Causal Graphs for Basic Epidemiologic Data

Part 1—Overview of 2-way tables

Tomás J. Aragón

2019-11-22

Overview and case studies

Consider cohort studies with binomial (risk or prevalence) data (Table 1):

Table 1: Cohort study with binomial (risk or prevalence) data

	Exposed	Unexposed	Total
Cases	a	b	M_1
Noncases	c	d	M_0
Total at risk	N_1	N_0	T_i

$$R_0 = \frac{a}{N_1}, \qquad R_1 = \frac{b}{N_0}$$

Case study 1

Consider an observational study where 700 patients were given access to a new drug for an ailment. A total of 350 patients chose to take the drug and 350 patients did not. The patients were assessed for clinical recovery (Table 2).

Table 2: Recovery outcomes of 700 patients given access to a new drug

	Drug	No drug	Total
Recovered	273	289	562
No recovery	77	61	138
Total at risk	350	350	700

```
jpdrugrx <- read.csv("~/git/phds/data/jp-drugrx-p2.txt", header = TRUE, sep = ",")
tab <- xtabs(~Recovered + Drug, data = jpdrugrx)
r1 <- tab["Yes", "Yes"]/sum(tab[, "Yes"])
r0 <- tab["Yes", "No"]/sum(tab[, "No"])
```

```
list(data = tab, results = c(recovered.drug = r1, recovered.nodrug = r0))
```

```
## $data
##          Drug
## Recovered  No Yes
##       No   61  77
##       Yes 289 273
##
## $results
##   recovered.drug recovered.nodrug
##        0.7800000        0.8257143
```

Based on these results, patients who took the drug had worse recovery. However, another data analyst shows the following stratified table and analysis.

Table 3: Recovery outcomes of 700 patients given access to a new drug, Stratified by sex

	Men			**Women**		
	Drug	No drug	Total	Drug	No drug	Total
Recovered	81	234	315	192	55	247
No recovery	6	36	42	71	25	96
Total at risk	87	270	357	263	80	343

```
tab2 <- xtabs(~Gender + Drug + Recovered, data = jpdrugrx)
r1.men <- tab2['Men','Yes','Yes']/sum(tab2['Men','Yes',])
r0.men <- tab2['Men','No','Yes']/sum(tab2['Men','No',])
r1.wom <- tab2['Women','Yes','Yes']/sum(tab2['Women','Yes',])
r0.wom <- tab2['Women','No','Yes']/sum(tab2['Women','No',])
results.men = c(recovered.drug = r1.men, recovered.nodrug = r0.men)
results.wom = c(recovered.drug = r1.wom, recovered.nodrug = r0.wom)
list(data = ftable(tab2), men = results.men, women = results.wom)
```

```
## $data
##             Recovered  No Yes
## Gender Drug
## Men    No              36 234
##        Yes              6  81
## Women  No              25  55
##        Yes             71 192
##
## $men
##   recovered.drug recovered.nodrug
##        0.9310345        0.8666667
##
```

```
## $women
##   recovered.drug recovered.nodrug
##         0.730038         0.687500
```

We have potentially conflicting results. Men who took the drug recovered better than men who did not take the drug. Likewise, women who took the drug recovered better than women who did not take the drug.

> **Exercise:** Assuming all results are "statistically significant," which results do you believe and why? What recommendation would you make?

Case study 2

Consider a treatment study with 700 patients, half of whom were assigned a new drug for their ailment. At the end of the study the patients were assessed for clinical recovery, and their blood pressure was measured.

Table 4: Recovery outcomes of 700 patients treated with a new drug

	No Drug	Drug	Total
Recovered	273	289	562
No recovery	77	61	138
Total at risk	350	350	700

```
jpdrugrx2 <- read.csv("~/git/phds/data/jp-drugrx-p4.txt", header = TRUE, sep = ",")
tab <- xtabs(~Recovered + Drug, data = jpdrugrx2)
r1 <- tab["Yes", "Yes"]/sum(tab[, "Yes"])
r0 <- tab["Yes", "No"]/sum(tab[, "No"])
list(data = tab, results = c(recovered.drug = r1, recovered.nodrug = r0))
```

```
## $data
##           Drug
## Recovered  No Yes
##       No   77  61
##       Yes 273 289
##
## $results
##   recovered.drug recovered.nodrug
##        0.8257143        0.7800000
```

Based on these results, patients who took the drug had better recovery. However, another data analyst shows the following stratified table and analysis.

Table 5: Treatment study recovery outcomes of 700 patients, of which 350 received a new drug, Stratified by blood pressure

	Low BP			**High BP**		
	No drug	Drug	Total	No drug	Drug	Total
Recovered	81	234	315	192	55	247
No recovery	6	36	42	71	25	96
Total at risk	87	270	357	263	80	343

```
tab2 <- xtabs(~BP + Drug + Recovered, data = jpdrugrx2)
r1.lbp <- tab2["Low", "Yes", "Yes"]/sum(tab2["Low", "Yes", ])
```

```r
r0.lbp <- tab2["Low", "No", "Yes"]/sum(tab2["Low", "No", ])
r1.hbp <- tab2["High", "Yes", "Yes"]/sum(tab2["High", "Yes", ])
r0.hbp <- tab2["High", "No", "Yes"]/sum(tab2["High", "No", ])
results.lbp = c(recovered.drug = r1.lbp, recovered.nodrug = r0.lbp)
results.hbp = c(recovered.drug = r1.hbp, recovered.nodrug = r0.hbp)
list(data = ftable(tab2), Low.BP = results.lbp, High.BP = results.hbp)
```

```
## $data
##           Recovered  No Yes
## BP   Drug
## High No              71 192
##      Yes             25  55
## Low  No               6  81
##      Yes             36 234
##
## $Low.BP
##   recovered.drug recovered.nodrug
##        0.8666667        0.9310345
##
## $High.BP
##   recovered.drug recovered.nodrug
##         0.687500         0.730038
```

We have potentially conflicting results. When stratified by blood pressure, patients that did not take the drug recovered better. In fact, lower blood pressure was associated with better recovery. These are opposite results compared to the combined table.

> **Exercise:** Assuming all results are "statistically significant," which results do you believe and why? What recommendation would you make?