STAT 518 --- Section 4.4 --- Measures of Dependence for Contingency Tables

• We have seen measures of dependence for two numerical variables: for example, ___________ and ____________ correlation coefficient.

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

• Suppose we wish to measure the <u>degree</u> (or perhaps <u>nature</u>) 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 ________________________.

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

where *q* =

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

<u>Interpretations</u>: Cramér's coefficient takes values between ____ and _____.

• A value near 0 indicates

• A value near 1 indicates

• Cramér's coefficient is <u>scale-invariant</u>: 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, Sec. 4.2:

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

***T* was** **_N_ was** **_q_ is**

Cramér's coefficient =

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

Example 2, Sec. 4.2:

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

***T* was** **_N_ was** **_q_ is**

Cramér's coefficient =

The Phi Coefficient

- **While Cramér's coefficient measures the <u>degree</u> of association, it cannot reveal the <u>type</u> 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 ____________ correlation coefficient between the row variable and the column variable, if the categories are coded as numbers.**

- **For a 2 × 2 table,**

using

<u>Interpretations</u>: The phi coefficient takes values between ____ and _____.

• A value near 0 indicates

• A value near +1 indicates

• A value near –1 indicates

<u>Example 3</u> (Page 233-234 data tables):

Table A: Phi =

Table B: Phi =

Table C: Phi =

<u>Example 4</u>: Hair Color / Eye Color:

Phi =

• For a 2 × 2 table, Phi equals Cramér's coefficient *V* times the sign of

Section 4.6 --- Cochran's Test

• In Sec. 5.8 we learned that a block design is simply an extension of a matched-pairs design.

• Instead of each of a pair of similar subjects receiving one of two treatments, we have each of a block of similar subjects receiving one of *c* treatments.

• When the measurements can be ranked (ordinal or stronger data), we have studied nonparametric analyses of both paired and blocked designs.

• When the measurements are binary, we have studied nonparametric analyses of paired designs.

Recall:

• Now we study block designs with binary measurements. The data are arranged as:

• Since the data are binary, all X_{ij} are either:

Hypotheses of Cochran's Test:

H_0:

where p_j =

H_1:

Development of Cochran's Test Statistic

• Note that for large *r*, by the Central Limit Theorem, the *j*-th column sum C_j =

and so

we estimate $E(C_j)$ by

and estimate $\mathrm{var}(C_j)$ by

since under H_0,

So the test statistic is

• By estimating E(C_j) and var(C_j), we lose 1 degree of freedom, so the null distribution is χ^2 with _______ d.f.

• We reject H_0 when T is excessively _________.

Decision rule:

• The P-value is found through interpolation in Table A2 or using R.

Note: For c = 2 treatments, Cochran's Test is equivalent to __________________________.

<u>Example</u>: We test whether three rock climbs are equally easy. Five climbers attempted each of the three climbs, and their outcomes were recorded as 0 (failure) or 1 (success). **Data:**

H_0:

H_1:

Test statistic

Decision Rule and Conclusion:

P-value