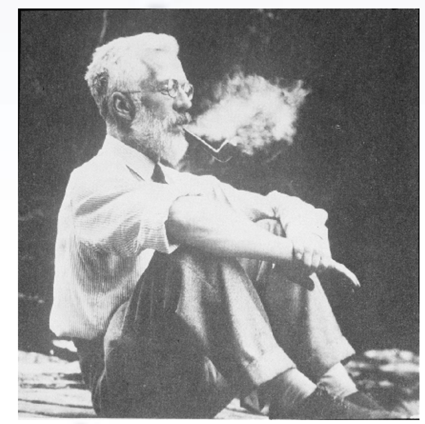
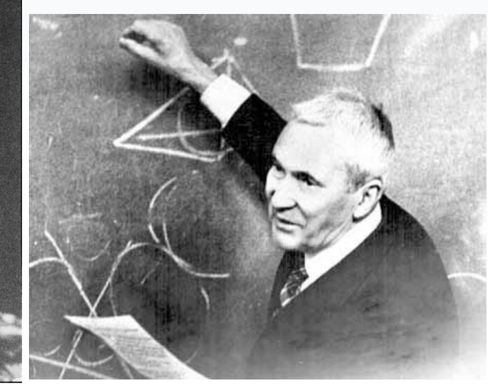
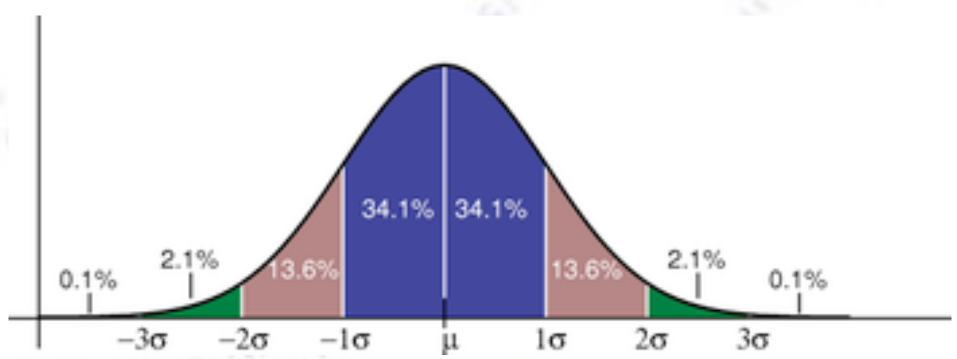
Applied Statistics

Simpson's Paradox

Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women are admitted.
Is that really the case, or is there more to the data than first glance reveals?

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women are admitted.
Is that really the case, or is there more to the data than first glance reveals?

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

https://homepage.stat.uiowa.edu/~mbognar/1030/Bickel-Berkeley.pdf

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women are admitted.
Is that really the case, or is there more to the data than first glance reveals?

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

https://homepage.stat.uiowa.edu/~mbognar/1030/Bickel-Berkeley.pdf

Table 1. Decisions on applications to Graduate Division for fall 1973, by sex of applicant—naive aggregation. Expected frequencies are calculated from the marginal totals of the observed frequencies under the assumptions (1 and 2) given in the text. $N = 12{,}763$, $\chi^2 = 110.8$, d.f. $= 1$, $P = 0$ (*18*).

Applicants	Outcome: Observed Admit	Outcome: Observed Deny	Outcome: Expected Admit	Outcome: Expected Deny	Difference Admit	Difference Deny
Men	3738	4704	3460.7	4981.3	277.3	− 277.3
Women	1494	2827	1771.3	2549.7	− 277.3	277.3

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women are admitted.
Is that really the case, or is there more to the data than first glance reveals?

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

https://homepage.stat.uiowa.edu/~mbognar/1030/Bickel-Berkeley.pdf

Table 1. Decisions on applications to Graduate Division for fall 1973, by sex of applicant—naive aggregation. Expected frequencies are calculated from the marginal totals of the observed frequencies under the assumptions (1 and 2) given in the text. $N = 12{,}763$, $\chi^2 = 110.8$, d.f. $= 1$, $P = 0$ (*18*).

Applicants	Outcome: Observed Admit	Outcome: Observed Deny	Outcome: Expected Admit	Outcome: Expected Deny	Difference Admit	Difference Deny
Men	3738	4704	3737 / (3738+4704) = **44.3%**		277.3	– 277.3
Women	1494	2827	1494 / (1494+2827) = **34.6%**		277.3	277.3

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women are admitted.
Is that really the case, or is there more to the data than first glance reveals?

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

P. J. Bickel, E. A. Hammel, J. W. O'Connell

https://homepage.stat.uiowa.edu/~mbognar/1030/Bickel-Berkeley.pdf

Table 1. Decisions on applications to Graduate Division for fall 1973, by sex of applicant—naive aggregation. Expected frequencies are calculated from the marginal totals of the observed frequencies under the assumptions (1 and 2) given in the text. $N = 12{,}763$, $\chi^2 = 110.8$, d.f. $= 1$, $P = 0$ (*18*).

Applicants	Outcome: Observed Admit	Observed Deny	Expected Admit	Expected Deny	Difference Admit	Difference Deny
Men	3738	4704	3737 / (3738+4704) = **44.3% ± ???**			– 277.3
Women	1494	2827	1494 / (1494+2827) = **34.6% ± ???**			277.3

Case: Berkeley admission

In 1973, University of California, Berkeley, were considering which of their applicants got admitted.
As can be seen below, there is seemingly a **bias against women**, as a smaller fraction of women

Is that rea

the data t

Sex Bias in Graduate Admissions: Data from Berkeley

Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation.

hel, J. W. O'Connell

)30/Bickel-Berkeley.pdf

As already noted, we are aware of the pitfalls ahead in this naive approach, but we intend to stumble into every one of them for didactic reasons.

Table 1. D … pplicant— naive aggregation. Expected frequencies are calculated from the marginal totals of the observed frequencies under the assumptions (1 and 2) given in the text. $N = 12{,}763$, $\chi^2 = 110.8$, d.f. $= 1$, $P = 0$ (*18*).

Applicants	Outcome: Observed Admit	Observed Deny	Expected Admit	Expected Deny	Difference Admit	Difference Deny
Men	3738	4704			277.3	− 277.3
Women	1494	2827			277.3	277.3

3737 / (3738+4704) = **44.3%**
1494 / (1494+2827) = **34.6%**

Case: Berkeley admission

Bickel et al. goes on to analyse the data further with several interesting findings:

> sex. Our computations, therefore, except where otherwise noted, will be based on the remaining 85. For a start let us identify those of the 85 with bias sufficiently large to occur by chance less than five times in a hundred. There prove to be four such departments. The deficit in the number of women admitted to these four (under the assumptions for calculating expected frequencies as given above) is 26. Looking further, we find six departments biased in the opposite direction, at the same probability levels; these account for a deficit of 64 men.

Out of 85 departments with relevant data, a few seem to show a bias… in both directions, and mostly agains men!!! What!

This seems counter intuitive to what we found to begin with. Where did the bias of 277 women less than expected go?

Case: Berkeley admission

Bickel et al. goes on to analyse the data further with several interesting findings:

sex. Our computations, therefore, except where otherwise noted, will be based on the remaining 85. For a start let us identify those of the 85 with bias sufficiently large to occur by chance less than five times in a hundred. There prove to be four such departments. The deficit in the number of women admitted to these four (under the assumptions for calculating expected frequencies as given above) is 26. Looking further, we find six departments biased in the opposite direction, at the same probability levels; these account for a deficit of 64 men.

Out of 85 departments with relevant data*, a few seem to show a bias… in both directions, and mostly agains men!!! What!

This seems counter intuitive to what we found to begin with. Where did the bias of 277 women less than expected go?

*Here you should ALWAYS ask, what this involves!
In this case, 16 departments either had no women applying, or did not deny any students admission.

Case: Berkeley admission

In order to illustrate the point, Bickel et al. gives a hypothetical (and fun!) case:

Table 2. Admissions data by sex of applicant for two hypothetical departments. For total, $\chi^2 = 5.71$, d.f. $= 1$, $P = 0.19$ (one-tailed).

Applicants	Outcome: Observed Admit	Observed Deny	Expected Admit	Expected Deny	Difference Admit	Difference Deny
Department of machismatics						
Men	200	200	200	200	0	0
Women	100	100	100	100	0	0
Department of social warfare						
Men	50	100	50	100	0	0
Women	150	300	150	300	0	0
Totals						
Men	250	300	229.2	320.8	20.8	− 20.8
Women	250	400	270.8	379.2	− 20.8	20.8

The two (very hypothetical) departments are clearly very fair regarding gender, but still a difference appears between the overall resulting observation and expectation.

Case: Berkeley admission

The "apparent conclusion" (Berkeley discriminates against applications from women) is a result of Simpson's Paradox (my text):

"Effect for group, which disappears or reverses, when considering subgroups".

It is effects such as this, which makes statistics difficult, yet at the same time **very important**.

> different degree. *The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into.* Moreover this phenomenon is more pronounced in departments with large numbers of applicants. Figure 1

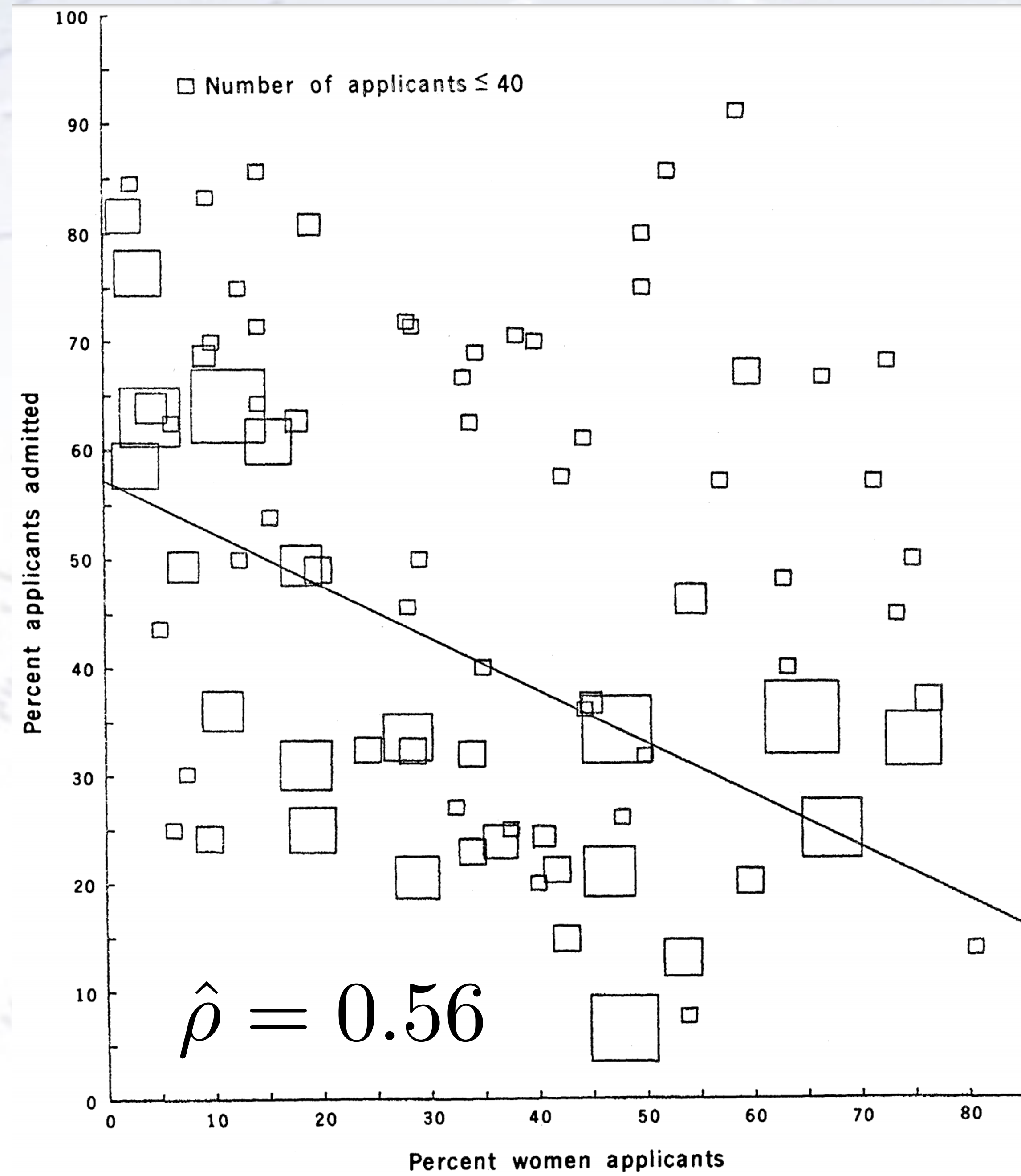


Fig. 1. Proportion of applicants that are women plotted against proportion of applicants admitted, in 85 departments. Size of box indicates relative number of applicants to the department.

Case: Berkeley admission

The "apparent conclusion" (Berkeley discriminates against applications from women) is a result of Simpson's Paradox (my text):

"Effect for group, which disappears or reverses, when considering subgroups".

It is effects such as this, which makes statistics difficult, yet at the same time **very important**.

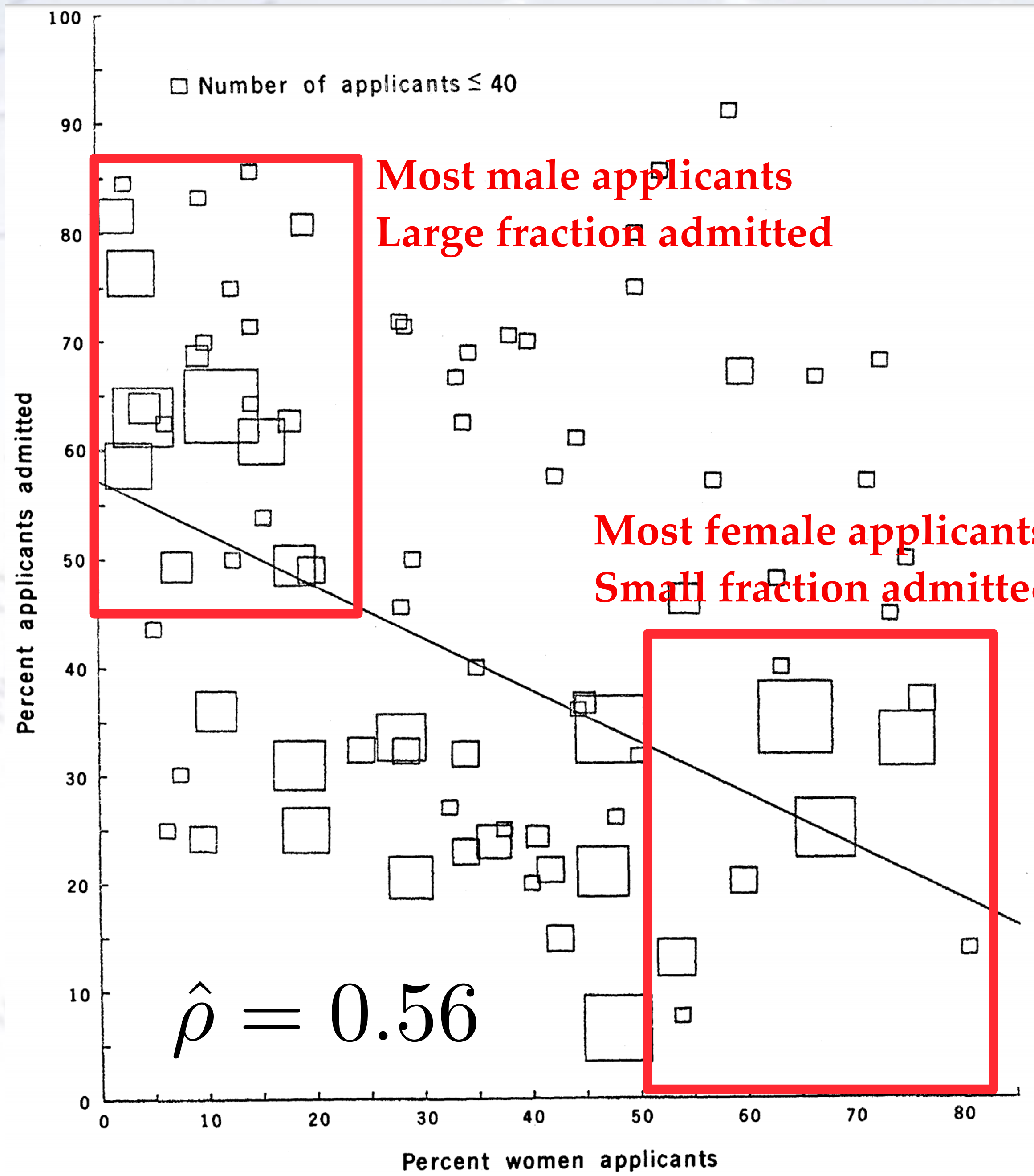


Fig. 1. Proportion of applicants that are women plotted against proportion of applicants admitted, in 85 departments. Size of box indicates relative number of applicants to the department.

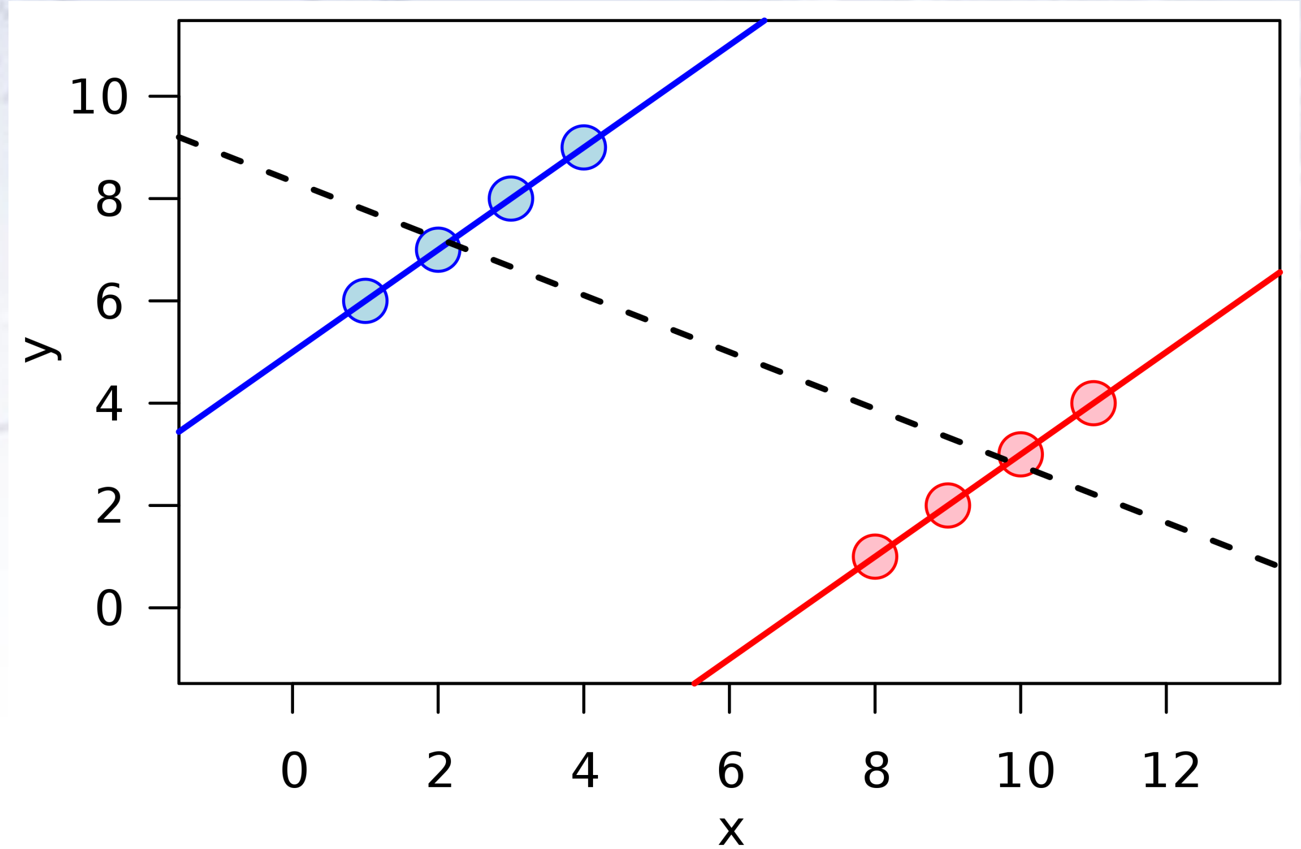
different degree. *The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into.* Moreover this phenomenon is more pronounced in departments with large numbers of applicants. Figure 1

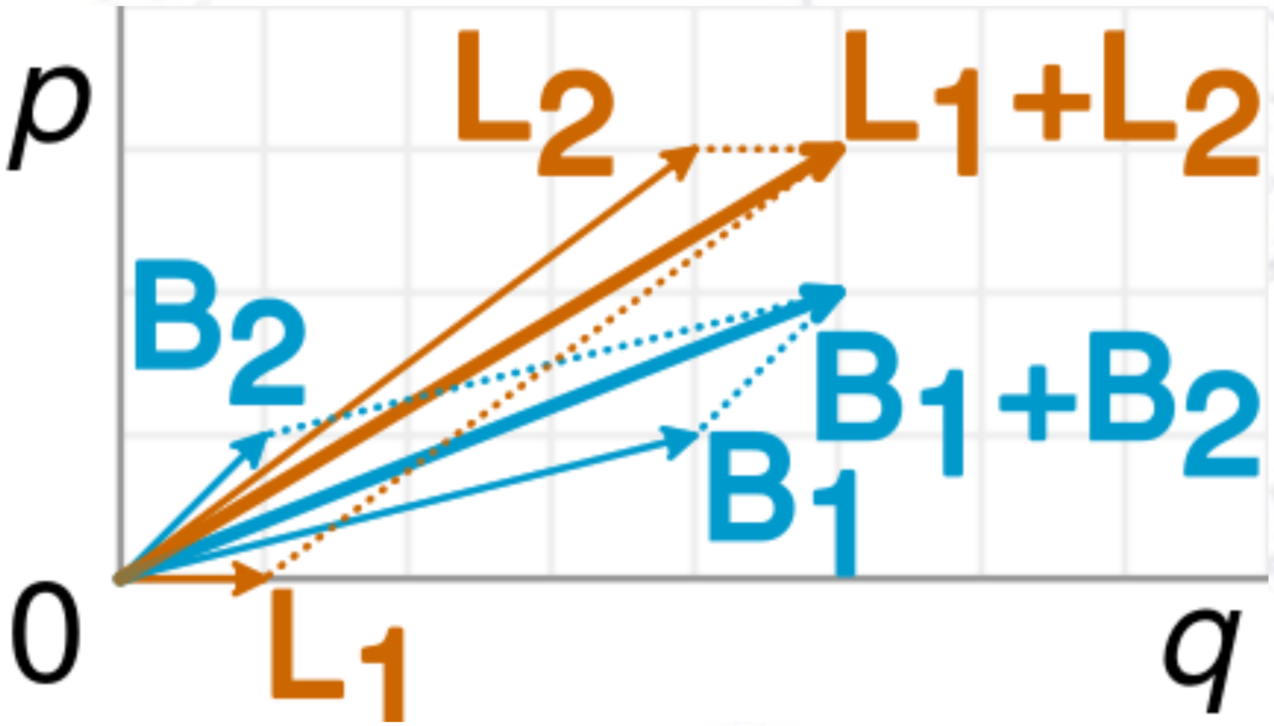
Simpson's Paradox explained

The reason for the **apparent** paradox arise when frequency data is unduly given causal interpretations.

The figure on the right illustrates the "paradox" nicely.

The situation can be illustrated with 2D vectors, as shown below.





A succes rate p/q (successes / attempts) can be represented by vectors with a slope. Higher slope = higher succes rate.

But though B1 is steeper than L1, and B2 is steeper than L2, then B1+B1 is not as steep as L1+L2.