PSY30100-03 -- Assignment 10

Two-Way Tables

TA: Laura Lu

April 26, 2010

A joke

- Here is a joke I heard days ago.

  "*Every human being's brain has two parts: the left and the right. The left brain has nothing right. The right brain has nothing left.*"

Problem

There are a two-way table:

	Smoker	Not
Dead	139	230
Alive	438	502

Questions

1. Row variable? Column variable?
2. Marginal distributions in percentage for both variables.
3. Conditional distributions of survival given smoking status.
4. Carry out a test on relationship between the two variables.
5. Conclusion.

Ans. of Q.1-2

		Smoking status		marg. dist.	
		Smoker	Not	Total	Per.
Survival status	Dead	139	230	369	28.2%
	Alive	438	502	940	71.8%
marg. dist.	Total	577	732	1309	
	Per.	44.1%	55.9%		100%

Ans. of Q.3

Conditional distributions of survival given smoking status

	Smoker		Not
Dead	24.1%	<	31.4%
Alive	75.9%	>	68.6%
total	100%		100%

Ans. of Q.4

$$Expected\ Cell\ Count = \frac{rwo\ tatal \times column\ total}{total\ count}$$

	Smoker	Not	Total
Dead	139 (E=162.7)	230 (E=206.3)	369
Alive	438 (E=414.3)	502 (E=525.7)	940
Total	577	732	1309

Ans. of Q.4

$$\sum_{all\ cells} \frac{(Observed - Expected)^2}{Expected}$$

$$= \frac{(139-162.7)^2}{162.7} + \frac{(230-206.3)^2}{206.3}$$

$$+ \frac{(438-414.3)^2}{414.3} + \frac{(502-525.7)^2}{525.7}$$

$$\approx 8.6$$

The table is 2×2 , so there are (2-1)(2-1)=1 degrees of freedom.

And the critical value for χ_1^2 at alpha=0.05 is 3.84.

Therefore, we reject H0.

Ans. of Q.5

- Conclusions:

- 1. Smoking status is significantly associate with survival status at the alpha level of 0.05.
- 2. Smokers have a lower death rate than non-smokers.

Ans. of Q.6: Age 18-44

		Smoking status		marg. dist.	
		Smoker	Not	Total	Per.
Survival status	Dead	19	13	32	5.1%
	Alive	269	327	596	94.9%
marg. dist.	Total	288	340	628	
	Per.	45.9%	54.1%		100%

Ans. of Q.6: Age 18-44

Conditional distributions of survival given smoking status

	Smoker		Not
Dead	6.6%	>	3.8%
Alive	93.4%	<	96.2%
total	100%		100%

Ans. of Q.6: Age 45-64

		Smoking status		marg. dist.	
		Smoker	Not	Total	Per.
Survival status	Dead	78	52	130	29.6%
	Alive	162	147	309	70.4%
marg. dist.	Total	240	199	439	
	Per.	54.7%	45.3%		100%

Ans. of Q.6: Age 45-64

Conditional distributions of survival given smoking status

	Smoker		Not
Dead	32.5%	>	26.1%
Alive	67.5%	<	73.9%
total	100%		100%

Ans. of Q.6: Age 65+

		Smoking status		marg. dist.	
		Smoker	Not	Total	Per.
Survival status	Dead	42	165	207	85.5%
	Alive	7	28	35	14.5%
marg. dist.	Total	49	193	242	
	Per.	20.2%	79.8%		100%

Ans. of Q.6: Age 65+

Conditional distributions of survival given smoking status

	Smoker		Not
Dead	85.7%	>	85.5%
Alive	14.3%	<	14.5%
total	100%		100%

Ans. of Q.7

- Conclusion:

- For each age subgroup, non-smokers have a lower death rate than smokers.

Ans. of Q.8

- Simpson's Paradox!

- The variable age is a lurking variable in this problem.

Ans. of Q.9

- Take-home message:
- Lurking variables can strongly influence relationships between categorical variables. An association or comparison that holds for all subgroups can reverse direction when lurking variables are ignored and the data are combined to form a single group.
- In this problem, when age is ignored, smokers seem to have a lower death rate, even though non-smokers do have a lower death rate in each of three age subgroups.

Another take-home message:

- Stay away from smoking, drugs, or other unhealthy habits. I hope you have a healthy body, and enjoy your life.

- Although we don't have perfect brains, I do hope you have perfect scores on your final examination.