z}_1$, $\mathbf{Z}_2$, ..., $\mathbf{Z}_k$ hold their (same) values for *every* population element when **X** = 0 and **X** = 1 (sometimes phrased as: *Hold all the* $\mathbf{Z}_i$ ***fixed*** for.....).

(2) **FOCAL VARIATE:** observe the population **Y**-values and calculate an appropriate attribute, under *two* conditions:
⊙ with *all* the elements having **X** = 0;
⊙ with *all* the elements having **X** = 1.

(3) **ATTRIBUTE:** The **X**-**Y** relationship is *causal* if:
$$\text{Attribute}(\mathbf{Y}, \text{perhaps some of } \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_k \,|\, \mathbf{X} = 0) \neq \text{Attribute}(\mathbf{Y}, \text{perhaps some of } \mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_k \,|\, \mathbf{X} = 1),$$
provided those of $\mathbf{Z}_1$, $\mathbf{Z}_2$, ..., $\mathbf{Z}_k$ *included* in the attribute have the *same* values when **X** = 0 and **X** = 1. (5.32)

Cause-and-effect diagrams: see **Ishikawa**.

Census: an investigation using *all* the respondent (or study) population elements/units/blocks. (5.26, 5.53, 5.55, 5.76)
A census is to be contrasted with an investigation based on a **sample** (the *usual* situation). (5.21, 5.50, 5.51, 5.53, 5.59, 5.84)

Central Limit Theorem (CLT): if the (probabilistically independent) random variables $Y_1, Y_2, Y_3, \ldots, Y_n$ each have mean μ and standard deviation σ, and if the random variable $T = Y_1 + Y_2 + Y_3 + \ldots + Y_n$, then:

- the standardized form of T, $(T - n\mu)/(\sqrt{n}\sigma)$, has a *standard normal p.d.f. in the limit as* $n \to \infty$,
- the standardized form $\overline{Y} \equiv T/n$, $(\overline{Y} - \mu)/(\sigma/\sqrt{n})$, has a *standard normal p.d.f. in the limit as* $n \to \infty$. (5.13)

The CLT is a key component of the theory for **interval estimating** in introductory statistics courses.

Centre: an *in*formal term for the 'middle' of a distribution, like the **median** or the **average**.

Chance: for a process with two or more possible outcomes, 'chance' usually refers to the unpredictability of which outcome will occur in a particular execution of the process. See Section 5 on pages HL94.3 and HL94.4 in Statistical Highlight #94.
'Chance' is generally avoided as terminology in these Materials

Check sheets: see **Ishikawa**.

Cherry pie paradox: see Statistical Highlight #47.

Chi squared distribution: if $Z_1, Z_2, \ldots, Z_\nu$ are probabilistically independent $N(0, 1)$ random variables, the *sum of their squares* has a χ^2 distribution with ν degrees of freedom, denoted χ^2_ν ('chi' rhymes with 'hi' – it is pronounced 'ki'). (13.14 to 13.16, Ap. 7, Ap. 8)

Clinical trial: see **Blind, Blinding**. (5.39, 5.49)

CI: see **Confidence interval**.

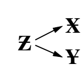
Cluster: a (natural) group of elements/units of a population.
The clusters which make up a population can be of:

- **equal size:** for example, cardboard cartons of 24 cans in a population of cans of soup; OR:
- ***un*equal size:** for example, households in a population of people.

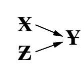
(5.24, 5.55, 5.57, 5.85, 5.86, 5.96)
Cluster sampling is discussed in Figures 2.14 and 2.16 of STAT 332.

Cold: we routinely perceive 'cold(ness)' and its degree can be quantified by (low) temperature.
Science identifies temperature with energy levels (*e.g.*, translational, rotational, vibrational) at an atomic level; a consequence is a lower limit for 'cold' ('absolute zero', 0°K) when all energy levels are in their ground state. (See also **Postulates of Impotence**.)
This model is incongruous to our senses in that it seems possible, no matter how 'cold' something is, to imagine it being colder.
The physical experience of 'cold' can thus remind us that perception may be at odds with (models of) 'reality' – see also the more abstract situations in Statistical Highlights #47 to #51.

Common cause: the situation where variate **Z** (say) causes *both* variates **X** and **Y**. (5.34, 5.35, 5.42, 5.73, 5.76, 5.77)



Common response: the situation where variate **Y** (say) is the response to *both* variates **X** and **Z**. (5.35, 5.42, 5.46, 5.76)



Comparative Plan: a Plan involving *changing* and *comparing*.
Changing and *comparing* are the basis for investigating a **relationship**. (a Question with a **causative aspect**). (5.28)

Comparing includes the processes of **assigning** and **estimating**. (5.45 to 5.49)

Comparison error: see **Error**.

Complement: see **Event**.

Comprehending: see **Sensing**.

Confidence interval (CI)**:** an expression for an interval estimate of a **model parameter**, derived from the distribution of an estimator; the interval *covers* the value of the model parameter with a specified probability called the **confidence level** (*e.g.*, 95%). (13.1 to 13.18)
A **realized** confidence interval is the expression evaluated from data and is usually given in the Analysis stage of the FDEAC cycle. (13.3)

- ***Informal*** interpretation: *a range of* ***plausible*** *values for a respondent population attribute or a model parameter representing it.*
- ***Formal*** interpretation: *under repetition of the selecting and measuring processes and of calculating the CI from the relevant expression, approximately the confidence level proportion of these intervals will contain the value of the population attribute or model parameter.*

University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 4)

A CI quantifies uncertainty in terms of behaviour under repetition.

Confidence level: see **Confidence interval**.

Confounded: variates involved in confounding can be said to be *confounded* (under the Plan). (5.30)

Confounder an explanatory variate involved in confounding (a 'confounding variate'). (5.30, 5.43, 5.70) See also **Lurking variate**.

Confounding: ***informally***, when non-focal explanatory variates $\mathbf{Z}_i$ do *not* all hold their same values as $\mathbf{X}$ changes to make apparent its relationship to $\mathbf{Y}$, *their* effects on $\mathbf{Y}$, and that of $\mathbf{X}$, are confused or mixed up in such a way that they cannot be distinguished.
Formally, differing distributions of values of one or more *non*-focal explanatory variate(s) among two (or more) groups of elements/units [like (sub)populations or samples] with different values of the focal variate.
When $\mathbf{Z}$ is a confounder, not taking account of $\mathbf{Z}$ values may make the Answer to a Question about a (*causal*) relationship between $\mathbf{X}$ and $\mathbf{Y}$ meaningfully different from the *correct* Answer.
The four types of confounding *we* distinguish, in order of *de*creasing importance for introductory statistics, are:
Type 2 (as defined above for $\mathbf{Z}$, $\mathbf{X}$ and $\mathbf{Y}$) is the primary concern of these Course Materials, specifically with reference to comparison error in comparative Plans.
Type 1 (of two or more *focal* variates): inability, under the Plan, to separate the effects of two (or more) focal variates on a response variate. (Type 1 confounding may be *exploited* in **Design of Experiments**.)
Type 3 reflects disagreement among statisticians as to how broadly 'confounding' is to be interpreted; for example, whether a phenomenon like Simpson's Paradox should be regarded as an instance of 'confounding'.
Type 4 is unique to these Course Materials and is *solely* to provide statistical insight from recognizing common themes of **probability assigning** and **probability selecting**. (5.30, 5.70 to 5.72)
For introductory statistics teaching, Types 1, 3 and 4 counfonding are optional enrichment. See Statistical Highlight #3.

Confounding effect: in an **observational** Plan, for a focal variate with q values, we think of the respondent population as being made up of q *sub*populations; each subpopulation is those elements which have a particular value of the focal variate. (5.49, 5.51, 5.54)
When $q = 2$ and the two subpopulation average responses are $\overline{\mathbf{Y}}_0$ and $\overline{\mathbf{Y}}_1$, we have:

$\overline{\mathbf{Y}}_1 - \overline{\mathbf{Y}}_0$ = effect of change in $\mathbf{X}$ + effect of change in $\mathbf{Z}_1$, ..., $\mathbf{Z}_k$
= treatment effect + **confounding effect**. -----(HL91.2)

The 'confounding effect' is terminology specific to these Materials.

Confusion: confusion of terminology (or its ideas) [including with ordinary English useage] can be a source of obscurity in explanations of statistical ideas; forty-two pairs of words easily confused are:

Accuracy	*Precision*	Mean	*Average*
Association	*Causation*	Minus	*Negative*
Average	*Mean*	Mistake	*Error*
Bias	*Error*	Model	*Real world*
Causation	*Association*	Negative	*Minus*
Causation	*Correlation*	Observation	*Observational*
Correlation	*Causation*	Observational	*Observation*
Data	*Ran. var. values*	Population	*Sample*
Data s.d.	*Probabilistic s.d.*	Precision	*Accuracy*
Error	*Bias*	Probabilistic s.d.	*Data s.d.*
Error	*Mistake*	Probability	*Likelihood*
Estimate	*Estimator*	Ran. var. values	*Data*
Estimator	*Estimate*	Real world	*Model*
Estimated s.d.	S.d.	Repetition	*Individual case*
Experiment	*Experimental*	Sample	*Population*
Experimental	*Experiment*	Significance test	*Hypothesis test*
Hypothesis test	Significance test	S.d.	*Estimated s.d.*
Individual case	*Repetition*	Uncertainty	*Variation*
Interact	*Interaction*	Variability	*Variation*
Interaction	*Interact*	Variation	*Uncertainty*
Likelihood	*Probability*	Variation	*Variability*

Some word pairs involve the *same* issue, like the estimate-estimator or real world-model confusion.

Contingency table: a rectangular table of frequencies arising from categorizing each element in a group (like a population or a sample) according to its values of two variates.
The simplest case is a 2×2 ('two by two') table when the variates are both binary (*e.g.*, each with values of *Yes* or *No*): the four cells of the table show the number of elements with combined values *Yes-Yes*, *Yes-No*, *No-Yes* and *No-No*.
Two variates with m and n values yield an m×n contingency table.
Also usually shown, in additional cells at the right and bottom of the table, are the total frequency of each row and each column and, sometimes, at the bottom right-hand corner, their sum, the number of elements that produced the data in the table.
A contingency table **test of (probabilistic) independence** of the two variates that are the basis of the element categorization is a formal statistical method of data analysis, based on a χ^2 distribution probability model with $(m-1)\times(n-1)$ degrees of freedom (for an m×n table) and the definition of probabilistic independence.
Contingency tables are not discussed in STAT 231 but are described in Figure 12.26 of the STAT 221 Course Materials.

Continuity correction: a way to improve accuracy when a continuous distribution is used to approximate a discrete distribution (or when selecting *with* replacement is used to approximate selecting with*out* replacement).
We encounter this idea mainly in the normal approximation to the binomial and the Poisson distributions (or the binomial approximation to the hypergeometric distribution).
See equation (2.10.19) and Tables 2.10.4 to 2.10.6 in Figure 2.10 of the STAT 332 Course Materials.

Continuous: see **Random variable** and **Sample space**.
We perceive much of the (macroscopic) material world as continuous and 'solid'), characteristics integral to the function of many entities, like pieces of string, lampposts and highways.
The (highly successful) atomic model of the (microscopic) world is different – it is inherently *discrete* and the atom, with a nuclear diameter around 10^{-5} of the atomic diameter, has about 15 orders of magnitude more 'empty' space than matter, although this space is permeated by effects of the nuclear and electronic charges and is modelled as the 'quantum vacuum'. How to reconcile these disparate views of 'reality' has (of course) been debated for decades.
In *our* routine use of *mathematical* continuity, it is easy to forget that the real world may actually be **quantized** with, for example, a lower limit for the smallest possible length. Continuity may be a mathematical *extrapolation* from a discrete material world but it is conceivable that continuity is *achieved* in the *non*-material world. *If* the material world *is* discrete, using continuous mathematics in models could be a source of model error.

Contrast: see **Effect**.

Control chart: see **Ishikawa**.

Control group: in an experimental Plan, the part of the sample assigned $\mathbf{X} = 0$; in practice, this may mean receiving a **placebo**.
(5.38, 5.39, 5.40, 5.45, 5.47 to 5.50, 5.53, 5.61, 5.80, 5.82)

Convenience selecting: see **Non-probability selecting**.

Correlation: a numerical measure of *tightness of clustering* of the points on a scatter diagram about a straight line – correlation is denoted r (c would be better, leaving r for a ratio) and its values lie in the interval [–1, 1]. (4.9 to 4.24, 5.29, 5.30, 5.34)
See Statistical Highlight #66.

Counterfactual: a variate value *not* observed under the Plan – for instance, an element's response if it were to have been assigned a focal variate value *different* from the value *actually* assigned. (5.50, 5.74)
Such *hypothetical* variate values may arise in statistical theory. (5.51)

Covariance: a measure of association; (4.16) for random variables X and Y: $cov(X,Y) = E(XY) - E(X)\bullet E(Y)$; -----(HL91.3)
i.e., covariance is *the mean product minus the product of the means.*

Covering: to try to manage of sample error, the values of explanatory variates of the units in the sample are chosen to cover the range of values that occur among (most of) the elements/units of the re-

(*continued overleaf*)

spondent (or study) population/process. (5.24)
Covering is relevant to implementing **judgement selecting**.

Critical value: see **Hypothesis testing**.

Cross-over design: a clinical trial in which *all* participants are assigned *both* values of the focal variate. (5.51 to 5.52)
For example, the trial starts with half the participants having $X = 0$ and half $X = 1$; after an appropriate time, the values of X are interchanged for the two groups at the *cross-over* point of the trial.

Data: values (often numerical) of:
– *variate(s)* in statistics, – *outcome(s)* in probability.
The FDEAC cycle involves the sequence from data to information to knowledge. See also **Wisdom** and **Sensing**.

Deciles: see **Quantiles**.

Deduction: reasoning from *general* information to a *particular* conclusion (or answer), using the rules of logic and relevant theory.
Steps of deductive reasoning are considered to embody logical necessity, as in mathematical 'proofs', for example. This implies a 'correct' answer (or at least an answer consistent with the starting 'axioms'), to be contrasted with *un*certain answers from **induction**.
A danger on mathematical models and their associated deductive reasoning is that the latter's logical necessity makes it easy to overlook the effects of **model error** on answers.

Degrees of freedom (D.f.): the (*un*evocative) name given (for historical reasons) to one or more parameters of the t, K, χ^2 and F distributions; the phrase is also used in other contexts like ANOVA. (13.14 to 13.17)
Like its name, its 'explanations' are often obscure.

Deming: W. Edwards Deming was a pioneer in statistical methods of quality improvement. See Statistical Highlight #96.

Dependence: a characteristic of the material world is the *dependence* among its components, a matter of great complexity. This dependence impedes (but does not prevent) our understanding the real world and developing useful models for its behaviour, two processes aided by introducing the multi-faceted idea of *in*dependence, of which independent measurements, functional independence and probabilistic independence are instances we encounter in statistics.
See also **Independence** and the Appendix on page HL89.18 in Statistical Highlight #89.

Deterministic: this term arises in these Materials mainly in the context of models for non-response, the source of one of our six error categories. *We* contrast 'deterministic' with 'stochastic', although the distinction is complicated and (perhaps) equivocal. (5.26)
The determinism of Laplace's 'demon' – that knowledge at some instant of the forces on and positions (and velocities) of all the particles in the universe, plus prodigious ability to calculate, would reveal with certainty the past and the future – has been shown to be untenable for reasons involving measurement, classical physics, quantum theory, time and computation. See **References**.
Some of these reasons involve ideas available to Laplace.

Deviation: see **Difference**.

Difference: calculating a data standard deviation involves 'deviations' from the average; calling them 'differences' would avoid (unnecessary) duplication of terminology, but thus is *un*likely to happen.

Discrepancy measure: see **Significance testing**.
Elsewhere, a *discrepancy measure* may be called a **pivotal quantity** or a **test statistic**; *we* avoid the latter two terms in these Materials.
It is regrettable that statistics has three (arcane) synonyms for one idea: ***quantifying the disagreement*** (on the basis of a suitable probability *model*) between what is *observed* and what would be *expected* if the (null) hypothesis is true.
Obviously, the greater the disagreement, the more likely it is that the null hypothesis is *not* true – that is, the more lilely the null hypothesis is *false* (or the probability model is *not* suitable).
This parade of negatives – not true = false, null hypothesis is false = the alternative hupothesis is true = there *is* an effect – can engender confusion and is a disadvantage of argument by contradiction.

Discrete: see **Continuous**, **Random variable** and **Sample space**.

Disjoint: see **Event**. (5.29)

Dispersion: how spread out ('dispersed') a data set or probability distribution is – a similar idea to **spread**.

Distinctions: for distinctions important in statistics and its teaching, see the lower half of page HL91.3 and page HL91.4.

Distribution: for a quantity that can take two or more values which each may occur one or more times, its distribution is the set of values and their frequences, usually with the values arranged in an appropriate order (*e.g.*, in ascending numerical order).

DOE: acronym for **Design of Experiments**; this would be more evocative as *Designing Experimental Plans* (DEP), to indicate its role as a *process* [like the Design (and other) stages of the FDEAC cycle]. See also **Experiment**.

Ecological fallacy: using correlations among **attributes** (like averages) to answer a question about correlations among *individuals*, without recognizing that the former are typically *higher* in magnitude than the latter.

Eddington: Arthur S. Eddington (1882-1944) was a British astronomer and mathematician; his extensive writings include philosophy of science and popularizing science. See also **Quantized**.
His posthumous book *Fundamental Theory* presents in detail his claim that, because science investigates the world by *measuring*, it is possible to deduce *a priori* the values of some dimensionless quantities like the total number of atomic charged particles in the universe, the ratio of the electrostatic and gravitational force between two electrons, and the proton-to-electron mass ratio. See **Sensing**.

Effect: the **effect** of X on Y (usually) refers to the change in the *average* of Y for *unit* change in X and:
– implies the X-Y relationship is (believed to be) *causal* – a change in X *causes* (brings about) a change in Y;
– includes both the *magnitude* and *direction* of the relationship – for example, the *slope* and its *sign* for a *linear* relationship;
– requires that all non-focal explanatory variates Z_i hold their (same) values when X changes;
– is defined (the 'true' effect) over the elements of the *respondent* (or *study*) *population*. (5.43, 5.44)
- **Main effect:** the effect of a factor *individually*. (5.44)
- **Treatment effect:** a more explicit term for *effect*. (5.50, 5.51)
Treatment effect is also a broader term for main effects and interaction effects.
- **Contrast:** any *linear combination* of treatment effects where the coefficients sum to *zero*. (5.44)

Effron's dice: see Statistical Highlight #48.

Eikosogram: a pictorial display involving a unit square subdivided into areas in a way that illustrates combinations of events and their probabilities; eikosograms do this more effectively than **Venn diagrams**. See Statistical Highlight #5.

Element: the population entity of interest to the Question(s) to be answered by an investigation and for which variate values could be obtained. [***Informally***, an 'element' is an 'individual'.]
An *element* is to be distinguished from a **unit**, which is determined by the sampling **frame**. (5.55, 5.86)
An illustration is a Question about *people* as elements but a frame of *households* used to select the units at the first stage of sampling.
Elsewhere, elements may be called **elementary units** or **observation units**; units may be called **sampling units**.
In STAT 231, the unit-element distinction is ignored and only 'unit' is used as terminology. See also **Frame** and **Unit**.
In probability, a *set* is made up of elements.
See also Appendix 2 on page HL77.9 in Statistical Highlight #77.

Empirical: based on **data**.

EPA or **EPS:** see **Equiprobable assigning** or **Equiprobable selecting**.

EPSWIR: equiprobable selecting *with* replacement. See pages HL94.9 and HL100.10 (Note 4) in Statistical Highlights #94 and #100.

EPSWOR: equiprobable selecting with*out* replacement – our *default* meaning of EPS.

Equations: equations, which involve *equality* of their right- and left-

University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 5)

hand sides, occur throughout mathematics; two familiar categories are algebraic equations and differential equations. A common task is *solving* an equation to exhibit its 'solution', a process which appears to be **deterministic**, regardless of whether this process is analytic or numerical, exact or **approximate**.

If solving equations *is* deterministic, it may be that the source of the **uncertainty** inherent in quantum mechanics lies in the concept of a *wave function*, which has the (curious) property that its square is usually interpreted as a *probability*. Deciding how (or if) *this* uncertainty differs from that in statistics needs its careful definition, avoiding confusing it with **variation**.

The (unfortunately named) concept of *chaos* – that some (relatively simple, differential) equations are so sensitive to initial conditions that their approximate numerical 'solution' cannot be specified regardless of the number of significant figures carried in the calculations – *may* cast doubt on whether 'solving' equations *is* (always) deterministic.

Dirac's model for the electron, a second-order differential equation, allowed for *two* solutions, the second being physically realized when the positron was (later) identified. *Two* solutions here might be the analogue of a property of second-degree *algebraic* equations. In the *model*, positrons can be associated with (physically bizarre) ideas like electrons moving backwards in time or having negative energy.

Dirac's equation also yields the concept of electron 'spin', with dimensions of angular momentum. Equations are linked inextricably to **Modelling**.

Equiprobable assigning [EPA]: 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.

EPA is a special case (with *equal* assignment probabilities) of **probability assigning**.

EPA is usually called **random assigning** or **randomization** elsewhere, but EPA is more evocative of the assigning process.

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.

Equiprobable selecting [EPS]: all samples of size n units from a **study** population of size $\mathbf{N}_S$ units have probability $1/\binom{\mathbf{N}_S}{n}$ of being selected. This definition can also be stated in terms of the $\mathbf{N}$ units of the **respondent** population. (5.23, 5.56, 5.57, 5.86)

EPS is a special case (with *equal* inclusion probabilities) of the process of **probability selecting**. (5.48, 5.56)

EPS is usually called **simple random selecting (or sampling) (SRS)** elsewhere, but EPS is more evocative of the selecting process.

We distinguish two useages of 'EPS'.

– **EPS from an unstratified population**: a protocol for selecting units which is seldom used in practice but which is the basis of sampling theory; FROM:
– **EPS (unqualified)**: *part* of a protocol for selecting units which involves *other* statistical ideas like stratifying and/or clustering and/or systematic selecting – the more *common* useage of 'EPS'.

See Statistical Highlights #21 (and #84) for further discussion.

Error: the difference between what is stated [*e.g.*, in an Answer] or assumed [*e.g.*, in a response model] and the *actual* state of affairs.

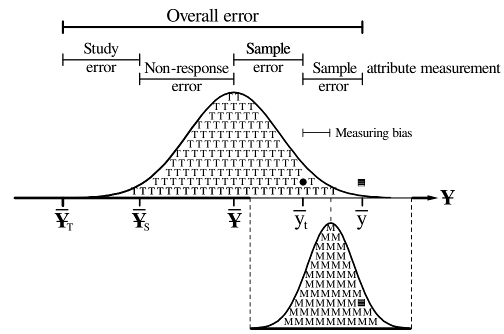
We distinguish six categories of error. (5.19, 5.25, 5.52 to 5.54, 5.84)

- **Study error:** the difference between [the (true) values of] the study population/process attribute and target population/process attribute. (5.20, 5.22, 5.23, 5.37, 5.40, 5.50, 5.84)
- **Non-response error:** the difference between [the (true) values of] the respondent population/process attribute and the study population/process attribute. (5.25, 5.26)
- **Sample error:** the difference between [the (true) values of] the sample attribute and the respondent population/process attribute. (5.20, 5.22, 5.50, 5.80 to 5.82, 5.84)
- **Measurement error:** the difference between a measured value and the true (or long-term average) value of a variate. (5.22)
  – **Attribute measurement error:** the difference between a measured value and the true (or long-term average) value of a [population/process or sample] attribute. (5.20)
- **Model error:** the divergence of the modelling assumptions from the *actual* state of affairs in the real world. (5.27, 5.28, 5.43)
- **Comparison error:** for an Answer about an **X**-**¥** relationship that is based on comparing attributes of groups of units 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 units 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. (5.30, 5.36 to 5.39, 5.42, 5.45, 5.46, 5.50, 5.51, 5.54, 5.70, 5.75, 5.80 to 5.82, 5.84)

Table HL91.5: SYMBOL DEFINITIONS

¥	Response variate
$\overline{\text{¥}}_T$	(True) target population average
$\overline{\text{¥}}_S$	(True) study population average
$\overline{\text{¥}}$	(True) respondent population average
$\overline{y}_t$	True average for sample selected
$\overline{y}_m \equiv \overline{y}$	Measured average for sample selected
T	True value of a sample average
M	Measured value of a sample average



We need to include both true values and long-term average values in two error definitions because:

– 'true' values for quantities like length, mass and time (and the many quantities derived from them) can be invoked because **standards** for measuring for such quantities are defined; BUT:
– long-term average values may be all we have available when, for instance, investigating for a particular questionnaire the effect of question wording and/or question order on the distribution of responses.

Our useage of *error* as the difference between an Answer and the true state of affairs is evocative of the central concern of data-based investigating and, more generally, of scientific enquiry. *We* avoid distraction from this central concern by our useage of **mistake** and **residual**.

We never use 'error' to refer to the **residual** in a regression model.

The concept of error and its categorization are important in statistics for several reasons.

– Error leads to recognizing the ideas of imprecision, inaccuracy and uncertainty and to their succinct definitions – we then see why statistical methods aim to *manage* ***imprecision*** *(by managing* ***variation****) and* ***inaccuracy.***

  It is unfortunate that the more familiar (and seemingly more straight forward) **accuracy** and **precision** are the *inverses* of what statistical methods manage directly.
– Error is the source of **limitations** imposed on Answer(s).
– When estimating an **average** to answer a Question with a descrip-

 (*continued overleaf*)

tive aspect, a convenient breakdown of the *overall* error is:

overall error = *study error* + *non-response error*
+ *sample error* -----(HL91.4)
+ *sample attribute measurement error.*

In the diagram overleaf at the top of the right-hand column (from page HL18.4 of Statistical Highlight #18), the black filled circle and square (• and ▪) represent the true and measured average for the *actual* sample from among the sets of all possible samples and measured values (modelled by **normal** distributions). For the true and measured averages of the *actual* sample, desirable as the subscripts t and m are as a reminders of a vital distinction, retaining the latter for such a widely-used symbol is unrealistic; *users* must take responsibility for remembering that *un*subscripted $\overline{y}$ is a *measured* sample average.

Licence on two matters improves the clarity of the diagram:

– all four error components are *positive* – in practice, overall error may involve some *cancellation* among error components of *opposite* sign;
– the distribution of *measured* sample attribute values has been moved *down*.

It is rare in data-based investigating to need to manage fewer than *three* categories of error – typically study, sample, measurement error.

+ If formal statistical methods of data analysis are used, model error must be managed – up to 4 categories to manage.
+ With *human* units (and sometimes with *in*animate units), non-response error must be managed – up to 5 categories to manage.
+ Comparative Plans require managing comparison error – up to 6 categories to manage.

Overall error in an investigation refers to the net effect of *all* relevant categories of error on the Answer(s) from the investigation.
See also **Residual**. (5.19, 5.25, 5.54)

Estimate: *numerical value(s)* for a population attribute or model parameter:

– derived from (the distribution of) the corresponding *estimator,* AND:
– calculated from *data*.
• **Point estimate:** a *single* value for an estimate.
• **Interval estimate:** an *interval* of values for an estimate, usually in a form that quantifies **variability** (representing imprecision). (5.21)

Estimated (or **realized**) **residual:** see **Residual**.

Estimating: a process which uses statistical theory to derive the distribution of an **estimator** and data to calculate an (interval) **estimate.**
In contrast to its statistical meaning, 'estimating' in ordinary English usually means 'approximating', based perhaps on intuition or common sense. A statistical estimate is, in a sense, an approximate value, but:

– is obtained by a defined process of inductive reasoning, AND:
– under appropriate modelling assumptions, (some) sources of error are quantified under repetition.

Estimator: a **random variable** whose distribution *represents* the possible values of the corresponding **estimate** under repetition of the selecting, measuring and estimating processes. (5.21)

Even: see **Odd**.

Event: an event (A, say) is a subset the points in the sample space S.

• **Complement:** the complement of event A is the set of points in S but *not* in A; *we* denote it ^{c}A.
• **Disjoint events** have no points in common. (5.29)

Evocativeness: a commitment to choosing symbols and terminology that are **evocative**: bring to mind the entity represented or named.
Examples are random variables R for residual or ratio, S for standard deviation and T for time; EPS for equiprobable selecting.
*Un*evocative choices are:

– ANOVA for *Analysis of Variance*, which does not involve what would usually be thought of as either 'analysis' or 'variance'.
An evocative acronym would be PASSDI for *Partitioning Sums of Squared Differences* – see **ANOVA**.
– Degrees of freedom.
– Disorder as used to describe increasing entropy – whatever its technical merits, for most people it evokes an image of the *opposite* of increasing *uniformity.*
– r as the symbol for correlation – historically, c would have been preferable, leaving R and r for residual and ratio.

Exchange Paradox: see **Ali-Baba Paradox**.

Experiment: in ordinary English, 'experiment' is often associated with investigative activity in a biological or chemical laboratory.
In the ten statements of the law of large numbers on the upper half of page HL91.3, five use the word 'experiment' although the (implied) context is repetition of a process involving probability – '**trial**' (as in the seventh and tenth statements) is then a better term.
Useage like this shows it is best to avoid 'experiment' in statistics because of the key statistical meaning of **experimental**.
Use of 'Experiments' in 'DOE' is regrettable – see **DOE**.

Experimental: to be contrasted with **observational** – it indicates a comparative Plan where the *investigators* (*actively*) assign the value of the focal variate to each unit in the sample or in each block.
(5.36, 5.38, 5.39, 5.45 to 5.50, 5.54)

Exploratory data analysis: see **Quantiles**.

External validity: social science terminology for **study error**.
See also **Internal validity** and **Validity**.

Factor: an explanatory variate; (5.43) we distinguish a factor that is:

– a **focal** variate;
– a *non*-focal variate used as a **blocking factor**; (5.36, 5.37)
– a *non*-focal variate value managed for other reasons. (5.45)

Factor level: see **Level** and also **Confidence level**.

Factorial treatment structure: *all* combinations of the levels of the (two or more) factors. (5.43, 5.44)

• **Fractional factorial treatment structure:** a subset of (the runs of) a (full) factorial treatment structure.
For instance, 8 (properly-chosen) runs from a full factorial structure of 16 runs is a *half fraction*. (5.44, 5.45)

Falsehood: in data-based investigating, an answer whose difference from the true state of affairs is large enough to be practically important in the question context. See also Appendix 2 on page HL91.24.
Such an answer may even be harmful – see Appendix 8 on pages HL77.13 and HL77.14 (and HL79.2) in Statistical Highlight #77.
Two reasons falsehoods arise are:

– bad luck, as in the roughly 5% of 95% confidence intervals that do *not* cover the value of the population attribute or model parameter, OR, MORE COMMONLY:
– inadequate error management.

Proper application of statistical methods of error management can make falsehood unlikely in a particular investigation but cannot preclude it. Conversely, ignoring statistical precepts does not guarantee falsehood – a sample of size one obtained by convenience selecting *may* have a variate value close to the population average. See also Appendix 1 on pages HL77.8 and HL77.9 in Statistical Highlight #77.

False positive: when there is an **X-Y association** for which *coincidence* can be ruled out as the reason, a partial or complete false positive is **causation** of **Y** that is *not* only, or even in part, by **X**. (5.74, 5.75)

False negative: absence of association of **X** and **Y** even though they have a **causal** relationship. (5.74, 5.75) See also
Section 4 on pages HL60.3 and HL60.4 in Statistical Highlight #60.
Our definitions of false positives and false negatives are for the specific context of association/correlation and causation but the *same* idea arises on other contexts, although differences in terminology may obscure the underlying common theme. Other contexts include:

– *measuring with a* ***binary*** *outcome* (*e.g.*, 0/1, Yes/No, Pass/Fail, Guilty/Not guilty, Present/Absent): these occur in student assessment and quality testing of manufactured items, verdicts in criminal trials, medical testing for a disease.
– **hypothesis testing** (a *decision* rule) where (unhelpfully) accepting a false null hypothesis is called a **Type I error** and rejecting a true null hypothesis is a **Type II error**.
There is the same underlying issue in **significance testing**, but it is less obvious because of the emphasis on 'strength of evidence' (as opposed to making a decision).

The *logical* structure of these matters is analogous to constructing a 2×2 contingency table.

FDEAC cycle: acronym for a 5-stage *structured process* for data-based investigating (as in Table HL91.4 on page HL91.5); the stages are:

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 6)

- **Formulation stage:** formulating clearly the Question(s) for which the investigation is intended to provide Answer(s).
- **Design stage:** drawing up a Plan for the processes that will generate Data that will provide Answer(s) to the Question(s).
- **Execution stage:** carrying out the Plan – collecting (selecting, measuring), examining, monitoring and storing the Data.
- **Analysis stage:** summarizing and analyzing the Data in ways that effectively provide Answer(s) to the Question(s).
- **Conclusion stage:** giving Answer(s) to the Question(s) in *context*, their limitations, and (if appropriate) recommendations.

The FDEAC cycle is described in detail in Statistical Highlights #88 and #89. See also **PPDAC cycle**.

Feedback: see **Causal chain**.

Fishbone diagram: a schematic display (reminiscent of a fish skeleton) for *organizing* the names of **explanatory** variates which may affect a particular **response** variate; there can be up to *six* main branches on the diagram, with labels like *measurement*, *person*, *environment*, *method*, *material* and *machine*. (5.23, 5.57 to 5.59)
See also **Ishikawa**.

Focal (explanatory) variate: for a Question with a **causative aspect**, the **explanatory** variate whose relationship to the **response** variate is involved in the Answer(s) to the Question(s). (5.28, 5.72)

Frame: a list [real or conceptual (*e.g.*, a rule that would, if implemented, generate the list)] of the units that can be selected from the respondent (or study) population. (5.56, 5.57, 5.64, 5.84, 5.86)

Functional dependence: see **Independence**.

Fundamental Theorem of Statistics: under equiprobable selecting (EPS), the distribution of the sample average in the set of all possible samples of size n is the distribution of the random variable $\bar{Y}$ representing the sample average in a (response) model. (5.91, 13.8)

Gamma function: The definition is:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx. \quad \text{-----(HL91.5)}$$

Two properties of the gamma function are:

- $\Gamma(\alpha) = (\alpha-1)!$ if α is a positive integer;
- $\Gamma(½) = \sqrt{\pi}$.

We can think of the gamma function as a generalization of the idea of a factorial. (5.9)

Gauge: a synonym for *measuring instrument*, often used, for example, in manufacturing industries. See also **Measuring process**.

Gauge R&R investigation: an investigation to quantify the **repeatability** and the **reproducibility** of a gauge. (5.62)
See also **Measuring process**.

Gaussian distribution: used in STAT 231. See **Normal distribution**.

Generality: a word to be avoided as duplication of existing statistical terminology; generality (of an Answer) usually refers to **sample error**.
In DOE, *generality* (or a **wider inductive basis**) may refer to using a factorial treatment structure so that interaction effect(s) can be estimated. (5.85) See also page HL91.6.

Generalization: another word to be avoided as statistical terminology; like *generality*, it may refer to **sample error**. (5.85)

Generalizability: another word to be avoided as statistical terminology; it may refer to **study error**. (5.85) See also page HL91.6.

Haphazard selecting: see **Non-probability selecting**.

Heisenberg: the **Uncertainty Principle** is a staple of modern physics; its origin in mathematics (presumably) makes it a *model* statement.
It has profound implications for the model of 'reality' provided by quantum mechanics.
Discussions of the Uncertainty Principle often leave it unclear whether *measuring* limitations are incidental to it or inherent in it and whether 'knowing' and 'measuring' are to be distinguished.
See also **Uncertainty** and **Variation**.

Histogram: a bar chart used to display the *distribution* of the values of a variate; the horizontal axis defines appropriate *intervals* of variate values and the *area* of a bar is the *proportion* of values that fall in the interval covered by the bar.
A histogram has a *density* scale on its vertical axis; many displays called 'histograms' are only *bar graphs*, which have a *frequency* scale on their vertical axis.
Replacing the 'stepped' profile of a histogram with a smooth curve can ease for students:

- the transition from discrete to continuous distributions;
- the idea of using a continuous model (like the normal distribution) to approximate a discrete model (like the binomial distribution or the Poisson distribution).

See also Statistical Highlight #26 and **Ishikawa**.

Hot: unlike the *lower* limit on 'cold', there seems to be no (hard) *upper* limit on 'hot'.

Hypothesis: see **Hypothesis testing** and **Significance testing**.

Hypothesis testing: this term should be used *only* to refer to statistical testing used as a *decision rule*, to emphasize the distinction from ***(statistical)*** **significance testing** for assessing *strength of evidence*.
When using statistical testing as a decision rule:

- the **null hypothesis** is a model parameter value corresponding to *no effect*;
- the **alternative hypothesis** is a model parameter value corresponding to there *being* an effect;
  [only a positive effect or only a negative effect is a **one-sided alternative**; accepting *both* directions is a **two-sided alternative**.]
- with a low enough *P*-value (*e.g.*, below .05), the null hypothesis is **rejected**, in favour of **accepting** the alternative hypothesis.
  [This point is obfuscated by additional unnecessary terminology: *A **critical value** (like .05 or .01) is the boundary between the **acceptance region** (where the null hypothese is* not *rejected) and the **rejection region** (the null hypothesis* is *rejected).*]

Such double negatives (like rejecting no effect meaning there *is* an effect), phrased in unfamiliar terminology, invite confusion.
We can think of rejecting the null hypothesis as (a claim of) a signal being detectable above the noise.
As a probabilistic argument by contradiction with specialized (and *un*evocative) terminology, hypothesis testing has *not* been helpful to the image of statistics as a souce of important and useful ideas.
A **critical value** (like .05 or .01) is a probability [the tail area(s) of a probability distribution] and is an idea that arises in three contexts:

- as the boundary between accepting or rejecting the null hypothesis in hypothesis testing used as a decision rule,
- as the boundary between what is not or is statistically (or highly statistically) significant in (statistical) significance testing used to assess strrength of evidence,
- to determine the confidence *level* in a **confidence interval** as a way to quantify uncertainty under repetition, although the concern is now with the *central* (not the tail) area of a distribution.

Hypothesis testing as a *decision rule* under uncertainty has been (over)sold as an appealing idea but *strength of evidence* in (statistical) significance testing is a better approach to the same problem.
See also **False negative** and **Significance testing**.

Ignorance: in these Materials, **uncertainty** is ignorance of error.
Ignorance of *subject-matter* can lead to impaired decision-making.

Impotence: see **Postulates of Impotence**.

Imprecision: **standard deviation** of error (*i.e.*, its *haphazard* component, exhibited as **variation**) under **repetition**. (5.21)

- **Sampling imprecision:** standard deviation of sample error under repetition of selecting and estimating. (5.25, 5.37, 5.39, 5.86)
- **Measuring imprecision:** standard deviation of measurement error under repetition of measuring the *same* quantity. (5.25, 5.50, 5.60)
- **Comparing imprecision:** standard deviation of comparison error under repetition of assigning and estimating. (5.37, 5.40, 5.46, 5.50, 5.52)

See also **Precision**.

Imputing: the process of assigning values for missing observations

– *e.g.*, assigning a value for the response of a non-respondent on the basis of its values for known explanatory variates (like sex, age, location) that (it is hoped) are reasonable 'predictors' of the response variate.

- The purpose of imputing (or **Imputation**) is to simplify the data analysis; it *rarely* meaningfully increases the completeness of the information in the data.

Inaccuracy: *average* error (*i.e.*, its *systematic* component) under **repetition**. (5.20)

- **Sampling inaccuracy:** average sample error under repetition of selecting and estimating. (5.25)
- **Measuring inaccuracy:** average measurement error under repetition of measuring the *same* quantity. (5.25, 5.60)
- **Comparing inaccuracy:** average comparison error under repetition of assigning and estimating. (5.50)

Inclusion probability: see **Selecting probability**.

Independence, Independent: a dictionary definition is: *not subject to the control, influence or determination of another or others.*

- **Independent measurements:** measurements are independent when the operator's knowledge of the value arising from one execution of the measuring process does *not influence* the value from any other execution. (5.28, 5.60)
- **Independent events (Probabilistic independence): events** A and B are independent when the probabilities of *events* A and B are such that $\Pr(A|B) = \Pr(A)$ and $\Pr(B|A) = \Pr(B)$. (5.60)
- **Independent random variables:** two random variables are independent when their **joint** probability (density) function is the *product* of their **marginal** probability (density) functions.

The idea of independence is a (mathematical) *idealization* – the usual state of affairs in the real world is one of ***dependence*** (*i.e.*, *lack* of independence). (5.11, 5.12, 5.29, 5.47, 5.75, 5.79)
See also the Appendix on page HL89.18 in Statistical Highlight #89.
In the response model (HL91.1) at the upper right of page HL91.4:

- the random variables R_j are taken as being *probabilistically independent,* SO THAT:
- the random variables Y_j are *probabilistically independent*, BUT:
- Y_j and μ are **functionally dependent** in the sense that Y_j values are affected by the value of μ; the possibility of functional dependence should be kept in mind when interpreting a scatter diagram.

Indicator variate: a binary variate which takes only values of 0 or 1. See also **Binary (response) variates**. (5.32, 5.74)

Induction: reasoning from *particular* cases, investigations, or data to a more *general* conclusion, using the rules of logic and relevant theory. (5.23) Induction is to be contrasted with **deduction**.
We think of inductive reasoning in terms of incomplete information leading to *uncertain* Answers to (statistical) Questions.

Infinity: a difficult concept (made more so by its mathematical properties).
We encounter infinity in, for example, the seeming endlessness of the sequence of (positive) integers or real numbers.
It seems reasonable that there would be *more* real numbers than integers and hence Cantor's *different* infinities $\aleph_0$ and $\aleph_1$ (ordinal numbers), but more than one infinity seems paradoxical.
Likely more familiar is the infinity that results from dividing by zero, which can be troublesome in mathematical modelling.

Influential observation: in regression, an observation whose x-value differs substantially from that of the other (bivariate) observations.
We distinguish an influential observation from an **outlier** which has a deviant y-value; 'outlier' is used in a broader context than regression.
See data sets 4 and 3 on page HL53.2 in Statistical Highlight #53.

Interact: a word with a different meaning in ordinary English from **interaction** in statistics. See **Interaction**.

Interaction of two factors $\mathbf{X}_1$ and $\mathbf{X}_2$ is said to occur when the effect of one factor on a response variate $\mathbf{Y}$ depends on the level of the other factor. Interaction means the combined effect of two factors is *not* the sum of their individual effects. (5.44, 5.45, 5.75 to 5.79)
Interaction may involve effects of *more* than two factors. (5.76, 5.81)
Thus, three (or more) factors are involved in interaction in statistics; in normal English, ordinarily only *two* 'factors' interact.

Internal validity: social science terminology for **comparison error**. See also **External validity** and **Validity**.

Interquartile range (IQR): see **Quantile**.

Intersection: the intersection [denoted $A \cap B$ (or AB)] of events A and B is the event comprising the set of all points in A *and* in B.

- **Union:** the union (denoted $A \cup B$) of events A and B is the event comprising the set of all points in A *or* in B *or* in *both*.

Interval estimate: an *interval* of values for an estimate, usually in a form that quantifies imprecision. (5.21)
See also **Confidence interval** and **Estimate**.

Investigation: a *data-based* investigative undertaking involving one (or a few) Questions to be answered.
An investigation is to be contrasted with a **project**.
See Section 2 starting on page HL88.2 in Statistical Highlight #88.

Invalidity: a word to be avoided as duplication of existing statistical terminology; invalidity (of an Answer) means **inaccuracy**. (5.85)

Ishikawa: see Note 27 on page HL88.18 in Statisticsl Highlight #88 and Statistical Highlight #97.

Joint probability function: see **Probability function**.

Judgement selecting: human judgement is used to select n units from the N (or N_S) elements/units of the respondent (or study) population.

- Judgement selecting [implemented to achieve proper **covering**] is commonly used for a Question with a causative aspect investigated using an experimental Plan.
  Usually, judgement selecting is used because probability selecting would be infeasible to implement.

(5.22, 5.23, 5.24, 5.38 to 5.39, 5.52, 5.56, 5.80 to 5.82)
See also **Non-probability selecting** and Statistical Highlight #83.

K_ν distribution: the distribution of a random variable which is the square root of the average of the *squares* of ν independent $N(0,1)$ random variables; the parameter ν is called the **degrees of freedom**. (13.15 to 13.17) See also Statistical Highlight #105.

Least squares: a process for estimating response model parameters, based on minimizing the sum of the squared residuals. (8.1, 8.2)
See also Statistical Highlight #73.

Level: factor levels are the set of value(s) assigned to a factor; that is, (usually) the set of values assigned to the (or a) focal variate. (5.43)

Likelihood function: the probability of the data as a function of the model parameter(s).

- **Relative likelihood function:** a function of the model parameter(s) defined as the likelihood function *divided* by the *maximum* value of the likelihood function.
  The denominator is the likelihood function evalued at the maximum likelihood estimate and is a *number.*

'Likelihood' in its statistical meaning is *not* a synonym for 'probability.'

Limitations: apply to Answer(s) to the Question(s) and must:

- assess the likely importance of each category of error;
- be expressed in the language of Question *context.*

Limitations are inherent in answers obtained from incomplete information (*e.g.*, from sampling and measuring).
Error management aims to reduce limitations to a level acceptable in the investigation context; ideally, this reduction is only as much as is *needed*, so as to conserve resources.

Location: refers to where (the 'centre' of) a distribution is positioned; in this sense, averages and means are *measures of location.* (5.28)
The same idea called a *measure of central tendency* is best avoided.
See Table HL91.2 on page HL91.4.

Location-scale transformation refers to a mathematical operation that alters the centre and/or spread of a data set or probability distribution – for instance, taking the logarithm.
Such transformations are used, for example, in managing model error.

Lurking variate/confounder: a non-focal explanatory variate whose differing distributions of values (over groups of elements/units) for

University of Waterloo W. H. Cherry

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 7)

different values of the focal variate, if taken into account, would meaningfully change an Answer about an **X**-**Y** relationship.
Lurking variates/confounders are responsible for the limitation imposed by comparison error on an Answer about a relationship.
(5.30 to 5.33, 5.43, 5.65 to 5.76)

Main effect: see **Effect**.

Marginal probability function: (5.68) see **Probability function**.

Margin of error: wording used in polling to describe the half-width of a confidence interval, typically for a (population) proportion.
A qualifier of *19 times out of 20* indicates a confidence *level* of 95% for the confidence interval; the behaviour under repetition implied by this phrase is likely obscure to most readers.
It is also seldom made clear that the 'margin of error' is *only* for **sample error** – 'margin of error' *sounds* more like **overall error**.

Matching in an *observational* Plan: forming groups of units with the *same* (or similar) values of one or more non-focal explanatory variates but *different* values of the **focal variate**. [See also **Blocking**] THUS:
Matching meets 'lurking variates' criterion (1) [on the lower half of the left-hand column on page HL91.8] 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**.
(5.37 to 5.40, 5.43, 5.49, 5.52)

Maximum likelihood: a process for estimating the values of model parameter(s), based on maximizing the likelihood function.

Maximum likelihood estimate [mle]: the value of the model parameter(s) which *maximize(s)* the value of the likelihood function.

Mean: a measure of **location** of a **random variable**. (5.5, 5.28)
A (model) *mean* is to be distinguished from a (real world) **average**.
See Table HL91.2 near the middle of page HL91.4.

Meaning: what a sentient being can extract from information to generate knowledge.
The distinction between *information* and *meaning* is useful in separating matters that *Information Theory* can and can*not* address.
It is curious that *probabilistic* structures, which comprise neural networks, coupled with immense computing power, can make it appear that a computer has extracted meaning from information.

Meaningful: see **Practical importance**.

Measurement error: see **Error**.

Measuring: the process used to determine the value of a variate.
(5.28, 5.59 to 5.62) See also **Sensing**.

Measuring instrument: see **Measuring process**.

Measuring process: a process for *quantifying* a variate value.
The *components* of a measuring process are:
- the *measuring instrument* or *gauge*;
- the *operator(s)*;
- the *measuring protocol*: the instructions for how to measure;
- the *element/unit measured*. (5.59 to 5.62)

See also Statistical Highlight #38.
An experienced measurer should maintain a healthy skepticism about the values of *all* measurements, especially those from automated and/or complex measuring processes; skepticism of the results of calculations, particularly complex ones, is also useful.

Measuring protocol: see **Measuring process**.

Median: the half-way point;
- for a *probability distribution*, the median divides the area under the probability (density) function in half;
- for a *data set* with an *odd* number of observations, the median is the central observation of the *ordered* data set;
- for a *data set* with an *even* number of observations, the median is half way between (*i.e.*, the average of) the two central observations of the *ordered* data set. (5.64)

See also **Quantiles**.

Milne: Edward A. Milne (1896-1950) was a British mathematician and astrophysicist.
His writings include a theory known as Kinematic Relativity (different from Einstein's relativity) based on an *a priori* principle of equivalence of "fundamental" observers and stressing *communicability* among observers, different from **Eddington's** emphasis on *measurability*.
See **Sensing**.

Minus: familiarity from an early age with minus signs in mathematis makes it easy to overlook their different roles:
- indicating subtraction,
- arising from deductive reasoning as a precursor to a symbol; when the symbol represents a physical entity (like time or energy), the physical implications of a minus sign may be equivocal,
- part of a lower limit of an interval – such a limit may be unrealizable physically (*e.g*, in a confidence interval for weight),
- the square root of −1 (usually denoted i) is mysterious in itself (perhaps less so as $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, etc), although its use opens up a valuable area of mathematics and it arises in quantum mechanics (*e.g.*, in the Schrödinger equation).
  $e^{i\pi} = -1$ compounds the mystery but, if t is time, e^{it} represents a *circle* (of radius 1 in the complex plane); resolving e^{it} into its cosine and sine components brings in the idea of a *wave*.
  The dictum: ... *multiplication by* ***i*** *in a physical diagram merely means turning the picture from horizontal to vertical* ... may be an oversimplification.

See also **Equations**.

Missing data: (5.24, 5.26, 5.52, 5.53) see **Non-respondent**.

M.l.e.: see **Maximum likelihood estimate**.

Model: a model in statistics is a mathematical structure that tries to describe the properties or behaviour of a real-world phenomenon. Statistical models usually involve probability distributions.
In *any* situation where an Answer is based, in whole or in part, on a mathematical model, we should bear in mind a maxim of the late Dr. George E. P. Box, a respected U.S. statistician:
All model are wrong, some are useful.
See also **Response model**.

Model error: see **Error**.

Modelling assumptions: *we* only assess how well *five* modelling assumptions appear to be met – whether: (5.27, 5.28, 5.43, 5.50, 13.8)
- the selecting process for units *is* (equivalent to) ***EPS***;
- the response model **structural component** form is appropriate;
- a normal model *is* appropriate for the distribution of the **residuals**;
- there is equality among **standard deviation(s)** [or they vary in a *known* way, such as dependence on an explanatory variate];
- the residuals can be taken as **probabilistically independent.**

More generally, recognizing the modelling assumptions underlying a particular data-based investigation and assessing how well they are met is an *onerous* (and vital) task.

Modelling: an early example is the association of the five Platonic solids – tetrahadron, octahedron, dodecahedron, cube and icosahedron – with the four 'elements' – fire, air, water, earth and the aether. The harmonies exemplified by the Platonic solids were seen as reflections of harmonies in the structure of the universe. This modelling implied that it might be possible to achieve interconversions among the 'elements' by appropriate manipulations of the relevant regular polygons – triangles and squares – but *how* to do so remained elusive.
About two other historic models, David Bentley Hart has commented:
... the Aristotelian model of the universe was an object of rare beauty, with its immense etherial machineries, its imperishable splendors, its innumerable wellsprings of harmony and synchrony; and the Ptolemaic system, with its intricate coils and spirals and elaborately exact actions, was as exquisitely glittering a cage as any reasoning mind could hope to inhabit. Practically every educated intellect was in thrall to that model and confined to that cage; a few perceptive souls were aware that the two systems did not perfectly coincide, but were still

(*continued overleaf*)

more or less condemned to circle back and forth between them.

These illustrations of older 'descriptive' modelling may indicate that such 'modelling' is like Lakoff's 'framing', a term he uses to describe how thought processes rationalize our experience of 'reality.'

More recent overtly *mathematical* modelling involving equations (like the theory of relativity and quantum theory) are notable for how closely their calculated values or predictions align with observation (like the relativistic correction to the advance of the perihelion of mercury's orbit). However, modelling like this raises the matter of the *un*reasonable effectiveness of mathematics – a *logical* structure – in describing 'reality' – a *material* structure.

Model parameter: a constant (usually denoted by a *Greek* letter) in a response model that *represents* a respondent **population attribute**. (5.27, 5.28) See also **Attribute**.

We avoid the term **population parameter**.

Monty Hall: see Statistical Highlight #49.

Muddled thinking: Should it (or, worse, gobbledygook) masquerading as 'explanation' in statistical materials provoke statisticians to rueful laughter, tears or a commitment to do better? [See pages HL94.4 and the bottom of HL94.10 in Statistical Highlight #94.]

n − 1 *vs.* n saga: Unprofitable debate about the 'correct' divisor for calculating (under constraints) the average deviation from the average for a (data) standard deviation – see Appendix 3 on pages HL100.7 to HL100.9 in Statistical Highlight #100. See also **Average**.

Negative: see **Minus**.

Non-probability selecting: a process for seleting a sample in which the unit inclusion probabilities are *un*known. (5.56, 5.57)

In addition to **judgement selecting**, such methods include:

- **Accessibility selecting:** selecting units (easily) *accessible* to the investigator(s) – for instance, the *top* layer in a basket of fruit or a truckload of potatoes or the *front* pallets or cartons in a large stack in a warehouse.
- **Convenience selecting:** selecting units that are *conveniently* available to the investigator(s) – *e.g.*, people with a medical condition of interest at a hospital or clinic *nearby* to the investigator(s).
- **Haphazard selecting:** selecting units with*out* (conscious) preference by the investigator(s) – shoppers who pass the location of an interviewer in a mall or rats in a cage which are more easily caught for a laboratory test.
- **Quota selecting:** selecting units according to values of specified explanatory variates (like sex, age, income for human units) so the sample distribution of each variate will (approximately) match that of the study population.
- **Volunteer selecting:** asking for (human) volunteers, usually after a brief explanation of what the investigating will entail for units in the sample.

These names do not necessarily specify a *unique* selecting method – the first two methods overlap and all five involve some degree of 'accessibility' and/or 'convenience'.

Haphazard selecting is sometimes ***wrongly*** equated with 'random' selecting; *i.e.*, with our **equiprobable selecting**.

Quota selecting is a similar idea to **covering**.

Volunteer selecting is *not* to be confused with **volunteer** (or **voluntary**) **response**, a phrase sometimes used to indicate that *human* units can (usually) *choose* whether to respond, *i.e.*, whether to provide the requested data; a separate (measuring) issue is whether these responses are correct or truthful – see **Randomized response**.

Non-respondent: an element/unit with some or all of its data missing at the end of the Execution stage of the **FDEAC cycle**.

Missing (sample) data may be due to:

- *non-contact* with the unit, OR TO:
- non-response (partial or complete) when the unit *is* contacted.

'Non-respondent' usually refers to a *human* unit, whereas 'missing data' is more commonly applied to an *in*animate unit and may arise from measuring instrument malfunction. (5.24, 5.26, 5.52, 5.53)

The *non-respondents* plus the **sample** comprise the **selection** in our terminology – see **Selection**.

Non-respondent population: see **Respondent population**.

Non-response error: see **Error**.

Normal distribution: a probability distribution with a symmetrical bell-shaped probability density function on the interval $(-\infty, \infty)$; this p.d.f. is of the form e^{-y^2}. It is also called a **Gaussian distribution**.

The normal distribution is denoted $N(\mu, \sigma)$ [or $G(\mu, \sigma)$], where:

- μ is the **mean** AND:
- σ is the (probabilistic) **standard deviation**. (5.15 to 5.18, 5.28)

Elsewhere, the notation may be $N(\mu, \sigma^2)$, in which the second parameter is (*un*helpfully) the *variance* – see also **Variance**.

The **standard normal distribution** is $N(0, 1)$; values for its probabilities (areas under its p.d.f.) are tabulated in many texts.

See Statistical Highlight #104.

Nothingness: a difficult concept, easily confused with:

- 'nothing', meaning zero or zero magnitude;
- 'absence of difference', as in undifferentiated sameness.

Null hypothesis: see **Hypothesis testing**.

Observation: an individual datum.

Best *avoided* because of the key statistical meaning of **observational**.

Observational: to be contrasted with **experimental** – it indicates a comparative Plan where, for each unit selected, the focal explanatory variate (*passively*) takes on its 'natural' value *uninfluenced* by the investigator(s). (5.36 to 5.43, 5.49, 5.51, 5.54, 5.82 to 5.84)

Some statisticians use *observational* as the adjectival form of *observation*, a fundamental component of the scientific method; because observation is *always* involved in data-based investigating, this *non*-technical (tautological) use of 'observational' is to be avoided in statistics.

Odd: the distinction between 'odd' and 'even', as exhibited by the integers 2 and 3, appears profound – for instance, no power of 2 is (exactly) divisible by 3 (and *vice versa*?).

Operator: see **Measuring process**.

Order of magnitude: a factor of 10, so 'two orders of magnitude' is a factor of 10^2 or 100.

The phrase is typically used to convey succintly (and approximately) a difference of large magnitude – for instance, the atomic nucleus is around 5 orders of magnitude smaller than the atom.

Orthogonal: a generalization of the geometric concept of perpendicular; 'orthogonal' and '**perpendicular**' are often (carelessly) used interchangeably.

Outcome: the result generated by one execution of a *probabilistic* process. See also **Data** and page HL94.3 in Statistical Highlight #94.

Outlier: an observation whose variate value(s) differ substantially from those of other observations in the data set.

See also **Influential observation**.

Overall error: see **Error**.

Pairing: A *matched pair* is a **block** of size *two* units.

Pairwise difference: see **Standard pairwise difference**.

Parameter: see **Model parameter**.

Pareto diagram: see **Ishikawa**.

Perceiving: see **Sensing**.

Percentages: see the bottom of page HL46.2 in Statistical Highlight #46.

Percentiles: see **Quantiles**.

Perpendicular: an angle of 90° (or $\pi/2$ radians) between two directions (or axes); this (seemingly) allows entities associated with the two directions to be treated as though they were 'independent'; we exploit perpendicularity in:

- Cartesian axes, conventionally labelled x and y and referred to as 'horizontal' and 'vertical',
- the real and 'imaginary' axes of the complex plane,
- squaring and adding under a square root to calculate the length of the hypotenuse of a right-angled triangle, which is also the way statistics combines (independent) standard deviations; a distraction from this idea is the fact that variances *add*.
- obtaining the 'least' in least squares minimization using linear algebra – see step 4 in three places in the right-hand columns of Statistical Highlight #73.

Pivotal quantity: see **Discrepancy measure**.

(*continued*)

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 8)

Placebo: see **Blind**.

Pleonastic fallacy: The belief that an absolute *qualitative* difference can be overcome by a successive accumulation of extremely small and entirely relative *quantitative* steps.
See page HL46.2 in Statistical Highlight #46.

Plus: The (long familiar and seemingly straight forward) operation of addition. See also **Minus**.

Point: see **Sample space** and page HL94.3 in Statistical Highlight #94.

Point estimate: a single value for an estimate.
See also **Interval estimate**.

Poisson distribution: Conventionally (but *un*helpfully) described as being used to model 'random events in space or time'.
It is useful to recognize that the Poisson probability function is successive terms of the power-series expansion of e^{μ}, divided by e^{μ}.
This (of course) is why the (infinite) sum of the Poisson probability function terms is 1.
See also pages HL4.4 and HL94.11 to HL94.12 Statistical Highlights #4 and #94.

Population: a well-defined group of *elements*, ***other than*** the sample. An infinite 'population' is a (sometimes useful) *model*. (5.19, 5.55)

Population parameter: see **Model parameter** and **Attribute**.

Postulates of Impotence: Statements about the impossibility of achieving certain conditions in the material world, like a speed greater than that of light.
See pages HL9.9 and HL9.10 in Statistical Highlight #9.

PPDAC cycle: the acronym for a original (early 1990s to early 2000s) 5-stage *structured process* in STAT 231 for data-based investigating; this Glossary uses instead the FDEAC cycle because:
- for the first three stages, *Formulation*, *Design* and *Execution* are more evocative names than *Problem*, *Plan* and *Data*; also, our Design stage corresponds to the meaning of 'Design' in 'DOE'.
- these three revised names avoid confusing the stage *name* with its *input* or *output*.

Some later versions of PPDAC became QPDAC, naming the first stage for the *Question*, but this still invites confusion of the first stage name with its input.

Practical importance: a practically important numerical Answer is one whose value is *not negligible* in the Question context in relation to a prescribed value.
The issue of practical importance arises most commonly with an Answer that is a *difference* or a *ratio*, when the respective prescribed values are often 0 and 1.
A practically important *change* (or *difference*) (*e.g.*, in an Answer) can also be called a **meaningful** change (or difference).
We *must* distinguish practical importance from statistical significance – see **Significance testing**. (5.30, 5.46, 5.47, 5.71)

Precision: the inverse of *im*precision. See also **Reliable**. (5.21)
'Precision', which involves behaviour under repetition, arises when:
- a numerical answer to a (statistical) question is being sought,
- inductive reasoning (based on *in*complete information) is used to obtain the answer which likely involves (overall) error (and so has limitations).

We consider how the answer would *vary* if the processes which generated it (like selecting, estimating and measuring) were to be repeated over and over; *small* variation means a *precise* answer, increasing variation represents increasing *im*precision of the answer.
- + Assessing the effect of repetition is, in practice, rarely achieved by repeating an investigations but rather by **modelling**; we see this in answers which are **interval estimates**.
- + The goal of *precise* answers (*i.e.*, answers with *fewer* limitaions) is sought by adequate management of relevant error categories.
- + Precision's concern with variation under repetition is (of course) *distinct* from accuracy's focus on closeness to the 'truth' – the actual state of affairs in the real world.
  However, **accuracy** may become involved with precision when there is (estimating and/or measuring) bias so average error is *not* zero – see page HL77.13 in Statistical Highlight #77.

Precision (as, say, the proportion of answers in each category) is less commonly assessed for **categorical** answers, like 'Yes' or 'No' to the Question: *Is* **X** *a cause of* **Y**?
Precision and accuracy are easily confused;
- the standard deviations which quantify the precision of estimating a population average or proportion under EPS are sometimes (wrongly) referred to as 'the accuracy of averages' or 'the accuracy of proportions' – see equations (HL77.9) and (2.10.4) on pages HL77.5 of Statistical Highlight #77 and 2.81 in Figure 2.10 of the STAT 332 Course Materials,
- in ordinary English, a 'precise' answer often means an *accurate* answer, *different* from its statistical meaning.

To experience precision (or *im*precision) first-hand, the reader can stand on their bathroom scales three times 15 seconds apart, or ask their health professional to give them the results of measuring their blood pressure *twice* (instead of once).

Prisoner's Dilemma: see Statistical Highlight #49.

Prediction interval: an expression for an interval estimate of a **random variable** representing a response variate, derived from a response model and the distribution of an estimator; the interval covers the value of the random variable with a specified **coverage probability**.
A **realized** prediction interval is the expression evaluated from data and is usually given in the Analysis stage of the FDEAC cycle.
(16.5, 16.10, 16.13, 16.14)

Probability (density) function: a table or function which gives the probability distribution of a discrete (continuous) random variable.
- **Joint probability (density) function:** a probability (density) function for *two or more* discrete (continuous) random variables.
- **Marginal probability (density) function:** a probability (density) function for *one* of two or more discrete (continuous) random variables. (5.5, 5.65, 5.68)

Probability (or probabilistic) assigning: using a *probabilistic* mechanism, described in the **protocol for choosing groups**, in an *experimental* Plan to generate a *known* (*e.g.*, equal) probability of assigning the value of the focal variate to the units:
- within each block in a **blocked** Plan;
- in the sample in an *un*blocked Plan.

Probability assigning is the basis of statistical theory for:
- unbiased estimating of **model parameters** representing treatment effects,
- the relationship of **comparing imprecision** to degree of **replicating** in the groups being compared,
- an expression for a **confidence interval** for estimating a difference of averages.

Probability assigning *reduces the risk* of (our type 2) *confounding* of the focal variate with unblocked, unknown or unmeasured non-focal explanatory variates; the greater the degree of **replicating**, the greater the reduction in risk. (5.37, 5.47, 5.48, 5.49, 5.72) See also **EPA**.

Probability (or probabilistic) selecting: using a probabilistic mechanism, described in the **protocol for selecting units**, to generate a *known* (*e.g.*, equal) probability of selecting each respondent population unit for the sample. (5.23, 5.48, 5.52, 5.56)
Probability selecting is the statistical ideal but is not always feasible in practice (5.79 to 5.82); it is the basis of statistical theory for:
- unbiased estimating of **model parameters**,
- the relationship of **sampling imprecision** to degree of **replicating**,
- the expression for a **confidence interval** for estimating an average.

Probability selecting manages our type 4 **confounding**. (5.72)
See also **Equiprobable selecting** and **Selecting probability**.

Probable error: Unhelpful terminology to be avoided.
A 'definition' of it – *in statistics, probable error defines the half-range of an interval about a central point for the distribution, such that half the values from the distribution will lie within the interval and half will lie outside* – sounds (sort of) like the **interquartile range**.
If this is so, 'probable error' is neither 'probable' nor 'error.'

(*continued overleaf*)

Process: ● a set of *operations* that produce or affect elements, OR:
● the *flow* on an entity (like water or electrons).

Thus, in statistics, a process may involve **elements**. WHEREAS:

In *probability*, a process is any set of *operations* from which there are at least two possible *outcomes*; observing *which* outcome occurs in any execution of the process generates *data*.

Such data may yield values for probabilities associated with, or for other characteristics of, the process. (5.55)

Process thinking: Thinking of an activity as a series of *steps* which form a process; the aim is usually to *improve* the process by, for example, eliminating unnecessary steps.

A 'process' in this context is also called a 'system'.

Process thinking is a basis for **statistical thinking**.

Project: a *broad* investigative undertaking involving *many* questions. A project is to be contrasted with an **investigation**.

Proportionality refers to a straight-line **X**-**Y** association *through the origin* – see also **Association**. (5.31)

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 by comparison error or sample error. (5.36 to 5.38)

Protocol for selecting units [(unwisely) also called the **sampling protocol**] is (a description of) the process (to be) used to select, from the respondent population, the units that comprise the sample. (5.56)

Protocol for setting levels specifies the *values* to be taken by relevant explanatory variate(s). (5.43 to 5.45)

***P*-value:** see **Significance testing**.

Qualitative an adjective applied to 'variate' and denoting a *nominal categorical* variate (like marital status or skin colour).

We use 'qualitative' as a synonym for **categorical**. (5.61, 5.62)

Quantile: For the components of a data set ordered by magnitude (say, from smallest to largest), if the data set is divided into a specified number of parts with *equal* numbers of components, the two ends and the points of division are the quantiles of the data set.

For example, for an ordered data set with 29 components divided into four parts each with six components, the **quartiles** (signifying *four* divisions) are the 1st, 7th, 15th, 21st and 29th components, called respectively the minimum, the first quartile, the median or second quartile, the third quartile and the maximum.

- These names comprise the **five number summary** of the data set.
- The difference between the third and first quartiles is the **interquartile range**.
- Having 29 components in the data set in this illustration means the five-number summary components are *all* components of the data set. Had there been 30 components, the four divisions would each have 7½ components and the first, second and third quartiles would conventionally be taken as the average of the two components either side of the three division points (and so would *not* be data set components).

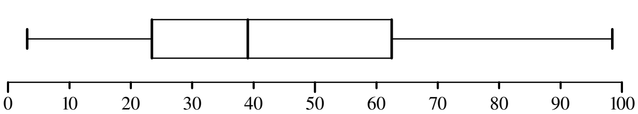
Having seven division points for an ordered data set would yield its **octiles**; similarly, nine division points give **deciles**.

Ninety-nine division points yield **percentiles**.

We hear percentiles used to indicate where a person in a population stands in relation to its other members – for example, the 40th percentile means 40% of others are below the person, 60% are above.

Quantiles can also refer to distributions; we are familiar with this useage from looking up, say, the 2.5 and 97.5 percentiles of a *t* distribution to use in a 95% confidence interval for a population average.

A pictorial display of the *five number summary* is a **box plot** which consists (as shown below) of a (usually long thin) rectangle whose length is the IQR and with a line across it at the median; at its two ends are 'T'-shaped 'whiskers' from the minimum to the first quartile and from the third quartile to the maximum. The five values can be shown on the box plot or (as below) on an adjacent scale.



Box plots can be oriented horizontally (as here) or vertically, and are often used side by side to provide a *comparison* of key features of two or more data sets. They are one tool for **exploratory data analysis** and their use was promoted by statistician John Tukey.

Quantized: restricting the magnitude of the units ('subdivisions') of a variable representing a physical quantity, so it can take only certain discrete values – for example, the *energy levels* of electrons in atoms.

Curiously, Eddington (cited by Mascall – see **References**) claimed that Dirac's theory of the electron and Kummer's quadric surface are homologous mathematical structures. This raises the (tantalizing but seemingly unlikely) possibility that pictorial or physical realizations of the members of Kummer's surface might represent the quantized electron orbital model of atoms that 'explains' the discrete lines of atomic emission spectra and chemical bonding in molecules – for example, the first 'disc-like' Kummer surface could represent the innermost 1s orbital.

Quartile: see **Quantile:**.

Quantitative an adjective applied to 'variate' and denoting a *measured* or *counted* variate (like length or number of instances). (5.61, 5.62)

Quota selecting: see **Non-probability selecting**.

Random: Means *equiprobable* in a selecting (or sampling) context or an assigning context. See also **Simple random sampling**.

'Random' is best avoided as statistical terminology, but it is (unfortuately) perpetuated in **random variable** and **randomized response**.

Random assigning: see **Equiprobable assigning**.

Randomization: a synonym for **equiprobable assigning**. (5.37, 5.48) The latter term is preferred as being more evocative of the process.

Randomized response: a measuring process in which an interviewer (*e.g.* in a sample survey) can ask a sensitive question but cannot with certainty interprert the unit's response in terms of the question. (5.62)

Random sampling: a (careless) synonym for **equiprobable selecting**. The latter term is preferred as being more evocative of the process.

Random variable (r.v.): ***informally***, a variate that takes on different values according to chance.

Formally, a random variable is a *function* which assigns a real number to each point of the sample space (S); *i.e*., a random variable is a function with domain S and range $\mathbb{R}$ – it is a mapping from the sample space to the real numbers. (5.17, 5.5 to 5.9)

Usually, a **discrete** r.v. is used to model a variate with *counted* values, a **continuous** r.v. to model a variate with *measured* values. (5.61)

Ratio and regression estimating: using information about the values of an explanatory variate, over the elements of the respondent population, to decrease imprecision of estimating a population attribute like an average or total; to accomplish this, the explanatory variate must have a (strong) positive association with the response variate whose attribute is of interest – the stronger the association, the greater the decrease in imprecision. (5.22, 5.24, 5.53)

Ratio estimating is discussed in Figure 2.17 of STAT 332.

Realized: see **Confidence interval**, **Prediction interval**, **Residual**.

References:

Barnett, V. *Sample Survey Principles and Methods.* Second edition, Edward Arnold, London, 1991.

Cochran, W.G. *Sampling Techniques.* John Wiley & Sons, Inc., New York, 3rd Edition, 1977.

Bentley Hart, D. *Atheist Delusions.* Yale University Press, New Haven & London, 2009, page 59.

Lakoff, G. *Moral Politics: How Liberals and Conservatives Think*. The University of Chicago Press, 1996.

Mascall, E.L. *Christian Theology and Natural Science.* Longmans, Green and Co., London, New York, Toronto, 1957, pages 65-76 (models), 104-131 (Eddington and Milne), 129 (Kummer's quadric surface), 181-195 (determinism).

The last three authors provide a broader perspective on statsitical issues like causation, knowing, modelling and uncertainty.

Regression effect: see Statistical Highlight #55.

Regression estimating: see **Ratio estimating**.

Relationship: a relationship in statistics is cast in terms of **variates** – there are *two* variates in the simplest case: an **explanatory** variate **X**

(*continued*)

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 9)

(the **focal** variate) and a **response** variate **Y**.

Data-based investigating of a relationship (to answer a Question with a causative aspect) involves *change* and *comparing* – these activities are therefore essential in the components of a proper comparative Plan. (5.28, 5.29, 5.35, 5.36)

Relative frequency: an obscure term to be avoided as duplication of 'proportion'.

Relative likelihood function: see **Likelihood function**.

Reliable or reliability: words to be avoided as statistical terminology; reliability (of an Answer) may mean:

- adequate **precision**; OR:
- (**overall**) **error** that is likely to impose *acceptable* limitations on the Answer(s) in the Question context. (5.85)

Repeatability of a gauge is the **variation** [expressed as an appropriate (data) standard deviation] of repeated measurements on each of a sample of (10, say) parts by *one* operator using the gauge;

- **Reproducibility** of a gauge is the *between-operator* **variation** [expressed as an appropriate (data) standard deviation] of two measurements, one by each operator using the gauge, on each of a sample of (10, say) parts. (5.62)

We might think of 'variation' in these two definitions as **variability**.
See also **Gauge R&R investigation**.

Repetition: repeating over and over (usually *hypothetically*) one or more of the processes of selecting, measuring and estimating – see also **CI**, **estimator**, **inaccuracy** and **imprecision**. (5.20, 5.21)

Replicating: selecting more than one unit/block from the respondent (or study) population for the sample. (5.23, 5.50)

Under **probability selecting**, increased replicating reduces **sampling imprecision**, thus reducing the *likely* magnitude of sample error.

- **Adequate replicating:** selecting *just* enough units/blocks from the respondent population to make the likely magnitude of *sample* error [and, hence, the limitation it imposes on Answer(s)] *acceptable* in the Question context. (5.24, 5.50)

Representative sample: a sample which has sample error (and corresponding limitation) that is *accceptable* in the Question context.
Because the representativeness of a sample can *rarely* be known, this phrase should be *avoided* as statistical terminology. (5.21)
See also Appendix 3 on page HL77.9 in Statistical Highlight #77.

Reproducibility: see **Repeatability**.

Residual: the **stochastic component** of a response model; it models variation about the structural component of the model.

In STAT 231, we denote the residual by R_j or R_{ij}, a **random variable** with a $N(0, \sigma)$ distribution. (5.88 to 5.91, 5.97, 5.98)

- **Estimated** (or **realized**) **residual**: a *number* derived from data and a model parameter estimate; it is *not* a value of the residual random variable *R*.

  In STAT 231, we denote the estimated residual by $\hat{r}_j$ or $\hat{r}_{ij}$; an (*un*known) value of *R* is denoted r_j or r_{ij}. (7.1, 7.2)

Elsewhere, the residual may be called the *error* term in the model; we (rigorously) *avoid* this terminology because we use 'error' for a different (pervasive) statistical (and scientific) concern.

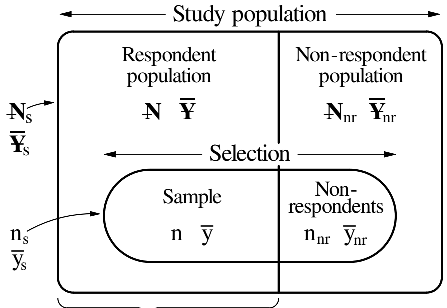
In a regression context, use of (the misleading) 'error' instead of (the evocative) 'residual' compounds the problem when 'error' carries over into *analysis of variance* – see also **ANOVA**.

Respondent population: those elements of the study population that *would* provide the data requested under the incentives for response offered in the investigation; (5.24)

- **Non-respondent population:** those elements of the study population that would *not* provide the data requested under the incentives for response offered in the investigation. (5.25)

These two populations can also be **processes** in some (rare) contexts.
The diagram at the top of the right-hand column shows these two populations (with relevant notation) in relation to the **study population**.

Response model: a mathematical description, including modelling assumptions, of the relationship between a response variate and explanatory variate(s); the form of the relationship is contingent, in part,



Statistical theory, particularly of survey sampling, is developed mainly in the context of the *respondent* population, often with*out* recognizing it explicitly.

'Selection' is terminology specific to these Materials; it draws attention to (human) non-response as an instance of the pervasive issue of missing data.

on the Plan. (5.97, 5.98)

- The **structural component** models the effect of specific explanatory variate(s) on the response variate.
- The **stochastic component** models variation about the structural component. (5.27, 7.1, 7.2, 8.1, 8.2)

See Statistical Highlights #71 to #73.

Response modelling of the behaviour of **error** under (hypothetical) repetition, with*out* doing *actual* repetition, is the basis for **estimating**.
We rigorously avoid 'error term' for the stochestic component of a (statistical) model – see **Residual**.

Root mean square: an adjectival phrase (abbreviated rms and often applied to 'error') denoting the three processes of taking the *square root* of a *mean* of entities *squared*. (5.63) See Statistical Highlight #7.
In *real-world* contexts, 'root ***average*** square' is preferable. (5.28)

Root n law: a 'law' said to be familiar to physicists, but a clear description is elusive; the 'definition' – *the* ***probable error*** *involved in asserting that the number of individuals in an aggregate is n, is itself of the order of the square root of n* – is of little help.
Possible *statistical* candidates are:

- for n probabilistically independent random variables with mean μ and standard deviation σ:
  - their *sum* has standard deviation $\sqrt{\text{n}}\sigma$,
  - their *average* has standard deviation $\sigma/\sqrt{\text{n}}$;
- for the Poisson distribution, the standard deviation is the square root of the mean. [This *may* be the '**root n law of random counts**' – *the s.d. of a random count variable is equal to the square root of the expected value or mean of that variable.*]

Rounding: see **Significant figures**.

Run: part of the Execution stage of an experimental Plan in which all the data are collected for *one* treatment. (5.43, 5.44)

R.v., r.v.: short form for **random variable**.

Sample: the group of units/blocks selected from the respondent population *actually used* in an investigation. (5.19)
A sample is a *sub*set of the respondent population, a *census* uses *all* the respondent (or study) population units/blocks. (5.20, 5.24)
A common (implicit) assumption is that there are *no* missing data for the units of the sample; our term **selection** makes this assumption *ex*plicit. (5.25) See also **Census**.

Sample error: see **Error**.

Sample of convenience: using as the sample units *conveniently* available to the investigator(s) – *e.g.*, parts still at a manufacturing site or patients at a hospital or clinic close to the research site. (5.40, 5.48)

Sample size: the number of units/blocks in the sample.
Notation for sample size is n (Roman, not *italic*). (5.25)
(5.22 to 5.24, 5.39, 5.40, 5.48, 5.50 to 5.52, 5.57, 5.59, 5.63, 5.72, 5.80, 5.85, 5.86) See also **Replicating**.

Sample space: the set of all possible outcomes of one execution of a **process** ('process' is used here with its *probability* meaning).
As a term in *probability*, the adjective 'sample' in *sample space* has

(*continued overleaf*)

none (or few) of its *statistical* connotations.

Our notation of (Roman) S for sample space tries to minimize confusion with the symbols s, *s*, *S* and $\mathbf{S}$ used for standard deviations.

A **discrete** sample space has *countably* many points, a **continuous** sample space has *un*countably many.

- A **point** is *one* outcome of one execution of a process.

Sample statistic: means a *sample attribute* – it implies **probability selecting** of the sample although this is rarely recognized by its users. It is terminology to be *avoided* because it:

- is *one* ambiguous and redundant term which runs together *two* existing terms: **attribute** and **estimator**;
- invites confusing *estimator* with ***estimate*** and both with *attribute*.

Sample survey: see **Survey**. (5.84)

Sampling: the processes of **selecting** and **estimating**. (5.20)

Scale: an (unnecessary) *un*evocative term with the same meaning as **dispersion** and **spread**. See also **location-scale transformation**.

Scatter diagram (or scatter plot): a Cartesion plot with a response variate or estimated residual on the vertical axis, an explanatory variate on the horizontal axis. (5.29, 5.31, 5.65) See also **Ishikawa** and Statistical Highlights #29 to #31.

Selecting: the process by which the units/blocks of the sample are obtained from the respondent population – it is described in the **protocol for selecting units** in the Design stage of the FDEAC cycle. (5.20, 5.22, 5.26 to 5.28, 5.37 to 5.40, 5.48, 5.49, 5.52, 5.53, 5.56, 5.57, 5.64, 5.71, 5.72, 5.79 to 5.82, 5,84 to 5.86)

Selecting can involve more than one **stage**; in *two-stage* selecting:

- at the *first* stage, *clusters* (groups of *elements*) are selected from the respondent (or study) population;
- at the *second* stage, one (or more) element(s) are selected from each cluster selected at the first stage. (5.57)

See also **Equiprobable selecting** and **Probability selecting**; see also **Judgement selecting** and **Voluntary response**.

Selecting probability: for any unit selecting process, we distinguish:

- the probability a particular *sample* is selected, FROM:
- the probability a particular *unit* is selected. (5.85)

The latter is also called the unit **inclusion probability**. (5.56)
See pages HL21.6 to HL21.8 in Statistical Highlight #21.

Selection: the group of units selected from the *study* population, comprising the **sample** and the **non-respondents**. (5.25)

The relationships among the numbers of elements/units are:

$$\left.\begin{array}{ccccc} \text{Study population} & = & \text{Respondent population} & + & \text{Non-respondent population} \\ \mathbf{N}_s & = & \mathbf{N} & + & \mathbf{N}_{nr} \\ \text{Selection} & = & \text{Sample} & + & \text{Non-respondents} \\ n_s & = & n & + & n_{nr} \end{array}\right\} \text{(HL91.6)}$$

See also the diagram at the upper right overleaf on page HL91.19.

Sensing: an illustration of how **process thinking** might approach the complex (and much debated) topic of how information from our senses becomes our personal experience of 'reality' is:

- *sensing* – stimulation of one or more of the senses,
- *perceiving* – interpreting the information the sense(s) provide,
- *comprehending* – recognizing the intelligibility of what is perceived,
- *measuring* – in a (scientific) investigative context, quantifying the intelligible 'structure' comprehended,
- *communicating* – sharing with others knowledge of the quantified 'structure'.

The purpose of this illustration here is to suggest a reconcilliation of **Eddington's** and **Milne's** seemingly incompatible starting points for their *a priori* theories, which critics have used as evidence against these theories. See also **Data**.

Sensitivity: a word to be avoided as duplication of existing statistical terminology; sensitivity (ability to detect an effect) refers to adequate **precision** (attained by managing ***im*precision**). (5.85)

Short form: see **Acronym**.

Significance level: a term to be avoided – a synonym for ***P*-value**, although it may then be called the ***observed* significance level**.

Significance testing: a *five*-step probabilistic **argument by contradiction** for assessing the *strength of evidence* provided by data *against* a hypothesized value of a model parameter;

1. State the **hypothesis** in terms of a **model parameter**.
2. **Estimate** relevant model parameters.
3. Choose a **discrepancy measure** (*D*), a random variable whose value (*d*) measures the 'distance' between what is *observed* and what is *expected* if the hypothesis is *true*.
4. Find the ***P*-value**, the probability of a value of *D* at least as extreme as the value (*d*) actually obtained.
5. *Interpret* the *P*-value in (formal) statistical language. ALSO: Give the *Answer* in the language of the Question context.

Significance testing is to be distinguished from **Hypothesis testing**; unfortunately, the two terms are (carelessly) used interchangably.

The useful process of (statistical) significance testing is *not* well served by its arcane terminology, some of it in more than one version.
See also **Hypothesis testing** and **Discrepancy measure**.

Significant figures: the number of digits in a numerical value, not including leading and trailing zeros; for example:

- 356 has three significant digits,
- 0.6024 has four significant digits,
- 600 has one, two or three significant digits – we can't tell whether the last two zeros are actual values or are due to rounding.

Concerns about significant figures usually arise in the context of a numerical answer and knowing how many of its digits are meaningful; the dilemma arises because people sometimes (unthinkingly) give a number of digits based on the number given by the calculator or computer that produced the answer.

The meaningful digits in such an answer are (usually) determined by the meaningful digits in the input to the calculations. In data-based investigating where this input is numerical data, three (say) significant figures in the data suggest *at most* three such figures in an answer derived from them, perhaps one *fewer* (two significant figures).

+ This reminds us that users of data should assess the reasonableness of their number of significant figures – too many or too few can be a warning sign of poor data quality.
  For example, people's weights in kilograms to three decimal places or lengths in metres to five decimal places are suspect. The constants of nature, like the speed of light, which have been measured innmuerable times by experts over decades, are given to *nine* to *twelve* significant figures; an exception to these numbers is the higher number of known figures for some *frequencies* associated with cesium atomic clocks.

The number of significant figures given in an answer also depends on its use – for example, an industrial standard has more stringent requirements than, say, a public health or dietary guideline.

For calculations involving several or many steps, significant digits also serve as a reminder not to compromise final answers by inadvertent *rounding* part-way through the calculations; this is why the statistical tables in these Materials are unusual in giving *six* decimal places.

The tables of the standard normal distribution in Statistical Highlight #104 contain two useful notational ideas:

- a table that is read naturally to four decimal places can, with only minor extra effort, be read to five or to six places,
- compact presentation of decimal numbers with (many) leading nines.

Simple random sampling (SRS): a synonym for **equiprobable selecting** and, like **random sampling**, is best *avoided* as statistical terminology – 'simple' and 'random' are *un*clear, 'sampling' is really selecting. (5.56)

Simpson's Paradox: a name that refers to the (surprising) behaviour of proportions when they are used to assess the *direction* of an $\mathbf{X}_1$-$\mathbf{Y}$ relationship for units in groups based on variate $\mathbf{X}_1$ and when these groups are subdivided on the basis of variate $\mathbf{X}_2$. (5.65 to 5.70)
See also Statistical Highlight #51.

Skepticism: see **Measuring process**.

Spread: refers to the *width* of a distribution – for example, the distance between the minimum and maximum values of a data set.

Stage: one (of possibly two or more) steps in a selecting process – see also **Selecting** and **Unit**. (5.57, 5.86).

Standard: an element with a *known* variate value, used to **calibrate**

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 10)

a measuring process. (5.20, 5.60)

Standardization: for a random variable, the process of subtracting its mean and dividing this difference by its standard deviation, so the *standardized* random variable has mean 0 and standard deviation 1.

Standard deviation (s.d.): a measure of **variation** or **variability** for:
- a set of **data** (or *numbers*) – a 'data' s.d.; (8.1, 8.2) OR:
- a **random variable** – a 'probabilistic' s.d. (5.6)

Recall Table HL91.2 on page HL91.3 and the related average-mean (real world-model) distinction for a measure of **location**. (1.20, 5.28) See also **Variance**.

Standard pairwise difference (s.p.d.): a measure of *data* **variation**, unfamiliar even to statisticians, with no agreed-on name.
Our 'standard pairwise difference' in these Materials uses 'standard' to indicate that its calculation involves the *same* three operations as standard deviation: squaring, averaging and taking the square root.
Where the s.d. is based on deviations (differences) of the data values from their average, the s.p.d. is based on pairwise differences among the data values, a more 'natural' measure of data variation.
The s.d. and the s.p.d. are (algebraically) *equivalent* measures of variation because of equation (HL91.7); they merely differ in magnitude by a factor of $\sqrt{2}$: *i.e.*, the s.p.d. is about 40% larger.

$$\text{s.p.d.} = \sqrt{2}\times\text{s.d.} \qquad \text{-----(HL91.7)}$$

*Un*like deviations from the average which add to zero for *any* data set, the set of pairwise differences has *no* (gereral) constraint that is independent of the *particular* data values; hence, its averaging involves a divisor of $\binom{n}{2}$, the number of possible (*un*ordered) pairwise differences of n data values. Because the s.d. and the s.p.d. are equivalent *algebraically*, the s.p.d. justifies the $n-1$ divisor for s.d.
In principle, the s.p.d. could be calculated from the set of *all* n^2 (ordered) pairwise differences that includes the n *self*-differences of value zero; this statistically bizarre measure of (data) variation corresponds to calculating the (data) s.d. with a divisor of n (or $\mathbf{N}$).
The s.p.d. is discussed in Statistical Highlight #100.

Statistic: see **Sample statistic**.

Statistical significance: a statistically significant numerical Answer is one with a ***P*-value** below 0.05 in an appropriate test of significance.
A **highly** statistically significant numerical Answer is one with a ***P*-value** below 0.01 in an appropriate test of significance.

Statistical thinking: building on **process thinking**, statistical thinking sees activities as processes (or collections of processes), which exhibit **variation**; identifying the sources of variation and managing them so as to reduce variation *improves* the process(es).

Stemplot: see Statistical Highlight #32.

Stochastic: varying unpredictably or erratically – probabilistic.
See also **Response model**.

Stratum, strata: see **Stratifying**.

Stratification, stratifying: subdividing the respondent (or study) population into groups (called **strata**) so that elements with*in* a stratum have *similar* response variate values and elements in different strata *differ* as much as feasible in the investigation context; the sample is obtained by selecting units from *each* stratum. (5.85)
Stratifying (properly implemented) can manage **sampling imprecision**. (5.24, 5.38, 5.40, 5.59, 5.86, 5.96) See also **Ishikawa**.

Strength: means **precision** (of an Answer); it is to be avoided as duplication of existing statistical terminology. (5.85)

Structural component: see **Response model**.

Student's t_ν distribution: the distribution of a random variable which is the *quotient* of independent random variables with $N(0,1)$ and K_ν distributions; the parameter ν is called the **degrees of freedom**. (13.14, 13.15, Ap. 3, Ap. 4)

Study: see **Investigation**.

Study error: see **Error**.

Study population: a group of elements *available* to an investigation. (5.22, 5.52)

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. (5.37, 5.38) See also Table HL89.6 and its discussion on pages HL89.13 and HL89.14 in Statistical Highlight #89.
We can think of *subdividing* as *matching* at an *aggregate* (rather than an *individual*) level; subdividing therefore has the *same* statistical benefit (managing confounding and comparing imprecision) as **matching** for the non-focal explanatory variate(s) that are the basis for the subdividing – see page HL91.15, upper half of left-hand column.
- 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. (5.43, 5.49)

Survey: an investigation to answer a Question(s) with a *descriptive* **aspect**.
- The term **sample survey** makes it explicit that the survey involves a *sample*, not a census; **survey sampling** is the area of statistics dealing with the theory and practice of sample surveys.
- Elsewhere, a survey may be described as an investigation with a Plan that is **observational** (used in its *non*-statistical sense).

Survey (unqualified) is better *avoided* as statistical terminology.

Systematic selecting: selecting one unit by EPS from the first k respondent (or study) population units [$k \ll \mathbf{N}$ (or $k \ll \mathbf{N}_S$)] and then selecting every k*th* unit. (5.56)

Target population/process: the group of elements/operations to which the investigator(s) want Answer(s) to the Question(s) to apply.
(5.19 to 5.23, 5.25, 5.26, 5.29, 5.32, 5.33, 5.36, 5.38, 5.52, 5.54, 5.55, 5.58, 5.63, 5.83, 5.84)

***t*-distribution:** see **Student's t_ν distribution**.

Teaching introductory statistics: to do it effectively is *not* an undertaking for the faint-hearted.
As well as the comments in Sections 3 to 6 on pages HL91.2 to HL91.6, there is discussion on pages HL94.10 and HL94.11 and HL100.7 to HL100.9 in Statistical Highlights #94 and #100.

Test of hypothesis: an instance of **hypothesis testing**.

Test of significance: an instance of **significance testing**; there are *many* such tests – context determines which (if any) test is appropriate.
A more evocative phrase is *test of **statistical** significance*, to emphasize the (easily overlooked) difference between **statistical significance** and **practical importance.**

Test statistic: a term to be avoided. See **Discrepancy measure**.

Treatment: a *combination* of the levels of the factor(s) that can be applied to a unit (in the sample/blocks). (5.43)

Treatment effect: see **Effect**.

Treatment group: in an experimental Plan, the part of the sample assigned $\mathbf{X} = 1$; in practice, this often means receiving the 'treatment'. (5.38, 5.39, 5.40, 5.45, 5.47 to 5.50, 5.53, 5.61, 5.80, 5.82)

Trial: one execution of a probabilistic process. See **Experiment**.

Trustworthiness: means **accuracy** (of an Answer); it is to be avoided as duplication of existing statistical terminology. (5.85)

Two-stage selecting: see **Selecting**.

Type I error, **Type II error:** see **False negative**.

Uncertainty: ignorance (incomplete knowledge) of error; for example:
- for a numerical Answer, ignorance of the magnitude and/or the sign/direction of error;
- for a categorical Answer (like *Yes* or *No*), ignorance of whether the Answer is the correct category. (5.20)

See also **Variation** and §11 on page HL94.12 in Statistical Highlight #94.
More broadly in statistics, the 'amount' of *sampling* uncertainty is determined by two matters.
- **variation** among the elements of the population – this cannot be altered but can be managed by **stratification**,
- the **sample size**, (weakly) in relation to the population size.

Given the Plan components which deal with these two matters in a particular investigation, investigators can decide how to present the resulting uncertainty in an answer which is a confidence interval by their choice of trade-off between **confidence level** and *interval width*.

Measuring uncertainty is also determined by two matters.

- **variation** among repeated measurements of the same variate, which can (sometimes?) be influenced by the resource-intensiveness of the measuring process,
- the **number** of measurement repetitions, although repeated measuring is *rare* in data-based investigating; also, the 'population' of repetitions is, in principle, no longer bounded in size.

If measurement error were to be modelled like sample error and with given Plan components, the resulting confidence interval answer would presumably exhibit the same type of trade-off as with sampling uncertainty. [See also Section 5 on page HL74.3 and the comment at the top of page HL74.5 in Statistical Highlight #74.]

This *statistical* background makes the Heisenberg Uncertainty Princile in physics look like an analogue of the confidence interval trade-off; a difference is that its components are cannonical *physical* variables like position and momentum or (the more curious) time and energy. The analogue in physics of the source of sampling uncertainty in statistics might be wave-particle duality; measuring might be a *common* source of uncertainty in physics and statistics. Clarification of such matters would be useful. See **Heisenberg**.

Union: see **Intersection**.

Unit: an entity which can be selected for the sample – it may contain one or more than one **element**. (5.86)

Multistage sampling Plans have **primary sampling units**, **secondary sampling units**, etc., at their successive stages. (5.57, 5.86)

In STAT 231, a unit is a basic entity for which variate values could be obtained; '*element*' is better terminology than 'unit'. (5.55)

Untrustworthiness: means **inaccuracy** (of an Answer); it is to be avoided as duplication of existing statistical terminology. (5.85)

Urn: an (imaginary) container for coloured balls, used to demonstrate properties of EPSWIR and EPSWOR. See Statistical Highlight #94.

Universe: referring to a population as a 'universe' is *un*helpful in statistics because of its:

- unnecessary duplication of terminology;
- wider meaning in ordinary English.

Validity: validity means **accuracy** (of an Answer); it is to be avoided as duplication of existing statistical terminology. (5.47, 5.85)

See also **External validity** and **Internal validity**.

Variability: the *model* quantity representing **imprecision**; we see it arising under repetition (*e.g.*, for error or a sample average).

Unfortunately, in ordinary English, 'variability' and '**variation**' seem to be used interchangeably. See also **Standard deviation** and **Variance**.

Variance: a measure of **variability** – it is the *square of the (probabilistic)* ***standard deviation.*** (5.11)

From the perspective of introductory statistics, the widespread use of variance (instead of standard deviation) in presenting statistical theory is unfortunate because:

- its units are *un*natural (*e.g.*, variance of length has units of *area*);
- it is (much) easier to *visualize* standard deviation than variance;
- some expressions involving variance imply that variabilities *add*, whereas (natural) variations *add like Pythagoras.*
- standard deviation, *not* variance, is used in confidence intervals and control charts. (5.11 to 5.14)

Emphasis on variance produces the (unhelpful) $N(\mu, \sigma^2)$ parameterization of the normal distribution. See also **Standard deviation**.

Variate: a characteristic associated with each **element** of a population/process.

- **Response variate (¥):** a variate defined in the Formulation stage of the FDEAC cycle; an Answer gives some attribute(s) of the response variate over the target population/process.
- **Explanatory variate (X̶ or Ƶ):** a variate, defined in the Formulation stage of the FDEAC cycle, that accounts, at least in part, for changes from element to element in the value of a response variate.

Variation: differences in (variate or attribute) values:

- across the elements/units/items in a group, such as;
  - a target population/process,
  - a study population/process,
  - a respondent population,
  - a non-respondent population,
  - a sample, and
  + repeated measurements.

We distinguish variation from **variability**

Variation is commonly *quantified* by (data) **standard deviation**, much less commonly in practice by **standard pairwise difference**. (5.21, 5.28). See also Statistical Highlight #100.

Variation is easily confused with **Uncertainty**. A reason may be that, in a population where a variate of interest has *no* variation (every element has the same value), a sample of size *one* yields this value with *zero* sample error [no *sampling* uncertainty], although *measurement* error (and its attendant uncertainty) may still be involved.

Such populations are rarely of *statistical* interest, although it is startling to realize that the population of, say, the electrons in the universe – a number Eddington claimed was around 10^{79} – are regarded as identical so that any *one* is suitable for determining a universal constant like the electron charge or mass.

See also Section 6 on pages HL94.6 to HL94.8 and Note 4 on page HL100.10 in Statistical Highlights #94 and #100.

Venn diagram: originally developed to illustrate overlap among (universal) *propositions* in logic, Venn diagrams were (unfortunately) adopted more recently to illustrate combinations of *events* and their probabilities – **Eikosograms** are better suited to the latter task.

See also Statistical Highlight #5.

Voluntary response: an *un*necessary term which acknowledges that it is not feasible to *compel* a person to respond (to a question); instead, we use *incentives.* Willingness to respond is a *separate* statistical issue from accuracy or truthfulness. See **Randomized response**.

Voluntary response is not to be confused with **volunteer selecting** – asking for volunteers, usually after a brief explanation of what the investigating will entail for units of the sample. (5.52, 5.56)

Weakness: means **imprecision** (of an Answer); it is to be avoided as duplication of existing statistical terminology. (5.85)

Wider inductive basis: a (dubious) phrase sometimes used to describe the advantage of Answers about relationships obtained from a *full factorial* treatment structure – see also **Generality**. (5.45, 5.85)

Wisdom: the last (extra-statistical) step in the progression from data to information to knowledge, although 'data' may now need to be understood more broadly than is customary in a statistical context.

David Bentley Hart, an Eastern Orthodox scholar of religion, defines wisdom as *the recovery of innocence at the far end of experience.*

Acquiring wisdom is usually an arduous life-long endeavour which, it seems, is not commonly taken up and is rarely brought to fruition.

Wrong answer: see **Falsehood**.

8. Appendix 1: The Notation in Use – Illustrations from the STAT 332 Course Materials

The following tables (with STAT 332 page references) illustrate the notation advocated and used in these Materials.

Figure 2.3, page 2.37

Table 2.3.1: SYMBOL

Random variable	Value	Respondent population	DESCRIPTION
Y	y	¥	Response variate
–	j	i	Summation index
–	x	X̶	Focal explanatory variate
–	z	Ƶ	Explanatory variate
–	n	N̶	Number of units/elements
$\overline{Y}$	$\overline{y}$	$\overline{¥}$	Average (sum ÷ number)
R	r	R̶	Residual [or Ratio]
S	s	S̶	Standard deviation

Figure 2.3, page 2.38

Table 2.3.2: SUMMARY OF STANDARD DEVIATIONS

Response Models	Survey Sampling
σ Model parameter	S̶ Respondent population standard deviation – an attribute
$\hat{\sigma}$ Estimate of σ	s Sample standard deviation – we take s = s to estimate S̶
$\tilde{\sigma}$ Estimator of σ	S Estimator corresponding to s – a random variable

STATISTICS and STATISTICAL METHODS: Glossary for Introductory Stat ..... (continued 11)

Figure 2.3, page 2.38

Table 2.3.3

Respondent population standard deviation	$\mathbf{S}$	*data* standard deviation
Sample standard deviation	s	
Standard deviation of the sample average	$s.d.(\overline{Y})$	*probabilistic* standard deviation
Estimated standard deviation of the sample average	$\widehat{s.d.}(\overline{Y})$	

Figure 2.3, page 2.40

Table 2.3.5:

....QUANTITY......	RESPONDENT POPULATION	.......SAMPLE [MODEL].....................
Size (elements/units)	$\mathbf{N}$	n
Response	$\mathbf{Y}_i$ (i = 1, 2, ..., $\mathbf{N}$)	y_j (j = 1, 2, ..., n) [r.v.s are Y_j with value y_j]
Average	$\overline{\mathbf{Y}} = \frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\mathbf{Y}_i = \frac{1}{\mathbf{N}}{}_T\mathbf{Y}$	$\overline{y} = \frac{1}{n}\sum_{j=1}^{n}y_j$ [r.v. is $\overline{Y}$ with value $\overline{y}$]
Total	${}_T\mathbf{Y} = \mathbf{N}\overline{\mathbf{Y}} = \sum_{i=1}^{\mathbf{N}}\mathbf{Y}_i$	${}_Ty = \mathbf{N}\overline{y}$ [r.v. is ${}_TY$ with value ${}_Ty$]
Standard deviation	$\mathbf{S} = \sqrt{\frac{1}{\mathbf{N}-1}SS_{\mathbf{Y}}} \equiv \sqrt{\frac{1}{\mathbf{N}-1}\sum_{i=1}^{\mathbf{N}}(\mathbf{Y}_i - \overline{\mathbf{Y}})^2}$	$s = \sqrt{\frac{1}{n-1}SS_y} \equiv \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2}$ [r.v. is S with value s]

Figure 2.3, page 2.45

Table 2.3.7:

	Elements	Clusters	Relationships
Respondent population	$\mathbf{N}$	$\mathbf{M}$	$\mathbf{N} = \mathbf{M}L$, $L = \mathbf{N}/\mathbf{M}$
Sample	n	m	n = mL, L = n/m

Also: $\overline{\mathbf{Y}}_i = \frac{1}{L}\sum_{k=1}^{L}\mathbf{Y}_{ik}$ is the average response of the ith population cluster,

$\overline{y}_j = \frac{1}{L}\sum_{k=1}^{L}y_{jk}$ is the average response of the jth sampled cluster,

the subscript *ec* in Table 2.3.8 at the right denotes 'equal-sized clusters'.

Figure 2.3, page 2.45

Table 2.3.8

EPS of elements	Page	EPS of clusters	Page
$\overline{y} = \frac{1}{n}\sum_{j=1}^{n}y_j$	2.40	$\overline{y}_{ec} = \frac{1}{m}\sum_{j=1}^{m}\overline{y}_j$	2.105
$s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2}$	2.40	$s_{ec} = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}(\overline{y}_j - \overline{y}_{ec})^2}$	2.106
$\mathbf{S} = \sqrt{\frac{1}{\mathbf{N}-1}\sum_{i=1}^{\mathbf{N}}(\mathbf{Y}_i - \overline{\mathbf{Y}})^2}$	2.40	$\mathbf{S}_{ec} = \sqrt{\frac{1}{\mathbf{M}-1}\sum_{i=1}^{\mathbf{M}}(\overline{\mathbf{Y}}_i - \overline{\mathbf{Y}})^2}$	2.105
$s.d.(\overline{Y}) = \mathbf{S}\sqrt{\frac{1}{n} - \frac{1}{\mathbf{N}}}$	2.41	$s.d.(\overline{Y}_{ec}) = \mathbf{S}_{ec}\sqrt{\frac{1}{m} - \frac{1}{\mathbf{M}}}$	2.105
$\overline{y} \pm {}_{\alpha}t^{*}_{n-1}s\sqrt{\frac{1}{n} - \frac{1}{\mathbf{N}}}$	2.42	$\overline{y}_{ec} \pm {}_{\alpha}t^{*}_{m-1}s_{ec}\sqrt{\frac{1}{m} - \frac{1}{\mathbf{M}}}$	2.106

Figure 2.10, page 2.81

Table 2.10.1:

.......QUANTITY.......	RESPONDENT POPULATION	...............SAMPLE [MODEL]................
Size (elements/units)	$\mathbf{N}$	n
Indicator variate	$\mathbf{V}_i$ (i = 1, 2, ..., $\mathbf{N}$)	v_j (j = 1, 2, ..., n) [r.v.s are V_j with value v_j]
Proportion	$\mathbf{P} = \frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\mathbf{V}_i = \overline{\mathbf{V}}$	$p = \frac{1}{n}\sum_{j=1}^{n}v_j = \overline{v}$ [r.v. is P with value p (= $\overline{v}$)]
Number (or frequency)	$\mathbf{NP}$	(number in sample = np)

Figure 2.14, page 2.112

Table 2.14.7:

		Figure 2.3	Figure 2.10	Figure 2.14
Population	Response	$\mathbf{Y}_i$	C_i or ${}^cC_i$	$\overline{\mathbf{Y}}_i$
	Attribute(s)	$\overline{\mathbf{Y}}$, $\mathbf{S}$	$\mathbf{P}$	$\overline{\mathbf{Y}}$, $\mathbf{S}_{ec}$
	Size	$\mathbf{N}$	$\mathbf{N}$	$\mathbf{M}$
Model	Random variable(s)	$\overline{Y}$, S	P	$\overline{Y}_{ec}$, (S_{ec})
	Value(s)	$\overline{y}$, s	p	$\overline{y}_{ec}$, s_{ec}
Sample	Response	y_j	C_j or ${}^cC_j$	$\overline{y}_j$
	Attribute(s)	$\overline{y}$, s	p	$\overline{y}_{ec}$, s_{ec}
	Size	n	n	m

Figure 2.10, page 2.88

Table 2.10.9

QUANTITY	SYMBOL
Population proportion	$\mathbf{P}$
Sample proportion	p
Random variable	P
A value of P	p
Model parameter	π
Probability	Pr

Figure 2.15, page 2.113 and **Figure 2.17, page 2.131**

Table 2.15.1:

Quantity	Respondent Population		Sample (Estimate) (= model *value*)		Model (*Estimator*)	
Average	$\overline{\mathbf{Y}}$	$\overline{\mathbf{X}}$	$\overline{y}$ (= $\overline{y}$)	$\overline{x}$ (= $\overline{x}$)	$\overline{Y}$	$\overline{X}$
Total	${}_T\mathbf{Y}$	${}_T\mathbf{X}$	${}_Ty = \mathbf{N}\overline{y}$ (= ${}_Ty$)	${}_Tx = \mathbf{N}\overline{x}$ (= ${}_Tx$)	${}_TY$	${}_TX$
Ratio	$\mathbf{R} = \overline{\mathbf{Y}}/\overline{\mathbf{X}} \equiv {}_T\mathbf{Y}/{}_T\mathbf{X}$		$r = \overline{y}/\overline{x} \equiv {}_Ty/{}_Tx$ (= r)		R	

(*continued overleaf*)

Figure 2.16, page 2.125

Figure 2.15	$\mathbf{R} = \dfrac{\bar{\mathbf{Y}}}{\bar{\mathbf{X}}} = \dfrac{{}_{\mathrm{T}}\mathbf{Y}}{{}_{\mathrm{T}}\mathbf{X}} = \dfrac{\sum_{i=1}^{\mathbf{N}}\mathbf{Y}_i}{\sum_{i=1}^{\mathbf{N}}\mathbf{X}_i}$	$\mathbf{R}$	$\mathbf{Y}_i$	$\mathbf{X}_i$	$\bar{\mathbf{Y}}$	$\mathbf{N}$
Table 2.16.1:	↓	↓	↓	↓	↓	↓
Figure 2.16	$\bar{\mathbf{Y}} = \dfrac{\sum_{i=1}^{\mathbf{M}}{}_{\mathrm{T}}\mathbf{Y}_i}{\sum_{i=1}^{\mathbf{M}}\mathbf{L}_i} = \dfrac{\text{sum of cluster totals}}{\text{sum of cluster sizes}}$	$\bar{\mathbf{Y}}$	${}_{\mathrm{T}}\mathbf{Y}_i$	$\mathbf{L}_i$	${}_{\mathrm{T}}\bar{\mathbf{Y}}$	$\mathbf{M}$

Figure 2.16, page 2.125

Table 2.16.2:

......QUANTITY......	**RESPONDENT POPULATION**	**................SAMPLE [MODEL]................**
Size: units ≡ clusters	$\mathbf{M}$	m
elements	$\mathbf{N} = \sum_{i=1}^{\mathbf{M}}\mathbf{L}_i$	$\mathrm{n} = \sum_{j=1}^{\mathrm{n}}\mathrm{l}_j$
Cluster size	$\mathbf{L}_i$ $(i = 1, 2, \ldots.., \mathbf{M})$	l_j $(j = 1, 2, \ldots.., \mathrm{m})$ [r.v. is L_j with value l_j]
Average cluster size	$\bar{\mathbf{L}} = \mathbf{N}/\mathbf{M} = \frac{1}{\mathbf{M}}\sum_{i=1}^{\mathbf{M}}\mathbf{L}_i$	$\bar{\mathrm{l}} = \mathrm{n}/\mathrm{m} = \frac{1}{\mathrm{m}}\sum_{j=1}^{\mathrm{m}}\mathrm{l}_j$ [r.v. is $\bar{L}$ with value $\bar{l}$]
Response	$\mathbf{Y}_{ik}$ $(i = 1, 2, \ldots.., \mathbf{M};\ k = 1, 2, \ldots.., \mathbf{L}_i)$	y_{jk} $(j = 1, 2, \ldots.., \mathrm{m};\ k = 1, 2, \ldots.., l_j)$ [r.v. is Y_{jk} with value y_{jk}]
Cluster average	$\bar{\mathbf{Y}}_i = \frac{1}{\mathbf{L}_i}\sum_{k=1}^{\mathbf{L}_i}\mathbf{Y}_{ik} = \frac{1}{\mathbf{L}_i}{}_{\mathrm{T}}\mathbf{Y}_i$	$\bar{\mathrm{y}}_j = \frac{1}{\mathrm{l}_j}\sum_{k=1}^{l_j}\mathrm{y}_{jk} = \frac{1}{\mathrm{l}_j}{}_{\mathrm{T}}\mathrm{y}_j$ [r.v. is $\bar{Y}_j$ with value $\bar{y}_j$]
Cluster total	${}_{\mathrm{T}}\mathbf{Y}_i = \sum_{k=1}^{\mathbf{L}_i}\mathbf{Y}_{ik} = \mathbf{L}_i\bar{\mathbf{Y}}_i$	${}_{\mathrm{T}}\mathrm{y}_j = \sum_{k=1}^{l_j}\mathrm{y}_{jk} = \mathrm{l}_j\bar{\mathrm{y}}_j$ [r.v. is ${}_{\mathrm{T}}Y_j$ with value ${}_{\mathrm{T}}y_j$]
Average cluster total	${}_{\mathrm{T}}\bar{\mathbf{Y}} = \frac{1}{\mathbf{M}}\sum_{i=1}^{\mathbf{M}}{}_{\mathrm{T}}\mathbf{Y}_i$	${}_{\mathrm{T}}\bar{\mathrm{y}} = \frac{1}{\mathrm{m}}\sum_{j=1}^{\mathrm{m}}{}_{\mathrm{T}}\mathrm{y}_j$ [r.v. is ${}_{\mathrm{T}}\bar{Y}$ with value ${}_{\mathrm{T}}\bar{y}$]
Average	$\bar{\mathbf{Y}} = \sum_{i=1}^{\mathbf{M}}{}_{\mathrm{T}}\mathbf{Y}_i \Big/ \sum_{i=1}^{\mathbf{M}}\mathbf{L}_i$	$\bar{\mathrm{y}}_{uc} = \sum_{j=1}^{\mathrm{m}}{}_{\mathrm{T}}\mathrm{y}_j \Big/ \sum_{j=1}^{\mathrm{m}}\mathrm{l}_j$ [r.v. is $\bar{Y}_{uc}$ with value $\bar{y}_{uc}$]
Total	${}_{\mathrm{T}}\mathbf{Y} = \sum_{i=1}^{\mathbf{M}}{}_{\mathrm{T}}\mathbf{Y}_i = \mathbf{N}\bar{\mathbf{Y}} = \mathbf{M}{}_{\mathrm{T}}\bar{\mathbf{Y}}$	${}_{\mathrm{T}}\mathrm{y} = \mathbf{N}\bar{\mathrm{y}}_{uc}$ (if $\mathbf{N}$ is known) ${}_{\mathrm{T}}\mathrm{y} = \mathbf{M}\,{}_{\mathrm{T}}\bar{\mathrm{y}}$ (if $\mathbf{N}$ is *un*known) [r.v. is ${}_{\mathrm{T}}Y$ with value ${}_{\mathrm{T}}y$]

9. Appendix 2: Notation – Managing Falsehood

The foregoing twelve tables from the STAT 332 Course Materials occur near the start of the relevant Figure to summarize the notation that will be used in that Figure. The tables also have a broader purpose – they provide continuing emphasis on maintaining the three distinctions, listed on the lower half of page HL91.3, which are vital for effective statistics teaching. The tables have the following features:

- ∗ the terms 'respondent population', 'sample' and 'model' are prominent in the column and row headings,
- ∗ the sample is taken as the source of real-world data, as is the case in most data-based investigating,
- ∗ the respondent population columns notation is upper-case **bold** letters with high visual impact,
- ∗ the sample columns have low-visual-impact lower-case Roman letters, emphasizing the population-sample distinction
- ∗ the model columns are distinguished by their *italic* letters and (usually) 'random variable' (r.v.) label(s),
- ∗ as illustrated in equations (HL91.8) and (HL91.9) at the right [taken from the upper half of page 2.125 in Figure 2.16], a quantity and its estimate are notationally *different* in appearance, a useful reminder (*not* dependent on knowing the *meanings* of the symbols) of the distinction between behaviour under repetition and the individual case.

$$s.d.(R) \simeq \sqrt{\frac{1}{(\mathbf{N}-1)\bar{\mathbf{X}}^2}\sum_{i=1}^{\mathbf{N}}(\mathbf{Y}_i - \mathbf{R}\mathbf{X}_i)^2\left(\frac{1}{\mathrm{n}} - \frac{1}{\mathbf{N}}\right)} \qquad \text{-----(HL91.8)}$$

$$\widehat{s.d.}(R) \simeq \sqrt{\frac{1}{(\mathrm{n}-1)\bar{x}^2}\sum_{j=1}^{\mathrm{n}}(y_j - r x_j)^2\left(\frac{1}{\mathrm{n}} - \frac{1}{\mathbf{N}}\right)} \qquad \text{-----(HL91.9)}$$

This illustration is for standard deviation, but the same is true of estimating bias, with multiple examples in Figures 2.15, 2.16 and 2.17; the difference in visual impact is higher in the more complicated expressions in these three later Figures than in the first instance encountered as equations (2.3.9) and (2.3.17) on page 2.41 in Figure 2.3.

Distiguishing *observed* (Roman) y from a *model* (italic) *y* to distinguish the observed sample average and standard deviation, $\bar{\mathrm{y}}$ and s, from the model quantities $\bar{y}$ and *s*, reminds us of the importance in the Design stage of the FDEAC cycle of developing a Plan that makes it *reasonable* to treat $\bar{\mathrm{y}}$ as $\bar{y}$ and s as *s* (as in, for example, Table 2.3.5 near the top of page HL91.23).

Failure to observe the foregoing distinctions makes it easier to overlook the question to be asked about *all* data: *How were they generated?* If this question is *not* asked, it becomes more likely that statistical methods will be applied in situations where their underlying assumptions (like probability selecting, accurate measuring processes, independent measuring, distributional assumptions) are violated, leading to answers with such severe limitations that they are likely to embody falsehoods.

- ● Our error categories (for example, study error, non-response error, sample error, measurement error and model error in the case of sample surveys) provide a framework for assessing the severity of the limitations imposed by error on answers.