

Monday, Apr 25

Cramer's V

Consider the following data where a sample of 1398 children were classified with respect to tonsil size and carrier status of *Streptococcus pyogenes*.¹

Size	Carrier		Total
	yes	no	
small	19 (0.04)	497 (0.96)	516
medium	29 (0.05)	560 (0.95)	589
large	24 (0.08)	269 (0.92)	293
Total	72 (0.05)	1326 (0.95)	1398

The numbers in parentheses are the *proportions* of children of each tonsil size who are or are not carriers. The value of the test statistic for a test of independence is $X^2 \approx 7.88$. We might decide tonsil size and carrier status are dependent. But can we *measure* the amount of dependence?

Cramer's V is a *measure of association* between two categorical variables. It is defined as

$$V = \sqrt{\frac{X^2/n}{\min(r-1, c-1)}},$$

and is bounded such that $0 \leq V \leq 1$. It effectively measures the degree to which the observed counts deviate from the expected counts under the assumption of independence, and thus can be viewed as a measurement of the degree of dependence.

Example: For the data on tonsil size and carrier status, the value of Cramer's V is

$$\sqrt{\frac{7.88/1398}{\min(3-1, 2-1)}} \approx 0.08,$$

which is a relatively *weak* association.

¹Holmes, M. C. & Williams, R. E. O. (1954). The distribution of carriers of *Streptococcus pyogenes* among 2413 healthy children. *Journal of Hygiene*, 52, 165–179.

Here are some hypothetical observed counts showing a *stronger* association.

Size	Carrier		Total
	yes	no	
small	25 (0.05)	475 (0.95)	500
medium	90 (0.15)	510 (0.85)	600
large	75 (0.25)	225 (0.75)	300
Total	190 (0.14)	1210 (0.86)	1400

The test statistic is $X^2 \approx 65.77$ and the measure of association is $V \approx 0.22$.

And here are some hypothetical observed counts showing an *even stronger* association.

Size	Carrier		Total
	yes	no	
small	25 (0.05)	475 (0.95)	500
medium	300 (0.5)	300 (0.5)	600
large	285 (0.95)	15 (0.05)	300
Total	610 (0.44)	790 (0.56)	1400

The test statistic is $X^2 \approx 635.36$ and the measure of association is $V \approx 0.67$.

Here is an example of the *maximum degree of association*.

Size	Carrier		Total
	yes	no	
small	0 (0)	500 (1)	500
medium	0 (0)	600 (1)	600
large	300 (1)	0 (0)	300
Total	300 (0.21)	1100 (0.79)	1400

The test statistic is $X^2 = 1400$ and the measure of association is $V = 1$.

What about the *minimum* degree of association?

Size	Carrier		Total
	yes	no	
small	25 (0.05)	475 (0.95)	500
medium	30 (0.05)	570 (0.95)	600
large	15 (0.05)	285 (0.95)	300
Total	70 (0.05)	1330 (0.95)	1400

The test statistic is $X^2 = 0$ and the measure of association is $V = 0$. Here the observed counts would be *equal* to the expected counts.

Example: Consider the following data from a randomized experiment comparing two strategies for chemotherapy.²

Strategy	Tumor Response				Total
	progressive disease	no change	partial remission	complete remission	
sequential	32 (0.21)	57 (0.38)	34 (0.23)	28 (0.19)	151
alternating	53 (0.36)	51 (0.34)	23 (0.16)	21 (0.14)	148
Total	85 (0.28)	108 (0.36)	57 (0.19)	49 (0.16)	299

The test statistic is $X^2 \approx 8.62$ and the measure of association is $V \approx 0.17$.

Here is a weaker association with some hypothetical observed counts.

Strategy	Tumor Response				Total
	progressive disease	no change	partial remission	complete remission	
sequential	40 (0.2)	80 (0.4)	50 (0.25)	30 (0.15)	200
alternating	50 (0.25)	84 (0.42)	42 (0.21)	24 (0.12)	200
Total	90 (0.22)	164 (0.41)	92 (0.23)	54 (0.14)	400

The test statistic is $X^2 \approx 2.57$ and the measure of association is $V \approx 0.08$.

Here is a stronger association with some hypothetical observed counts.

Strategy	Tumor Response				Total
	progressive disease	no change	partial remission	complete remission	
sequential	60 (0.3)	100 (0.5)	20 (0.1)	20 (0.1)	200
alternating	20 (0.1)	40 (0.2)	100 (0.5)	40 (0.2)	200
Total	80 (0.2)	140 (0.35)	120 (0.3)	60 (0.15)	400

The test statistic is $X^2 \approx 105.71$ and the measure of association is $V \approx 0.51$.

Here is a very strong association with some hypothetical observed counts.

Strategy	Tumor Response				Total
	progressive disease	no change	partial remission	complete remission	
sequential	90 (0.45)	100 (0.5)	6 (0.03)	4 (0.02)	200
alternating	2 (0.01)	18 (0.09)	100 (0.5)	80 (0.4)	200
Total	92 (0.23)	118 (0.3)	106 (0.26)	84 (0.21)	400

The test statistic is $X^2 \approx 293.28$ and the measure of association is $V \approx 0.86$.

²Holtbrugge, W. & Schumacher, M. (1991). A comparison of regression models for the analysis of ordered categorical data. *Applied Statistics*, 40, 249–259.

McNemar's Test for Matched Pairs

Example: A retrospective case-control study was used to investigate the theory that tonsils protect the body against the invasion of the lymph nodes by the virus responsible for Hodgkin's disease. The study compared patients with Hodgkin's disease (the cases) with their siblings without the disease (the controls) with respect to whether or not they had a tonsillectomy in the past.³

Pair	Patient (Case)	Sibling (Control)
1	no tonsillectomy	tonsillectomy
2	no tonsillectomy	no tonsillectomy
3	no tonsillectomy	no tonsillectomy
4	tonsillectomy	tonsillectomy
5	no tonsillectomy	no tonsillectomy
6	tonsillectomy	no tonsillectomy
7	no tonsillectomy	no tonsillectomy
8	tonsillectomy	no tonsillectomy
9	no tonsillectomy	no tonsillectomy
10	tonsillectomy	tonsillectomy
\vdots	\vdots	\vdots
85	tonsillectomy	tonsillectomy

Each *pair* can be classified in terms of whether or not the patient (case) had a tonsillectomy, and whether or not the sibling (control) had a tonsillectomy.

Patient (Case)	Sibling (Control)		Total
	tonsillectomy	no tonsillectomy	
tonsillectomy	26	15	41
no tonsillectomy	7	37	44
Total	33	52	85

Let p_p be the probability that the patient (case) had a tonsillectomy, and let p_s be the probability that the sibling (control) had a tonsillectomy. How can we test the null hypothesis $H_0: p_p = p_s$ versus $H_a: p_p \neq p_s$? It would be tempting to use the test statistic

$$z = \frac{\hat{p}_p - \hat{p}_s}{\sqrt{\hat{p}(1 - \hat{p})(1/n_p + 1/n_s)}},$$

where $\hat{p}_p = 41/85$, $\hat{p}_s = 33/85$, $n_p = 85$, $n_s = 85$, and $\hat{p} = (41 + 33)/(85 + 85)$. However this test statistic assumes that the samples are *independent*, but they are *not independent*.

³Johnson, S. K. & Johnson, R. E. (1972). Tonsillectomy history in Hodgkin's disease. *New England Journal of Medicine*, 287, 1122–1125.

Derivation of McNemar's Test Statistic

Let p_a , p_b , p_c , and p_d denote the probabilities of each of the four possible sibling pairs.

Patient (Case)	Sibling (Control)	
	tonsillectomy	no tonsillectomy
tonsillectomy	p_a	p_b
no tonsillectomy	p_c	p_d

So the probability that the *patient* had a tonsillectomy is

$$p_p = p_a + p_b,$$

and the probability that the *sibling* had a tonsillectomy is

$$p_s = p_a + p_c.$$

If the null hypothesis is true then

$$p_p = p_s \Rightarrow p_a + p_b = p_a + p_c \Rightarrow p_b = p_c.$$

Now we don't know p_b or p_c , but we can estimate them from the observed counts,

Patient (Case)	Sibling (Control)		Total
	tonsillectomy	no tonsillectomy	
tonsillectomy	26	15	41
no tonsillectomy	7	37	44
Total	33	52	85

The estimates of p_b and p_c are obtained by averaging the corresponding proportions because we assume that $p_b = p_c$ under the null hypothesis so that

$$\hat{p}_b = \frac{7/85 + 15/85}{2}, \quad \hat{p}_c = \frac{7/85 + 15/85}{2}.$$

Now these estimates can be used to compute two of the (estimated) expected counts corresponding to the bottom-left and top-right cells.

$$n \times \hat{p}_b = 85 \times \frac{7/85 + 15/85}{2} = \frac{7 + 15}{2} = 11, \quad n \times \hat{p}_c = 85 \times \frac{7/85 + 15/85}{2} = \frac{7 + 15}{2} = 11.$$

The null hypothesis does not imply anything about the expected counts for the top-left and top-right cells, so we just use the observed counts as estimates of the expected counts in those cells. The expected counts for each cell are shown in the table below.

Patient (Case)	Sibling (Control)		Total
	tonsillectomy	no tonsillectomy	
tonsillectomy	26	$(7+15)/2$	41
no tonsillectomy	$(7+15)/2$	37	44
Total	33	52	85

Now plugging the observed and expected counts into the formula for X^2 gives us

$$X^2 = \frac{(26 - 26)^2}{26} + \frac{[15 - (7 + 15)/2]^2}{(7 + 15)/2} + \frac{[7 - (7 + 15)/2]^2}{(7 + 15)/2} + \frac{(37 - 37)^2}{37} \approx 2.91.$$

Some algebra will show that this can be simplified considerably to

$$X^2 = \frac{(7 - 15)^2}{7 + 15} \approx 2.91.$$

In general, we can write the test statistic as

$$X^2 = \frac{(O_{bl} - O_{tr})^2}{O_{bl} + O_{tr}}$$

where O_{bl} and O_{tr} denote the bottom-left and top-right *observed counts*, respectively. The degrees of freedom for computing the p -value is always 1.

Example: An enzyme-linked immunosorbent assay (ELISA) is an analytical biochemical procedure that can be used to detect the presence of antigens or antibodies, and so so it can be used to detect the presence of specific infections. A study applied two kinds of ELISA — a standard version and the ABC-ELISA — to each of 101 patients with hydatidosis (i.e., an infestation with *echinococcus*, a genus of tapeworms).⁴ Each test will give a *positive* or *negative* test result for the presence of the disease. Let p_{abc} and p_s be the probability that a ABC-ELISA and standard ELISA, respectively, will produce a positive result when applied to someone with the disease (this probability is called the *sensitivity* of the test). To determine if the two tests differ with respect to their sensitivity we could test the hypotheses $H_0: p_{abc} - p_s = 0$ versus $H_a: p_{abc} - p_s \neq 0$ using the test statistic

$$z = \frac{\hat{p}_{abc} - \hat{p}_s}{\sqrt{\hat{p}(1 - \hat{p})(1/n_{abc} + 1/n_s)}}.$$

This is what was done in the original analysis, but it was later pointed out that this analysis is incorrect because the two samples are *dependent* because both assays were applied to the same patients.⁵ So how can we test the hypotheses $H_0: p_{abc} - p_s = 0$ versus $H_a: p_{abc} - p_s \neq 0$?

ABC-ELISA	Standard ELISA		Total
	positive	negative	
positive	82	13	95
negative	6	0	6
Total	88	13	101

⁴Shen, Z. Q., Feng, X. H., Qian, Z. X., Liu, R. L., & Yang, C. R. (1988). Application of biotinadvin system, determination of circulating immune complexes, and evaluation of antibody response in different hydatidosis patients. *American Journal of Tropical Medicine and Hygiene*. 39, 93–96.

⁵Cruess, D. F. (1989). Review of use of statistics in The American Journal of Tropical Medicine and Hygiene for January–December 1988. *American Journal of Tropical Medicine and Hygiene*, 41, 619–626.

Example: In educational testing, a simple measure of the “easiness” of an item is the proportion of examinees that get the question correct (similarly, a measure of the “difficulty” of an item is the proportion of examinees that get the question incorrect). Suppose a test has two items, A and B , and let p_A and p_B denote the probability that a randomly selected examinee will get each item correct. The test was administered to $n = 1000$ examinees. The responses of the examinees to these two items are summarized in the table below.

Item A	Item B		Total
	correct	incorrect	
correct	400	125	525
incorrect	175	300	475
Total	575	425	1000

Now consider a test of the hypotheses $H_0: p_A - p_B = 0$ versus $H_a: p_A - p_B \neq 0$. However the two samples of responses — i.e., the samples of responses to item A and the sample of responses to item B — are not independent. How then do we test these hypotheses?