STAT 333
Section 4.4
Measures of Dependence for Contingency Tables

• We have seen measures of dependence for two numerical variables: for example, Pearson's and spearman's and Kendall's tau correlation coefficient.

• For categorical data summarized in a contingency table, we have seen how to test for dependence between rows and columns.

• Suppose we wish to measure the degree (or perhaps nature) of the dependence?

• The size of the chi-square test statistic T tells us something about the degree of dependence, but it is only meaningful relative to the degree of freedom .

Cramér's Contingency Coefficient

• A more easily interpretable measure of dependence than T is obtained by dividing T by its maximum possible value (for a given r and c).

• This **maximum** is N(q-1)

where q = the smaller of r or c

• The **square root of this ratio** is called Cramér's coefficient:

$$V = \sqrt{\frac{T}{N(q-1)}}$$

Interpretations: Cramér's coefficient takes values between 0 and 1.

• A value **near 0** indicates little association between row and column variables.

• A value **near 1** indicates strong dependence between row and column variables.

- Cramér's coefficient is scale-invariant: If the scope of the study were increased such that every cell in the table were multiplied by some constant, Cramér's coefficient remains the same.

Example 1, from Sec. 4.2:

	Score			
	Low	Marginal	Good	Excellent
Private	6	14	17	9
Public	30	32	17	3
				N=128

T was =17.29 N was =128 q is =2 (smaller c=4 ,r=2)

Cramér's coefficient =

$$V = \sqrt{\frac{T}{N(q-1)}} = \sqrt{\frac{17.29}{128\,(2-1)}} = \sqrt{\frac{17.29}{128\,(1)}} = 0.368$$

- We conclude there is moderate association between school type and score category.

- We can easily verify that Cramér's coefficient is unchanged if every cell count were multiplied by 10 (or any number).

Example 2, From Sec. 4.2:

		Snoring Pattern		
		Never	Occasionally	≈Every Night
Heart	Yes	24	35	51
Disease	No	1355	603	416
				N=2484

T was = 71.75 N was = 2484 q is =2 (smaller c=3,r=2)

Cramér's coefficient =

$$V = \sqrt{\frac{T}{N(q-1)}} = \sqrt{\frac{71.75}{2484\,(2-1)}} = \sqrt{\frac{71.75}{2484(1)}} = 0.17$$

-We conclude a mild association between heart disease and snoring pattern.

The Phi Coefficient

• While Cramér's coefficient measures the degree of association, it cannot reveal the type of association (positive or negative).

• The type of association is only meaningful when the two variables have corresponding categories.

• The table must be set up so that the row category ordering "matches" the column category ordering.

• Phi is calculated as the Pearson correlation coefficient between the row variable and the column variable, if the categories are coded as numbers.

• For a 2 × 2 table, phi $= \frac{ad - bc}{\sqrt{r_1 r_2 c_1 c_2}}$

Using

		Column		
		1	2	
Row	1	a	b	r_1
	2	c	d	r_2
		c_1	c_2	

Interpretations: The phi coefficient takes values between -1 and 1.

• A value **near 0** indicates :little association between row and column variable.

• A value **near +1** indicates :a strong tendency for observation to fall in "a like" categories for both rows and columns.

• A value **near –1** indicates : a strong tendency for observation to fall in "un like" categories for both rows and columns.

Example 3 **(Page 110 data tables):**

Table A:

Phi $=\frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{28x7-0x5}{\sqrt{28x12x33x7}} = 0.7035$

(strong tendency for mothers and fathers to have "alike" hair color)

		Father		
		Dark	Light	Total
Mother	Dark	28	0	28
	Light	5	7	12
	Total	33	7	40

Table B:

Phi $=\frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{21x0-7x12}{\sqrt{28x12x33x7}} =$ - 0.3015

(moderate tendency for mothers and fathers to have "unlike" hair colors.

		Father		
		Dark	Light	Total
Mother	Dark	21	7	28
	Light	12	0	12
	Total	33	7	40

Table C:

Phi $=\frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{23x2-5x10}{\sqrt{28x12x33x7}} =$ - 0.0144

(little association between mothers and fathers' to have "unlike" hair colors)

		Father		
		Dark	Light	Total
Mother	Dark	23	5	28
	Light	10	2	12
	Total	33	7	40

Example 4: Hair Color / Eye Color:

Phi = 0.341

(Moderate tendency for people with light eyes to have "like" light hair, and dark eyes to have "like" dark hair)

Proof that :

• For a 2×2 table, Phi equals Cramér's coefficient V times the sign of (ad-bc)

Proof:

For r = c =2 ,the χ^2 test statistic can be written as

$$T = \frac{N(ad-bc)^2}{r_1 r_2 c_1 c_2}$$

So,

$$V = \sqrt{\frac{T}{N(q-1)}} = \sqrt{\frac{\frac{N(ad-bc)^2}{r_1 r_2 c_1 c_2}}{N(q-1)}}$$

Since

q = 2 , q-1=2-1=1

$$V = \sqrt{\frac{(ad-bc)^2}{r_1 r_2 c_1 c_2}} = \frac{\sqrt{(ad-bc)^2}}{\sqrt{r_1 r_2 c_1 c_2}} = \frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}} = \text{phi}$$
