OFFERING STAT 102: SOCIAL STATISTICS FOR DECISION MAKERS

Milo Schield
Augsburg College, Minneapolis Minnesota USA
Schield@Augsburg.edu

Statistical educators should support offering three introductory statistics courses: STAT 100 (Statistical Literacy for non-quantitative majors), STAT 101 (Traditional inferential statistics) and STAT 102 (Social statistics for decision makers). The support for the STAT 102 claim includes the needs of most students taking introductory statistics, the different kinds of decisions being made, the growing importance of big data, the limited amount of free time in the current STAT 101 course, the 2016 update to the GAISE guidelines, the importance of confounding in influencing statistical associations and the ability of confounding to influence statistical significance. This paper provides student-tested ways of showing and explaining confounding, statistical significance and the influence of confounding on statistical significance. Indeed 100% of the IASE respondents agreed that students should be shown how confounding can influence statistical significance, 84% agreed that failure to illustrate this confounder-significance connection constituted "professional negligence" and 69% agreed that statistical educators should support offering STAT 102 along with STAT 100 and STAT 101. In a separate survey of the Augsburg students taking this STAT 102 type course, 61% agreed or strongly agreed that a STAT 102 course should be required by all students for graduation. With most of these statistical educators supporting the existence of a STAT 102 course and most Augsburg students seeing significant value in such a course, the door is now open for a new generation of courses, textbooks and teachers.

INTRODUCTION

The theme of the 2016 IASE Roundtable Conference in Berlin was "Promoting understanding of statistics about society." "The Roundtable aims to advance current knowledge about ways to improve the understanding of data and statistics related to key social phenomena (such as trends in migration, employment, equality, demographic change, crime, poverty, access to services, energy, education, human rights, and others). Understanding of such issues is essential for civic engagement in modern societies, but involves statistics that often are open, official, multivariate in nature, and/or dynamic, which are usually not at the core of regular statistics instruction. The conference goal is to contribute to the development of conceptual frameworks, teaching methods, technology solutions, and curricular materials (especially for learners at the tertiary/college and high-school/secondary levels) that can support and promote learning and understanding of statistics about such social phenomena."

This paper is based on a workshop presentation at that conference. It argues that (1) most of our students need STAT 102: an introductory statistics course that focuses on social statistics, (2) such a course must focus on observational studies, multivariate thinking and confounding, (3) students see significant value in such a course, and (4) statistical educators support offering this kind of course alongside STAT 101 (inferential statistics) and STAT 100 (Statistical Literacy).

The phrase, social statistics, may seem ill-chosen: lacking in essence. Social statistics certainly includes those statistics involving humans in a social environment, but it could include those statistics of interest to humans. But if social statistics are primarily observational (hence multivariate and subject to confounding) then observational statistics might be a more useful phrase. But the distinction between observational and experimental is a technical distinction. Aren't all statistics observational?

Multivariate (multivariable) statistics is also a useful description, but multivariate is a technical term and multivariate statistics better describes the second course in statistics than an introductory course. Confounded (confounding) statistics really gets to the core of a STAT 102 course, but very few students have ever used or even heard those words. Statistical educators seem much more comfortable with multivariable statistics than they do with confounding. Correlated statistics is also a possibility, but the STAT 101 research course also deals with correlated statistics.

Social statistics are the opposite of research statistics in three ways. Social statistics connote everyday statistics such as those tabulated by official agencies; research statistics connote the

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),
July 2016, Berlin, Germany. ©2016 ISI/IASE iase-web.org/Conference_Proceedings.php

opposite. Social statistics connote being uncontrolled and field-based (statistics involving humans in their natural setting); research statistics connote being obtained in controlled laboratory settings. Social statistics are typically the basis for social decisions; research statistics are the basis for scientific decisions. For these reasons, social statistics is a useful name – but not necessarily the best – to serve as an appropriate antonym to research statistics and to describe STAT 102.

Given these features, one could start by examining those teaching techniques needed to explain these features. But if the overall goal is to increase students' appreciation for the value of statistics, we should start with our students: their interests, aptitudes and attitudes.

Appendix A presents some data on these matters. To summarize, most students taking introductory statistics are in majors that focus primarily on social statistics obtained from observational studies. The students in these majors typically have lower math skills and tend to have difficulty appreciating the value obtained in taking an inference-based statistics course.

A better way to classify student interests is by how they will use statistics after graduation.

NEED FOR A NEW COURSE

College students can be classified into three groups based on how they will use statistics after graduation.

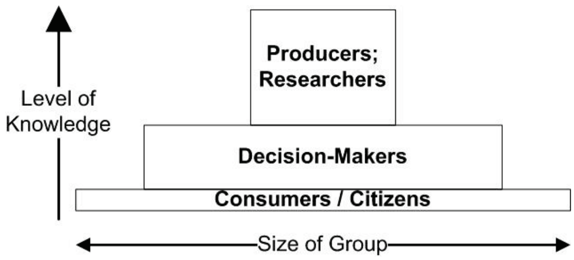


Figure 1 Student Needs for Statistics

Those students who will do research and will produce statistics require the most statistical knowledge. Those who will use only everyday statistics to make personal decisions need the least statistical knowledge. Those who will use statistics to make social decisions or to inform social decision-makers are in-between.

All of the students required to take introductory statistics are expected to be in the top two groups. Most of these are expected to be (or to inform) decision makers. Given the importance of this goal, statistical educators should drop the idea of "the" introductory statistics course.

Statistical educators should support offering three courses to match these different needs:

- STAT 100: Statistical Literacy for students in non-quantitative majors,
- STAT 101: Research Statistics for students who plan to conduct research, and
- STAT 102: Social Statistics for Decision Makers for students in quantitative majors.

Statisticians have always argued that statistics should be used in making good decisions. Offering three courses recognizes three different kinds of decisions: every day personal decisions (STAT 100), critical scientific decisions (STAT 101) and socially-important business and policy decisions (STAT 102).

A core reason for offering another course (rather than including new material in STAT 101) is the lack of free time. When asked how much free time was available in their course, 20% of IASE respondents said none, 35% said 2% or less while 70% said 5% or less. Schield (2016b)

The difference in needs leads to differences in these courses. STAT 101 is inference-based while STAT 100 and 102 are more correlation-based. All three courses deal with variation. In STAT 101, the variation is random: random selection or random assignment. In STAT 100 and 102, the variation tends to be systematic and often influenced by confounding.

TEACHING CONFOUNDING

Confounding has been conspicuously absent from introductory statistics. Confounding was not even on the McKenzie (2004) survey of 30 "possible core concepts" in statistical education. Confounding is not even listed in the index of many of today's introductory statistics textbooks.

Confounding involves multivariate statistics; all too-many introductory statistics textbooks present just univariate and bivariate statistics.

Recently there have been signs of change. Tintle et al (2013) noted that "multivariable methods dominate modern statistical practice but are rarely seen in the introductory course. Instead these methods have been, traditionally, relegated to second courses in statistics for students with a background in calculus and linear algebra." They noted that confounding and variation are "two substantial hindrances to drawing conclusions from data" but viewed this as second course material.

Statistical education has a long history of ignoring any calls for change. (Schield 2013) But this year – 2016 – may mark a sea change.

2016 UPDATE TO THE US ASA GAISE GUIDELINES

In 2005, the American Statistical Association endorsed the first Guidelines for Assessment and Instruction in Statistics Education at the college level (GAISE 2005).

The 2016 update to the US GAISE guidelines is arguably the biggest change in statistics education in decades. (GAISE, 2016). It includes two new areas of emphasis. (1) “Teach statistics as an investigative process of problem-solving and decision making. Statistics is a problem-solving and decision-making process, not a collection of formulas and methods.” (2) “Give students experience in multivariable thinking. The world is a tangle of complex problems with inter-related factors. Let's show students how to explore relationships among many variables.”

This update argued that statistical educators should help students “identify observational studies” and understand “the possible impact of ... confounding". Specifically, statistical educators should “teach multivariate thinking ‘in stages’” and use "simple approaches (such as stratification).” “Multivariable thinking is critical to make sense of the observational data around us. The real world is complex and can’t be described well by one or two variables.”

Appendix B summarizes ways that these guidelines recommend showing students how confounding influences statistical associations. Note that this report does not call for a new course (STAT 102), but this report certainly support teaching confounding. Appendix C summarizes ways that not only show but also explain how confounding influences statistical associations.

In summary, selection, stratification and standardization are three basic techniques for showing, explaining and controlling for confounding. These techniques are fairly easy to demonstrate. The hardest part is for students to think hypothetically about what could confound an observed association. This may be the first time they have ever had to think hypothetically about what could have – or should have – been taken into account. (Isaacson, 2005).

How long will all this take? Showing multivariable thinking could be done in one or two hours. Using tables and graphs that explain confounding could take three or four hours. Identifying what study designs block what kinds of confounders takes another two-to three hours. But getting students to think hypothetically about a given association with a given study design requires repeated exposure week after week. There is nothing in the student's prior math exposure that prepares them for this hypothetical thinking in various contexts. This takes a lot of time; that is another reason why a separate course, STAT 102, is required.

WHY TEACHERS MAY CHOSE NOT TO TEACH CONFOUNDING

There are several reasons why teachers may choose not to teach confounding. (1) Teaching confounding may require statistical educators to become subject-matter experts. That is outside our domain expertise and should be avoided. (2) Teaching confounding certainly means less time for teaching statistical inference. (3) Teaching confounding requires more time for hypothetical (inductive) thinking. (4) Teaching confounding may support using observationally-based associations as evidence for causal connections. In the early 1900s, statisticians got burned on that kind of thinking. Karl Pearson was adamant when he said, causation is “a fetish amidst the inscrutable arcana of modern science.” (5) But a common reason raised among attendees was this: Teaching confounding may result in students having less trust in statistics. Do we want to train our students to be statistical cynics? Unless statistical educators can feel comfortable on this issue, they have excellent reasons not to spend much time on observational studies or confounding.

HOW TEACHERS CAN TEACH CONFOUNDING WITHOUT CREATING CYNICS

Statistical educators need to revisit the historic exchange between Cornfield and Fisher in the 1950s. US statisticians wanted to support the claim that smoking 'caused' lung cancer (smoking caused smokers to have a high chance of lung cancer). Fisher, a smoker, wrote a short article reminding statisticians that in observational studies association is not causation! Fisher also provided data from a German twins study that showed an association between the degree of twinship (identical vs. fraternal) and smoking preference.

When the world's foremost statistician provides data showing that genetics could confound the observed relationship between smoking and lung cancer, most statisticians would quietly "walk away". But Jerome Cornfield (Cornfield, 2015), the creator of the odds ratio and relative risk, drafted the reply. In the appendix to his reply, he proved that in order to nullify or reverse the observed association, the relation between the confounder and the outcome must be greater than that between the predictor and the outcome. The relationship between the predictor and outcome (between smoking and lung cancer was around a factor of 10. The relationship between Fisher's confounder (twinship) and the outcome was less than a factor of three. Fisher never replied.

Schield (1999) reviewed this debate in detail and included a copy of Cornfield's derivation. Schield asserted that "Cornfield's minimum effect size is as important to observational studies as is the use of randomized assignment to experimental studies."

Appendix D presents various ways of presenting this material. The main take-away is this: the larger the predictor effect size, the less likely there are confounders that can nullify or reverse the observed association. In observational studies, a small effect size is like a confounder magnet. A decision-maker may not be a subject-matter expert, they may not know what confounders are plausible in a given observational study, but they have good reason to conclude that associations with small effect sizes give only weak support for a causal connection.

TEACHING RANDOMNESS

If STAT 102 is going to focus primarily on confounding in large samples from observational studies, then less time is available for teaching about randomness. This is true. But randomness takes new forms. As the size of the data set increases, coincidence becomes increasingly likely – indeed it is to be expected. Appendix E presents this topic and introduces the Law of Very Large Numbers as a simple way to understand it.

Statistical decision making is an important topic – even with big data. A/B testing is common in web-based venues. Appendix F introduces a somewhat controversial Bayesian suggestion for teaching statistical decision making. Regardless of whether one agrees with this approach, teachers should be aware of the recent commentary by Gigerenzer and Marewski (2014) and the recent ASA statement on p-values (ASA 2016).. Statistical educators should consider saying that the traditional approach for statistical decision making holds only in those cases where the alternate is more likely to be true than is the null.

Appendix G presents some statistical shortcuts for statistical significance. Although these are not necessary conditions, they introduce students quickly and easily to some of the big ideas in statistics.

Appendix H presents the use of non-overlapping confidence intervals to show statistical significance and to show the influence of confounding on statistical significance. Showing the effect of confounding on statistical significance links two important topics. Statistical educators should show students how a statistically-significant association can become statistically-insignificant after controlling for a confounder and vice versa. Failing to show students how a statistically-significant association can be made insignificant (or vice versa) is professional negligence.

SURVEY RESULTS

By design this presentation involved a large number of contentious assertions. IASE Roundtable statistical educators attending Schield's workshop were asked to complete a survey on their conclusions. Details are in Appendix I. Here are the statements with which the attendees most strongly agreed and disagreed:

A unanimous 100% of respondents agreed or strongly agreed that "students should see how control for confounder can influence statistical significance" (Q16).

Most respondents (69%) agreed that "statistical educators should support offering three courses: STAT 100, STAT 101 and STAT 102." (Q18) Prior to this roundtable, less than 30% were expected to agree.

Most (76%) of respondents agreed that "we should use multivariate regression without assumptions or diagnostics." (Q10). But only 29% agreed that "we should use multivariate regression only when significant; skip diagnostics." (Q11)

Respondents expressed the greatest disagreement in the following areas. Only 52% agreed that context – not variability – is what distinguishes statistics from mathematics (Q10). Only 53% agreed that "Frequentists should use Bayesian thinking to justify the normal decision-making rules." (Q14). Only 61% agreed that "Frequentists should not support decision making based solely on statistical significance." (Q13).

Of the six people who disagreed with the statement that "statistical educators should support offering three courses: STAT 100, STAT 101 and STAT 102." (Q18), half said either drop STAT 100 or combine 100 and 102. The other three mentioned "not enough faculty", "everyone needs Stat 101" and "Don't separate the concepts (into separate courses)."

Combining STAT 100 and STAT 102 is a most interesting idea. Having taught both, combining them makes a lot of sense. The difference between STAT 101 and either of these is much greater than the difference between 100 and 102. This idea is definitely worth exploring.

STUDENT FEEDBACK ON STAT 102

Augsburg College has offered all three of these courses. A STAT 100 course since 1998 and a STAT 102 course since 2015. Schield (2016) summarized student feedback on Augsburg's STAT 102 equivalent. Of the 105 business majors taking this class in the last two years, 87% said the course helped them read and interpret everyday statistics; 79% said this course improved their critical thinking skills, 86% said they would recommend this course to a friend, while 61% agreed that this course should be required by all college students for graduation.

This last statistic is the most telling. Most students (61%) see significant value in a STAT 102 course by the time they finish the course.

CONCLUSION

Of the attendees who had an opinion, 83% agreed with the 2016 GAISE update in saying that students should be introduced to multivariate thinking. But these attendees went beyond the 2016 GAISE update. Most (64%) agreed that statistical educators should offer a STAT 102 course alongside STAT 100 (statistical literacy) and STAT 101 (traditional research statistics); 81% agreed that not showing confounder influence on statistical significance is negligence while 95% agreed that we should show how controlling for confounding influences statistical significance.

Meanwhile, most (61%) of the Augsburg students taking a STAT 102 type course agree that this kind of course should be required by all college students for graduation.

Statistical education has successfully resisted most of the calls for change by the leaders in statistical education. But these two outcomes, teacher willingness and student appreciation, for offering STAT 102 are synergistic. Together they open the door for a new generation of courses, textbooks and teachers. Statistical education may finally measure up to the MacNaughton (2004) goal for an introductory statistics course: “to give students a lasting appreciation for the vital role of statistics."

ACKNOWLEDGEMENTS
Thanks to Joachim Engel and Marc Isaacson for commenting on a draft of this paper. Thanks to the W. M. Keck Foundation and Mercedes Tally for supporting these ideas in 2001-2005.

REFERENCES
Broers, Nick (2006). Learning Goals: The Primacy of Statistical Knowledge. ICOTS-7. Copy at www.stat.auckland.ac.nz/~iase/publications/17/6G2_BROE.pdf

College Board (2014). SAT® Percentile Ranks for Males, Females, and Total Group: 2014 College-Bound Seniors — Critical Reading + Mathematics. https://secure-media.collegeboard.org/digitalServices/pdf/sat/sat-percentile-ranks-composite-crit-reading-math-2014.pdf

Cornfield, Jerome (2015). Jerome Cornfield page at www.StatLit.org/Cornfield.htm

GAISE (2005). Guidelines for Assessment and Instruction in Statistical Education. ASA. Copy at www.amstat.org/asa/files/pdfs/GAISE/2005GaiseCollege_Full.pdf

GAISE (2016). Guidelines for Assessment and Instruction in Statistical Education. American Statistical Association. Copy at www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf

Gigerenzer, Gerd and Julian Marewski (2015). Surrogate Science: The Idol of a Universal Method for Scientific Inference. *Journal of Management*, Vol. 41, No. 2, February 2015, 421–440. Copy at http://www.dcscience.net/Gigerenzer-Journal-of-Management-2015.pdf

Isaacson, Marc (2005). Statistical Literacy – An Online Course at Capella University. ASA Proceedings of the Section on Statistical Education. P. 2244-2252. Copy at www.statlit.org/pdf/2005IsaacsonASA.pdf

MacNaughton, Donald (2004). The Introductory Statistics Course: The Entity Property Approach. Posted at www.MatStat.com/teach

McKenzie, John, Jr. (2004). Conveying the Core Concepts. ASA Proceedings of the Section on Statistical Education. P. 2755-2757. Copy at www.statlit.org/pdf/2004McKenzieASA.pdf

Nuzzo, Regina (2014). Scientific method: Statistical errors. *Nature* magazine. Vol 506, Issue 7487. 12 February. Copy at www.nature.com/news/scientific-method-statistical-errors-1.14700

Schield, Milo (1999). Simpson's Paradox and Cornfield's Conditions, *ASA Proceedings of the Section on Statistical Education*, pp. 106-111. See www.StatLit.org/pdf/1999SchieldASA.pdf.

Schield, Milo (2004). Statistical Literacy Curriculum Design. *IASE Curricular Development in Statistics*. IASE Roundtable, Sweden. Copy at www.statlit.org/pdf/2004SchieldIASE.pdf

Schield, Milo (2006). Presenting Confounding Graphically Using Standardization. ASA *Stats* Magazine. Copy at www.statlit.org/pdf/2006SchieldSTATS.pdf.

Schield, Milo (2012). Coincidence in Runs and Clusters. Mathematical Association of America JMM Copy at www.statlit.org/pdf/2012Schield-MAA.pdf

Schield, Milo (2013). Statistics Education: Steadfast or Stubborn? *ASA Proceedings of the Section on Statistical Education*. P. 1005-1017. See www.StatLit.org/pdf/2013-Schield-ASA.pdf

Schield, Milo (2015). Statistical Inference for Managers. *ASA Proceedings of the Section on Statistical Education*. P. 3032-3046. Copy at www.StatLit.org/pdf/2015-Schield-ASA.pdf

Schield, Milo (2016a). IASE Survey Results. www.StatLit.org/pdf/2016-Schield-IASE-Survey.pdf

Schield, Milo (2016b). Augsburg Student Evaluations of STAT 102. *ASA Proceedings of the Section on Statistical Education*. [In press] Copy: www.StatLit.org/pdf/2016-Schield-ASA.pdf

Sharpe, Norean, Richard DeVeaux and Paul Velleman (2014). *Business Statistics.* Pearson Pub.

Tan (2012): www.statlit.org/pdf/2012-Arjun-Tan-Simpsons-Paradox.pdf

Tintle, Nathan, Beth Chance, George Cobb, Allan Rossman, Soma Roy, Todd Swanson and Jill VanderStoep (2013). Challenging the state of the art in post-introductory statistics. 59th World Statistics Congress (ISI).. Copy at www.statistics.gov.hk/wsc/IPS032-P1-S.pdf

Tintle, Nathan, Beth Chance, George Cobb, Allan Rossman, Soma Roy, Todd Swanson and Jill VanderStoep (2016). *Introduction to Statistical Investigations*. Wiley.

Utts, Jessica (1995). An Assessment of the evidence for psychic functioning. *Journal of Parapsychology*, 59, 4, 289-320. Copy at https://www.ics.uci.edu/~jutts/air.pdf

Utts, Jessica (2014). Seeing Through Statistics. Fourth edition. Wadsworth Publishing.

Wainer, Howard (2002). The BK-Plot: Making Simpsons' Paradox Clear to the Masses. *Chance* 2002 15(3). Copy at www.statlit.org/CP/Cornfield/2002-Wainer-Visual-Revelation-BK-plot.pdf

APPENDIX A: STUDENT APTITUDES, INTERESTS AND ATTITUDES

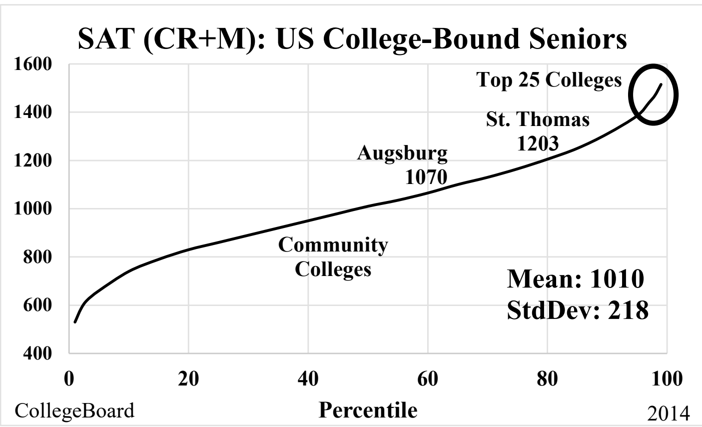
College students taking statistics have a wide variety of interests. These are reflected in their choice of major. Of the 1.6 million US students graduating with a BA/BS in 2009, 57% (910,605) graduated in majors that typically require statistics (list below). If all the students taking introductory statistics were in these majors and if all students in these majors took introductory statistics, they would be distributed as follows: (US Statistical Abstract, 2012, Table 302):

Table 1: Distribution of Majors in Stat 101

%	Major
38%	Business or Economics
19%	Social Science or History
13%	Health
10%	Psychology
9%	Engineering
9%	Biological Science
2%	Math or Statistics
100%	All students in these majors

Based on their choice of majors, those most interested in social statistics are in the top four groups (an estimated 80% of those taking introductory statistics). This high level of interest provides strong support for offering a separate course focusing on social statistics.

The College Board (2014) provided the mean and standard deviation of SAT scores (Critical Reading and Math) among college-bound students. The graph in Figure 2 was created assuming this distribution was normal. Colleges select students based on their SAT scores. Many of the faculty interested in statistics education tend to teach at the more selective schools. Their students are more likely to go on to graduate school and research. Thus, their choice of topics may not be that useful to faculty teaching at the less selective schools.



SAT MATH	PERCENTILE	MAJOR
613	80%	Math/Stats
585	72%	Physical Sciences
579	70%	Engineering
554	62%	Comp. Science
551	61%	Biological
550	61%	Social Sciences
522	51%	Business
522	51%	English Lang/Lit
506	46%	History
498	43%	Communication
489	40%	Psychology
482	38%	Education

Business Insider (2014), College Board (2015)

Figure 2 Student Aptitudes by College and by Major

As shown in the right-side table, those students most interested in social statistics (education, psychology, communication, business and the social sciences) rank lower in their math aptitude while those least interested in social statistics (biological sciences, computer science, engineering, physical sciences and math/stats) rank higher in their math aptitude. Business Insider (2014) Again, those faculty interested in statistics education tend to be in math/stat departments. They may be unaware of the differences in math aptitudes among those taking introductory statistics.

Many – if not most – students taking introductory statistics see less value after taking the course than before they started. (Schield, 2008). This may explain why student loose almost half of the course gain within fourth months after the course. (Tintle et al, 2013) Statistical educators may attribute students' negative views of mathematics to poor teaching by unqualified teachers when it may be due to the teachers' insistence on a higher level of mathematical abstraction and formalism than is needed to present the key concepts to those with lower levels of math aptitude.

APPENDIX B: 2016 UPDATE TO THE US ASA GAISE GUIDELINES

Appendix B of this GAISE (2016) report presented three approaches to showing confounding. The first used Ekisograms: charts that show probabilities as areas. My students have difficulty decoding these area diagrams.

The second technique for showing confounding involves the use of an X-Y plot.

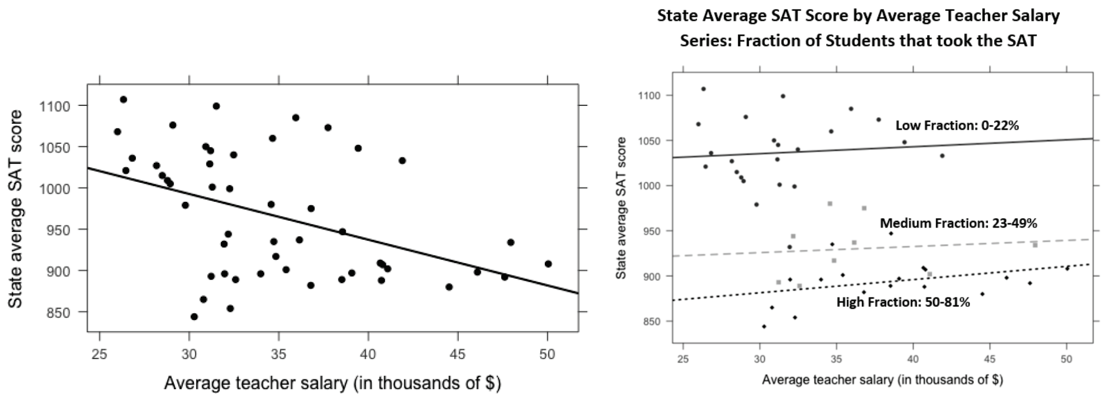


Figure 3 Using XY Plots to Show Confounding

As teacher salary increase, the average state SAT score decreases. This association supports the idea that cutting teacher salaries could increase student SAT scores. But when the data is broken into separate series based on the percentage of students taking the SAT in each state, the negative association is nullified or reversed. Unfortunately this example involves a complex confounder. One must understand the relation between geography, incomes and SAT test taking.

The third technique for showing confounding involves multivariable regression output. The output in Figure 4 models the Women's record time (in seconds) on various Scottish hill climbs. The linear model (left side of Figure 4) calculates the association between the time for the female winner to complete the race and the Climb (the vertical distance in meters) as 1.755 seconds per meter.

```
Response variable is:   Women's Record
R squared = 85.2%     R squared (adjusted) = 84.9%
s = 1126  with  70-2 = 68  degrees of freedom

Variable       Coefficient     SE(Coeff)    t-ratio    P-value
Intercept      320.528         222.2         1.44       0.1537
Climb            1.755           0.088      19.8      < 0.0001
```

```
Response variable is:   Women's Record
R squared = 97.5%     R squared (adjusted) = 97.4%
s = 468.0  with  70 - 3 = 67  degrees of freedom
Variable     Coefficient  SE(Coeff)    t-ratio   P-value
Intercept     -497.656     102.8        -4.84    < 0.0001
Distance       387.628      21.45       18.1     < 0.0001
Climb            0.852       0.0621     13.7     < 0.0001
```

Figure 4 Using Regression Output to Show Confounding

But if one takes into account the length of the race (Distance in kilometers), then the linear association between the female winner's time and the Climb is becomes 0,852 seconds per meter: a reduction of 51%. See the right side of Figure 4. Moreover, the percentage of variation explained increases from 85% to 97%. What one controls for can influence an observed association.

The 2016 update to the GAISE guidelines made these closing thoughts:

> "Multivariable thinking is critical to make sense of the observational data around us. This type of thinking might be introduced in stages
> 1. learn to identify observational studies
> 2. explain why randomized assignment … improves things
> 3. learn to be wary of cause-effect conclusions from observational data
> 4. learn to consider potential confounding factors and explains why they might be confounding factors, and
> 5. use simple approaches (such as stratification) to show confounding."

"The real world is complex and can't be described well by one or two variables. If students do not have exposure to simple tools for disentangling complex relationships, they may dismiss statistics as an old-school discipline only suitable for small sample inference of randomized studies." "This report recommends that students be introduced to multivariable thinking, preferably early in the introductory course and not as an afterthought at the end of the course."

APPENDIX C: OTHER WAYS TO SHOW OR EXPLAIN CONFOUNDING

Selection is arguably the simplest way to see confounding. Just select on a given value of the confounder. For example, suppose that the patient death rate is higher at a city hospital than at a rural hospital. Patient condition is a plausible confounder. By selecting just those patients in very poor condition, it may be that the observed association is reversed and the patient death rate is higher at the rural hospital than at the city hospital.

Stratification is also an easy way to show confounding. Consider the death rates of patients shown in Table 2. Overall, death is more likely among non-smokers (31%) than among smokers (24%). But when age is taken into account, the reverse is true. Among the young, death is more likely among (18%) than among non-smokers (12%); among the old, death is more likely among smokers (88%) than among non-smokers (86%).

By stratifying these death rates on these age groups, students can see how age confounds the original association between smoking status and overall death rate.

Table 2: Using Death Rates Tables to show Confounding

DIED	YOUNG	OLD	TOTAL
NON SMOKER	12%	86%	31%
SMOKER	18%	88%	24%
TOTAL	15%	86%	28%

Even if each of these approaches 'shows' the influence of confounding, they may not help students 'understand' confounding. And if these presentations don't allow students to work problems with numerical answers (so they won't be on the final exam), teachers and students may well agree to spend just a modest amount of time on confounding in the last class period before the final exam. Here are four ways that "explain" confounding – that help students 'understand' confounding.

#1: One way to explain confounding is model based. In year one, 10% of the students are disadvantaged and average 80%. The other students average 90% for a class average of 89%.

	Year 1		Year 2		
	Number	Score	#	Score	
Disavantaged	10	80%	50	81%	⇧
Advantaged	90	90%	50	91%	⇧
TOTAL	100	89%	100	86%	⇩

Table 3: Modeling Outcomes Before/After Change in order to Explain Confounding

After Year 1, other disadvantaged student discover this teacher helps them get higher scores, so they switch to this teacher increasing their prevalence from 10% to 50%. The disadvantaged students average 81% while the rest average 91% for a class average of 86%. Average scores for both groups increased, but the overall class average decreased. Students can readily see why: "It's the mix". Students need a simple explanation to make sense of this apparent paradox.

#2: A second way to explain the confounding of an association of totals involves ratio standardization. Prison expenses in California are 50% higher than those in New York. But when we calculate these expenses as ratios per inmate, prison expense per inmate is 40% lower in California than in New York.

#3: A third way to explain confounding involving the graphical standardization of ratios. It was publicized by Wainer (2002). This graphical technique requires that the predictor and confounder both be binary. In the left side of Figure 5, the average death rate of patients was higher for City hospital (5.5%) than for Rural hospital (3.5%). But notice the big difference in patient condition. 90% of City patients were in poor condition (30% of Rural patients). Notice that patients in poor condition are much more likely to die than those in good condition.

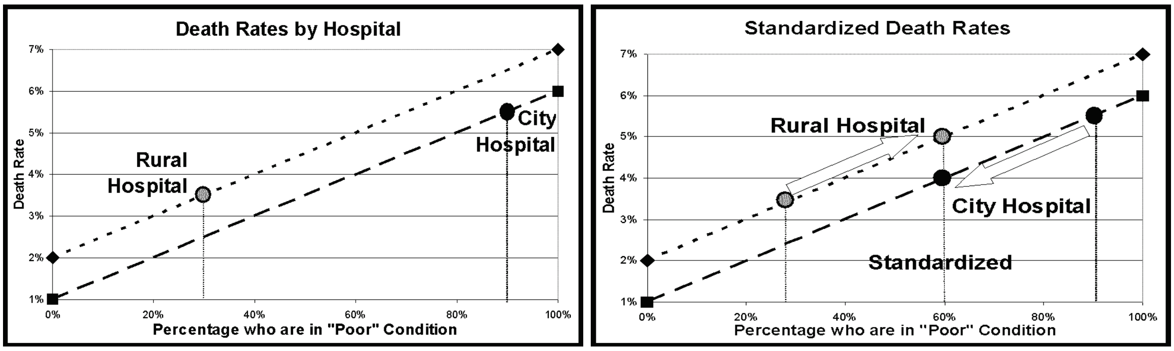


Figure 5 Using Wainer Plots to Show Confounding

In the right side of Figure 5, the hospital averages are standardized by giving both hospitals the same mix of patients. In this case, the association is reversed – a clear example of Simpson's paradox. Again, students can readily see that the difference in mix confounded the initial association. Peter Holmes said that seeing this graph was the first time he "really understood" confounding. Music and art majors find this graph easy to read. They can easily work problems with numerical answers. (Schield 2004 and 2006) For the origin and details of this diagram, see Tan (2012).

#4: A fourth way to explain confounding of an association of ratios involves algebraic standardization. Consider the same rate table that was used to 'show' confounding. Each of the rows is an average – not a sum. But these are weighted averages. Each row average can be shown as the answer to single equation with one unknown.

Students should be able to determine that 74% of top row are young, that 91% of Row 2 are young, and that 82% of Row 3 are young. Realizing that the different rows (groups) have different mixtures of young and old, and that old are much more likely to die than young, students should realize these group averages form a mixed-fruit comparison: comparing apples and oranges.

Repeat of Table 2: Patient Death Rates by Smoking Status and Age

DIED	YOUNG	OLD	TOTAL
NON SMOKER	12%	86%	31%
SMOKER	18%	88%	24%
TOTAL	15%	86%	28%

To avoid this mixed-fruit comparison, the groups need to be standardized. The top two rows should be standardized with 82% of each row as young as calculated for the total row. This gives the following: the non-smoker standardized death rate is 25% = 0.82*12 + 0.18*86. The smoker standardized death rate is 31% = 0.82*18 + 0.18*88. The standardized death rate for smokers (31%) is greater than that for non-smokers (25%). Standardizing on age reversed the observed association.

APPENDIX D: THE CORNFIELD CONDITIONS

To prevent students from becoming cynics after showing them how easily a confounder can influence an association, students need exposure to the Cornfield conditions. (Schield, 1999)

The Cornfield conditions establish minimum effect sizes that are needed to nullify or reverse an observed association. The larger the observed predictor-outcome effect size, the larger the confounder effect size must be. Students can quickly see this in a two-way rate table.

Table 4: Patient Death Rates by Hospital and Patient Condition

Patients	Rural	City	ALL
Good	2.0%	1.0%	1.875%
Poor	7.0%	6.0%	6.250%
ALL	3.5%	5.5%	4.5%

Patients at City hospital are 60% (2 percentage points) more likely to die than those at Rural hospital. Patients in Poor condition are 230% (4.35 percentage points) more likely to die than are those in Good condition. The association between the confounder (patient condition) and the outcome (death) is greater than that between the predictor (hospital) and the outcome (death). Thus taking into account patient condition has the ability – the potential – to nullify or reverse the observed association between hospital and death. Figure 6 shows this visually using rounded values:

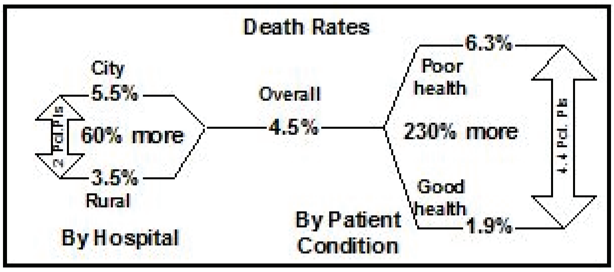


Figure 6 Showing the predictor-outcome and confounder-outcome associations

Cornfield identified a necessary condition: the association between the predictor and the confounder (highest value) must exceed the association between the predictor and the outcome.

This association is tricky since the association between the predictor and the confounder does not involve the outcome. The common part must be the confounder value that has the highest outcome rate. In this case, that is patients in poor condition.

In Table 4, this association can be obtained from the margin values which are weighted averages. First consider the rows. In the top row, the percentage of patient in good condition who are rural is 87.5%: 100%*(1.875-1)/(2-1). In the middle row, the percentage of patients in poor condition who are rural is 25%: 100%*[(6.25-6)/7-6)]. Note that the rural hospital is the common part for these two percentages.

But this Cornfield condition must have patients in poor condition as the common part. In this table, that involves the columns – not the rows. The values in the bottom row are averages for their respective columns. Thus, the percentage of City patients who are in poor condition is 90%: 100%*[(5.5-1)/(6-1)], while the percentage of rural patients who are in poor condition is 30%: 100%*[(3.5-2)/(7-2)].

Thus as a difference, the effect size of the predictor-confounder relationship is 60 percentage points: (90% - 30%). As a percentage difference, the effect size is 200% more: 100% * (90% - 30%)/30%.

Figure 7 illustrates these associations shown as a simple difference (percentage points) on the left side and as a percentage difference on the right side.

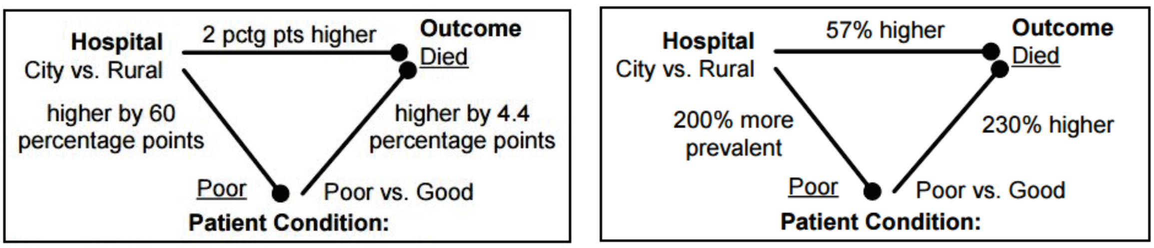


Figure 7 Showing all three associations between predictor, confounder and outcome

To repeat, An observationally-based association can be nullified or reversed only if the confounder (patient condition) has a stronger association with the outcome (death) than does the predictor (hospital), AND if the predictor (hospital) has a stronger association with the confounder (patient in poor condition) than with the outcome (death).

Note that all the analysis can be done without any reference to the underlying counts. While mathematicians may want to see the underlying counts, all too often the data presented to decision makers is already summarized by rates and percentages.

Being able to evaluate and compare effect sizes without needing access to the underlying counts is of great value to decision makers.

APPENDIX E: TEACHING COINCIDENCE

Coincidence is seldom – if ever – covered in introductory statistics. But as the size of the data increases, coincidence becomes increasingly likely – indeed it is to be expected.

A simple way to explain this involves the law of Very Large Numbers. This law has two forms. Qualitative: The unlikely becomes all but certain as the number of tries increases. Quantitative: If the chance of a rare event is 1/N, then one such event is expected – at least one such event is more likely than not – in N tries.

If the chance of heads if 50%, then the chance of 10 heads in a row is one chance in 1,024. In 1,024 flips of a coin, one sequence of 10 heads is expected – at least one sequence is more likely than not.

This approach helps students understand von Mises Birthday problem. If the chance of a match in birth month and day is one in 365, then at least one match is expected if there are more than 365 ways to pair the N participants. Schield (2012) showed graphically that with 28 people there are more than 365 ways of picking two people at a time. Showing students visually that there are more than 365 possible pairs is more helpful for some than proving it mathematically.

APPENDIX F: TEACHING STATISTICAL DECISION MAKING

Telling users they should reject the null and accept the alternate for a statistically-significant outcome is NOT justified by Frequentist theory. First, it totally ignores the context – the likelihood that the research hypothesis is true. Second, frequentist can only say the outcome is extremely unlikely OR the alternate hypothesis is true. There is nothing in frequentist theory by which rejecting the null can be deduced. As Frequentists, statistical educators should never allow statistical significance to be sufficient for decision-making in any statistics courses.

Recognizing this, some statistical educators recommend shifting from statistical significance to p-values and requiring that subject-matter experts decide whether to reject or fail to reject the null. But shifting from statistical significance to p-values as a way of avoiding the decision-making transforms statistics from decision making to bean-counting. Statistical educators may "win the battle but lose the war."

Another variation, is to connect ranges of z-scores or p-values with ordinal labels on the strength of evidence against the Null. The information in Table 5 was obtained from Tintle et al, (2015).

Table 5: Z-scores, p-values and Evidence against the Null

Z-score (p. 54)	p-value (p. 44)	Evidence against the Null
Between -1.5 & 1.5	0.10 < p-value	Little or no evidence
Below -1.5 or above 1.5	0.05 < p-value < 0.10	Moderate evidence
Below -2 or above 2.	0.10 < p-value < 0.05	Strong evidence
Below -3 or above 3.	p-value < 0.01	Very strong evidence

The authors correctly note that "the smaller the p-value, the stronger the evidence against the null hypothesis (the "by chance alone" explanation)." But this statement is ambiguous. It is true for a given alternate, but that does not make it true when comparing different alternates. The labels provided by Tintle et al (2015), seem to imply that they are intrinsically true for all alternates.

Their "intrinsic" approach seems to be an example of what Gigerenzer and Marewski (2015) called "the idol of a universal method for scientific inference." The "intrinsic" approach in Table 5 also seems to be in conflict with the spirit of the ASA Statement on Statistical Significance and P-Values (2016) which states:

> **3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.**
> Practices that reduce data analysis or scientific inference to mechanical "bright-line" rules (such as "$p < 0.05$") for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making. A conclusion does not immediately become "true" on one side of the divide and "false" on the other. Researchers should bring many contextual factors into play to derive scientific inferences …"
>
> **4. Other Approaches**
> "In view of the prevalent misuses of and misconceptions concerning p-values, some statisticians prefer to supplement or even replace p-values with other approaches. These include methods that emphasize estimation over testing, such as confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, such as likelihood ratios or Bayes Factors; and other approaches such as decision-theoretic modeling and false discovery rates."

The aforementioned "intrinsic" approach also seems to be in conflict with one of the basic tenants of critical thinking: the stronger (the more outlandish) the claim, the stronger the evidence needed to support that claim. There is nothing in the "intrinsic" approach that involves the nature of the alternate hypothesis. Nuzzo (2014) noted "The more implausible the hypothesis — telepathy, aliens, homeopathy — the greater the chance that an exciting finding is a false alarm."

So what can statistical educators do? Statistical educators should embrace Bayes-light. Schield (1996) showed that under certain assumptions if the alternate (Ha) is more likely to be true than the null (Ho), then a statistically-significant result gives at least a 95% confidence that Ho is False and Ha is true. Nuzzio (2010) using a more robust model of the null and alternate hypothesis found that if the alternate were as likely to be true as the null, then there was a 71% chance that the alternate was true (the null was false).

Both Schield and Nuzzio will be satisfied by saying if the alternate is at least as likely as the null and the outcome is statistically significant, then the alternate is more likely than not to be true (the null is more likely than not to be false), so there is good reason to reject the Null and accept the Alternative. .

In many cases, there are strong reasons for believing the research hypothesis to be true. In such cases the normal rejection policies are applicable. But in those cases where the research hypothesis is unlikely to be true, decision makers must be alerted to take care: a smaller p-value is needed to justify rejecting the null.

Figure 2 in Schield (1996) shows the relationship between p(Alternate) and the p-value needed to be 95% confident when accepting the alternate. A simple necessary condition is obtained by drawing a straight line between the end points. Thus, when p(Alternate) is less than 50%, the 95% p-value must be less than p(Alternate) by a factor of ten: 0.05 / (½).

Perhaps the best example of a most-unlikely research hypothesis was when Utts (1995, page 11) argued that a p-value of 10^{-20} was extremely strong evidence in favor of clairvoyance (remote viewing) But if our prior probability that clairvoyance is real is also 10^{-20}, then this incredibly small p-value is too large to justify accepting this particular alternative.

APPENDIX G: SHORTCUTS FOR STATISTICAL SIGNIFICANCE

Note that two leading statistics textbook (Utts 2015 and Sharpe et al, 2014) have all but skipped the derivation of the sampling distribution.

Statistical educators have long used 1/sqrt(n) to estimate the maximum 95% margin of error. Schield (2015) identified some other shortcuts that provide sufficient conditions for statistical significance as shown in Figure 8.

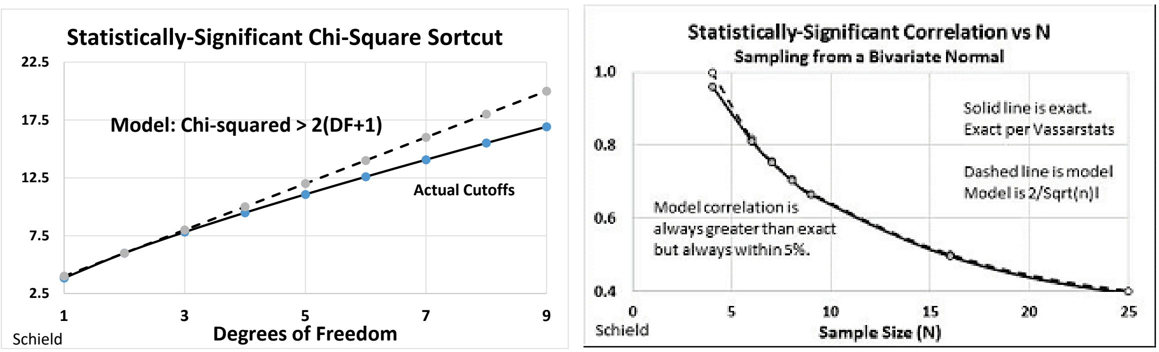


Figure 8 Simple Sufficient Short-Cuts for Statistical Significance

The chi-squared sufficient condition for statistical significance is easy to remember. Chi-squared, used as a test of independence or homogeneity, often involves a 2 by k table. In that case, df = (2-1)*(k-1) = k-1, so 2*(df+1) always equals 2k for k < 10. So if chi-squared is greater than 2k in a 2 by k table, that statistic is statistically-significant.

For Pearson correlation, a sufficient condition for statistical significance is given by 2/sqrt(n) for n < 25. This too is easy to remember. Pearson correlation is sometimes used to compare two time series. See www.tylervigen.com But this test for significance only applies when X and Y values are not determined in advance. This excludes comparing two time series.

APPENDIX H: SHOWING CONFOUNDER INFLUENCE ON STATISTICAL SIGNIFICANCE

Another way to illustrate statistical significance is to note that if two 95% confidence do not overlap, then the difference in their sample means is statistically significant. Students find this approach to be easily understood and remembered. While this condition is certainly sufficient, it is far from necessary. If the decision is important and the confidence intervals barely overlap, then users are recommended to consult with a statistician.

Figure 9 shows how a statistically-significant association can become statistically-insignificant after taking into account a confounder.

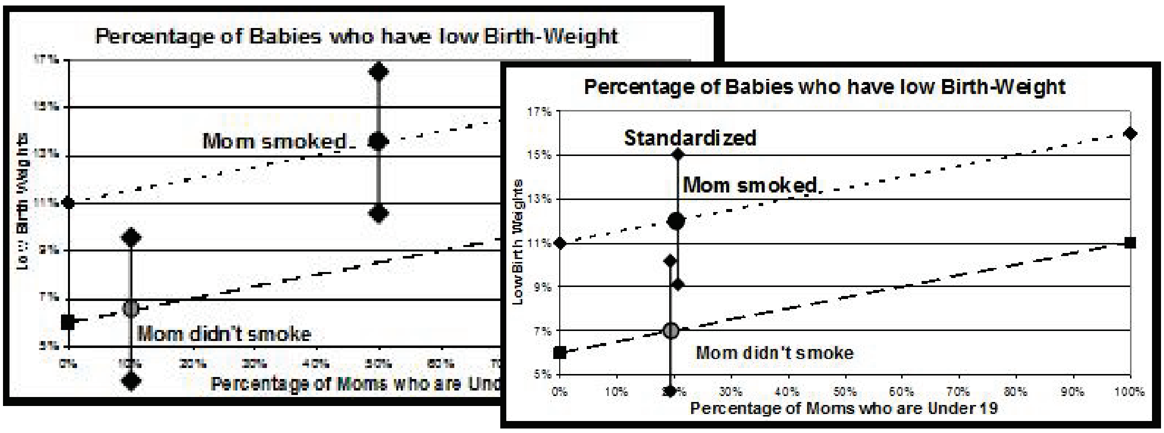


Figure 9: Confounder Influence on Statistical Significance: Before vs. After

Showing how statistical significance can be influenced when controlling for a confounder is arguably the keystone of STAT 102. Students need to see this and understand it so they don't treat statistical significance – or the lack thereof – as something that is constant for a given association, something set in "stone", something that is "intrinsic."

APPENDIX I: ATTENDEE SURVEY RESULTS

The source for this data is at www.StatLit.org/pdf/2016-Schield-IASE-Survey.pdf

1. Omit
2. How much free time is available in the typical intro stat course for new material?
   None (20%); 2% (15%); 5% (35%); 10% (10%); 20% (15%); 30% (5%); 40% or more (0%)

In cumulative terms,
   None (20%); ≤2% (35%); ≤5% (70%); ≤10% (80%); ≤20% (95%); ≤30% (100%)

The following questions all have the same choice of answers:

a. Strongly disagree b. Disagree c. Neutral d. Agree e. Strongly agree
The percentages are of those answering with an opinion.

3. Schield's *presentation was informative:* 0; 13%; 13%; 50%; 25%.
4. *Schield's presentation was accessible (easy to follow):* 0%; 13%; 0; 50%; 38%.
5. Context – not variability – is what distinguishes statistics from math: 16; 20; 12; 32; 20
6. The various influences on statistics should be classified into a few groups. 0; 19; 19; 56; 6
7. Teachers should show how statistics can be influenced by definitions: 9; 9; 9; 55; 18
8. *Educators should introduce students to multivariate thinking.* 0; 4; 9; 26; 61
9. *Educators should use tools that "show confounding."* 0; 5; 0; 55; 41
10. *Use multivariate regression without assumptions or diagnostics.* 0; 6; 18; 41; 35
11. *Use multivariate regression only when significant; skip diagnostics*: 21; 29; 21; 21; 7
12. *Use tools that show AND EXPLAIN why confounding gives unexpected results:* 0, 9; 9;45; 36
13. *Do not support decision making based solely on statistical significance*: 6; 28; 6; 39; 22
14. *Use Bayesian thinking to justify the normal decision-making rules*: 0; 16; 32; 42; 11.
15. *Students should be exposed to Law of Very Large Numbers and coincidences*: 0; 5; 14; 50; 32
16. *Show how control for confounder can influence statistical significance*: 0; 0; 0; 62; 38
17. *Not showing confounder effect on stat. significance is professional negligence*: 0; 5; 11; 68; 16
18. *Statistical educators should support offering* ***three courses****: STAT 100 Statistical Literacy, STAT 101 Traditional Research Statistics & STAT 102 Stats for Decision Makers:* 8; 15; 8; 38; 31

****

OPEN-ENDED ESSAY:

What are the strongest reasons FOR offering a separate course (STAT 102) that focuses on confounding and assembly (how groups and measures are defined, combined, summarized and presented)?

What are the strongest reasons AGAINST offering a separate course (STAT 102) that focuses on confounding and assembly (how groups and measures are defined, combined, summarized and presented)?