11

Assessing statistical literacy: Take CARE

Milo Schield

11.1 Statistical literacy: A new goal for statistical education

In 2006, statistical literacy was adopted as a goal by the ASA in endorsing the ASA Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (ASA, 2007). This goal is stated in the first sentence in the PreK-12 portion of the GAISE report:

> The ultimate goal: Statistical Literacy'. It is the first recommendation of the college section of the report: 'introductory courses in statistics should, as much as possible, strive to emphasise statistical literacy and develop statistical thinking' and it is in the ASA Strategic Plan (2008) for Education: 'Through leadership in all levels of statistical education, the ASA can help to build a statistically literate society . . . '

The PreK-12 section of the GAISE report underscores the importance of statistical literacy. 'Statistical literacy is essential in our personal lives as consumers, citizens, and professionals'. 'Statistical literacy is required for daily personal choices'. 'Statistical literacy involves a healthy dose of scepticism about "scientific findings"'. 'An investment in statistical literacy is an investment in our nation's economic future, as well as in the well-being of individuals'.

Assessment Methods in Statistical Education Edited by Penelope Bidgood, Neville Hunt and Flavia Jolliffe

11.2 Identifying statistical literacy

Broers (2006: 1) notes that the diverse attempts at defining statistical literacy (SL) 'usually take two different turns'. One stipulates 'what it is exactly, that a statistically literate citizen should know of statistics . . . These discussions tend to be clear, although not very consistent'. The other focuses 'on additional requirements for becoming statistically literate; requirements other than SK [Statistical Knowledge] elements. Here, the discussion of what it means to be statistically literate becomes far less clear'.

The first approach begins by linking statistical literacy with 'for whom' (all liberally-educated adults) and 'for what' (to be good citizens and decision makers). The second links statistical literacy with cognitive skills that are selected based on expert insight.

Moore (1998: 1) follows the first approach when he focuses on the needs of individuals in different roles to distinguish statistical literacy from statistical competence. 'What is statistical literacy, what every educated person should know? What is statistical competence, roughly the content of a first course for those who must deal with data in their work?'

Gal (2002: 2) also follows the first approach when he defines statistical literacy by focusing on the needs of adults in modern societies. From their needs, he concluded that statistical literacy refers to two interrelated concepts, primarily (a) their ability to 'interpret and critically evaluate statistical information' which they may encounter in diverse contexts, and when relevant (b) their ability to 'discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions'. Gal elaborates on this consumer-producer distinction, linking reading contexts with statistical literacy, and enquiry contexts with data producers. Although statistical literacy is considered a key educational goal, Gal (2003: 81) concludes that many instructors 'neither teach for statistical literacy nor assess it', that 'serious attention to statistical literacy issues (in terms of both skills and dispositions) cannot be accomplished within an introductory course focused on core statistical topics' and that 'separate courses focused on statistical literacy will have to be planned'. Utts (2003: 74) follows the first approach, in relating statistical literacy to what every educated citizen should understand about probability and statistics.

The college section of the GAISE report follows the first approach, linking statistical literacy with art appreciation and consumers of data. Some courses focus on teaching students to become 'statistically literate and wise consumers of data; this is somewhat similar to an art appreciation course'. Some courses focus on teaching students to become 'producers of statistical analyses; this is analogous to the studio art course'. The mixture of consumer and producer components 'will determine the importance of each recommendation'. This report suggested assessing statistical literacy by students ‘interpreting or critiquing articles in the news and graphs in media’.

Ben-Zvi and Garfield (2004a) follow the second approach in trying to distinguish statistical literacy (SL) from statistical reasoning (SR) and statistical thinking (ST) in terms of the types of understanding or cognitive outcomes. Broers (2006: 4) questions the grounds of these distinctions and argues that in looking for learning goals, ‘SL, SR and ST were postulated as constructs’: constructs that could help us in searching for new directions in statistics education; constructs that could provide a better view of what we should be teaching and how best to teach it. ‘Their existence is not dictated by empirical observations, but rather empirical observations are sought in order to justify their creation. It is the individual researcher who decides what will be included in a definition and what will be left out’.

Broers’ psychological critique may be unjustified, but if a non-empirical classification is to avoid being arbitrary it must clearly distinguish the constructs and have strong support among subject-matter experts.

Despite the contrast between a user-needs approach and an expert-insight approach – between an empirical approach and an idealistic approach – they may give similar results.

A somewhat similar tension exists in the mathematics community in defining quantitative literacy. One approach focuses on finding applications of mathematical ideas in everyday life (Gilman, 2006), another on quantitative literacy as part of effective citizenship in a modern democracy (Madison and Steen, 2008; Steen, 2001).

In this chapter, the user-needs approach of Moore and Gal is combined with the effective citizenship approach of Steen and Madison to argue that statistical literacy should be empirically based on the statistical needs of educated adults in a modern society.

11.3 Statistical literacy: For whom

Statistical competence is the ability to produce, analyse and summarise detailed statistics in surveys and studies. Statistical competence is needed by data producers – students in quantitative majors that have a statistics requirement, such as business, psychology, sociology, economics, biology and nursing – and possibly majors that have a calculus requirement such as those in science, technology, engineering and mathematics (STEM).

Statistical literacy is the ability to read and interpret summary statistics in the everyday media: in graphs, tables, statements and essays. Statistical literacy is needed by data consumers – students in non-quantitative majors: majors with no quantitative requirement such as political science, history, English, primary education, communications, music, art and philosophy. About 40% of all US college students graduating in 2003 had non-quantitative majors (Schield, 2008b).

11.4 Statistical literacy: The goal

Once statistical literacy is generally defined, the next step is to identify what activities would serve as the best indicator of proficiency. Two closely related approaches are:

- Make a decision based on multiple statistics. Consider the Collegiate Learning Assessment (CLA) organised by the Council for Aid to Education (CAE). Students are asked to present a reasoned conclusion given either a simple performance task (analyse some summary data) or a 'Make an Argument' task (given some related factors). See www.cae.org/content/pro_collegiate.htm.
- Evaluate the statistics in the everyday media. Tables, graphs, headlines, news stories, press releases and research reports present statistical associations as evidence for causal connections (Schield, 2008a; Budgett and Pfannkuch, 2007; Madison, 2006; Lutsky, 2006; Hayden, 2004; Moreno, 2002; Snell, 1999; Watson, 1997).

Unfortunately, choosing between deciding and evaluating gives little guidance on what ideas, tools and skills should be assessed. Doing this requires input from subject-matter experts. Once these topics are identified, they can be ranked in two ways:

- Ranked by subject-matter experts. For such a ranking in statistics, see McKenzie (2004).
- Ranked by their prevalence in the everyday media.

11.5 Identifying relevant topics by subject-matter experts

Moore (1998: 1) thought that statistical literacy should involve two clusters of 'big ideas':

1. 'The omnipresence of variation, conclusions are uncertain, avoid inference from short-run irregularity, [and] avoid inference from coincidence'.
2. 'Beware the lurking variable, association is not causation, where did the data come from? [and] observation versus experiment'.

Best (2008) claims that the dominant influence on all statistics is human choice since all statistics are socially constructed. Not that reality is subjective, but that human beings decide what and how to count and measure, what to summarise and how to model, and what to compare and how to communicate.

Utts (2003) identifies seven topics that every educated citizen should understand about probability and statistics. These include distinguishing causation from association, experiment from observational study, statistical significance from practical importance, and no-effect from no-difference.

11.6 Take CARE

Schield (2008a) reviews the topics – the sources of influence on the value of a statistic – identified by subject matter experts. The goal is to classify these influences into a small number of categories that are exhaustive, exclusive and fundamental. Now, all too often there is something omitted so the categories are not exhaustive, there are borderline cases so the categories are not exclusive and identifying what is essential is certainly contextual so that in a different context the categories are not fundamental. Nevertheless consumers of statistics – people who are not working with statistics regularly – can benefit from focusing on a smaller number of categories, even if they are not logically pristine, provided they are fundamentally different.

All the factors that influence a statistic have been classified into four categories:

- **Context** The influence of factors taken into account (1) by counts, averages, ratios and comparisons of counts, averages and ratios; (2) by epidemiological models (cf., deaths attributable to obesity); (3) by regression models; and (4) by the study design (cf., controlled vs. uncontrolled; longitudinal vs. cross-sectional; experiment vs. observational study) or by selection (cf., in tables and graphs). The influence of related factors (confounders) that were not taken into account in the study and were not blocked by the study design.
- **Assembly** The influence of choices (1) in defining groups or measures, (2) in selecting the summary measure (e.g. mean vs. median), the type of comparison (e.g. simple difference versus times more), and the type of ratio (e.g. the confusion of the inverse or the prosecutor's fallacy), (3) in selecting the group in forming an average, the base in a comparison of numbers and the denominator in a ratio (e.g. rate or fraction) and (4) in selecting the graph, table or statistic in presenting statistical results and summaries.
- **Randomness** The influence of chance on averages and coincidences (e.g. hot hand, too unlikely to be due to chance and regression to the mean). The difference between statistical significance and practical significance in large samples or between 'no statistical effect' and 'no effect' in small samples. The influence of a confounder on statistical significance.
- **Error** (or Bias) The influence of any factor that generates a systematic difference between what is observed and the underlying reality: subject bias (people can lie), measurement bias (instruments can fail, questions may lead and researchers may manipulate) and sampling bias (the difference between the sampled and the target population influences the result).

Given the extensive influence of human choice on numbers, the W.M. Keck Statistical Literacy Project grouped these four sources of influence under the age-old admonition, ‘Take CARE’ where each of the four letters in ‘CARE’ signified a distinct source of influence on any statistic: Context, Assembly, Randomness and Error. If students were to remember to ‘Take CARE’ in analysing statistics, that would be a considerable achievement. The choice of ‘Context’ for the first category is based on the importance that context plays in the liberal arts and on the importance that statisticians place on context in distinguishing statistics from mathematics.

11.7 Relevant topics based on empirical evidence

But all this analysis is based on expert opinions. What about the empirical approach? Gal (2002: 19) noted that:

> What is basic knowledge . . . depends on the level of statistical literacy expected of citizens, [depends] on the functional demands of different contexts of action (e.g. work, reading a newspaper), and depends on the larger societal context of living . . . Unfortunately no comparative analysis has so far systematically mapped out the types and relative prevalences of statistical and probabilistic concepts and topics across the full range of statistically-related messages or situations that adults may encounter and have to manage in any particular society.

Here are the results of two empirical studies:

- An analysis (Schield and Schield, 2007) of 250 statistically-based news stories found that 75% were observational and thus vulnerable to confounder influence, 74% involved assembly (choices in definitions, comparisons or presentation) and over half involved the grammar of percent, percentage, rate or chance indicating the use of ratios.
- An analysis (Raymond and Schield, 2008) of 160 numerically-based stories found that 27% involve assembly in constructing categories or measures, and 62% present associations that imply causation. In terms of the four ‘take CARE’ categories, 42% of these stories involved statistics that were influenced by confounding, 17% by assembly, 11% by bias and 9% by random effects.

Although this is empirical data, these findings are very tentative. First, there is considerable latitude in identifying these influences in a news story. Second, these articles and studies are not very representative. They are based almost entirely on short health-related news stories. They do not include longer articles in magazines or the press releases or detailed reports of government statistical offices. Third, they involve the same author. Given the latitude in selecting the stories and in classifying these influence, having an independent analysis would be valuable.

Nevertheless, both these empirical studies uphold an important finding: that Context and Assembly are much more prevalent as statistical influences in the everyday news than are Randomness and Error/Bias. This finding is important in deciding what to assess and how to assess it.

This analysis of the definition, the nature, the goals and the content of statistical literacy is just theory – theory that is barren until it results in exercises that students can work and that are a useful basis for assessment. Without such exercises, statistical literacy will exist as a vague 'habit of mind' that cannot be readily assessed or taught. So let us turn to the assessment of statistical literacy.

11.8 Assessing statistical literacy: Introduction

Assessing statistical literacy can be done at four levels by having students:

1. Evaluate the use of statistics in a news story (Schield, 2008a).
2. Estimate a quantity or make a decision in an open-ended situation (Gilman, 2006).
3. Describe and compare statistics presented in graphs or tables.
4. Answer multiple-choice questions on specific aspects of statistical literacy.

The last two are presented below.

11.8.1 Describe and compare statistics presented in graphs or tables

To be statistically literate, one must be able to use ordinary English to describe and compare rates and percentages as presented in tables and graphs. This can be quite difficult for students – especially those for whom English is a second language. Most students do not see a difference between 'times as much as' and 'times more than'; most students do not see that 'the percentage of men who smoke' is the same as 'the percentage of smokers among men'; nor that 'the death rate of men' is the same as 'the rate of death among men'.

Assessing student classifications or writing can be done quickly using a grading template:

- *Part-whole classification.* When students are asked to identify the part and whole given a statistic in a description, graph or table, a three-, four- or five-mark scale is adequate (depending on the complexity of the ratio) with one mark deducted for each part-whole error. Consider this statement: 'Among women who were first married in 2003, the percentage who were ages 25–29 was 27%'. A student who classified 'first married' as part would lose one mark as would a student who classified 'ages 25–29' as whole.
- *Comparison of numbers.* When students are asked to write a statement comparing two counts, totals or averages, a four mark scale is adequate: one

for the proper comparison grammar (difference, ratio, or percent/times difference); one for the proper base indicator (as or than); one for the proper comparison amount; and one for the appropriate test and base. Suppose a student is asked to compare six and two as a ratio with the larger as the base, and they write, ‘Six is three times more than two’. The student would lose three marks: one for using two as the base, when using the larger (six) was specified; one for using ‘times more’ grammar when a ratio (times as much) was specified; and one for using ‘than’ to introduce the basis of the comparison when ‘as’ is appropriate.

Statements that describe and compare ratios such as percentages and rates are more difficult to assess. Table 11.1, a table of percentages with 100% column totals, is used to illustrate the assessment process in such cases. The P and W letters signify Part and Whole respectively.

Table 11.1 Table of percentages (hypothetical data).

Students [W]	----------------SEX-[W]--------		
Ⓟ MAJOR	[W] MALE	[W] FEMALE	[W] ALL
Ⓟ Business	↓ (60%)	↓ (20%)	↓ 40%
Ⓟ Economics	↓ 10%	↓ 50%	↓ 30%
Ⓟ MIS	↓ (30%)	↓ 30%	↓ 30%
ALL	100%	100%	100%

- *Description of a ratio.* When students are asked to write a statement describing a ratio in a graph or table, a four mark scale is adequate – one mark for the ratio keyword and the rest for proper part-wholes. When asked to describe the 60% in the upper left corner using percentage grammar, a student wrote, ‘60% of business majors are males’. This reply would lose four marks: one for omitting ‘college students’; one for describing ‘males’ as a part; one for describing ‘business majors’ as a whole; and one for using percent grammar instead of percentage grammar. A correct answer would be: ‘Among college students, the percentage of males who are business majors is 60%’.

- *Comparison of ratios.* When students are asked to compare two ratios given in statements, a table or graph, a six-mark scale is adequate, with one mark for the keyword (percentage, rate, etc); one for part; one for wholes; one for test/base and base indicator; and two for compare (numeric and grammar). When asked to compare the circled 60% and 20% as a simple ratio using the smaller as the base using percentage grammar, a student wrote: ‘Among college students, males are three times more prevalent than females among business majors'. This reply would lose five marks: one for describing 'business majors' as a whole; one for describing 'males' and 'females' as parts; one for using a 'times more' comparison when a simple ratio was requested; one for using 'than' to introduce the base in the comparison; and one for using likely/prevalent grammar when percentage grammar was requested. A correct answer would be: 'Among college students, the percentage of business majors is three times as big among males as among females'.

For more on the grammar for describing and comparing ratios, see Schield (2004a).

11.8.2 Answer multiple-choice questions on specific aspects of statistical literacy

The rest of this chapter deals with multiple-choice questions similar to those administered to students in non-quantitative majors at Augsburg College. These are taken from Moodle exercises involving over a thousand questions. While open-ended problems, essays and portfolios provide more comprehensive forms of assessment that may be needed to identify higher levels of statistical literacy, right-wrong exercises are useful in assessing basic levels of statistical literacy, for minimising instructor grading time and for handling large-enrolment courses. These questions are presented in four categories: context, assembly, randomness and error/bias.

11.9 Assessing statistical literacy: Context

To understand the influence of context on a statistic, a statistically literate person must understand the features and benefits of different types of study designs (observational vs. experimental, longitudinal vs. cross-sectional, controlled vs. uncontrolled, and randomly-assigned). They must also understand the simplest ways of taking related factors into account (comparing subgroups, using averages, comparisons, ratios, comparisons of ratios and relative risks), and the more complex ways of taking related factors into account (using weighted averages to adjust rates and percentages for the influence of a binary confounder). Statistically literate adults must understand association-generated measures involving epidemiological models: the percentage – and number – of cases that are attributable to an associated factor. They must also be able to describe and compare rates and percentages presented in tables and graphs. Here are some right/wrong questions involving context:

1. A controlled study means it is an experiment – not an observational study. [Answer: False. Explanation: A controlled study is any study having more than one group.]

2. Do these statements have the same meaning? (A) Widows are more likely among suicides than widowers. (B) Widows are more likely to commit suicide than widowers. [Answer: No. Explanation: Suicide is the common-whole in A, the common-part in B.]

3. What percentage of low-weight births are attributable to the mother smoking? In the US in 2002, the percentage of newborns that have low birthweight is 12.2% among mothers who smoke and 7.5% among non-smoking mothers. See Figure 11.1 (Table 84, 2004 US Statistical Abstract).

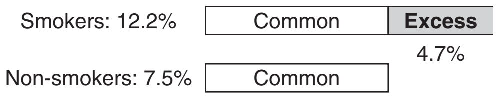

Figure 11.1 Percentage of newborns with low birth-weights.

[Answer: 39%. Explanation: The risk 'attributable to' membership in the group with the larger rate (the 'attributable risk' or 'excess risk') is the rate difference (the 4.7 percentage point excess in the larger) divided by the larger rate (12.2%) shown as a percentage: 100%(12.2%-7.5%)/12.2%.]

4. In 2002, of the low-weight births to US mothers who smoke, how many are attributable to their mother smoking? Of the 4.02 million US births in 2002, 8.0% were low weight so that 10% of these mothers reported smoking. [Answer: $\sim$19,000. Explanation: 400,000 (10%) babies born to mothers who smoke. 48,000 (12%) have low birth-weights. 19,000 (39%) are attributed to their mother smoking.]

5. At a given hospital, suppose that entering patients are in either good condition or poor condition and that the patient death rate is 4% for patients in good condition versus 12% for those in poor condition. What is the average death rate for this hospital if 75% of the patients are in poor condition? [Answer: 10%. Explanation. This average is a weighted average – not a simple average. $0.75*0.12 + 0.25*0.08 = 0.08 + 0.02 = 0.10$.]

6. If one of these percentages is bigger which is it? (A) The percentage of infant deaths which are due to birth defects. (B) The percentage of infants who die due to birth defects. [Answer: A. Explanation: Let X be 'due to birth defects', Y be 'death' and Z be 'infant'. $A = P(X|YZ)$ and $B = P(XY|Z)$. Since $XY \leq X$ and $YZ \leq Z$, it follows that $P(XY|Z) \leq P(X|YZ)$.]

7. What is the chance that a young adult who fails to graduate from high school will spend time in prison? Suppose that 72% of the young adults in prison did not graduate from high school whereas 12% of all young adults did not graduate from high school. Suppose that 5% of all young adults spend time in prison. [Answer: 30%. Explanation: Let X be 'did not graduate from high school' and Y be 'in prison'. $P(X|Y) = 72\%$; $P(X) = 12\%$, $P(Y) = 5\%$. Bayes Rule: $P(X|Y)/P(X) = P(Y|X)/P(Y)$. So $P(Y|X) = P(X|Y)[P(Y)/P(X)] = 0.72[.05/.12] = 0.3 = 30\%$.]

8. Do these statements say the same thing? (A) Smoking is twice as prevalent among women as among men. (B) Women are twice as likely to be smokers as are men. [Answer: Yes.]

9. Do these statements say the same thing? (A) Smoking is 50% more prevalent among women than among men. (B) Men are 50% less likely to be smokers than are women. [Answer: No. '50% more' is equivalent to '33% less'.]
   A statistically literate person should be able to envision what it means to 'control for' or 'take into account'. In the simplest case, it means recognising that a difference in ratios may be spurious after controlling for a related factor – a confounder.

10. Is this difference in state NAEP scores shown in Table 11.2 real or spurious?

Table 11.2 NAEP Mathematics Scores 2000, grade 8: MD vs AZ.

Average Score	State	Internet access at home?				
		YES		NO		ALL
274	Maryland (MD)	281	70%	258	30%	100%
271	Arizona (AZ)	281	55%	258	45%	100%

Source: US Dept of Education, National Assessment of Educational Progress (NAEP).

Since the three-point difference in scores between states vanishes for each of the two subgroups, a statistically literate student must recognise the original difference by state is spurious after taking into account the different mixtures of students having Internet access at home.

11.10 Assessing statistical literacy: Assembly

Recall that 'assembly' means the choices in defining, selecting or presenting statistical relationships. Consider these examples of 'assembly' from Schield (2007):

- OPEC countries supply 50% of US oil imports, but only 30% of US oil usage.
- The average US farm is 440 acres; the average US family farm is 326 acres.
- Annual income is $43K for households, $53K for families and $62K for married couples.
- In 2005, the world gained 2.3 people per second (over 74 million people per year).

To understand the influence of assembly on a statistic, statistically literate consumers should know that as the definition of a group becomes more restrictive, the size of the group will decrease, but that the size of a ratio involving that group as the whole may increase. They should know that arithmetic differences are typically bigger than percent differences or times ratios when comparing large numbers but that percentage differences or times ratios are typically bigger than arithmetic differences when comparing small numbers. They should know that the choice of the cut point in forming subgroups from a continuous distribution can strongly influence the difference and ratios of statistics for the subgroups.

The following items assess one's knowledge about the influence of 'assembly' on a statistic.

Example 1: Which definition gives the larger number?

1.1 Number of teens: 'those 13–6' vs. 'those 13–19'. [Answer: The less restrictive group (13–19).]

1.2 Smokers: those who smoked in the last month vs. in the last year. [Answer: Longer time period.]

1.3 Average incomes: those ages 20–65 vs. those ages 1–100. [Answer: Ages 20–65.]

Table 11.3 US Women (in millions) who had a child in 2004 by family income.

<10K	10–19.9K	20–24.9K	25–29.9K	30–34.9K	35–49.9K	50–74.9K	75K and up
4.2	6.2	3.4	3.8	3.6	8.9	10.6	12.5

Source: 2006 US Statistical Abstract, Table 88.

Example 2: Who had more babies: rich mothers or poor mothers? (see Table 11.3)

2.1 Define 'Rich' as 35K and up; define 'Poor' as under 35K. [Answer: Rich mothers.]

2.2 Define 'Rich' as 75K and up; define 'Poor' as under 25K. [Answer: Poor mothers.]

Table 11.4 Estimated US persons (thousands) living with AIDS by race/ethnicity for 2003.

ALL	Non-hispanic white	Non-hispanic black	Hispanic	Other
406	147	172	80	5

Source: 2006 US Statistical Abstract, Table 180.

Example 3: According to Table 11.4, which group had the most cases of AIDS? (Assume that 75% of Hispanics are white.)

3.1a. Non-Hispanic white	b. Non-Hispanic black	[B]
3.2a. White (including white Hispanics)	b. Black (including black Hispanic)	[A]

Allocating Hispanics by race gives 207,000 whites and 192,000 blacks.

Example 4: Which saves more gas in going 10,000 miles?

4.1 Improving car A's miles per gallon from 10 to 20, or

4.2 Improving car B's miles per gallon from 20 to 100? [Answer: 4.1 saves 500 gallons – half of the initial 1,000; 4.2 saves only 400 gallons – 80% of the initial 500.]

11.11 Assessing statistical literacy: Randomness

Even before learning anything mathematical about chance, a statistically literate person should recognise the law of Very-Large Numbers: the unlikely is almost certain, given enough tries. More specifically, a random event with 1 chance in N of occurring on the next try is expected – is more likely than not to occur at least once – in the next N tries. (Schield, 2005).

Even before they can calculate a margin of error, a statistically-literate person should recognise that statistical significance can be determined, given two sample means and their associated margins of error or confidence intervals. If two 95% confidence intervals do not overlap, that test indicates the difference between these sample means is statistically significant at the 5% level. When they do overlap, that test indicates the difference is not statistically significant. A more accurate test (e.g. a t-test) will find that some of these cases of statistical insignificance are actually statistically significant. But a more sophisticated test (resampling) may find that some of the t-test cases of statistical insignificance are actually significant. Statistical significance is determined by the test.

1. When an activity involves discrete outcomes with various probabilities, is the expected value always one of the outcomes? [Answer: No. An average need not be one of the outcomes.]
2. When flipping a fair coin, what is the chance that the next 5 flips are all heads? [Answer: 1 in 32.]
3. When flipping a set of five coins 32 times where a success is flipping all five heads, what is the expected value: the number of successes one can expect? [Answer: One.]
4. If the expected value is a possible outcome, does this mean that expected outcome is more likely than not? [Answer: No. When flipping a fair coin four times, the expected outcome is two heads. This is the most likely outcome, but the chance of two heads is less than 50%.]

5. If the expected value is a possible outcome, does this mean the expected value is the most likely outcome? [Answer: No. If the probability distribution for discrete outcomes is symmetric with a bowl shape, the expected value will be the middle which has the smallest probability.]
6. Is a rare coincidence – an event that is extremely unlikely if due to chance – therefore highly unlikely to be due to chance and thus highly likely to be due to some causal factor? [Answer: Not necessarily. If predicted before the fact, probably. If selected after the fact, probably not.]
7. Suppose that in a two-candidate race, a poll indicates that one candidate has 55% of the vote while the other has the remainder. The 95% margin of error is 4%. Is this difference statistically significant when using the simple overlapping confidence-intervals test? [Answer: Yes.]
8. Suppose that a survey found that the average income was $55,000 for one group – 10% more than that of a second group. The 95% margin of error for both groups was $3,000. Is this difference statistically significant when using the simple overlapping confidence-intervals test? [Answer: No. The average income for the second group must have been $50,000. With a $3,000 margin of error on each value, the $5,000 difference is not statistically significant.]
9. Suppose that subjects are randomly assigned to two groups and that the difference in their outcomes is statistically significant. Can we expect that taking into account a pre-existing condition will make the difference statistically-insignificant? [Answer: No, random assignment tends to allocate any pre-existing condition equally to the two groups, so taking into account this condition is not expected to change anything.]
10. If subjects are assigned to two groups without random assignment and the difference in their outcomes is statistically significant, can taking into account a related condition give a new difference that is statistically insignificant? [Answer: Yes if strong enough. See Schield (2004b).]

11.12 Assessing statistical literacy: Error/bias

1. Will having a larger sample mitigate the influence of error or bias? [Answer: Not generally.]
2. Can getting a larger sample (conducting a census) increase the chance of error? [Answer: Yes.]
3. Of those surveyed, 20% did not respond. Is this non-response bias? [Answer: No. Non-response causes bias only if non-respondents would have answered differently than respondents.]

11.13 Assessing the influence of confounding

Ridgway *et al*. (2008: 1) note that 'most interesting problems are multivariate', and so 'the curriculum (and ideas about statistical literacy) should encompass reasoning with multivariate data'. Statistical educators may question whether it is possible teach students about the influence of a confounder on a statistic without teaching multivariate regression and the associated diagnostics and assumptions. Schield (2006) demonstrates a simple graphical technique for a binary predictor and a binary confounder that bypasses the need to discuss the assumptions of linear regression. This graphical technique involves weighted averages and uses a statistical principle from the 1960s called 'standardising'.

The following exercise uses this technique to show the influence of a confounder on three things: the size of an association; the number of cases attributable to a related factor; and the statistical significance of a difference between two groups.

11.13.1 The size of an association

To see the influence of a confounder on an association, consider two hospitals: Rural and City. Patients in good condition can walk in; patients in poor condition are carried in. Suppose the death rates (hypothetical) are 2% and 7% for those in good and poor condition at the Rural hospital; 1% and 6% respectively at the City hospital. Suppose that 90% of the City patients are in poor condition (30% of the Rural).

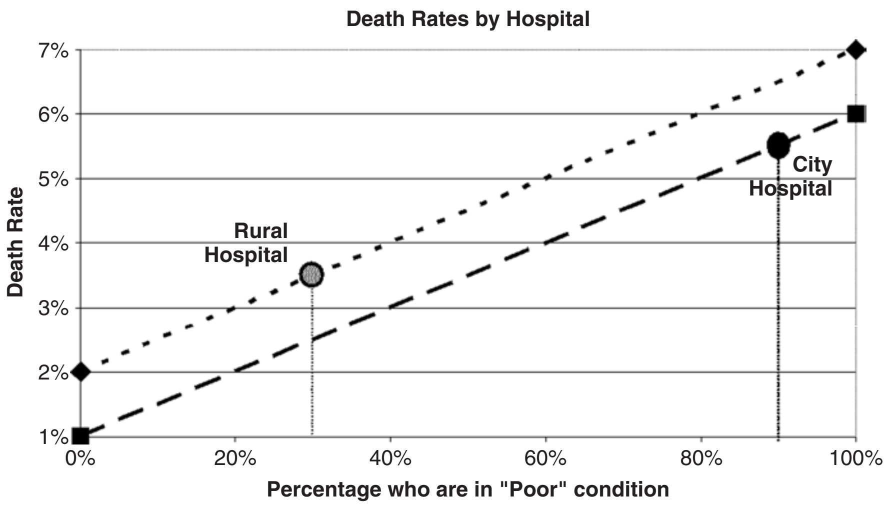

Figure 11.2 Raw hospital death rates.

1. What are the average deaths rates at these two hospitals? [Answer: City (5.5%), Rural (3.5%). Algebraic solution for a weighted average: City (0.9*0.06 + 0.1*0.01), Rural (0.3*0.07 + 0.7*0.02). Figure 11.2 shows how this weighted average can be solved graphically for each hospital.]

2. Which hospital has the higher death rate after taking into account the difference in patient mix? Figure 11.3 illustrates standardising: taking into account a difference in mix. Suppose the combined hospitals have 60% of their patients in poor conditions. If we standardise on 60%, we see that Rural has a higher standardised deaths rate (5%) than City (4%). This reversal is an example of Simpson's Paradox. This simple graphical technique allows students to work problems and thereby calculate the influence of a binary confounder on an association.

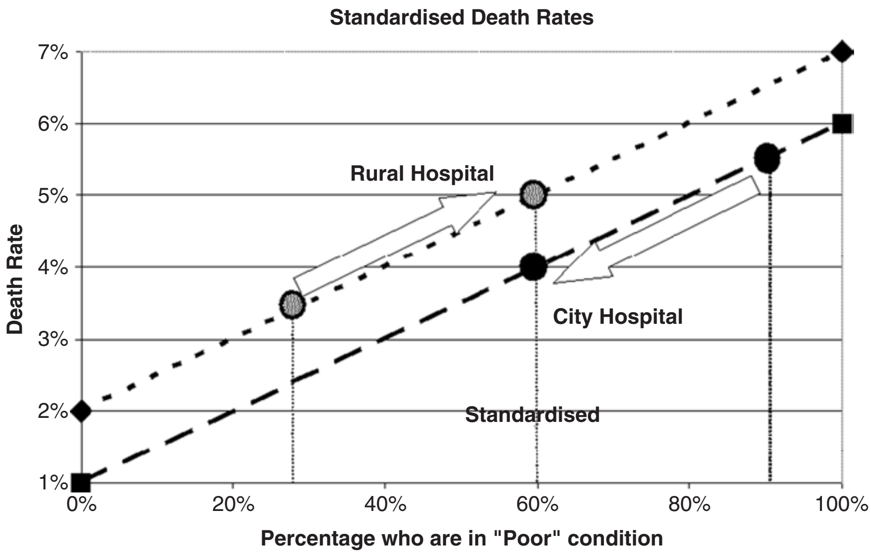

Figure 11.3 Standardised hospital death rates.

11.13.2 The number of cases attributable to a related factor

Confounding can influence speculative statistics based on epidemiological models. Suppose that among mothers who do not smoke the percentage of their babies who have low birth-weights is 6% and 11% among older and younger mothers; 11% and 16% among those who do smoke. Suppose that those under 19 are 10% of non-smokers (50% of the smokers).

1. Among non-smoking mothers, what percentage of babies have low birth-weights? [Answer: This problem in weighted averages can be solved

algebraically. Among non-smoking mothers: 0.10*0.11 + 0.90*0.06 = 0.065. Among smoking mothers: 0.5*0.16 + 0.5*0.11 = 0.135. Or it can be solved graphically as shown in Figure 11.4. Either way, the percentage is 6.5% among non-smoking mothers and 13.5% among mothers who smoke.]

2. What percentage of low birth-weight babies with mothers who smoke are attributable to their mother smoking? [Answer: 52%: (13.5%–6.5%)/ 13.5%. See prior discussion on excess risk.]

3. How many babies having low birth-weights are attributable to their mother smoking? Assume there were 3.5 million births. Assume 25% of these mothers smoked. Of the 875,000 babies whose mothers smoked, 13.5% (118,125) have low birth-weights. Of these 118,125 low birth-weight babies whose mothers smoked, 52% (61,250) are attributable to their mother smoking. [Answer: 61,250.]

4. After taking into account the influence of age, what are the standardised percentages of babies who have low birth-weights? Assume that 20% of all mothers are 19 or younger. [Answer: The standardised percentage of babies who have low birth-weight is 7.0% among non-smoking mothers, 12.0% among mothers who smoke. Algebraically: 0.2*0.11 + 0.8*0.6 = 7%; 0.2*0.16 + 0.8*0.11 = 12%. Figure 11.4 illustrates these standardised values graphically.]

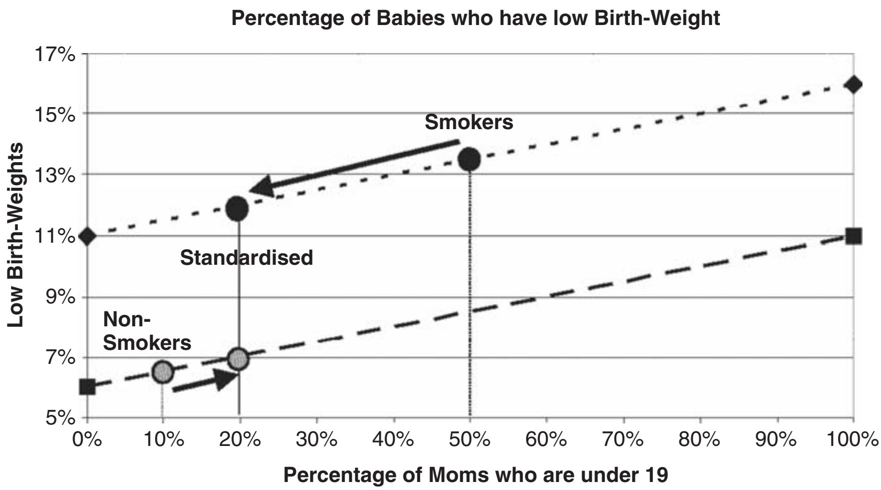

Figure 11.4 Influence of confounding on percentages and cases attributed.

5. Using the standardised values what percentage of low birth-weight babies whose mothers smoke are attributable to their mother being a smoker? [Answer: 42%: (12%–7%)/ 12%.]

6. After taking into account the influence of age, how many babies having low birth-weights are attributable to their mother smoking? Assume 3.5 million. Assume 25% of these mothers smoked. Of the 875,000 babies whose mothers smoked, 12% (105,000) have low birth-weights. Of these 105,000 low birth-weight babies whose mothers smoked, 42% (43,750) are attributable to their mother smoking. [Answer: 43,750.]

7. Compare the number of babies who have low birth-weights that are attributable to their mother smoking before and after taking into account the influence of age. In both cases, there were 875,000 babies whose mothers smoked.

   - Without taking age into account, 118,125 had low birth-weights. Of these 61,250 (52%) were attributable to their mother smoking.
   - After taking age into account, 105,000 had low birth-weights. Of these 43,750 (42%) were attributable to their mother smoking.

8. Analyse the difference in these two cases. Taking age into account reduced the number of low-weight births attributable to smoking by almost 30% – from 61,250 to 43,750. Figure 11.5 illustrates these differences.

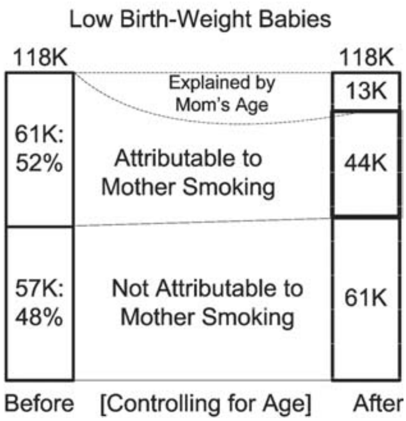

Figure 11.5 Low birth-weights attributed to smoking: influence of age.

Figure 11.5 preserves the original number of cases. Statistical educators can decide how best to make and explain these comparisons. Controlling for age decreased the number of low birth-weight births attributed to smoking from 61K to 44K – a reduction of almost 30%.

11.13.3 The statistical significance of a difference between two groups

Controlling for a confounder can influence whether a difference that is statistically significant becomes statistically insignificant – or vice versa.

1. Are these differences statistically significant if the 95% margin of error is three percentage points? Using the gap between confidence intervals as a simple – but conservative – test for statistical significance, the initial seven-point gap (13.5% vs. 6.5%) is statistically significant.

2. Are these differences statistically significant after controlling for the influence of age? [Answer: No. The standardised gap is five points: (12% vs. 7%). Using the simple overlapping confounder intervals test, this difference is not statistically significant.
   Both of these results can be seen graphically in Figure 11.6.

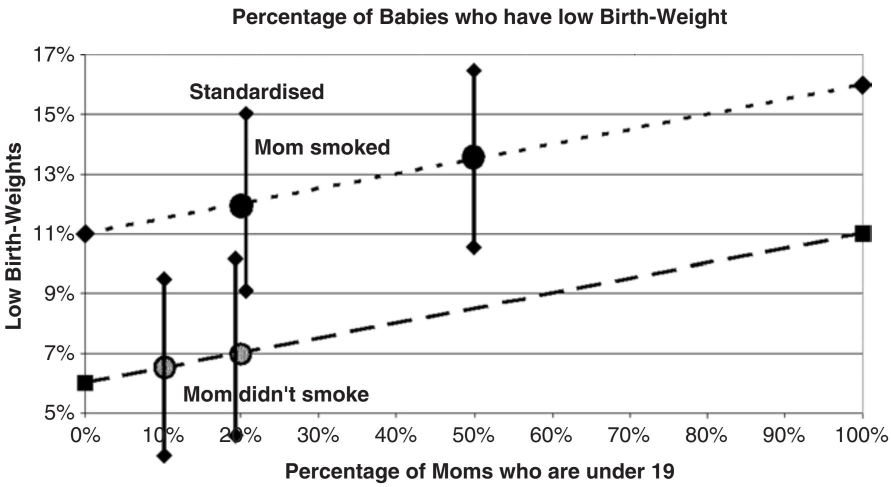

Figure 11.6 Influence of confounding on statistical significance.

11.14 Student feedback

Student feedback indicates they have found these exercises very helpful in understanding the key idea that association is not causation. Their primary request is for a wider variety of problems so they can strengthen their skills. Students are amazed at how the choice of a definition can so easily influence the size of a number. They are astounded that an arithmetic association can change direction after taking something else into account. Students are dismayed that risks 'attributed to' or 'due to' an associated factor are really speculative statistics that involve no causal claim. They are very dismayed to learn that these excess risk numbers can change after controlling for a related factor. And they are often bewildered when they learn that a statistically significant difference can become statistically insignificant (and vice versa) after controlling for a confounder. In summary they are amazed that statistics are so easily influenced by social construction, whereas the numbers in arithmetic are immune. As a result, they often say, 'I see statistics differently now'. They strongly agree that one must definitely 'take care' when using statistical associations as evidence for causal connections in arguments.

11.15 Assessment

These exercises provide a basis for assessing student knowledge of the various influences on a statistic. Assessing content knowledge is important, but assessing student attitudes is more important. Macnaughton (2004) argues that the primary goal of an introductory statistics course should be to give students 'a lasting appreciation for the value of statistics'. This may be difficult and not as important in teaching statistical competence when students have no idea of how valuable statistical inference is in certain situations. This should be easier and more important in teaching statistical literacy where students can readily find examples of statistical illiteracy in the everyday news.

11.16 Conclusion

Statistical literacy and statistical competence are related but different. Neither guarantees the other. Students in non-quantitative majors need statistical literacy. Students in some quantitative majors may need both.

A statistical literacy course should be designed to satisfy the needs of citizens in a modern, data-driven society, to help them think critically about statistics when used as evidence in arguments. There are many ways to design a course to achieve these goals. However, if a course is to carry a statistical literacy designation and meet these goals it should: (1) study all sources of influence on a statistic; (2) choose topics based – in large part – on their prevalence in the everyday media; and (3) inspire data consumers to see a positive value in the material presented.

Statistical literacy courses are becoming increasingly common in US four-year colleges. This increase allows the 40% of US college students in non-quantitative majors a better chance to think critically about numbers in the news. With empirical data on statistics in the media, with a distinct content and with new forms of assessment, statistical literacy is rapidly emerging as a new course in the liberal arts.