LUISS

Economics and business

Mathematics

The Simpson’s paradox

Relator:
Prof. Marco Isopi

Candidate: Lorenzo Stama
Matr. 215891

Thanks to my prof. M. Isopi for giving me this very
interesting paradox to study, proof that mathematics
sounds like an abstract issue, but in fact it is highly concrete.

To my Wonder Womom, who was always
by my side when I needed her. To my grandmother,
who kept preparing snacks for me every day
while I was studying. To Giorgia, Titti, Berardo, Tiziano and Ivana
who were ready to help me if I needed anything. To my
friendly-neighborhood Mr. Wolf, the main reason
why I know English well enough to write this
work, and who was ready to help me
once again with a proofreading.
To Altea, Michele, Pier Giorgio, Federica, Elena
and all my friends who I love so much, and
will have to call me doctor, which will be so funny.

INDICE

INTRODUCTION

Analytics projects often present us with situations in which common sense tells us one thing, while the numbers seem to tell us something quite different.
Such situations are often opportunities to learn something new by taking a deeper look at the data. Failure to perform a sufficiently nuanced analysis, however, can lead to misunderstandings and decision traps.
Starting from a generic definition of paradox**,**we'll come to talk about Simpson Paradox, giving extract of history of paradox in different disciplines to illustrate its danger, so that we can present several instances of Simpson's Paradox in business and non-business environments.
In the last 30 years there has been an exponential increase in the use of technology that has paved the way for statistical modeling to be at the front and center of decision making not just in business, but everywhere. Statistics is the means to interpret data and transform vast amounts of raw data into meaningful information.
As we demonstrate below, statistical tests and analysis can be confounded by a simple misunderstanding of the data. Often taught in elementary probability classes, Simpson's Paradox refers to situations in which a trend or relationship that is observed within multiple groups reverses when the groups are combined However, paradoxes and fallacies lurk behind even elementary statistical exercises, with the important implication that exercises in business analytics can produce deceptive results if not performed properly. This point can be neatly illustrated by pointing to instances of Simpson's Paradox.

Simpson’s Paradox is in a sense an arithmetic trick: weighted averages can lead to reversals of meaningful relationships, i.e., a trend or relationship that is observed within each of several groups reverses when the groups are combined. To show all of this, we’ll see in the first chapter the paradox in general and its evolution throughout time, then we denote the paradox in everyday life and see a little biography about Simpson and Yule, who studied it first.

In the second chapter we will see the Simpson Paradox in detail: it will be explained with the help of some examples of real life with real data.

The first example regards Berkley University, the second one is about two different ways which can be used to treat renal calculi, the third one is about a highly surprising observation in a healthcare study and the last one focus on the link between price and profits.

To all of you, a good reading.

CHAPTER 1

1.1 The “swift-footed” Achilles

It is said that the great swift-footed Achilles, one day decided to challenge a tortoise in a speed race. Obviously confident in himself and capable of defeating the tortoise since he was twice as fast (we’re saying Achilles is exactly twice as fast as the tortoise to simplify the reasoning), he made the mistake of giving the tortoise a little advantage. The opponent, with the help of Zeno’s sharp thinking, managed to outrun Achilles forever.

The hero in fact, as soon as he covered the initial distance “d” separating the two competitors, saw that the tortoise was a little more far away: while he was arriving to the point, she travelled “d/2” more. He kept going and ran that “d/2” more, but in the meantime the tortoise was another “d/4” away from him. “*Now I got her!”,* thought Achilles but, as soon as he covered that distance, the tortoise had gained “d/8” more.No matter how Achilles ran to the position of the tortoise, he never succeded on reaching her.

1.2 Paradox definition

A paradox, from greek παρά (*against*) and *δόξα* (*opinion*), is, generically, the description of an event that goes against the common opinion or the daily experience, resulting then surprising, astonishing or bizarre. More precisely, a paradox is a seemingly absurd or self-contradictory statement that, superficially, cannot be true but cannot be false either.

Further analysis of the statement or proposition may reveal a fallacious axiom or some obscure underlying truth. Not all paradoxes are fundamentally incongruous, as some may only appear so.

Many famous problems of this kind exist.

A paradox can show up in many different disciplines, mainly logic: the most famous one is probably the liar's paradox: *"This sentence is a lie".*

If the sentence is true, then it is a lie, as it says. But if it is a lie, how can it be true? A lie cannot also be the truth. So the sentence being true makes it a lie.

Usually disciplines such as logic, philosophy and economics, the paradox is seen as a synonym of antinomy: a genuine logical contradiction. Moreover, the paradox is usually caused by the wrong variable being taken into consideration: an important example is the missing dollar paradox:

"Three guests decide to stay the night at a lodge whose rate they are initially told is \$30 per night. However, after the guests have each paid \$10 and gone to their room, the proprietor discovers that the correct rate should actually be \$25. As a result, he gives the bellboy the \$5 that was overpaid, together with instructions to return it to the guests.

Upon consideration of the fact that \$5 will be problematic to split three ways, the bellboy decides to pocket \$2 and return \$1 each, or a total of \$ 3, to the guests.

Upon doing so, the guests have now each paid a total of \$ 9 for the room, for a total of \$ 27, and the bellboy has retained \$ 2.

So, 27 + 2 = 29, where has the remaining \$1 from the initial \$ 30 paid by the guests gone?!"

As we can see, the paradox is created by an error which force the mind of the reader to focus on the $ 27 spent by the guest, adding them to the $ 2 in the bellboy pocket.

The relevant numbers to consider are however the total amount spent and where it has gone, which are fully accounted for:

the guests have each paid $ 9 for a total of $ 27. Of this, $ 25 has gone to the cost of the room, and $ 2 has been pocketed by the bellboy.

Of all things, it is accountancy that supplies a concise answer:

"You must not add debits to credits."

Money flowing out is a debit, money flowing in is a credit, and they always balance over a transaction.

In mathematics though, it is different.

The paradox can be defined as

"an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises." (Mark **S**ainsbury)

1.3 Paradox in everyday life

Many are the paradoxes we can observe in the daily life.

For example, one of the "hottest" topics is the problem of global warming.

According to the accepted climatological patterns the Ar**c**tic warming, along with the problem of ice melting, is causing the cooling of Europe.

So the global rising in temperatures is causing local decrease in temperatures.

This phenomenon is known as the Ar**c**tic paradox.

Many other paradoxes are lying behind plots of famous movies like, for example, "Terminator":
A cyborg is sent in the past to kill Sarah Connor, in order to prevent her from giving birth to John Connor. As a consequence, a man is also sent in the past, to stop the droid. Unbelievably, the man sleeps with Sarah, becoming the father of the baby he's been sent to protect. So, through a series of consequences, a time travel sets off the event which was meant to be prevented.
This paradox is called causal loop.

1.4 Paradox in history

Zeno of Elea, Greek philosopher (5^{th} century AC), used the paradox as a tool to prove, "logically speaking", Parmenides' ideas (his mentor), even if those ideas were contrary to conventional wisdom and experience. If Parmenides wanted to prove that "being" was unique and immutable, Zeno built logical games to even deny movement.
In one of his paradoxes, he said that to cross the entire length of a stadium, we firstly have to walk a half of it, before that a quarter of it, even before that 1/8 of it, and so on. Zeno represented a distance as an infinite sum of fractions, or rather as the numerical series created by the powers of 1/2:
$1/2 + 1/4 + 1/8 + \ldots + (1/2)^n = 1/2 \sum (1/2)^n$.
Zeno's paradox went on saying that it was impossible to walk in a finite time an infinite quantity of parts of space: there will always be some part ahead of you. Verbally, the reasoning sounds acceptable but, keeping in mind the limit of the function and the formula that we use to calculate the sum of the geometric series, the series describing the infinite spaces travelled in space has for sum:

1/2 * 1/(1-1/2) = 1.   The paradox is finally solved.

Logically speaking we can in fact agree that the distance is finite, so it is viable in a finite time, even if it can be represented as the sum of infinite terms.

As well as this paradox, also Achilles' paradox and the arrow paradox speculate on the infinite divisibility of the space.

1.5 Statistical paradox

In the statistical field, one of the most interesting phenomena that can happen is called the Simpson paradox.

Let's look at an example of it: on a certain disease, Hospital X has 55% of solved cases, while Hospital Y has 60% of solved cases. Focusing only on those data, which we can see represented in Table 1.1, it seems logical to prefer Hospital Y.

Table 1.1	Total cases	Solved cases	Solved cases %
Hospital x	200	110	55%
Hospital y	200	120	60%

Searching deeper in the hospitals data though, we find out that in Hospital X 90% of the cases are severe, of which 50% are solved, while the other 10% of the minor cases are solved with an accuracy of 100%.

In Hospital Y, 60% of the cases are severe, of which 40% are solved, the other 40% of minor cases are solved with an accuracy of 90%. We can see those data in Table 1.2.

Table 1.2	Casi gravi		Casi lievi	
	Casi	Casi risolti	Casi	Casi risolti
Hospital x	180 (90%)	90 (50%)	20 (10%)	20 (100%)
Hospital y	120 (60%)	48 (40%)	80 (40%)	72 (90%)

Considering this more in-depth research, it seems better to prefer Hospital X, since it actually solves more cases than Y (50% against 40% of the sever cases, 100% against 90% in minor cases).

Basically, the interpretation of data is distorted by previously unconsidered parameters.

This kind of paradox is called Simpson paradox, or Yule-Simpson paradox, from the names of the researchers that helped discovering and studying it.

1.6 Edward Hugh Simpson's biography

Edward Hugh Simpson CB (born 10 December 1922) is a retired British civil servant and former statistician best known for describing Simpson's paradox along with Udny Yule.

Edward Simpson was introduced to the thinking of mathematical statistics as a cryptanalyst at Bletchley Park (1942–45). He wrote the paper *The*

Interpretation of Interaction in Contingency Tables while a postgraduate student at the University of Cambridge in 1946 with Maurice Bartlett as his tutor; and published it in the *Journal of the Royal Statistical Society* in 1951 at Bartlett's request because Bartlett wanted to refer to it.

The paradox is used in mathematical statistics teaching to illustrate the care statisticians need to take when interpreting data. It figured in a 2009 episode of the US television crime-solving series *Numb3rs*.

Simpson entered the civil service administrative class in the UK Ministry of Education in 1947 and subsequently worked also in the Treasury, the Commonwealth Education Liaison Unit, as Private Secretary to Lord Hailsham as Lord President of the Council and Lord Privy Seal, and in the Civil Service Department.

He was a Commonwealth Fund (Harkness) Fellow in the USA (1956–57). At one point a useful observation of his on the aggregate behaviour of teachers' pay was labelled "Simpson's Drift". He retired from the Department of Education and Science as a Deputy Secretary and Companion of the Order of the Bath in 1982 and now lives in Oxfordshire.

In 2017, Simpson contributed two chapters on the cryptanalytic process Banburismus, developed by Alan Turing at Bletchley Park during World War II.

1.7 George Udny Yule

Frank Yates culminated his 1952 obituary of Yule by saying:

“To summarize we may, I think, justly conclude that though Yule did not fully develop any completely new branches of statistical theory, he took the first steps in many directions which were later to prove fruitful lines for further progress… He can indeed rightly be considered as one of the pioneers of modern statistics”.

Yule made important contributions to the theory and practice of correlation, regression, and association, as well as to time series analysis. He pioneered the use of preferential attachment stochastic processes to explain the origin of power law distribution. The Yule distribution, a discrete power law, is named after him.

Although Yule taught at Cambridge for twenty years, he had little impact on the development of statistics there. M. S. Bartlett recalled him as a "mentor" but his famous association with Maurice Kendall, who revised the *Introduction to the Theory of Statistics*, only came about after Kendall had graduated.

CHAPTER 2

2.1 The Simpson paradox

Simpson's Paradox was first introduced by Yule (1903) as "the fallacies that may be caused by the mixing of distinct records''. Simpson (1951), without citing Yule, discussed the interpretation of interaction in contingency tables. In one of his examples, Simpson (1951) observed that "there is a positive association between treatment and survival both among males and among females; but if we combine the tables we again find that there is no association between treatment and survival in the combined population''. Blyth (1972) provided an excellent mathematical description of this:

For events A, B, and C (and the complements BC and CC) it is possible to have

$$P(A|B) < P(A|BC)$$

and simultaneously to have

$$P(A|BC) \geq P(A|BCC) \text{ and } P(A|BCC) \geq P(A|BCCC).$$

Blyth called this "Simpson's Paradox'' (rather than "Yule's Paradox''), and the name has stuck.

We'll deepen the theory presented with some real life and exhausting examples later.

Moore, McCabe, and Craig (2012) defined Simpson's Paradox as follows:

"An association or comparison that holds for all of several groups can reverse direction when the data are combined to form a single group. This reversal is called Simpson's paradox''.
Unfortunately, this definition does not fully explain the paradox. On the other hand, the mathematical description by Blyth (1972) is almost impossible to understand for students of introductory statistics courses and for non-statisticians. Thus, explaining Simpson's Paradox in an introductory statistics class is particularly challenging. More unfortunately, as Lesser (2002) pointed out, "some well-known introductory textbooks [...] do not mention Simpson's Paradox at all, some discuss it in a section marked 'optional' [...]''
Strictly speaking**, the Simpson's one is not actually a paradox, but a counterintuitive feature of aggregated data, which may arise when (causal) inferences are drawn across different explanatory levels: from populations to subgroups, or subgroups to individuals, or from cross-sectional data to intra-individual changes over time** (cf. Kievit et al., 2011).
Understanding when the Simpson paradox comes into play is not easy unless all the data can be checked accurately. To facilitate this process, several graphical applications exists in order to recognize it.
We can find one at *www.math.usu.edu/~schneit/CTIS/SP/*

2.2 Paradox illustration

To give an example of the work done by the said applications, we'll see a case regarding flights from 2 companies, AA and AW.
Firstly, we consider the two companies and the total number of flights delayed

in a year. (Data reported in table 2.1)

Table 2.1	flights	Flights delayed	Flights delayed %
Alaskan Airlines	792	74	9.3%
America West	6.066	532	8.8%

Looking at those data, it seems obvious to prefer AW, since its delayed flights are only 8.8%, against the 9.3% of AA.

To deepen the study of this case, we focus on the departure city of the plane, gathering some more data that makes the situation change again, as we can see from Table 2.2.

Table 2.2	Los Angeles		Phoenix	
	Flights	Flights deleyed	Flights	Flights deleyed
Alaskan Airlines	559 (70,6)	62 (11,1%)	233(29,4%)	12 (5,2%)
America West	811(13,4%)	117 (14,4%)	5.255 (86,6%)	415 (7,9%)

As we can see, adding more data made our choice different: it is in fact more convenient to book an Alaskan Airlines (AA) flight, since the flights delayed are fewer both from Los Angeles (11.1% against 14.4% of AW) and from

Phoenix (5.2% against 7.9% of AW).

The same data are used in the graph of the graphical app that analyzes all the variables and helps highlighting the paradox.

The First line of the table shows a count of Alaskan Airlines (AA) flights (792), a count of America West (AW) flights (6066), the number of each of these that was delayed (74 for AA And 532 for AW) and finally the percentage delayed for each airline. From this row, we see that Alaskan Airlines Had a greater percentage of delayed flights than did America West Airlines (9.3% versus 8.8%).

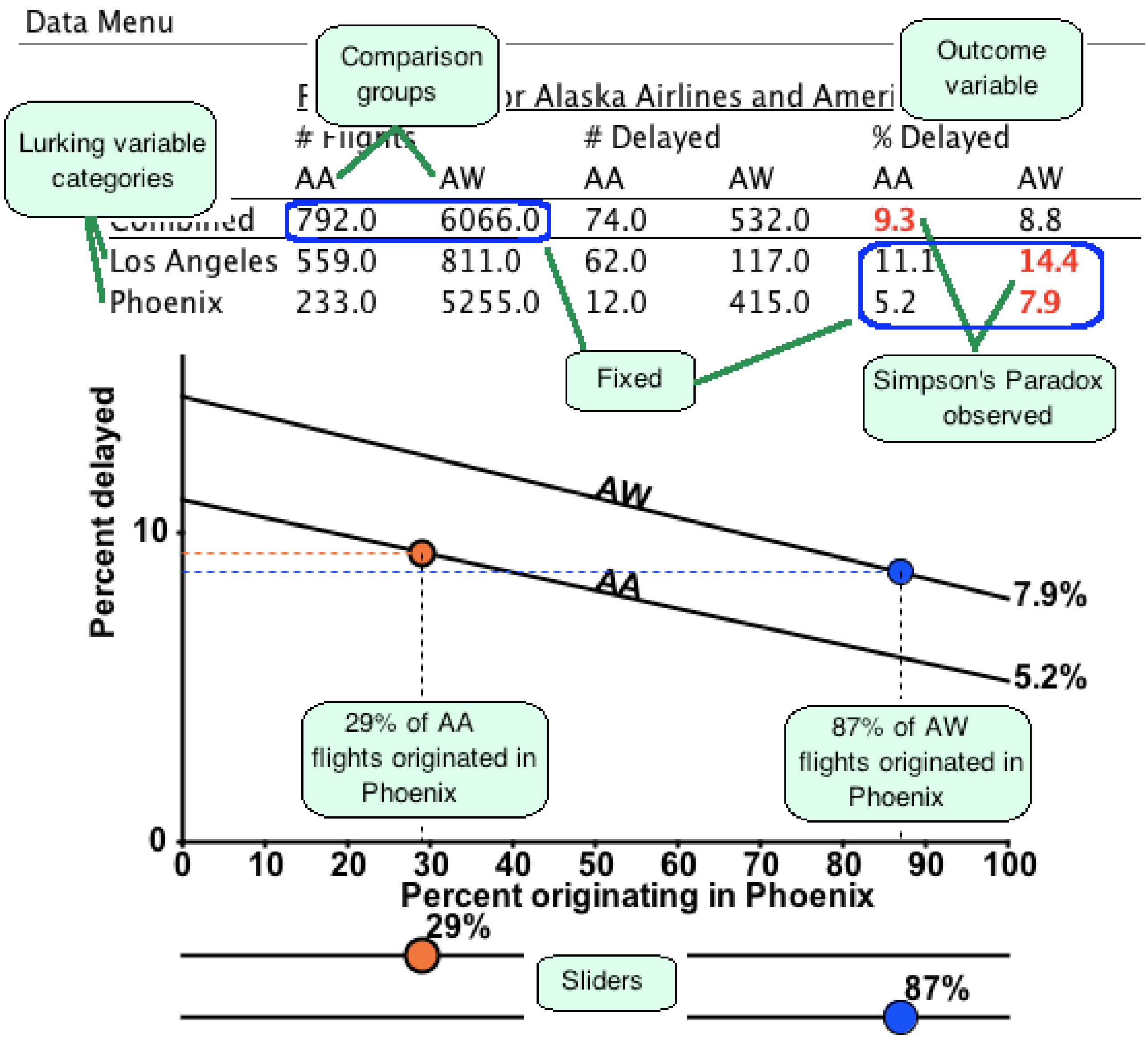


Image 1:
example of the work done by the graphical app that
analyzes all the variables and helps in highlighting the paradox.

In the second and third rows of the table, the data for each airline has been divided into subgroups based on the place of origin of the flight (e.g. 559 of the 792 AA flights originated in Los Angeles and 233 originated in Phoenix). For these subgroups, we see that a greater percentage of the AW flights originating in Los Angeles were delayed (14.4% versus 11.1 % for AA) and a greater percentage of The AW flights originating in Phoenix were delayed (7.9% versus 5.2% for AA).

Whereas AA had a greater number of delayed flights when the data were combined, the relationship is reversed when we divide the data into subgroups based on the lurking variable ‘Place of origin’. Why does this occur?
For each of the comparison groups (AA and AW) the plot shows the percentage of delayed flights as a function of the percentage of flights originating from Phoenix.
Colored dots on the lines indicate the actual percentages of flights that originated in Phoenix for each of the comparison groups. We see that 87% of AW flights originated in Phoenix whereas only 29% of AA flights began there. Since Phoenix flights were less likely to be delayed than Los Angeles flights (5.2% versus 11.1% for AA and 7.9% versus 14.4% for AW) and the vast majority of AW flights began there, AW’s overall percentage of delayed flights is lower than that of AA for which most of the flights originated in LosAngeles.
The sliders (on Image 1) allow the user to adjust the percentage of flights originating in Phoenix for each of the two airlines and to see how this affects the observed relationships. As a slider is adjusted, a circle on the corresponding line (circle color the same as slider dot color) moves. Dashed lines from the circles to the axes highlight the relationship between the variable values.
As the percentage of flights originating in Phoenix is changed, the data in the table is updated to reflect this. The ‘combined’ counts for each category are fixed as are the percentages of delayed flights for each airline from each origination point. When the Simpson’s Paradox is observed, i.e. the airline with the greater percentage delayed for the ‘combined’ data is different from the airline with the greater percentage delayed for each of the origination points, the percentages in the table are highlighted in red, otherwise they are green.
Obviously, the two methods, the table and the graph, have the same outcome.

2.3 A discriminating university?

One of the canonical examples of it concerns possible gender bias in admissions into Berkeley graduate school (Bickel et al., 1975; see also Waldmann and Hagmayer, 1995).

In 1973 in fact, Berkeley graduate school was one of the first universities to be reported for gender discrimination.

For the fall semester there were approximately 12.763 applications (8.442 males and 4321 females).

The results, represented on table 2.3, shows that males that applicated to the university had more opportunities than females of being accepted.

Table 2.3	applications	accepted	% accepted
Males	8.442	3.738	44,28%
Females	4.321	1.494	34,57%

The university, in order to make clear that there was no discrimination, decided to release more data, regarding the 6 most important *majors* offered. (represented in Table 2.4)

Table 2.4	male		female	
Faculty	Applications	% Admitted	Applications	% Admitted
A	825	62 %	108	**82 %**
B	560	63 %	25	**68 %**
C	325	**37 %**	593	34 %
D	417	33 %	375	**35 %**
E	191	**28 %**	393	24 %
F	272	6 %	341	**7 %**

As shown in Table 2.4, no faculty was actually discriminating in any way, on the contrary, in almost all *majors* the proportion of females admitted is greater than that of males.

This seems paradoxical: globally, there appears to be bias toward males, but when individual graduate schools are taken into account, there seems to be bias toward females. This conflicts with our implicit causal interpretation of the aggregate data, which is that the proportions of the aggregate data (44,28% males and 34,57% females) are informative about the relative likelihoods of male or female applicants being admitted if they were to apply to a Berkeley graduate school. In this example, SP arises because of different proportions of males and females attempt to enter schools that differ in their thresholds for accepting students.

This example is actually very easy to explain: we will in fact use a model to simplify the problem.

Given that there are only two *majors* to choose between (A and B) in the university X, let's suppose that 400 males and 200 females applicate for course A. The commission has to create an admission test which do not favors nor males nor females. Only a half of the applications can be accepted, and to simplify even more our problem, we'll say that half of the males and half of the females are selected. At faculty B, with 150 males and 450 females enrolled, due to more places available, 80% of the applications will be accepted.

Like we did before, for simplicity we'll say 80% of males and 80% of females are selected.

Table 2.5	Faculty A		Faculty B	
	applications	admitted	applications	admitted
Males	400	200	200	160
Females	200	100	400	320
Total	600	300	600	480

Although we know there has been no bias, from the data collected in Table 2.5 it seems that university X privileges females because, having the same number of applications (600 males and 600 females) only 360 males were selected against 420 females!

The solution lies in the fact that more boys applicated to *major* A, which accepted less people, while more girls applicated to *major* B, which granted more chairs.

2.4 Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy

This study was designed to compare different methods of treating renal calculi in order to establish which was the most cost effective and successful. Of 700 patients with renal calculi, 350 underwent open surgery, 350 percutaneous nephrolithotomy.

Data collected (represented in Table 2.6) seems to suggest the Percutaneous nephrolithotomy as the most effective.

Table 2.6	Total cases	success	success%
Underwent open surgery	350	273	78%
Percutaneous nephrolithotomy	350	289	83%

We can deepen the study dividing total cases in 2 groups (represented on Table 2.7): the former focusing on renal calculi which are greater than 2cm in diameter, the latter for smaller stones.

Doing so we can see that the most effective is actually the open surgery both with stones greater than 2cm (73% of solved cases against 69%) and with smaller stones (93% of solved case against 87%)

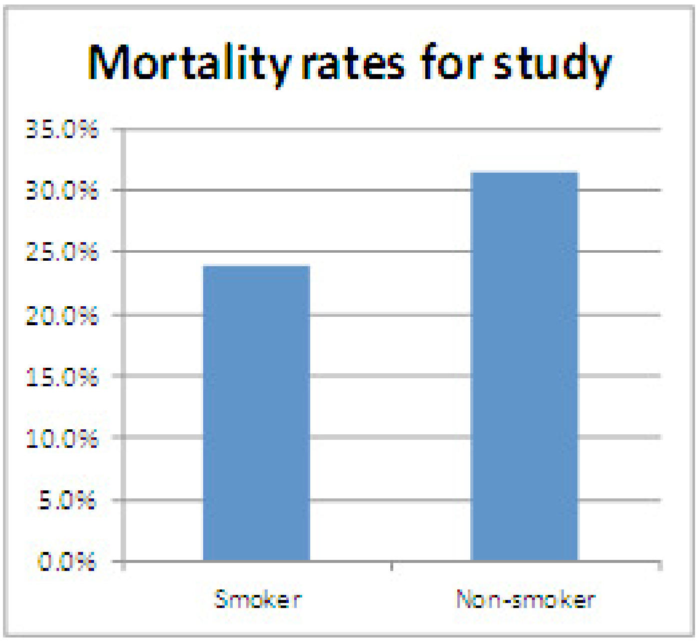
Table 2.7	Stones greater than 2cm		Smaller stones	
	Casi	Casi risolti	Casi	Casi risolti
Underwent open surgery	263 (75%)	192 (73%)	87 (25%)	81 (93%)
Percutaneous nephrolithotomy	80 (23%)	55 (69%)	270 (77%)	234 (87%)

2.5 Deaths of smokers and non-smokers

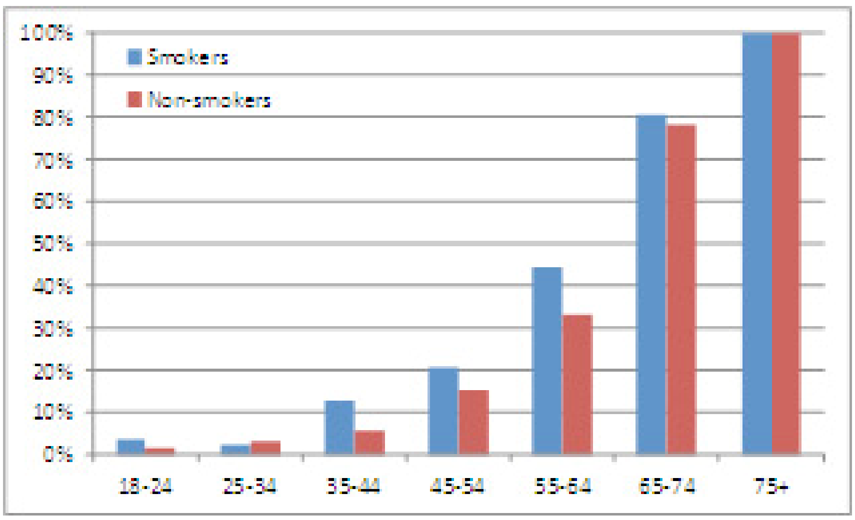
The data are taken from a 1996 follow-up study from Appleton, French, and Vanderpump on the effects of smoking. The follow-up catalogued women from the original study, categorizing based on the age groups in the original study, as well as whether the women were smokers or not. The study measured the deaths of smokers and non-smokers during the 20-year period.

The overall counts of the study are as follows (table 2.8 or in the graph)

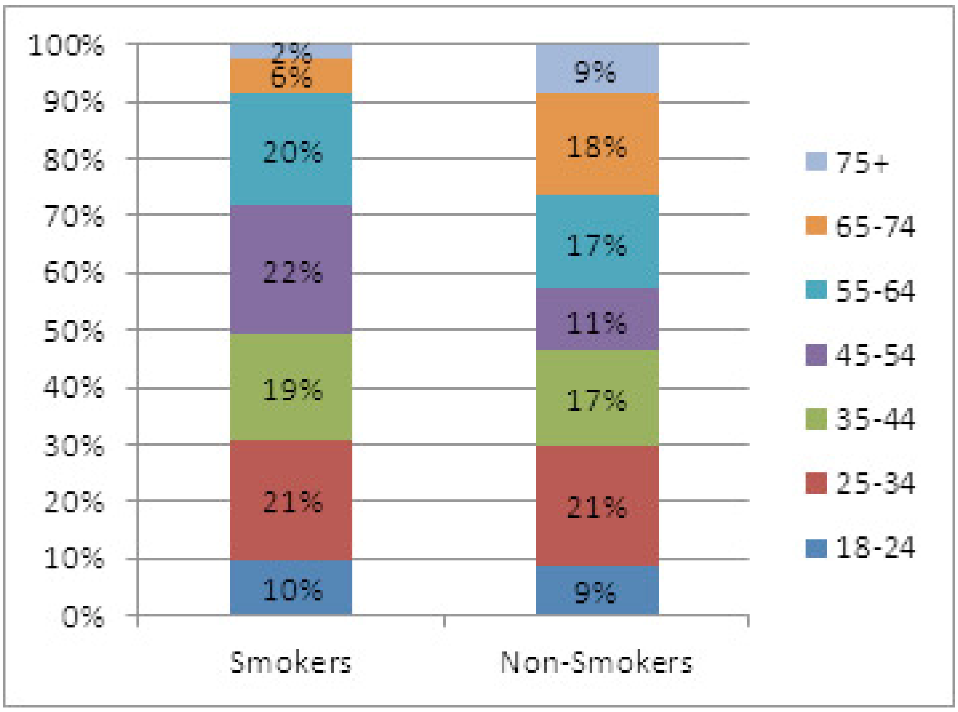
Table 2.8	died	survived	Mortality rate%
Smoker	139	443	582 (23,9%)
No-smoker	230	502	732 (31,4%)



This analysis suggests that non-smokers actually have higher mortality rates than smokers, certainly a surprising result and contrary to current medical teachings, and maybe a potential boon to the tobacco industry. But the numbers tell a much different story when mortality is examined by age group:



Now, we see that smokers have higher mortality rates for virtually every age group. What is going on here? This is a classic example of the Simpson's Paradox phenomenon; it shows that a trend present within multiple groups can reverse when the groups are combined. The phenomenon is well known to statisticians, but counter-intuitive to many analysts. To paraphrase Nassim Taleb, not only are people regularly "fooled by randomness"; they are also fooled by lurking multivariate relationships. Simpson's Paradox requires several things to occur. First, the variable being reviewed is influenced by a "lurking" variable. In our example, age is the lurking variable, with the population grouped into a discrete number of subcategories. Second, the subgroups have differing sizes. If both of these conditions are met, they conspire to obscure the salient relationships in the data due to the relative weighting attached to each subgroup.



The age distributions are substantially different for smokers and non-smokers. In particular, the non-smoking population is older on average. Twenty seven percent of non-smokers are in the two oldest groups, compared to approximately eight percent for smokers. Combined mortality rates are near 100% for both groups, but the greater proportion of older non-smokers pushes up the average for that group. Viewing the data by age leads to a more plausible theory – one that comports with long standing medical teaching: that long-term smoking shortened lifespans, thereby affecting the age distributions in the study's population.

2.6 Did price optimization increase or decrease profits?

The final case study comes from a price optimization program implemented at a major manufacturer of cosmetic products. The manufacturer had three different store brands, which in this example will be called A, B, and C. Although the product offering was similar across all stores, each store brand had a different consumer base and prices were optimized independently for each brand.

The objective of computing optimal prices was twofold: to improve profit margins for the business without negatively impacting aggregate profit, i.e., the three store brands together; and for each store brand separately.

After the optimization models were developed and implemented, the manager assisted the manufacturer in a series of live controlled experiments to quantify the potential benefit coming from optimized prices and assessed whether the financial objectives were being met. In order to analyze the results, profit margins were compared before the test (control period) and during the test period - a straightforward method to quantify benefit. It was found, however, that the average margin for the whole business before the test was 7.7% and during the test was 7.4%, implying that the program was unsuccessful in improving margin. Aside from any measure of statistical significance, the absolute improvement did not justify the costs to make the program changes. Is this enough to say that the optimization program was a failure?

On the surface, the answer would seem to be "yes": optimization apparently led to the unsatisfactory result of reduced margins. Nevertheless, when results were broken down by store brand, the story changed; and indeed, the program showed significant profit margin increases across all three brands.

Table 2.9	Control period	Test period	Change
Brand A	10	12	20%
Brand B	5	6	20%
Brand C	4	5	25%

How could it be that all three brands showed margin improvement and yet the total average margin had deteriorated?

In fact, both answers are correct. Profit margins improved in each brand and yet the weighted average of the margins dropped in total. The answer is hidden in Simpson's Paradox and its arithmetic illusion in which clear patterns are obscured when the mix of the groups being weighted averaged changes over time. Responsible for the strange and apparently contradictory results is the fact that the mix of store brands changed pre- and post-test. During the test period, the manufacturer underwent a major change in its branding strategy and had shifted the mix of store brands significantly.

Table 2.10	**Control period**			**Test period**		
Store brand	Number of stores	Mix of brands	Margin	Number of stores	Mix of brands	Margin
Brand A	400	57%	10%	200	29%	12%
Brand B	200	29%	5%	300	43%	6%
Brand C	100	14%	4%	200	29%	5%
Overall	700		7.7%	700		7.4%

As can be seen from the table above, although the total number of stores remained constant, the concentration of store brands B and C increased. Moreover, store brands B and C have lower margins than store brand A. The store mix shift to lower-margin brands created a Simpson's Paradox effect. It is important to point out that the optimization program was successful and did increase margins and dollars for the manufacturer. The percentage margin reduction for the whole business resulted not from problems with the price optimization strategy, but from the manufacturer's store brand strategy. Had the optimization results not been thoroughly understood and analyzed with an awareness of Simpson's Paradox, one could have easily come to the wrong business conclusion.

Findings

We have seen that Simpson Paradox might occur frequently, and that people are often poor at recognizing it. When Simpson Paradox goes unnoticed, incorrect inferences may be drawn and, as a result, decisions about resource allocations (including time and money) may be misguided. Interpretations may be wrong not only in degree but also in kind, suggesting benefits where there may be adverse consequences (Can smoking actually be good for your health?).

Fortunately, the data and computational power continue to grow exponentially, analysts have gained unprecedented power to build and promulgate data-driven decision models, shifting business practice away from traditional decision making practices rooted in industry knowledge and intuition. However to paraphrase Voltaire (or Uncle Ben from *Spiderman*), with great power comes great responsibility. It is obvious that poorly or naively performed statistical analyses can yield incomplete or ambiguous answers, but now we know that phenomena such as Simpson's Paradox can downright mislead and motivate misguided decisions. This should be borne in mind in situations where sophisticated analytical workbenches are adopted by users who do not possess commensurate statistical understanding.

Bibliography

Paradox definition:
https://it.wikipedia.org/wiki/Paradosso

Paradoxes in History:
Matematica.blu 2.0 5 Zanichelli

Edward Hugh Simpson:
https://en.wikipedia.org/wiki/Edward_H._Simpson

George Udny Yule:
https://it.wikipedia.org/wiki/George_Udny_Yule

The simpson paradox:
https://plato.stanford.edu/entries/paradox-simpson/
Mathematics for economists - Carl P. Simon- Lawrence Blume

paradox illustration:
www.math.usu.edu/~schneit/CTIS/SP/

sexist university:
http://www.sciencestorming.eu/index.php/2012-01-18-17-01-34/il-paradosso-delle-stelle/32-discriminazione-tra-uomo-e-donna-si-e-no

Simpson paradox in real life:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1339981/

Did price optimization increase or decrease profits?
https://www.statslife.org.uk/the-statistics-dictionary/2012-simpson-s-paradox-a-cautionary-tale-in-advanced-analytics