International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 3, Issue 3, ISSN No. 2455-2143, Pages 62-68
Published Online July 2018 in IJEAST (http://www.ijeast.com)

SIMPSONS PARADOX: A TALE AT DICHOTOMY WITH ITSELF

Palakh Gupta
Department of Computer Science & Engineering
Manipal University Jaipur, Jaipur, Rajasthan, India

***Abstract*—This paper facilitates exploration of Simpson's Paradox. It exercises the context of multiple real datasets to demonstrate this contradiction. This paradox has confounded the human population and occasionally even led to wrong allegations which were resolved by the expertise of statisticians. These instances will explicate how easy it is to plummet into a web of paradoxical interpretations when relying solely on intuition. They demonstrate a grave need of a flexible tool to detect this anomaly which can be exercised by non-statisticians. The paper provides a comprehensive yet concise history of the Paradox, it's mathematical significance. It elaborates it's meaning with the help of multiple examples. The analysis of these examples provides an in-depth knowledge of the anomaly and informs on how to avoid such complications.**

Keywords— **Simpson's Paradox, Instance of Simpson's Paradox**

I. INTRODUCTION

A. History

Simpson's Paradox or the **Yule–Simpson effect** was revealed by George Udny Yule (1871-1951) in his paper titled "Notes on the Theory of Association of Attributes in Statistics." He initially explained the elementary technique for statistical classification- "division applying dichotomy". The process comprised on dividing units (e.g. humans or animals) into either of two mutually exclusive groups based on presence or absence of that attribute (Yule, p. 121). Two independent attributes and their complements (e.g. the association between a and b), when considered wholly may be positively or negatively associated. The direction can be determined on the basis on the value of the related attributes i.e. if it is greater or lesser then the value placed on the attribute when it is independent (Heather-Goltz, 2010). In the presence of three attributes (A, B, C), the relationship becomes complex. In this state, relationship is considered in pairs in the presence of the third variable or it's contrary, i.e. Given C or it's complement, relationship between A and B or their complements is considered (Heather-Goltz, 2010). Even when the conclusion that these attributes are independent was an easy task, Yule mentioned that one must actually conclude whether these "attributes are independent, wholly or in part". These partial associations are critical in determination for the accuracy of the observed total associations.

A pair of attributes are not certainly independent within a universe even if they exhibit independence in every sub-universe or vice-versa. Given the association of A and B given C is zero does not always conclude that association of A and B equal to zero. One should also not infer the inverse. Yule explained that if A and B have a positive or negative value, the same might not hold true for A and B given C is present (Heather-Goltz, 2010). Yule illustrated these points using an example from population genetics.

	Fathers (n=100)		Mothers (n=100)		Parents (n=100 couples)	
	With attribute	Without Attribute	With attribute	Without Attribute	With attribute	Without Attribute
Sons w/ attribute	25%	25%	-	-	-	-
Sons w/o attribute	25%	25%	-	-	-	-
Daughters w/ attribute	-	-	1%	9%	-	-
Daughters w/o attribute	-	-	9%	81%	-	-
Offspring w/ attribute	-	-	-	-	13%	17%
Offspring w/o attribute	-	-	-	-	17%	53%

Fig. 1. Yule's Example of Paradoxical Phenomenon

Affected sons with an affected father = 25%
Affected offspring whose parents both had the attribute = 13% = 13/ (13+17) = 13/30 = 43.3% cases where both parents were affected.

Inference - Researchers erroneously assume that this attribute is more dominant or easily passed between generations justified by the pooled data. However, this attribute is not predominantly heritable through the mother or father's genetic line (Yule,1903). Yule described this phenomenon as "quite a large but illusory inheritance created simply by the mixture of the two distinct records" (i.e., maternal and paternal inheritance). Exploring this attribute inheritance by gender

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 3, Issue 3, ISSN No. 2455-2143, Pages 62-68
Published Online July 2018 in IJEAST (http://www.ijeast.com)

might help to explain the origins of this erroneous conclusion and others in similar situations. Yule explained "there will be

an apparent association between A and B in the universe unless either A or B is independent of C" (Heather-Goltz, 2010). However, variables A, B, and C fail to meet this requirement in Yule's example. When the data are pooled, a positive association exists between both parents with the attribute (A) and offspring with the attribute (B), and male gender (C).

A weighted average situation is formed when data is aggregated. The reason for this is unequal proportion of males possessing the attribute when compared to the number of females who show this attribute. The aggregated data exhibits an association which negates at least one or more of the sub-universe. Yule referred to this phenomenon as the "fallacy of mixing distinct records" (Yule,1903:pg132). This was later further strengthened by Simpsons and the Yule Paradox is more commonly referred to as the Simpson's Paradox.

B. Simpson's Influence

E.H. Simpson published a paper titled "The Interpretation of Interaction in Contingency Tables." in 1951.Simpson considered a 2x2x2 (attributes A, B, and C) contingency table (Simpson,1951) in his paper. Simpson explained that this relationship contained partial associations (i.e., first order interaction, 2x2), also interactions between all three variables (i.e., second order; Simpson,1951, pg238).

He established the boundary between first and second order interactions by asserting that if A and B have an association (first order interaction), an interaction between AB and C (second order) will cease to exist if the degree of association for AB given C is the same as AB given C's complement (Simpson,1951:pg 239).

Simpson explained that when there is no visible "second order association', there is a huge probability for error i.e. if AB show a positive association given C and a negative association given complement of C, AB may appear independent when the whole population is aggregated. Simpson explained the above phenomena exercising a heuristic example of clinic patients (Heather-Goltz, 2010). In the example, patients received treatment or no treatment and were monitored for survival over time.

TABLE 4

	Male		Female	
	Untreated	Treated	Untreated	Treated
Alive . . .	4/52	8/52	2/52	12/52
Dead . . .	3/52	5/52	3/52	15/52

Fig. 2. Fourth Table of Yule's Paradox

	Male (n=20)		Female (n=32)	
	Untreated	Treated	Untreated	Treated
Alive	7.69%	15.38%	3.85%	23.08%
Dead	5.77%	9.62%	5.77%	28.85%

Fig. 3. Reproduction of Figure 2

Data when examined on the basis of Gender, the conclusion was that both males and females responded more positively to the treatment and survived (i.e., 15.38% and 23.08%, respectively), compared to those who didn't receive treatment (i.e., 7.69% and 3.85%, respectively; Simpson, 1951). However, when the data were aggregated, these positive associations disappeared and treatment and survival appeared to be independent attributes showing no association (Simpson,1951; pg241).

C. Diagnosis of Simpson's Paradox

For some whole numbers we may have:

$a/b < A/B$, $c/d < C/D$, and
$(a + c)/ (b + d) > (A + C)/ (B + D)$.
This is known as the *Simpson's Reversal of Inequalities*. Mentioned below is a demonstration.

$1/5 < 2/8$
$6/8 < 4/5$
$7/13 > 6/13$

Simpson's Paradox 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 (Kievit et al, 2011).

II. INSTANCES OF SIMPSON'S PARADOX

A. Higher Education Admissions

Discrimination on the basis of Gender even though less prevalent now, dominated our society 40 years ago. One of such presumed cases was the admission process of University of California, Berkley. In the Fall of 1973, it was noted that out of all the female students who applied, only 35% were admitted whereas out of all the male students who applied, 44% of them were selected.

This data showed a clear inclination towards one race over another. This is identified as Gender Biasing. In fear of a potential lawsuit, the university committee sought to discern the culprit behind this. This led to comprehensive analysis of the data. The main task was to determine those departments who took part in this discrimination against females.

Data was then manipulated and segregated on the basis of departments. The inference from the data showed an anomaly. There was no gender discrimination of any behavior. Even if there was, it was diminutive partiality towards women. The next inquiry was on the data's credibility and the reason for its two explanations at dichotomy. There must have been a hidden factor which was responsible for this irregularity.

Described earlier is a coveted example of Simpson's Paradox, explaining in a pellucid manner, how if one or more variables are ignored, data can represent a false story. After accurate analysis, the hidden factor was unearthed.

Primarily, analyzing accepts and rejects department wise, proves in an impeachable manner that there was no Gender Biasing involved in the admission process.

Admitted	Rejected
1755	2771

The number of admits were 1775 out of 4526 applicants which shows a 38% acceptance rate for that term. Further the data was divided on the basis of gender.

	Admits	Rejects	% of Admits
Male	1198	1493	44.51
Female	557	1278	30.35

Clearly visible from the data, only 557 females out of a total of 1835 were accepted i.e. 30.35 % whereas for males, the acceptance rate was 44.51% showing a significant difference in the rates.

But when this data was further segregated department wise, it represented a different story which showed no biasing

> DepartmentA

	Admits	Rejects	% of Admits
Male	512	313	62.06
Female	89	19	82.40

> DepartmentB

	Admits	Rejects	% of Admits
Male	353	207	63.03
Female	17	8	68.00

> DepartmentC

	Admits	Rejects	% of Admits
Male	120	205	36.92
Female	202	391	34.06

> DepartmentD

	Admits	Rejects	% of Admits
Male	138	279	33.09
Female	131	244	34.93

> DepartmentE

	Admits	Rejects	% of Admits
Male	53	138	27.74
Female	94	299	23.91

> DepartmentF

	Admits	Rejects	% of Admits
Male	22	351	5.89
Female	24	317	7.91

Fig. 4. Department Wise Admissions

Department A has a significantly high acceptance rate—especially for women at 82%. However, only 108 of them applied to Department A. That's only 6% of all women who applied across departments. Whereas, 825 is the number of men who applied to Department A. That's 30% of all the male applicants. The difference in the number of applicants is vast.

Also in Department E, 393 females applied for the course whereas only 191 males applied, almost half of the number of females.

Now, let's consider Department F, the acceptance rate is 5-7% whereas in Department A is it more than 50%. The number of women who applied in Department F is very high considering the number who applied for Department A

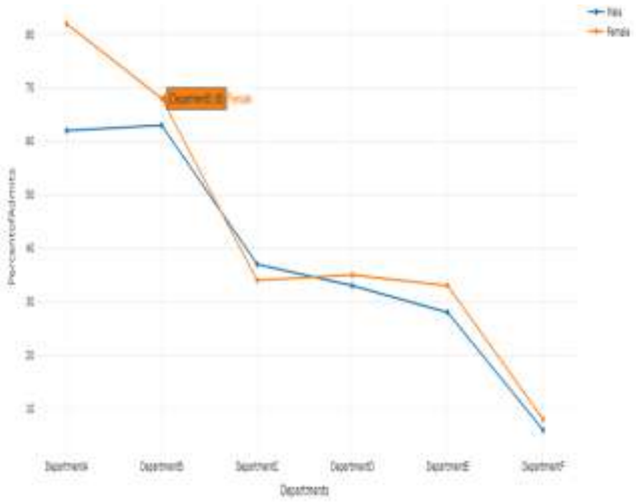

Fig. 5. Grouped Line Graph for Department Wise Admissions

This gives a very clear inference that women weren't being discriminated against but a large proportion of them were applying to a low-acceptance rate department while a large proportion of men were applying to high-acceptance rate department which skewed the overall results. The chart below represents similar acceptance rates for men and women in all the departments.

B. Survival Aftermath

Titanic was a British passenger liner was making its maiden voyage from Southampton to New York City which was unsuccessful due to a collision with an iceberg in North Atlantic Ocean after which it sank in the early hours of 15 April 1912. Titanic's passengers numbered approximately 1,317 people: 324 in First Class, 284 in Second Class, and 709 in Third Class. Of these, 869 (66%) were male and 447 (34%) female. There were 107 children aboard, the largest number of which were in Third Class. The ship was considerably under capacity on her maiden voyage, as she could accommodate 2,453 passengers—833 First Class, 614 Second Class, and 1,006 Third Class. There were an estimated 2,224 passengers including crew aboard, and more than 1,500 died, making it one of the marine disasters in history.

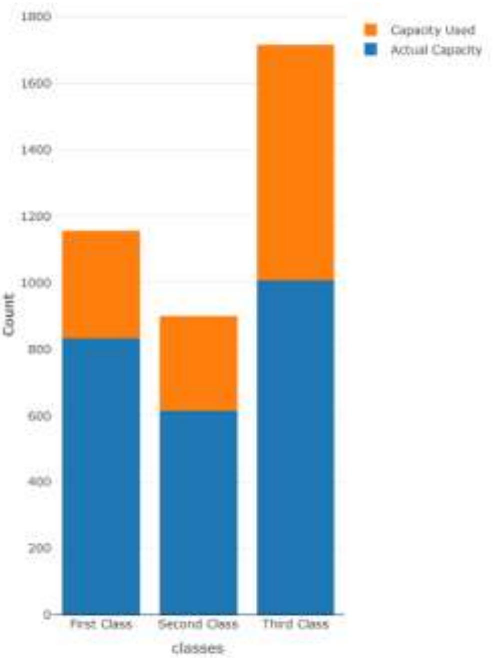

Fig. 6. Stacked Bar Chart Class wise

Although Titanic possessed advanced safety features such as watertight compartments and remotely activated watertight doors, *Titanic* only carried enough lifeboats for 1,178 people—about half the number on board, and one third of her total capacity. The reason for this was the outdated maritime safety regulations. The ship carried 16 lifeboat davits which could lower three lifeboats each, for a total of 48 boats for its maximum capacity. However, Titanic carried only a total of 20 lifeboats. Four of them were collapsible and proved hard to launch during the sinking leaving only 16 working boats for 2,224 people. Even when only 16 boats were present, many of them were launched only partially loaded. A disproportionate number of men were left aboard because of a "women and children first" protocol for loading lifeboats. Titanic broke apart at 2:20 a.m. and foundered with well over one thousand people still aboard. The number of casualties of the sinking is imprecise owing to multiple reasons. Mainly the confusion over the passenger list, some people cancelled their trip, some were travelling under alias names and were double counted on casualty lists.

The death toll was determined between 1,490 and 1,635 people. The number of people who survived the disaster was less than 1/3rd of those aboard titanic. Some survivors even died shortly after the tragedy because of injuries and the effects of exposure. When the survivors are segregated on the basis of different classes aboard titanic, patent differences are observed. It showed that the survival rate for Second Class Passengers was higher than the Crew. The survival rate for First Class Passengers was the highest – 61.75% followed by the Second Class then the Third Class and then the crew. As visible from the chart, the third class had a higher survival rate than the crew, but when these numbers were further segregated on the basis on numbers, an anomaly was represented.

	Men				Women			
Class	Saved	Lost	Total	Survival rate	Saved	Lost	Total	Survival Rate
First	57	118	175	32.57%	140	4	144	97.22%
Second	14	154	168	8.33%	80	13	93	86.02%
Third	75	387	462	16.23%	76	89	165	46.06%
Crew	192	693	885	21.69%	20	3	23	86.96%

Fig. 7. Division of life loss and survival

For Second Class and Crew, the survival rates overall are 36.01% for Second Class and 23.95% for the Crew. The survival rate for second class was considerably higher than that of the first class. But if we take a more detailed look at this data set, we notice that the survival rates for both men

Class	Saved	Lost	Total	Survival Rate
First Class	197	122	319	61.75%
Second Class	94	167	261	36.01%
Third Class	151	476	627	24.08%
Crew	212	673	885	23.95%

Fig. 8. Survival Rate Graph for Different Classes

and women of crew i.e. 21.69% and 86.96% is higher than that of the second class i.e. 8.33% and 86.02%. This brings the accuracy of both these tables in doubt because both are contradicting statements. One explains how the overall survival rate was, if anything, greater for the second class when paralleled to that of the crew. Whereas, the other represents that both male and female survival rates are higher for the crew. This apparent contradiction appears because the relationship between survival and class is influenced by a hidden variable. The confounding variable here is the gender which is ignored in the first table. The second table also emphasizes on the greater proportion of women in second class than in the crew. In Second Class, almost one-third of the passengers were male whereas less than 3% of people in the crew were female. Number of females in second class is almost 4 times the number of women in the crew. This significant difference has a substantial impact on the overall survival rate. It also disguises the overall passenger vs crew survival rate. This examples serves the need for prevention of Simpson’s Paradox through exhaustive analysis of the data before establishing conclusions.

C. Proper Representation in a Jury

New Zealand Department of Justice in the year of 1993 in Septembers encountered Simpson's Paradox while forming juries from the pool of jurors exercising surveys. The anomaly was noticed in the representation of Maori, the native of New Zealand. The survey included all potentials jurors who arrived at court during a period in September and October 1993. The results elucidated that 9.5 percent of people living within the jury districts were Maori compared to the 10.1 percent of Maori in the pool of potential jurors.
But these results were juxtaposed when analyzed district wise. Overall, Maori appeared to be slightly over-represented in the jury pool. But in every single local area, Maori were under-represented - often substantially.

Percentage Maori ethnic group

District	Eligible Population (aged 20-64)	Jury Pool	Shortfall
Whangarei	17.0	16.8	0.2
Auckland	9.2	9.0	0.2
Hamilton	13.5	11.5	2.0
Rotorua	27.0	23.4	3.6
Gisborne	32.2	29.5	2.7
Napier	15.5	12.4	3.1
New Plymouth	8.9	4.1	4.8
Palmerston North	8.9	4.3	4.6
Wellington	8.7	7.5	1.2
Nelson	3.9	1.7	2.2
Christchurch	4.5	3.3	1.2
Dunedin	3.3	2.4	0.9
Invercargill	8.4	4.8	3.6
All Districts	**9.5**	**10.1**	**-0.6**

Fig. 9. District Wise Distribution

The chart elucidates that for every district except for Overall Districts the Eligible Population is higher than
Jury Pool. Even though every district shows an underrepresentation in every district, the overall participation is higher.

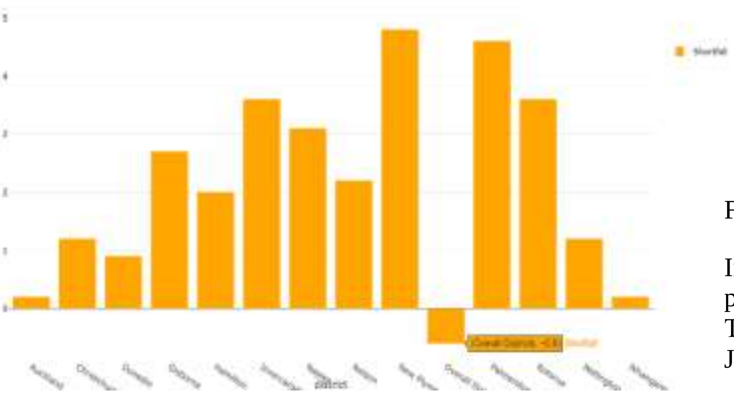

Fig. 10. Graph for Shortfall district wise

To delve a little further, let's consider a subset of the two districts- Rotorua and Nelson. Rotorua has around 20% more Maori population in both jury and population, whereas Nelson has a significantly low figure in both. Individually, both districts present an under-represented Maori in the jury pool. However, when the data is combined, Maori is revealed as over-representation.

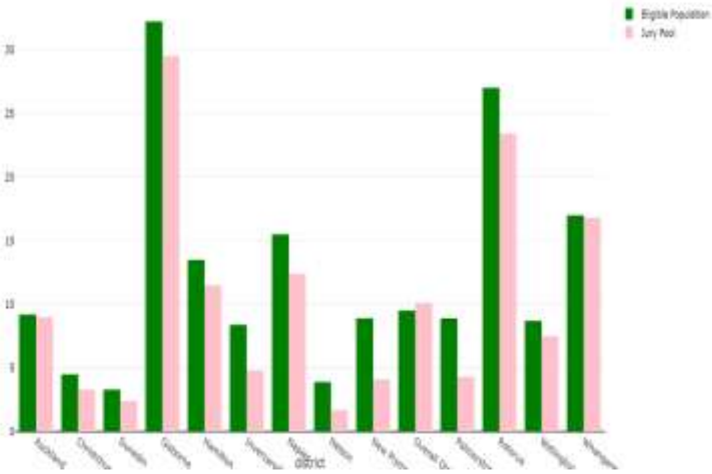

Fig. 11. Grouped Bar Chart of Elgible Population and Jury Pool

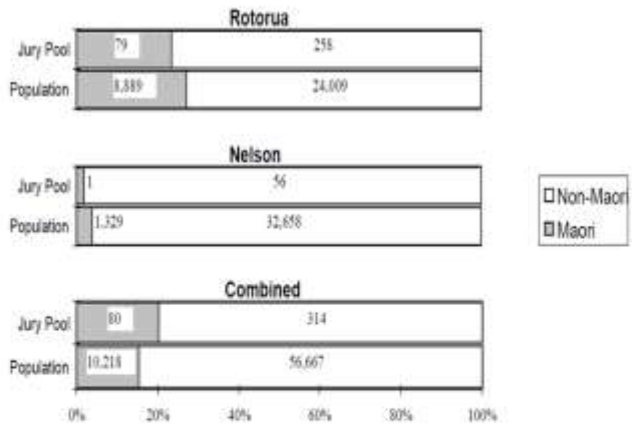

Fig. 12. Rotorua and Nelson Jury Pool and Population

In this case, the much larger size of the jury pool in Rotorua pulls up the proportion of Maori in the combined jury pool. The Simpson's Paradox takes place due to uneven number of Jury Pool proportion size.

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 3, Issue 3, ISSN No. 2455-2143, Pages 62-68
Published Online July 2018 in IJEAST (http://www.ijeast.com)

Considering overall population size, the number of people are overall similar, but Rotorua provides 337 to the jury pool, while Nelson contributes only 57, or under15 percent. This is the hidden confounding factor which causes the Simpsons' Paradox.

III. Experimental Study of the Risks of Simpson's Paradox

Medical Recommendations are pivotal for a patient therefore healthcare professionals need to be cautious regarding prescription of medicines.

	UTI	No UTI	% with UTI
Antibiotic prophylaxis	42	1237	3.3
No antibiotic prophylaxis	104	2136	4.6

Fig. 13. UTI & No UTI division on Antibiotic Prophylaxis

Let's consider a healthcare professional gauging whether to advise antibiotics or not in respect to UTI (Urinary Tract Infections) in hospital patients.
Data summarized shows eight hospitals
in the Netherlands for the number of patients who receive and do not receive prophylactic antibiotics (PAB). It is observed that patients who have received PAB have a lower rate of UTI compared to those who haven't received the treatment. (3.3% versus 4. 6%).According to the statistics, it seems that PAB is effective and should be used to prevent UTI. However, upon further inspection, the results can vary. The
hospitals can be divided into groups – low-incidence hospitals(LIH) and high- incidence hospitals(HIH), on an average value of 2.5%.

	UTI	No UTI	% with UTI
Low-incidence hospitals			
Antibiotic prophylaxis	20	1093	1.8
No antibiotic prophylaxis	5	715	0.7
High-incidence hospitals			
Antibiotic prophylaxis	22	144	13.3
No antibiotic prophylaxis	99	1421	6.5

Fig. 14. Reproduction of Fig 13 on Type of Hospitals

This chart shows that PAB is positively associated with
an augmented rate of UTI in both groups, and that it is
damaging to the patients who received it. The conclusion is that use of PAB shows negative of what was thought before. If further analysis wouldn't have been performed and the first conclusion would have deemed as correct. Implementation of sale of PAB to prevent UTI would have taken place at a larger scale in turn augmenting the chances of the same issue it was exercised to eradicate. This experimental study explains that an in-depth analysis is very crucial and avoiding impetuous decisions. Simpson's Paradox is one of the many, complex statistical anomalies which if undetected can create erroneous interpretations and mislead the readers. Thorough analysis can aid researchers in prevention of this. It is essential to consider the whole picture while imparting each aspect equal significance.

IV. CONCLUSION

Simpson's Paradox is a type of effect whose presence in because of confounding variables which distort the statistical findings in a way not visible at a macro level.
This makes it imperative for the researchers to implement analysis in a fashion where statistical assumptions which can uncover such buried factors that might mask or even alter the relationship between two or more variables (predictor and outcome). Researchers have even suggested to exercise randomized trials to isolate the effects of multiple variables. It is also very crucial to involve a statistician in the team to get expertise on the same if any inconsistencies are introduced.

This paper gives a comprehensive knowledge of Simpson's Paradox with real life instances elucidating the effects and presenting ways to discover the hidden variable falsifying the results. The paper starts with a comprehensive explanation of Simpson's Paradox, how it was discovered and how it can influence data. It then moves on to comprehensive analysis of three Simpson's examples presenting charts formed by exercising Plotly.
The conclusion is an experimental study which showed the deleterious effects if Simpson's Paradox is left undetected.

V. REFERENCE

[1] Simpson E.H. (1951). "The Interpretation of Interaction in Contingency Tables"; Journal of the Royal Statistical Society. Series B Vol. 13 No. 2 (pp. 238-241).

[2] Goltz Heather Honoré (2010). "Yule-Simpson's Paradox in Research";Practical Assessment, Research & Evaluation, 15(15).

[3] Bickel P. J., Hammel E.A. ,O'Connell J.W.(1975), "Sex Bias in Graduate Admissions: Data from Berkeley"; Science, New Series, Vol. 187, No. 4175 (pp. 398-404).

[4] Wikipedia Contributors.(2018)."Simpson's Paradox"; Wikipedia, The Free Encyclopedia.

[5] Marley Sarah,"Hidden Data and Surviving a Sinking Ship: Simpson's Paradox"; Select-Statistics.

[6] Yule G. Udny.(1903)“Notes on the Theory of Association of Attributes in Statistics”; Biometrika Vol. 2, No. 2 (pp. 121-134)

[7] Kievit RogierA.(2013), “Simpson's paradox in psychological science: a practical guide”; Front. Psychol. 4:513. doi: 10.3389/fpsyg.2013.00513

[8] Hernán Miguel A , Clayton David, Keiding Niels.(2011)“The Simpson's paradox unraveled”;International Journal of Epidemiology, Volume 40, Issue 3, (pp 780–785)

[9] Hershbein Brad .(2015).“When average isn’t good enough: Simpson’s paradox in education and earnings”;Brookings.

[10] Wikipedia contributors.(2018) "RMS Titanic." ;Wikipedia, The Free Encyclopedia.

[11] Westbrooke Ian,(1997).“Simpson’s Paradox-An example in a New Zealand Survey of Jury Composition”;Statistics New Zealand Te Tari Tatau Wellington, New Zealand

[12] Norton H. James .Divine George,(2015).“Simpson's paradox and how to avoid it”; The Royal Statistical Society

[13] Pearl Judea.(2013).“Understanding Simpson’s Paradox”; The American Statistician.