POLS 5377 Scope & Method of Political Science

Week 15 Measure of Association - 2

Correlation within SPSS

Key Questions:

* How to compute and interpret the following measures in SPSS
    * Ordinal Variable – Gamma (G)
    * Ordinal Variables – Spearman's rho (r_s)
    * Interval Variables – Pearson's r (r)

Outline

* SPSS: Ordinal Variables – Gamma (*G*)
* SPSS: Ordinal Variables – Spearman's rho (r_s)
* SPSS: Interval Variables – Pearson's *r* (*r*)

Conducting Correlation within SPSS

* In real world, we often have to make inference with the data contains hundreds or thousands of cases. It is difficult to compute the measures of association without using a computer.
* We will learn how to generate the statistic results of measures of association with SPSS and interpret the results
  * Ordinal variables with a few categories: **Gamma (*G*)**
  * Ordinal variables with a broad range of scores: **Spearman's rho (r_s)**
  * Interval/ratio variables: **Pearson's *r***
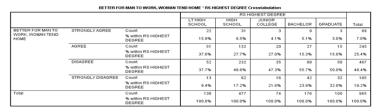
SPSS: Ordinal Variable – Gamma (*G*)

* When measure the association for ORDINAL variables that have a few categories: Gamma (*G*)

* SPSS procedures
  * [Analyze] → [Descriptive Statistics] → [Crosstabs]
  * Place dependent variables in the **Row** box
  * Place independent variables in the **Column** box
  * Click the **Statistics** button, select **Chi-square** and **Gamma**
  * Click the **Cells** button, select **column** in the **Percentage** box
  * Proceed the analysis

SPSS: Ordinal Variable – Gamma (*G*)

* Example: Is there an association between education level and attitude toward women working?
  * Dataset: GSS2010_POLS5377
  * DV (Row): attitude toward woman working (fefam)
  * IV (Column):Education Level (degree)
* Step 1 Meet requirements:
  * Ordinal variables with a few categories
* Step 2 Null hypothesis
  * H_0: There is no relationship between education level and attitude toward women working
* Step 3 Critical Region
  * α=0.05

SPSS: Ordinal Variable – Gamma (*G*)

BETTER FOR MAN TO WORK, WOMAN TEND HOME * RS HIGHEST DEGREE Crosstabulation

			RS HIGHEST DEGREE					
			LT HIGH SCHOOL	HIGH SCHOOL	JUNIOR COLLEGE	BACHELOR	GRADUATE	Total
BETTER FOR MAN TO WORK, WOMAN TEND HOME	STRONGLY AGREE	Count	22	31	3	9	3	68
		% within RS HIGHEST DEGREE	15.9%	6.5%	4.1%	5.1%	3.0%	7.0%
	AGREE	Count	51	132	20	27	15	245
		% within RS HIGHEST DEGREE	37.0%	27.7%	27.0%	15.3%	15.0%	25.4%
	DISAGREE	Count	52	232	35	98	50	467
		% within RS HIGHEST DEGREE	37.7%	48.6%	47.3%	55.7%	50.0%	48.4%
	STRONGLY DISAGREE	Count	13	82	16	42	32	185
		% within RS HIGHEST DEGREE	9.4%	17.2%	21.6%	23.9%	32.0%	19.2%
Total		Count	138	477	74	176	100	965
		% within RS HIGHEST DEGREE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

P value<0.05

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	63.418[a]	12	.000
Likelihood Ratio	61.385	12	.000
Linear-by-Linear Association	47.647	1	.000
N of Valid Cases	965		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.21.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	.296	.039	7.376	.000
N of Valid Cases		965			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

* Degree of freedom (df) = (R-1)*(C-1)=(5-1)*(4-1)=12
* χ^2(obtained) = 63.418
* Test result: significance value (p value) = .000 < α=0.05. Reject H_0
* **Gamma= 0.296** indicates the relationship is **positive** and "**weak to moderate**"

SPSS: Ordinal Variable – Gamma (*G*)

* Interpret the result:
  * At the significance level of 0.05, the test of chi square is significant (p value = 0.000 < 0.05), so we **reject** H_0. The data suggests there is a significant relationship between education level and attitude toward women working.
  * According to the Gamma statistic, +0.296, the education level is **positively** associated with the attitude toward women working. The strength of the association is ranked as **weak to moderate.** The people with higher education level are more likely to support women working (more likely to disagree the statement of "it's better for man to work, woman tend home").

If the value is	Strength of the relationship
Between 0.00 and 0.30	Weak
Between 0.31 and 0.60	Moderate
Greater than 0.60	Strong

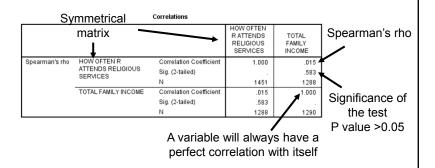# SPSS: Ordinal Variable – Spearman's rho (r_s)

* When measure the association for ORDINAL variables that have a broad scores: **Spearman's rho**
* SPSS procedures
  * [Analyze] → [Correlate] → [Bivariate]
  * Place both dependent and independent variables in the **Variables** box
  * Select **Spearman** in the box labeled "Correlation Coefficients" and **uncheck Pearson.**
  * Select between **one-tailed** and **two-tailed** test according to the research expectation
  * Proceed the analysis

SPSS: Ordinal Variable – Spearman's rho (r_s)

* Example: Is there an association between family income and the frequency of attending religious services?
  * Dataset: GSS2010_POLS5377
  * Variables:
    * frequency of attending religious services (attend)
    * total family income (income06)
* Step 1 Meet requirements:
  * Ordinal variables with a broad range of scores
* Step 2 Null hypothesis
  * H_0: There is no relationship between family income and the frequency of attending religious services
* Step 3 Critical Region
  * α=0.05 and Two-tailed

SPSS: Ordinal Variable – Spearman's rho (r_s)

Symmetrical matrix

Correlations

			HOW OFTEN R ATTENDS RELIGIOUS SERVICES	TOTAL FAMILY INCOME
Spearman's rho	HOW OFTEN R ATTENDS RELIGIOUS SERVICES	Correlation Coefficient	1.000	.015
		Sig. (2-tailed)	.	.583
		N	1451	1288
	TOTAL FAMILY INCOME	Correlation Coefficient	.015	1.000
		Sig. (2-tailed)	.583	.
		N	1288	1290

Spearman's rho

Significance of the test
P value >0.05

A variable will always have a perfect correlation with itself

* Test result: significance value (p value) = 0.583 > α=0.05.
* The relationship is not significant. We fail to reject H_0. The two variables are not associated.
* **Spearman rho = 0.015** indicates the relationship is **positive** and **weak** (No significant relationship anyways)

SPSS: Ordinal Variable – Spearman's rho (r_s)

* Interpret the result:
  * At the significance level of 0.05, the p value of 0.574 indicates that we **fail to reject** H_0. The data suggests there is no significant relationship between family income and the frequency of attending religious services.
  * According to the Spearman rho of 0.015 suggests that there is a **very weak** relationship between the family income and the frequency of attending religious services.

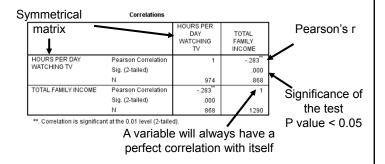| If the value is	Strength of the relationship
Between 0.00 and 0.30	Weak
Between 0.31 and 0.60	Moderate
Greater than 0.60	Strong

SPSS: Interval Variables – Pearson's *r*

* When measure the association for INTERVAL variables: **Pearson's *r***
* SPSS procedures
  * [Analyze] → [Correlate] → [Bivariate]
  * Place both dependent and independent variables in the **Variables** box
  * Make sure **Pearson** in the box labeled "Correlation Coefficients" is checked
  * Select between **one-tailed** and **two-tailed** test according to the research expectation
  * Proceed the analysis

SPSS: Interval Variables – Pearson's *r*

* Example: Is there an association between watching TV and being success?
  * Dataset: GSS2010_POLS5377
  * Variables:
    * Hours per day watching TV (tvhours)
    * Total family income (income06)
* Step 1 Meet requirements:
  * Interval variables and a ordinal variable with a broad range of scores
* Step 2 Null hypothesis
  * H_0: There is no relationship between watching TV and being success
* Step 3 Critical Region
  * $\alpha=0.05$ and two-tailed

SPSS: Interval Variables – Pearson's *r*

Symmetrical matrix   **Correlations**   Pearson's r

		HOURS PER DAY WATCHING TV	TOTAL FAMILY INCOME
HOURS PER DAY WATCHING TV	Pearson Correlation	1	-.283**
	Sig. (2-tailed)		.000
	N	974	868
TOTAL FAMILY INCOME	Pearson Correlation	-.283**	1
	Sig. (2-tailed)	.000	
	N	868	1290

**. Correlation is significant at the 0.01 level (2-tailed).

Significance of the test
P value < 0.05

A variable will always have a perfect correlation with itself

* Test result: significance value (p value) = 0.000 < α=0.05.
* The relationship is significant. We reject H_0 . The two variables are associated.
* **Pearson's r = - 0.283** indicates the relationship is **negative** and **weak to moderate**

SPSS: Interval Variables – Pearson's *r*

* Interpret the result:
    * At the significance level of 0.05, the p value of 0.000 suggests that the there is a significant relationship between two variables, we **reject** H_0. The data suggests there is a significant relationship between family income and number of hours per day watching TV.
    * According to the Pearson's r of **-0.283** suggests that there is a **negative** relationship between the family income and the hours of per day watching TV, and the strength of the relationship is **weak to moderate**.
    * The result suggests that people with higher family income level spend less time watching TV. However, it doesn't mean people who spend many hours watching TV will cause lower family income. The correlation statistic only suggests there is a relationship, doesn't suggest a causality between the two variables.

After this lecture:

You should learn the following key concepts:

* Know how to conduct correlation procedures with SPSS and how to interpret results