

Simulation of Simpson's Paradox With Hospital Data

Andy Grogan-Kaylor

18 Sep 2020 08:46:40

Background

Simpson's paradox occurs when a bivariate association is reversed in a multivariate model. This example uses simulated data from a hospital. Source TBD.

Setup

```
. clear all

. cd "/Users/agrogan/Desktop/newstuff/categorical/contingency-tables/simpsons-paradox-hos
> pital-data"
/Users/agrogan/Desktop/newstuff/categorical/contingency-tables/simpsons-paradox-hospital-
> data

. use "hospitaldata.dta"
```

Outcome By Hospital Type

```
. list
```

	hospital	severity	outcome	count
1.	better	less severe	success	18
2.	better	less severe	failure	2
3.	better	more severe	success	32
4.	better	more severe	failure	48
5.	normal	less severe	success	64
6.	normal	less severe	failure	16
7.	normal	more severe	success	4
8.	normal	more severe	failure	16

```
. tabulate hospital outcome
```

hospital	outcome		Total
	failure	success	
better	2	2	4
normal	2	2	4
Total	4	4	8

```
. tabulate hospital outcome [fweight = count], row col
```

Key

<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

hospital	outcome		Total
	failure	success	
better	50	50	100
	50.00	50.00	100.00
	60.98	42.37	50.00
normal	32	68	100
	32.00	68.00	100.00
	39.02	57.63	50.00
Total	82	118	200
	41.00	59.00	100.00
	100.00	100.00	100.00

```
. graph bar [fweight = count], over(outcome) by(hospital) title(Hospital Outcomes) scheme
> (michigan)
```

```
. graph export bivariategraph.png, width(500) replace
(file bivariategraph.png written in PNG format)
```

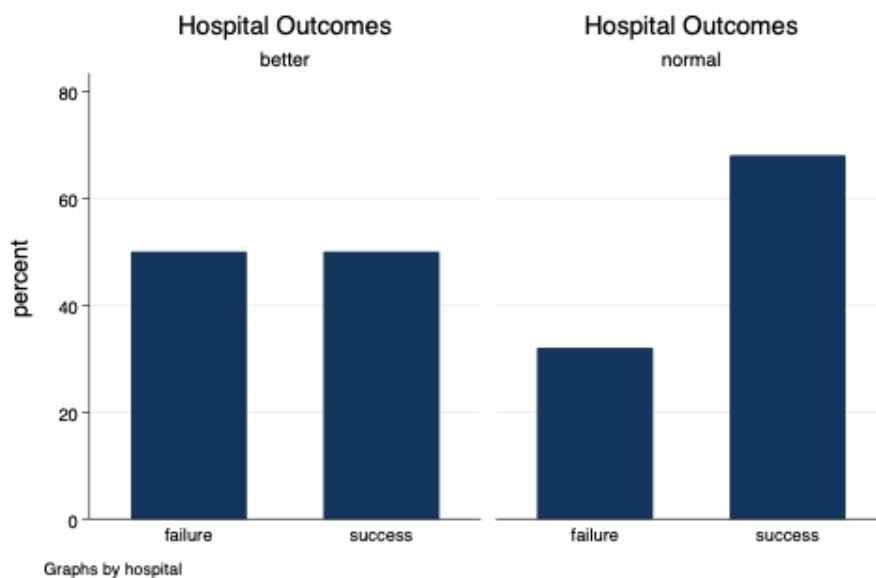


Figure 1: Bivariate Relationships

Outcome By Hospital Type by Severity

The Stata thinking and `command` for the table are relatively straightforward. We often think that visualization is an intuitive first step yet my first graphs were *wrong*, and I needed to *carefully look at the table* to make sure the graphs accurately reflected the numbers and story in the table!

```
. bysort severity: tabulate hospital outcome [fweight = count], row col
```

```
-> severity = less severe
```

Key
<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

hospital	outcome		Total
	failure	success	
better	2	18	20
	10.00	90.00	100.00
	11.11	21.95	20.00
normal	16	64	80
	20.00	80.00	100.00
	88.89	78.05	80.00
Total	18	82	100
	18.00	82.00	100.00
	100.00	100.00	100.00

-> severity = more severe

Key
<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

hospital	outcome		Total
	failure	success	
better	48	32	80
	60.00	40.00	100.00
	75.00	88.89	80.00
normal	16	4	20
	80.00	20.00	100.00
	25.00	11.11	20.00
Total	64	36	100
	64.00	36.00	100.00
	100.00	100.00	100.00

```
. graph bar [fweight = count] if severity == "less severe", ///
> title(Less Severe) ///
> over(outcome) ///
> by(hospital) ///
> scheme(michigan) ///
> name(lesssevere, replace)

. graph bar [fweight = count] if severity == "more severe", ///
> title(More Severe) ///
> over(outcome) ///
> by(hospital) ///
> scheme(michigan) ///
> name(moresevere, replace)

. graph combine lesssevere moresevere, title(Hospital Outcomes) scheme(michigan)

. graph export multivariategraph.png, width(1000) replace
(file multivariategraph.png written in PNG format)
```

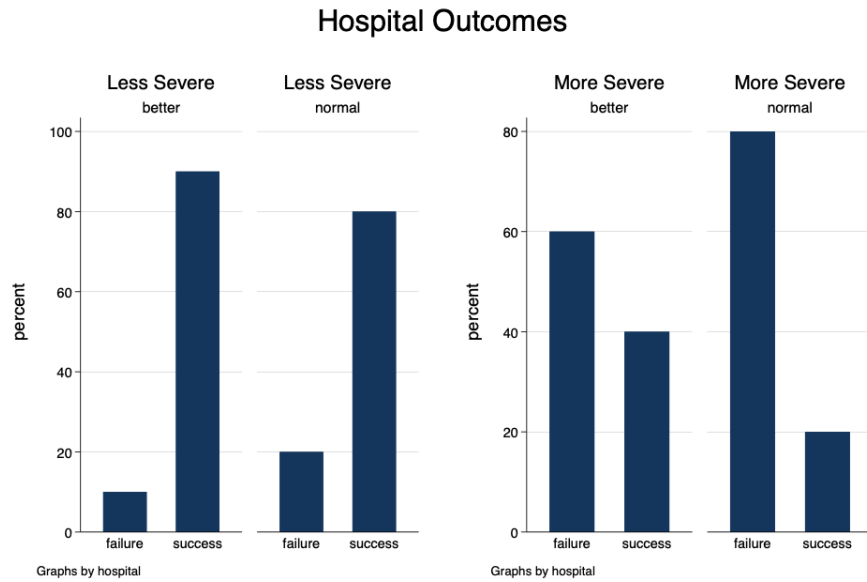


Figure 2: Multivariate Relationships

A Possibly Helpful Alternative (And More Compact) Table Format

```
. table hospital outcome [fweight=count], by(severity)
```

severity and hospital	outcome	
	failure	success
less severe		
better	2	18
normal	16	64
more severe		
better	48	32
normal	16	4