İST 292 STATISTICS

Sections: 05-06

For Department of Computer Engineering

LESSON 7 ***Chi-Square Test for Contingency Tables (R×C Crosstabs)***

Dr. Ayten Yiğiter and Dr. Esra Polat Lecture Notes

Introduction

- This procedure produces tables of counts and percentages for the joint distribution of two categorical variables. Such tables are known as ***contingency, cross-tabulation*****,** or ***crosstab*** tables.

- **Definition:** A *contingency table* is a tabular arrangement of count data representing how the row factor frequencies relate to the column factor. We call a contingency table with *“R” rows* and *“C” columns*, an *Rx C contingency table*. Each category in a contingency table is called a *cell.*

Types of Categorical Variables

- Note that we will refer to two types of categorical variables: ***Table*** variables and ***Grouping*** variables.

- *The values of the* ***Table*** *variables are used to define the rows and columns of a single contingency table.* Two *Table* variables are used for each table, one variable defining the rows of the table and the other defining the columns.

- ***Grouping*** *variables are used to split a data into subgroups.* A separate table is generated for each unique set of values of the *Grouping* variables. Note that if you only want to use one *Table* variable, you should use the *Frequency Table* procedure.

A contingency table of counts with R rows and C columns as in the table below. Let O_{ij} be the observed count/observed frequency for the i_{th} row (i = 1 to R) and j_{th} column (j = 1 to C).

Table 1. An R×C Contingency Table (R×C Crosstab)

		Column Variable					
Row Variable		Column 1	…	Column j	…	Column C	**Total**
	Row 1	O_{11}	…	O_{1j}	…	O_{1C}	$n_{1.}$
	⋮	⋮	⋱	⋮	⋱	⋮	⋮
	Row i	O_{i1}	…	O_{ij}	…	O_{iC}	$n_{i.}$
	⋮	⋮	⋰	⋮	⋰	⋮	⋮
	Row R	O_{R1}	…	O_{Rj}	…	O_{RC}	$n_{R.}$
	Total	$n_{.1}$	…	$n_{.j}$	…	$n_{.C}$	n

The row and column marginal totals be designated as $n_{i.}$ and $n_{.j}$ respectively. The total number of counts in the table be n

$$n_{i.} = \sum_j O_{ij} \qquad n_{.j} = \sum_i O_{ij}$$

$$n = \sum_i \sum_j O_{i\,j} = \sum_i n_{i.} = \sum_j n_{.\,j}$$

The table of associated proportions can then be written as in below.

				Column Variable			
		Column 1	...	Column j	...	Column C	**Total**
Row Variable	Row 1	p_{11}	...	p_{1j}	...	p_{1C}	$p_{1.}$
		$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	Row i	p_{i1}	...	p_{ij}	...	p_{iC}	$p_{i.}$
		$\vdots$	$\iddots$	$\vdots$	$\iddots$	$\vdots$	$\vdots$
	Row R	p_{R1}	...	p_{Rj}	...	p_{RC}	$p_{R.}$
	Total	$p_{.1}$	...	$p_{.j}$	...	$p_{.C}$	1

$$p_{ij} = \frac{O_{ij}}{n} \qquad p_{i.} = \frac{n_{i.}}{n} \qquad p_{.j} = \frac{n_{.j}}{n}$$

The expected counts for the i_{th} row and j_{th} column:

$$E_{ij} = \frac{n_{i.}n_{.j}}{n}$$

Following we will describe the various tests and statistics calculated by this procedure using the preceding notations.

Table Statistics

This section presents various statistics that can be output for each individual cell. These are useful for studying the independence between rows and columns. The statistics for the i_{th} row and j_{th} column are as follows.

Count: *The cell count $O_{i\,j}$ is the number of observations for the cell.*

Row Percentage: *The percentage for column j within row i:*

$$p_{j|i} = \frac{O_{i\,j}}{n_{i\,.}}$$

Column Percentage: *The percentage for row i within column j:*

$$p_{i|\,j} = \frac{O_{i\,j}}{n_{.\,j}}$$

Table Percentage: *The overall percentage for the cell:*

$$p_{i\,j} = \frac{O_{i\,j}}{n}$$

Expected Counts Assuming Independence: *The expected count/expected frequency, E_{ij} is the count that would be obtained if the hypothesis of row-column independence were true.*

$$E_{i\,j} = \frac{n_{i\,.}n_{.\,j}}{n}$$

Example: A typical cross-tabulation table comparing the two hypothetical variables **"City of Residence (İkamet Etme, Oturma)"** with **"Favorite Baseball Team"** is shown below. Are city of residence and being a fan of that city's Baseball team independent? The cells of the Table 2 report the frequency counts and percentages for the number of respondents in each cell.

Table 2. *Table Percentages.*

		What is Your Favorite Baseball Team?			
In What City Do You Reside (ikamet etmek, oturmak)?		Toronto Blue Jays	Boston Red Socks	New York Yankees	Totals
	Boston, MA	11	33	7	51
	Table Percentage	7. 53 %	22.60 %	4.80 %	34. 93%
	Montreal, Canada	23	14	9	46
	Table Percentage	15.76 %	9.59 %	6.16 %	31.51 %
	Montpellier, VT	22	13	14	49
	Table Percentage	15.07 %	8.90 %	9.59 %	33.56 %
	Totals	56	60	30	n=146
		38.36 %	41.10 %	20.55 %	100.00 %

In Table 2, for example:

$$p_{2.} = \frac{n_{2.}}{n} = \frac{46}{146} = 31.51\,\%$$

$$O_{13} = 7 \;\Rightarrow\; p_{13} = \frac{O_{13}}{146} = \frac{7}{146} = 4.80\,\%$$

$$p_{.3} = \frac{n_{.3}}{n} = \frac{30}{146} = 20.55\,\%$$

4.80% of persons are fan of NewYork Yankees and live in Boston, MA.

31.51% of persons live in Montreal Canada

20.55% of persons are fan of New York Yankees.

Table 3. *Row Percentages.*

		What is Your Favorite Baseball Team?			
In What City Do You Reside (ikamet etmek, oturmak)?	Cross tabulation Frequency Percent	Toronto Blue Jays	Boston Red Socks	New York Yankees	**Totals**
	Boston, MA	11	33	7	**51**
	Row Percentage	21.57 %	64.71 %	13.72%	**100 %**
	Montreal, Canada	23	14	9	**46**
	Row Percentage	50.00 %	30.43 %	19.57 %	**100 %**
	Montpellier, VT	22	13	14	**49**
	Row Percentage	44.90 %	26.53 %	28.57 %	**100 %**
	Totals	**56**	**60**	**30**	**146**

In Table 3, for example:

$$O_{23} = 9 \Rightarrow p_{3|2} = \frac{O_{23}}{n_{2.}} = \frac{9}{46} = 19.57\,\%$$

19.57% of persons who llve in Montreal, Canada are also fan of NewYork Yankees.

Table 4. *Column Percentages.*

		What is Your Favorite Baseball Team?			
In What City Do You Reside (ikamet etmek, oturmak)?	Cross tabulation Frequency Percent	Toronto Blue Jays	Boston Red Socks	New York Yankees	**Totals**
	Boston, MA	11	33	7	**51**
	Column Percentage	19.64 %	55 %	23.30%	
	Montreal, Canada	23	14	9	**46**
	Column Percentage	41.07 %	23.30 %	30%	
	Montpellier, VT	22	13	14	**49**
	Column Percentage	39.29 %	21.70 %	46.70 %	
	Totals	56	60	30	**146**
		100 %	**100 %**	**100%**	

In Table 4, for example:

$$O_{31} = 22 \Rightarrow p_{3|1} = \frac{O_{31}}{n_{.1}} = \frac{22}{56} = 39.29\,\% \quad (\text{Column Percentage})$$

39.29% of persons who are fan of Toronto Blue Jays are also live in Montpellier, VT.

Table 5. *Expected Counts/Frequencies.*

		What is Your Favorite Baseball Team?			
In What City Do You Reside (ikamet etmek, oturmak)?	Cross tabulation Frequency Percent	Toronto Blue Jays	Boston Red Socks	New York Yankees	**Totals**
	Boston, MA	19.56	20.96	10.48	**51**
	Montreal, Canada	17.64	18.90	9.45	**46**
	Montpellier, VT	18.79	20.14	10.07	**49**
	Totals	**56**	**60**	**30**	**146**

In Table 5, for example:

$$E_{12} = \frac{n_{1.}n_{.2}}{n} = \frac{51 \times 60}{146} = 20.96 \qquad E_{31} = \frac{n_{3.}n_{.1}}{n} = \frac{49 \times 56}{146} = 18.79$$

H_0: *City of residence and* being a fan of that city's Baseball team *are independent.*
H_1: *City of residence and* being a fan of that city's Baseball team *are not independent.*

$$\chi_P^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \frac{(11-19.56)^2}{19.56} + \ldots + \frac{(14-10.07)^2}{10.07} = 19.351$$

Since $\chi_P^2 = 19.351 > \chi_{0.05,4}^2 = 9.48773$, *$H_0$ is rejected. City of residence and being a fan of that city are not independent at the 0.05 significance level.*

The χ^2 (Chi-square) Distribution

The χ^2 *distribution allows us to test different hypotheses about frequencies (or proportions). Since χ^2 made up of a sum of squares it can never be negative, and the minimum value it can take on is 0.*

There are 3 types of χ^2 tests:

χ^2 Test for Differences among the Proportions of C Populations (Test of Homogeneity)

χ^2 Test for Independence – to determine whether two nominal variables are related (Test of Row-Column Independence)

χ^2 Goodness of Fit Test – to determine whether a sample may be regarded as a random sample drawn from a population with a specified distribution (e.g., normal, binomial, etc.)

Tests for Row-Column Independence

- **Many times, the n elements of a sample from a population may be classified according to two different criteria. It is then of interest to know whether the two methods of classification are statistically independent; for example, we may consider the population of graduating engineers, and we may wish to determine whether starting salary is independent of academic disciplines. Assume that the first method of classification has R levels and the second method has C levels. We will let O_{ij} be the observed frequency for level *i* of the first classification method and level *j* on the second classification method. The data would, in general, appear as shown in Table 1. Such a table is usually called an *R×C contingency table*.**

- **<u>We are interested in testing the hypothesis that the row-and-column methods of classification are independent. If we reject the hypothesis, we conclude there is some interaction between the two criteria of classification.</u>**

Pearson's Chi-Square Test

- **Pearson's chi-square statistic is used to test independence between the row and column variables. Independence means that knowing the value of the row variable does not change the probabilities of the column variable (and vice versa). Another way of looking at independence is to say that the row percentages (or column percentages) remain constant from row to row (or column to column).**

- **This test requires large sample sizes to be accurate. An often quoted rule of thumb regarding sample size is that none of the expected cell values should be less than five, $E_{ij} \nleq 5$. Although some users ignore the sample size requirement, you should also be very skeptical (şüpheci) of the test if you have cells in your table with zero counts.**

- **For 2 × 2 tables, consider using Fisher's Exact Test for small samples.**

- *Pearson's chi-square test statistic follows an asymptotic chi-square distribution with (R – 1)(C – 1) degrees of freedom when the row and column variables are independent (if the null hypothesis is true). It is calculated as:*

$$\chi_P^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

- Therefore, we would reject the hypothesis of independence if the observed value of the test statistic χ_P^2 exceeded $\chi^2_{\alpha,(R-1)(C-1)}$.

$$\chi_P^2 > \chi^2_{\alpha,(R-1)(C-1)} \text{ then } H_0 \text{ is rejected}$$

Example: A company has to choose among three pension(emeklilik) plans. Management wishes to know whether the preference for pension plans is independent of job classification and wants to use α=0.05. The opinions of a random sample of 500 employees are shown in Table 6.

Table 6. *Observed data*

	Pension Plan			
Job Classification	1	2	3	Totals
Salaried workers	160	140	40	340
Hourly workers	40	60	60	160
Totals	200	200	100	500

Firstly, the expected frequencies are computed as in below:

$$E_{11} = \frac{n_{1.}n_{.1}}{500} = \frac{340 \times 200}{500} = 136$$

$$E_{12} = \frac{n_{1.}n_{.2}}{500} = \frac{340 \times 200}{500} = 136$$

$$E_{13} = \frac{n_{1.}n_{.3}}{500} = \frac{340 \times 100}{500} = 68$$

$$E_{21} = \frac{n_{2.}n_{.1}}{500} = \frac{160 \times 200}{500} = 64$$

$$E_{22} = \frac{n_{2.}n_{.2}}{500} = \frac{160 \times 200}{500} = 64$$

$$E_{23} = \frac{n_{2.}n_{.3}}{500} = \frac{160 \times 100}{500} = 32$$

The expected frequencies are shown in Table 7.

Table 7. Expected Frequencies.

	Pension Plan			
Job Classification	1	2	3	**Totals**
Salaried workers	136	136	68	**340**
Hourly workers	64	64	32	**160**
Totals	**200**	**200**	**100**	**500**

Now hypothesis-testing procedure may now be applied to this problem as in below.

In this problem, the variable of interest is employee preference among pension plans. The null hypothesis (H_0) and alternative hypothesis (H_1) could be written as in below.

H_0: *Preference for pension plans is independent of job classification.*
H_1: *Preference for pension plans is not independent of job classification.*

Since R=2 and C=3, the degrees of freedom for chi-square are (R-1)(C-1)=(1)(2)=2 and α=0.05, we would reject H_0 if $\underbrace{\chi_P^2}_{\text{test statistic value}} > \underbrace{\chi_{0.05,2}^2}_{\text{table value}} = 5.99146.$

The test statistic is:

$$\chi_P^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\chi_P^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \frac{(160-136)^2}{136} + \frac{(140-136)^2}{136} + \frac{(40-68)^2}{68}$$

$$+ \frac{(40-64)^2}{64} + \frac{(60-64)^2}{64} + \frac{(60-32)^2}{32} = 49.63$$

Conclusions: *Since* $\chi_P^2 = 49.63 > \chi_{0.05,2}^2 = 5.99146$ *we reject the hypothesis of independence and conclude that the preference for pension plans is not independent of job classification. Further analysis would be necessary to explore the nature of the association between these factors. (Look at correlation statistics)*

Things to remember

- *The E_{ij}'s (expected counts) need not be integers and we do not round them.*

- *The row and column totals are the same for observed and expected tables (this is a good way to check your calculations!)*

- *The chi-squared approximation is not very good when expected cell counts are too small. Our rule of thumb will be this:The chi-squared approximation is good enough when;*

  --All the expected counts are at least 1.

  --Most (at least 80%) of the expected counts are at least 5.

(Notice that this depends on expected counts, not observed counts)

The Chi-square Test Of Homogeneity

Using the two-way contingency table to test independence between two variables of classification in sample from a single population of interest is only one application of contingency table methods.

*Another common situation occurs when there are R populations of interest and each population is divided into the same C categories. A sample is then taken from the ith population, and the counts are entered in the appropriate columns of the ith row. In this situation we want to investigate whether or not the proportions in the C categories are the same for all populations. The null hypothesis is in this problem states the populations are **homogeneous** with respect to categories.*

*For example, when there are only two categories, such as success and failure, defective and non-defective, and so on, the test for homogeneity is really a test of the equality of R binomial parameters. **Calculation of expected frequencies, determination of degrees of freedom, and computation of the chi-square statistic for the test for homogeneity are identical to the test for independence (GRUPLAR ARASI FARK KONTROLÜ)***

Yates' Continuity Corrected Chi-Square Test

Yates' Continuity Corrected Chi-Square Test (or just Yates' Continuity Correction) is similar to Pearson's chisquare test, but is adjusted for the continuity of the chi-square distribution. ***This test is particularly useful when you have small sample sizes. This test is generally calculated for 2 × 2 tables.***

Yates' continuity corrected test statistic follows an asymptotic chi-square distribution with (R– 1)(C – 1) degrees of freedom when the row and column variables are independent. It is calculated as:

$$\chi_Y^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(\left|O_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}}$$

Example: In 2000 the State of Vermont,—a slight editorial comment—approved a bill authorizing civil unions (sivil birlikler yetki veren bir yasa tasarısı) between gay and lesbian partners. The data shown below suggest that there was a difference on this issue between male and female legislators (parlamenter, meclis üyesi). The expected frequencies are shown in parentheses.

	Vote		
Legislators' Gender	Yes	No	**Total**
Women	7 (5.89)	2 (3.10)	**9**
Men	12 (13.10)	8 (6.89)	**20**
Total	**19**	**10**	**29**

H_0: How legislators voted is not associated with genders' of legislators (How legislators voted is independent of their gender)

H_1: How legislators voted is associated with genders' of legislators (How legislators voted is not independent of their gender)

$$\chi_Y^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(\left|O_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}} = \frac{\left(\left|7 - 5.89\right| - 0.5\right)^2}{5.89} + \frac{\left(\left|2 - 3.10\right| - 0.5\right)^2}{3.10}$$

$$+ \frac{\left(\left|12 - 13.10\right| - 0.5\right)^2}{13.10} + \frac{\left(\left|8 - 6.89\right| - 0.5\right)^2}{6.89}$$

$$= 0.0631 + 0.1161 + 0.0274 + 0.0540$$

$$= 0.2606$$

Since $\chi_Y^2 = 0.2606 < \chi_{0.05,1}^2 = 3.84146$ *we accept the hypothesis of independence. We can conclude that how legislators voted is independent of their gender.*

Fisher’s Exact Test (2×2 Tables)

This test was designed to test the hypothesis that the two column percentages in a 2×2 table are equal. It is especially useful when sample sizes are small (even zero in some cells) and the chi-square test is not appropriate.

Using the hypergeometric distribution with fixed row and column totals, this test computes probabilities of all possible tables with the observed row and column totals. This test is often used when sample sizes are small, but it is appropriate for all sample sizes because Fisher’s exact test does not depend on any large-sample asymptotic distribution assumptions. ***This test is only calculated for 2×2 tables.***

It is most useful when the total sample size and the expected values are small. The test holds the marginal totals fixed and computes the hypergeometric probability that O_{11} is at least as large as the observed value. ***Useful when cell counts < 5.***

Example: *2x2 table with cell counts a, b, c, d. Assuming marginal totals are fixed:*

2x2 Crosstab

	Column Variable A			
Row Variable B		A_1	A_2	**Total**
	B_1	a	b	**A**
	B_2	c	d	**B**
	Total	**C**	**D**	**n**

A= a+b, B= c+d, C= a+c, D= b+d.

Probability distribution of cell count a follows a hypergeometric distribution:

n = a + b + c + d = A + B= C + D

$$\Pr(X = a) = \frac{A!B!C!D!}{n!a!b!c!d!}$$

$$Mean(X) = \frac{AC}{n}$$

$$Var(X) = \frac{ABCD}{n^2(n-1)}$$

Fisher exact test is based on hypergeometric distribution.

Example: Is HIV Infection related to Hx of STDs (cinsel yolla bulaşan hastalıklar) in Sub Saharan African Countries? Test at 5% level.

	HIV Infection			
Hx of STDs		yes	no	**Total**
	yes	3	7	**10**
	no	5	10	**15**
	total	**8**	**17**	25

Since this is a 2 × 2 table and $O_{11} < 5$ we will use Fisher Exact Test.

- *Probability of observing this specific table given fixed marginal totals is:*

$$P_r(3,7,5,10) = \frac{10!\,15!\,8!\,17!}{25!\,3!\,7!\,5!\,10!} = \boxed{0.3332}$$

- *We will calculate* ***tail probability = sum of all values (a = 3, 2, 1, 0)***

$$P_r(2,8,6,9) = \frac{10!15!8!17!}{25!2!8!6!9!} = \boxed{0.2082}$$

	yes	no	Total
yes	2	8	10
no	6	9	15
total	8	17	25

$$P_r(1,9,7,8) = \frac{10!15!8!17!}{25!\,1!\,9!\,7!\,8!} = \boxed{0.0595}$$

	yes	no	Total
yes	1	9	10
no	7	8	15
total	8	17	25

$$P_r(0,10,8,7) = \frac{10!15!8!17!}{25!\,0!\,10!\,8!\,7!} = \boxed{0.0059}$$

	yes	no	Total
yes	0	10	10
no	8	7	15
total	8	17	25

Tail probability = *0.3332+0.2082+0.0595+0.0059 = 0.6068*

H_0: HIV Infection is not related to Hx of STDs
H_1: HIV Infection is related to Hx of STDs

If the tail probability (P-value) is equal or smaller than α ($P \leq \alpha$) null hypothesis H_0 is rejected.

In this example, since $p=0.6068>\alpha=0.05$, so H_0 cannot be rejected. HIV Infection is independent of Hx of STDs .

Final Notes on Chi-Square for Contingency Tables (Summary)

- Degrees of freedom for an R×C table is (R-1)(C-1)

- Pearson's χ^2 statistic for contingency tables uses the $\chi^2 \cong \chi^2_{df}$ approximation so in order to be a valid approximation, a standard rule of thumb is to require **all the expected counts are at least 1 and most (at least 80%) of the expected counts are at least 5.**

- If expected counts are small, and data forms a 2 x 2 table, Fisher's exact test may be appropriate

- $Pearson\ \chi^2$ statistic has good approximation for large sample sizes.

For R×C tables, we have the following two hypotheses

- C samples and we're checking for R levels of a row factor, then we're testing whether the distributions are the same (for the groups – your columns) **(Test of Homogeneity)**

- One sample and we're checking for R levels of a row factor, and c levels of a column factor, then we're testing for an association of the row and column factors. **(Test for Row-Column Independence)**

Association and Correlation Statistics

Row-column (nominal-nominal): ***Phi, Cramer's V,*** **Pearson's Contingency Coefficient**

Row-column (ordinal-ordinal): Kendall's tau-B, Kendall's tau-C, Gamma, Somers' d

If one of the row or column variable is nominal you can obtain correlation statistic for nominal-nominal.

Note: Since the calculations are very hard for ordinal-ordinal correlation statistics, we will not give formulas for these coefficients.

Association and Correlation Statistics

Phi

Appropriate when measuring degree of association between two binary variables.

Phi ranges: 0≤ φ ≤1; 0 (no relationship) and 1 (perfect relationship). Phi is designed for 2 × 2 tables as given formula:

$$\varphi = \frac{\left| O_{12} \times O_{21} - O_{11} \times O_{22} \right|}{\sqrt{n_{1.} \times n_{2.} \times n_{.1} \times n_{.2}}}$$

For larger tables, it has no upper limit and Cramer's V should be used instead. The formula is :

$$\varphi = \sqrt{\frac{\chi_{P}^{2}}{n}}$$

This measure of association commonly used to measure association between nominal variables.

Cramer's V

This statistic is a modification of the Phi statistic so that it is appropriate for larger than 2 × 2 tables. V ranges; ; 0 (no relationship) and 1 (perfect relationship). The formula is:

$$V = \sqrt{\frac{\chi^2_{\mathrm{P}}}{\mathrm{n}\left(min\left(R,C\right)-1\right)}}$$

This measure of association commonly is used to measure association between <u>nominal variables</u>.

<u>Pearson's Contingency Coefficient</u>

It ranges between 0 (no relationship) and 1 (perfect relationship). The formula is:

$$C = \sqrt{\frac{\chi^2_P}{\chi^2_P + n}}$$

This measure of association commonly is used to measure association between <u>nominal variables.</u>

Kendall's tau-B

This is a measure of correlation between two ordinal-level (rankable) variables. It is most appropriate for square tables (3×3, 4×4, etc.). This statistic is showed by τ_b.

Kendall' tau B ranges $-1 \leq \tau_b \leq 1$.

This measure of association commonly is used to measure association between <u>ordinal variables</u>.

Kendall's tau-C

Thsi is used in the case where the number of rows does not match the number of columns. This statistic is showed by τ_c.

Kendall' tau C ranges $-1 \leq \tau_c \leq 1$.

This measure of association commonly is used to measure association between <u>ordinal variables</u>.

Gamma

This statistic is showed by γ. Gamma ranges $-1 \leq \gamma \leq 1$.

This measure of association commonly is used to measure association between ordinal variables

Somers' d

Somers' $d(C|R)$ and Somers' $d(R|C)$ are asymmetric modifications of tau-b. $C|R$ represents that the row variable X is treated as an independent variable, whereas the column variable Y is treated as dependent. Similarly, $R|C$ represents the reverse interpretation. Sommers' d can be calculated only when both variables are ordered. It varies in the range $-1 \leq d \leq 1$. Formulas for Sommers' d is obtained according to the position of independent variable.