Simpson Paradox

2023-10-14

Simpson's paradox is a statistical phenomenon in which a relationship between two variables in an overall population is reversed when the data are divided into subgroups. In other words, a seemingly positive correlation may turn out to be a negative correlation when the data is looked at more closely. This can be confusing because it goes against our intuitive expectations. It is important to note that Simpson's Paradox can occur due to the nature of the data and the interactions between variables, and is not due to an error in the analysis.

Example: Bokai WANG, Pan WU, Brian KWAN, Xin M. TU and Changyong FENG1

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5936043/

Suppose a certain disease can be characterized as being less severe or more severe. The patients have an option to go to either one of two hospitals for treatment: better or normal hospital. The outcome of the treatment is binary: success or failure. Consider the following example.

```
data
```

```
## # A tibble: 200 x 4
##       id hospital severity outcome
##    <int> <chr>    <chr>    <chr>
##  1     1 normal   less     failure
##  2     2 normal   less     success
##  3     3 better   less     success
##  4     4 normal   severe   failure
##  5     5 normal   less     failure
##  6     6 better   severe   failure
##  7     7 better   severe   success
##  8     8 better   severe   failure
##  9     9 better   severe   success
## 10    10 better   severe   failure
## # ... with 190 more rows
```

```
data %>%
  count(hospital, severity, outcome) %>%
  spread(key = outcome, value = n) %>%
  mutate(total = failure + success)
```

```
## # A tibble: 4 x 5
##   hospital severity failure success total
##   <chr>    <chr>      <int>   <int> <int>
## 1 better   less           2      18    20
## 2 better   severe        48      32    80
## 3 normal   less          16      64    80
## 4 normal   severe        16       4    20
```

We can see that for less severe patients, the success rate in the better treatment hospital is much higher than the normal hospital. Similar results hold true for more severe patients.

Cross-classification of the hospital and outcome

The overall success rates of two types of hospitals are 50/100 and 68/100, respectively. This seems to show that the success rate in the normal hospital is higher than the better hospital. This is not what we have expected.

```
data %>%
  count(hospital, outcome) %>%
  spread(key = outcome, value = n) %>%
  mutate(total = failure + success)
```

```
## # A tibble: 2 x 4
##   hospital failure success total
##   <chr>      <int>   <int> <int>
## 1 better        50      50   100
## 2 normal        32      68   100
```

Cross-classification of severity and the outcome

The success rates of less severe and more severe patients are 82/100 and 36/100, respectively. This is reasonable.

```
data %>%
  count(severity, outcome) %>%
  spread(key = outcome, value = n) %>%
  mutate(total = failure + success)
```

```
## # A tibble: 2 x 4
##   severity failure success total
##   <chr>      <int>   <int> <int>
## 1 less          18      82   100
## 2 severe        64      36   100
```

Cross-classification of hospital and severity

We can see that proportion of more severe patients in the better hospital is much higher than that in the normal hospital.

```
data %>%
  count(hospital, severity) %>%
  spread(key = severity, value = n) %>%
  mutate(total = less + severe)
```

```
## # A tibble: 2 x 4
##   hospital  less severe total
##   <chr>    <int>  <int> <int>
## 1 better      20     80   100
## 2 normal      80     20   100
```

Explanation

From the last 2 tables we see that the success rate for more severe patients is much lower than the less severe patients, and the portion of more severe patients in the better hospital is much more than that in normal hospital. This imbalance reverses the direction of hospital effect.

Further example from Simpson's paradox in psychological science: a practical guide Rogier A. Kievit, Willem E. Frankenhuis, Lourens J. Waldorp and Denny Borsboom

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3740239/

Example 1

A higher dosage of medicine may be associated with higher recovery rates at the population-level; however, within subgroups (e.g., for both males and females), a higher dosage may actually result in lower recovery rates. Figure 1 illustrates this situation: Even though a negative relationship exists between "Treatment Dosage" and "Recovery" in both males and females, when these groups are combined a positive trend appears (black, dashed). Thus, if analyzed globally, these data would suggest that a higher dosage treatment is preferable, while the exact opposite is true.

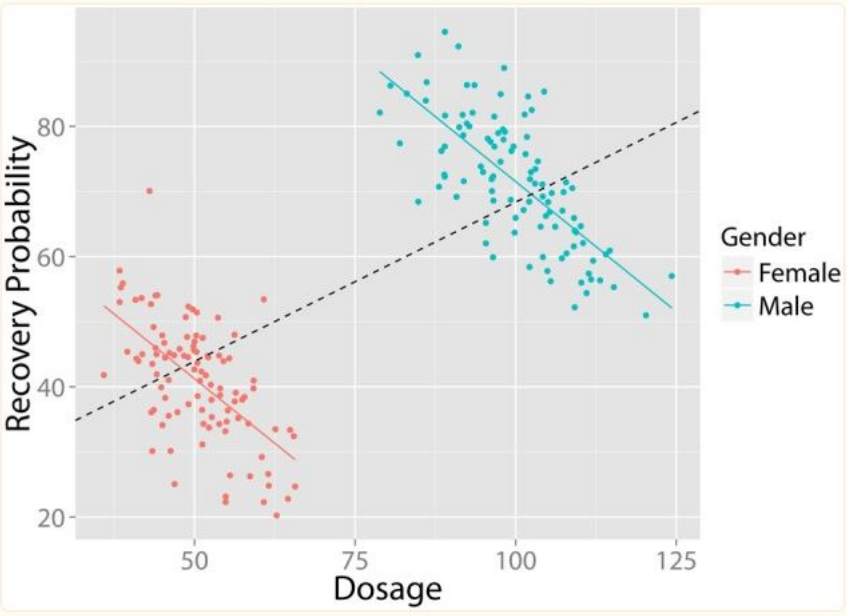

Figure 1

Figure 1: dosage and recovery

Example 2

Consider figure 3, which visualizes the relationship of data collected by a researcher studying the relationship between arousal (general level of activation of the central nervous system triggered by sensory stimuli and expressed as alertness, wakefulness, responsiveness or attentiveness) and performance on some athletic skill such as, say, tennis.

It seems that there is no significant association. However, imagine that the researcher now gains access to a large body of (previously inaccessible) additional data on the game statistics of each player: How many winning shots do they make, how many errors, how often do they hit with topspin or backspin, how hard do they hit the ball? Now imagine that using this new data, a cluster analysis on these additional variables, yielding two player types that we may label "aggressive" vs. "defensive". By including this additional (latent) grouping variable in our analysis, as can be seen in Figure Figure3B, we can see the value of latent clustering: In the aggressive players, there is a (significant) positive relationship between arousal and performance, whereas in the defensive players, there is a negative relationship between arousal and performance .

Figure 3

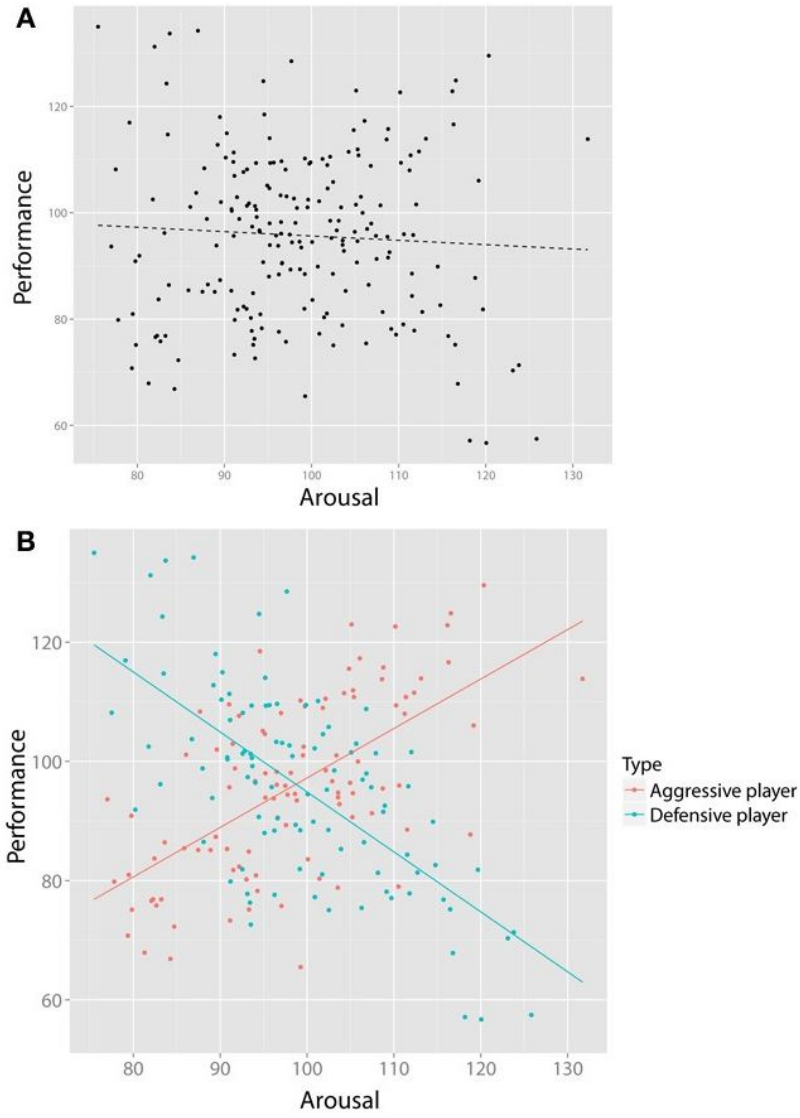

Figure 2: arousal and performance