

Stat 101L: Lecture 39

Categorical Data

- * In Chapter 3 we introduced the idea of categorical data.
- * In Chapter 15 we explored probability rules and when events are independent.
- * In Chapter 26 we put these two ideas together to compare counts.

1

Categorical Data

- * National Opinion Research Center's General Social Survey
- * In 1996 a sample of 1895 adults in the U.S. were asked the question "When is premarital sex wrong?" The participants were also asked with what religion they were affiliated.

2

Who?/What?

- * Who?
 - A sample of 1895 adults.
- * What?
 - Attitude towards premarital sex.
 - Religious affiliation.

3

Stat 101L: Lecture 39

What?

- * When is premarital sex wrong?
 - Categorical: Always, Almost Always, Sometimes, Never
- * What is your religious affiliation?
 - Categorical: Catholic, Jewish, Protestant, None, Other

4

When is Premarital Sex Wrong?

Religion	Always	Almost Always	Sometimes	Never	Total
Catholic	62	37	120	226	445
Jewish	0	3	14	34	51
Protestant	355	117	227	384	1083
None	20	13	45	147	225
Other	15	13	23	40	91
Total	452	183	429	831	1895

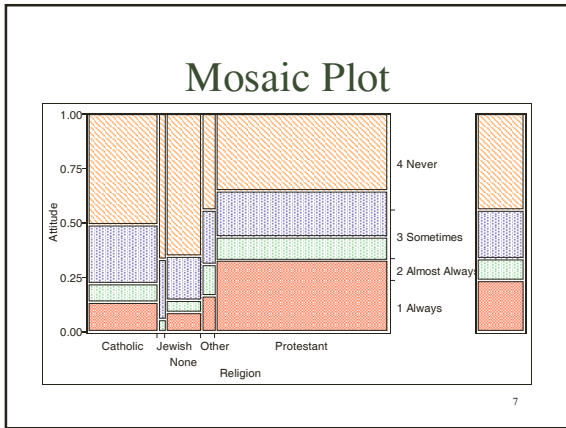
5

When is Premarital Sex Wrong?

Religion	Always	Almost Always	Sometimes	Never	Total
Catholic	13.9%	8.3%	27.0%	50.8%	100%
Jewish	0.0%	5.9%	27.4%	66.7%	100%
Protestant	32.8%	10.8%	21.0%	35.4%	100%
None	8.9%	5.8%	20.0%	65.3%	100%
Other	16.5%	14.3%	25.3%	43.9%	100%

6

Stat 101L: Lecture 39



Comparing Counts

- * People who have no religion or are Jewish are more likely to say premarital sex is never wrong.
- * Protestants are much more likely to say premarital sex is always wrong.

8

Comparing Counts

- * Are these differences statistically significant?
- * Or, are these differences due to chance variation that is religion and attitude towards premarital sex are independent?

9

Stat 101L: Lecture 39

Comparing Counts

*If religion and attitude towards premarital sex are independent then

$$\Pr(A \text{ and } B) = \Pr(A) * \Pr(B)$$

where A is a religion category and B is an attitude category.

10

Expected Count

*If religion and attitude toward premarital sex are independent we would expect to see

$$n * \Pr(A) * \Pr(B)$$

people in the religion category A and the attitude category B.

11

Religion	Always	Almost Always	Sometimes	Never	Total
Catholic					445
Jewish					51
Protestant					1083
None					225
Other					91
Total	452	183	429	831	1895

12

Stat 101L: Lecture 39

Expected Count

*Catholic and Always

$$E = 1895 * \frac{445}{1895} * \frac{452}{1895}$$

$$E = \frac{445 * 452}{1895} = 106.1$$

13

Expected Counts

Religion	Always	Almost Always	Sometimes	Never	Total
Catholic	106.1	43.0	100.8	195.1	445
Jewish	12.2	4.9	11.5	22.4	51
Protestant	258.3	104.6	245.2	474.9	1083
None	53.7	21.7	50.9	98.7	225
Other	21.7	8.8	20.6	39.9	91
Total	452	183	429	831	1895

14

Observed = Expected?

- *Take the difference between the observed and expected counts in a cell.
- *Square the difference.
- *Divide by the expected count.
- *Sum up over all the cells.

15

Stat 101L: Lecture 39

Chi-square Test Statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$
$$df = (r - 1) * (c - 1)$$

16

Cell contributions to χ^2

*Catholic and Always

$$\frac{(62 - 106.1)^2}{106.1} = \frac{(-44.1)^2}{106.1} = 18.33$$

17

Test of Independence

- * H_0 : Religion and attitude towards premarital sex are independent.
- * $\chi^2 = 157.017$, $df = (5-1)*(4-1) = 12$
- * P-value < 0.0001
- * Because the P-value is so small we reject the null hypothesis.
- * Religion and Attitude towards premarital sex are not independent.

18
