Figure 1.5. STAT 231 GLOSSARY

(about 310 entries, about 260 definitions – see also Statistical Highlight #91)

This Glossary gives definitions of terminology used in introductory statistics teaching, with emphasis on documenting the innovative usage of the 2004 STAT 231 Course Notes. Its wording reflects decisions which deal with four competing terminologies.

∗ The Glossary refers to the (more evocative) FDEAC cycle rather than the PPDAC cycle, which originated in an early version of the STAT 231 Course Notes around 1990 and, unfortunately, evaded needed reconsideration.

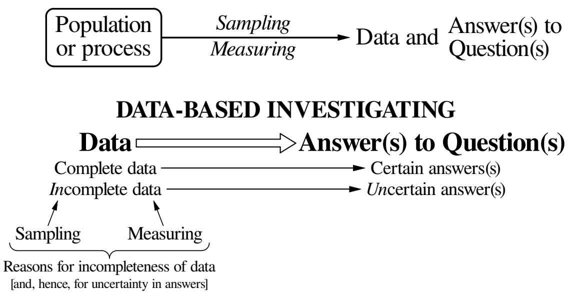
∗ As in Section 5 on pages 5.24 to 5.26 of Figure 5.7, this Glossary defines the study population as a combination of the respondent and non-respondent populations (see the lower left diagram below); this idea is *not* in the STAT 231 Course Notes but provides useful emphasis on the pervasive difficulty for statistics of non-response (or, more generally, missing data).

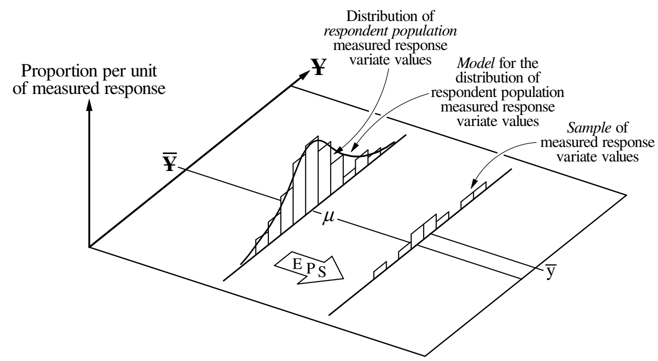
∗ The Glossary refers to the *normal* distribution whereas the STAT 231 Course Notes use the *Gaussian* distribution, although both use the *standard deviation* (*not* the variance) as the second parameter.

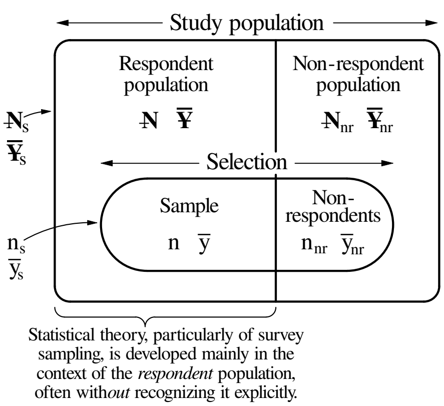
∗ Except for occasional reminders in brackets [‘(really elements)’], Glossary definitions ignore the element-unit distinction for consistency with the STAT 231 Course Notes. [The STAT 220 Course Materials *maintain* this distinction]

Number(s) in brackets in definitions (usually at their end) are page numbers (predominantly from the Chapter 5 Course Materials) where there is context for, and/or substantive discussion of matter(s) relating to, the term being defined.

As a reminder of the motivation for the terminology of STAT 231, four diagrams and a tabular summary of the components of the FDEAC cycle are given below; the diagrams are from pages 5.19, 5.25, 5.27 and 5.30 in Figure 5.7 of these Materials.







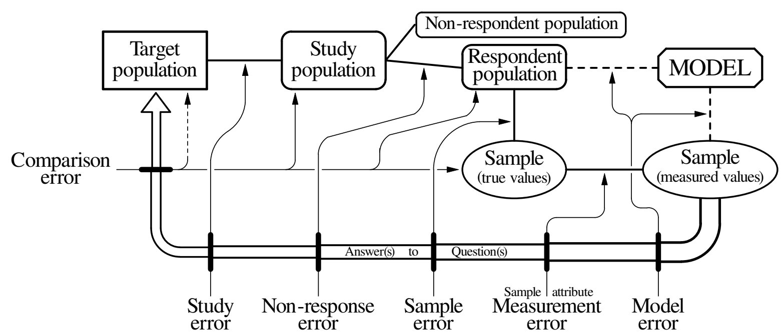


Table 1.5.1: The FDEAC cycle: a *structured process* for data-based investigating

Stage	Formulation stage	Design stage	Execution stage	Analysis stage	Conclusion stage
Input	***Question(s)***	***clear Question(s)***	***a Plan***	***Data***	***Information***
Components	Target unit (really element) Target population/process Variates: ● Response ● Explanatory Attributes Fishbone diagram Aspect: ● Descriptive ● Causative	Study unit (really element) 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 last side of the Glossary (page 1.20) deals with the related matters of statistical distinctions and notation.

(*continued overleaf*)

Accuracy: the inverse of *in*accuracy. (5.21)

Alternative hypothesis: see **Hypothesis testing**.

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.

Adequate replicating: see **Replicating**.

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*.

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*?

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 units, typically:
- the target population/process,
- the study population/process,
- the respondent population,
- the non-respondent population,
- the sample. (5.20)

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*.

Bar graph: see **Histogram**.

Bias: the *model* quantity representing *inaccuracy*.
(5.21, 5.46, 5.50, 5.63)

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 Zs 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 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)

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.
(5.32, 5.43, 5.76, 5.77)

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* unit (really 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 Z_1, Z_2, ..., Z_k hold their (same) values for *every* population unit (really element) when $X = 0$ and $X = 1$ (sometimes phrased as: *Hold all the* Z_i ***fixed*** for.....).

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

(3) **ATTRIBUTE:** The X-Y relationship is *causal* if:
 Attribute(Y, perhaps some of $Z_1, Z_2, \ldots.., Z_k \,|\, X = 0$) $\neq$
 Attribute(Y, perhaps some of $Z_1, Z_2, \ldots.., Z_k \,|\, X = 1$),
 provided those of Z_1, Z_2, ..., Z_k *included* in the attribute have the *same* values when $X = 0$ and $X = 1$. (5.32)

Census: an investigation using *all* the respondent (or study) population 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: if the random variables Y_1, Y_2, Y_3, ..., 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)

Chi squared distribution: if Z_1, Z_2, ..., Z_ν are probabilistically independent $N(0, 1)$ random variables, the *sum of their squares* has a χ^2 distribution with ν degrees of freedom ('chi' rhymes with 'hi' – it is

Figure 1.5. STAT 231 GLOSSARY (continued 1)

pronounced 'ki'). (13.14 to 13.16, Ap.7, Ap.8)

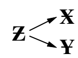
Clinical trial: see **Blind, Blinding**. (5.39, 5.49)

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

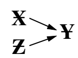
- **equal size:** for example, cardboard cartons of 24 cans in a population of cans of soup; OR:
- ***un*equal size:** for example, households in a population of people.

(5.24, 5.55, 5.57, 5.85, 5.86, 5.96)

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**.

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 the respondent population attribute represented by the model parameter.*
- ***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 model parameter.*

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 **X** changes to make apparent its relationship to **Y**, *their* effects on **Y**, and that of **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 esplanatory variate(s) among two (or more) groups of elements/units [like (sub)populations or samples] with different values of the focal variate.

When **Z** is a confounder, not taking account of **Z** values may make the Answer to a Question about a (*causal*) relationship between **X** and **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 **Z**, **X** and **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. (This type of 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)

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 units 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 $\bar{\mathbf{Y}}_0$ and $\bar{\mathbf{Y}}_1$, we have:

$$\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_0 = \text{effect of change in } \mathbf{X} + \text{effect of change in } \mathbf{Z}_1, \ldots, \mathbf{Z}_k = \text{treatment effect} + \textbf{confounding effect}. \qquad \text{-----(1.5.1)}$$

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

Contrast: see **Effect**.

Control group: in an experimental Plan, the part of the sample assigned **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)

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)

Counterfactual: a variate value *not* observed under the Plan – for instance, a unit'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: $cov(X,Y) = E(XY) - E(X)\bullet E(Y)$; -----(1.5.2)

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 units of the respondent (or study) population/process. (5.24)

Covering is relevant to implementing **judgement selecting**.

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.

Deduction: reasoning from *general* information to a *particular* conclusion (or answer), using the rules of logic and relevant theory.

Degrees of freedom: the (*un*evocative) name given (for historical reasons) to one or more parameters of the *t*, *K*, χ^2 and *F* distributions. (13.14 to 13.17)

Dependence: see **Independence**.

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 STAT 231.

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

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

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].

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 $\mathbf{Z}_i$ hold their (same) values when **X** changes;
- is defined (the 'true' effect) over the units of the *respondent* (or *study*) *population*. (5.43, 5.44)

(*continued overleaf*)

- **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)

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**.

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**.

Empirical: based on *data*.

Equiprobable assigning [EPA]: all possible assignments of equal numbers of units among the values of the focal variate(s) are equally probable; this probability is the reciprocal of the number of such assignments. (5.37, 5.45 to 5.50)
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 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 **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.

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**-**Y** 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)

Error is important as the source of *limitations* imposed on Answer(s).

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 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)

Estimating: a process which uses statistical theory to derive the distribution of an *estimator* and data to calculate an (interval) *estimate*. (5.21, 8.1, 8.2)

Estimated residual: see **Residual**.

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)

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 $\overline{A}$.
- **Disjoint events** have no points in common. (5.29)

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)

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**.

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)

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)

FDEAC cycle: acronym for a 5-stage *structured process* for data-based investigating; the stages are:

- **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.

See also **PPDAC cycle**.

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)

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 implemen-

Figure 1.5. STAT 231 GLOSSARY (continued 2)

ted, 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)

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:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \quad \text{-----(1.5.3)}$$

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**. (5.61)

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

Gaussian distribution: 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 factorial treatment structure so that interaction effect(s) can be estimated. (5.85)

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)

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.

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**; either direction for the effect is a **two-sided alternative**.]
- with a low enough *P*-value, the null hypothesis is **rejected**, in favour of **accepting** the alternative hypothesis.

The double negative (rejecting no effect meaning there *is* an effect), phrased in unfamiliar terminology, invites confusion.
We can think of rejecting the null hypothesis as (a claim of) a signal being detectable above the noise. See also **Significance testing**.

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)

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)

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**.

Interaction of two factors X_1 and X_2 is said to occur when the effect of one factor on a response variate 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 *more* than two factors. (5.76, 5.81)

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

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 **Confidence interval**.

Investigation: a *data-based* investigative undertaking involving one (or a few) Questions to be answered.
An investigation is to be contrasted with a *project.*

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

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

Judgement selecting: human judgement is used to select n units from the N (or N_S) 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)

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, Ap. 5, Ap. 6)

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

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

(*continued overleaf*)

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.*

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.*

Location: refers to where (the 'centre' of) a distribution is positioned; in this sense, averages and means are *measures of location.* (5.28)

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

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 *prevents confounding* of the focal variate with the non-focal explanatory variate(s) made the same within each group, thereby decreasing the *likely* magnitude of comparison error. SO THAT:
Matching *decreases* comparing imprecision, thus 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.*

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)

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 *unit measured*. (5.59 to 5.62)

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)

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

Model error: see **Error**.

Modelling asumptions: we 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.*

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

Non-respondent: a unit with some or all of its data missing at the end of the Execution stage of the FDEAC cycle.
Missing 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 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. (Ap. 1)

Null hypothesis: see **Hypothesis testing**.

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.

Operator: see **Measuring process**.

Outcome: the result generated by one execution of a *probabilistic* process. See also **Data**.

Overall error: see **Error**.

Pivotal quantity: see **Discrepancy measure**.

Placebo: see **Blind**.

Point: see **Sample space**.

Population: a well-defined group of *units* (really *elements*), ***other than*** the sample. (5.19, 5.55)

PPDAC cycle: the acronym for a original 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 confusion of 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.
Practical importance arises most commonly with an Answer that is a *difference* or a *ratio*, when the respective prescribed values may be 0 and 1.
A practically important *change* (or *difference*) (*e.g.*, in an Answer) can also be called a **meaningful** change (or difference).
Practical importance is to be distinguished from statistical significance – see **Significance testing**. (5.30, 5.46, 5.47, 5.71)

(*continued*)

Figure 1.5. STAT 231 GLOSSARY (continued 3)

Precision: the inverse of *im*precision. See also **Reliable**. (5.21)

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)

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 **Selecting probability**.

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

Thus, in statistics, a process may involve *units*. 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)

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, sometimes 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)

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

Random means *equiprobable* in a selecting (or sampling) context or an assigning context.
'Random' is better avoided as statistical terminology, except in the established phrase *random variable*.

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 synonym *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 units 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)

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** (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 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)

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 the definitions of a *CI*, an *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 *sam-*

(*continued overleaf*)

ple 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.
BUT: The representativeness of a sample can *rarely* be known.
This phrase should be *avoided* as statistical terminology. (5.21)

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 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 $\hat{\mathrm{r}}_j$ or $\hat{\mathrm{r}}_{ij}$; 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 *avoid* this terminology because of we use 'error' for a different (pervasive) statistical concern.

Respondent population: those units (really 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 units 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.

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, 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)

Response modelling of the behaviour of *error* under (hypothetical) repetition, with*out* doing *actual* repetition, is the basis for *estimating.*

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)
In *real-world* contexts, 'root ***average*** square' is preferable. (5.28)

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

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 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 **S** used for (data) 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 survey: see **Survey**. (5.84)

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

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)

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: under any protocol for selecting units, 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)

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 units are:

$$\left.\begin{array}{ccccc} \text{Study population} & = & \text{Respondent population} & + & \text{Non-respondent population} \\ \mathrm{N}_{\mathrm{s}} & = & \mathrm{N} & + & \mathrm{N}_{\mathrm{nr}} \\ \text{Selection} & = & \text{Sample} & + & \text{Non-respondents} \\ \mathrm{n}_{\mathrm{s}} & = & \mathrm{n} & + & \mathrm{n}_{\mathrm{nr}} \end{array}\right\} \text{-----(1.5.4)}$$

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)

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** (P), 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 sometimes used interchangably
The useful process of 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**.

Simple random sampling (SRS): a synonym for *equiprobable selecting* and, like *random sampling*, is better *avoided* as statistical terminology. (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 X_1-Y relationship for units in groups based on variate X_1 and when these groups are subdivided on the basis of variate X_2. (5.65 to 5.70)

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

Standard: a unit with a *known* variate value, used to *calibrate* 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 with *two* contexts:

- for a set of *data* (or *numbers*) – a 'data' s.d.; (8.1, 8.2)

(*continued*)

Figure 1.5. STAT 231 GLOSSARY (continued 4)

– for a *random variable* – a 'probabilistic' s.d. (5.6)

This distinction for a measure of *variation* parallels the average-mean (real world-model) distinction for a measure of *location.* (1.20, 5.28)

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.

Stratum, strata: see **Stratifying**.

Stratifying: subdividing the respondent (or study) population into groups (called **strata**) so that units with*in* a stratum have *similar* response variate values and units 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)

Strength: a word to be avoided as duplication of existing statistical terminology; strength (of an Answer) means *precision.* (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 error: see **Error**.

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

Subdividing: a form of *matching*, used in an *observational* Plan, in which the values of the response variate for the units of the sample are *subdivided* on the basis of the values of a non-focal *explanatory* variate that may be *confounded* with the focal variate. (5.37, 5.38)

Subdividing *prevents confounding* of the focal variate with the non-focal explanatory variate that is the basis for the subdividing. (5.39, 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*-technical 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 [$k \ll N$ (or $k \ll N_S$)] and then selecting every k*th* unit. (5.56)

Target population/process: the group of units(really 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**.

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**.

Treatement: 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 $X = 1$; 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)

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

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

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)

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

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

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

Variability: the *model* quantity representing *imprecision.*

Variance: a measure of variation – 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 variations *add*, whereas (natural) variations *add like Pythagoras.* (5.11 to 5.14)

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

Variate: a characteristic associated with each *unit* (really each *element*) of a population/process.

- **Response variate (Y):** 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 (X or Z):** a variate, defined in the Formulation stage of the FDEAC cycle, that accounts, at least in part, for changes from unit to unit (really, from element to element) in the value of a response variate.

Variation: differences in (variate or attribute) values:

- across the individuals (*e.g.*, 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;
- arising under repetition (*e.g.*, for error or a sample average).

Variation can be *quantified* by standard deviation. (5.21, 5.28)

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.

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

Wider inductive basis: a 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)

(*continued overleaf*)

Appendix: Statistical distinctions and Notation

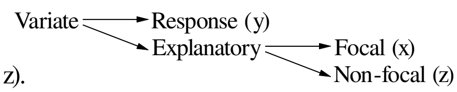
Understanding statistical methods requires an appreciation of three distinctions – between:

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

Disciplined use of terminology in this Glossary, together with notation like that discussed below, can enshrine maintaining these distinctions and so foster a mind-set which routinely recognizes what statistical methods can, and can*not*, accomplish.

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 using just *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 are used for response *model parameters* [*e.g.*, μ (mu), σ (sigma), π (pi)].

Table 1.5.2

Real world	Model
Y, y	Y, y
$\overline{\mathbf{Y}}$, $\overline{\mathrm{y}}$	$\overline{Y}$, $\overline{y}$, μ

As indicated for a response variate in the first line of Table 1.5.2 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 **Y** 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.

$$Y_j = \mu + R_j,\quad j = 1, 2, \ldots, n,\quad R_j \sim N(0, \sigma),\ \text{independent, EPS} \qquad \text{-----(1.5.5)}$$

The second line of Table 1.5.2 is for an *attribute* [of a *group* of units (really elements) like a population or a sample], specifically an *average* in the context of our simplest response model of equation (1.5.5). We say (or write): *the **population** average* $\overline{\mathbf{Y}}$ *is represented by the **model** mean* μ, *for which the **estimator** is the random variable* $\overline{Y}$ *and the **estimate** is the sample average* $\overline{\mathrm{y}}$.

- 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 5.7.5 on page 5.28 of Figure 5.7.
- Symbols are often subscripted; *e.g.*, Y_j and R_j in the model (1.5.5) and $\mathbf{Z}_1$, $\mathbf{Z}_2$, ..., $\mathbf{Z}_k$ for multiple non-focal explanatory variates.

Good places to see these conventions (and some extensions of them) in use are Figures 13.1 and 16.1 of these Course Materials.

We distinguish *four* populations by evocative adjectives:

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

These populations are key to developing the first two categories of **error** in Table 1.5.3 at the right below, and they appear, for example, in the two lower schemas on the first side (page 1.11) of this Glossary. [Some contexts involve a *process* rather than a population].

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

- *error*, which arises in an *individual* investigation, AND:
- a response model like the one in equation (1.5.5), because such models only describe behaviour under *repetition*.

The distinction is *explicit* in the terminology for our six categories of error; as shown in Table 1.5.3 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.

Table 1.5.3

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 error	Modelling	-----

- Studying and non-responding imprecision are de-emphasized in the right-hand column of Table 1.5.3 because these Course Materials assume a *deterministic* (*not* a stochastic) process for specifying the study population and, for people as 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 1.5.3) 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 1.5.3.

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