

Measures of Association

- A single number that summarizes the **strength of the relationship** between two variables.
 - It suggests the ability of one variable (the independent variable) to predict another (the dependent variable). It also, sometimes, shows the **direction of the relationship** (either positive or negative).

Measures of Association

"All good measures of association use a **proportionate reduction in error (PRE)** approach".

What is a **PRE measure**?

PRE measures compare:

(1) **the amount of error** that is made when trying to predict a variable and there is no information to help

to:

(2) **the amount of error** that is made when trying to predict a variable and there is some information to help, i.e., an independent variable.

For Example:
Take your best guess?

If you know nothing else about a person except that he or she lives in the United States and I asked you to guess his or her race/ethnicity, what would you guess?

The most common race/ethnicity for U.S. residents.

The model

Now, if we know that this person lives in San Diego, California

(i.e., we have **additional information** to help.

That is the **independent variable**: city of residence),

would you change your guess?

Since about 70% of the population in San Diego is Hispanic, you would predict each person is Hispanic and then be correct 70% of the time.

Summary

PRE measures are derived by comparing:

1. the number of errors made when predicting the dependent variable (DV) while ignoring all independent variables (IV) to:
2. the number of errors made when predicting the DV while using information about an IV.

Proportional Reduction of Error (PRE)

$$PRE = \frac{E1 - E2}{E1}$$

E1 = errors of prediction made when the independent variable is ignored

E2 = errors of prediction made when the IV is used to make the predictions

Proportional Reduction of Error (PRE)

- If the DV is related to the IV, then the IV will allow us to make a better prediction (fewer errors) than the prediction we would make without considering the IV.
- The better the ability of the IV to help us predict the DV, the "stronger" the relationship between the DV and the IV.

Two PRE Measures: Lambda & Gamma

- | | |
|--------------------|---|
| • Lambda λ | Appropriate for...
Nominal variables |
| • Gamma γ | Ordinal &
Dichotomous Nominal
variables |

Measure of Association: Lambda λ

- Provides us with an indication of the strength of an association between the independent and dependent variables.
- Suitable for use with nominal variables
- Ranges from 0.0 to 1.0
- A lower value represents a weaker association, while a higher value is indicative of a stronger association between the DV & IV

- Lambda is an asymmetrical measure of association.

- A measure whose value may vary depending on which variable is considered the independent variable and which the dependent variable.

The **size** of the **Lambda** is generally interpreted as follows:

.00 to .19	"little to no relationship"
.20 to .39	"weak relationship"
.40 to .59	"moderate relationship"
.60 to 1.00	"strong relationship"

Measure of Association: Gamma γ

- Gamma provides us with an indication of the **strength** and **direction** of the association between the variables (ranges from 0.0 to ± 1.0).

- Appropriate for **ordinal** variables or with **dichotomous nominal** variables (dichotomous variables have only two values such as female/male).

- Gamma γ is a **symmetrical** measure of association.

(A measure whose value will be the same when either variable is considered the independent variable or the dependent variable).

The **size** of the **Gamma** is generally interpreted as follows:

.00 to .19	"little to no relationship"
.20 to .39	"weak relationship"
.40 to .59	"moderate relationship"
.60 to 1.00	"strong relationship"

Four other measures include:

Yule's Q a PRE symmetric measure used with a **2 x 2 table**; **gives misleading information** when one of the four cells has a zero frequency

Phi a PRE symmetric measure used with a **2 x 2 table**; doesn't reach a maximum or minimum 1 so a "**phi adjusted**" has been developed

Four other measures include:

Tau C a non-PRE symmetric measure used with **two discrete ordered variables**; due to how it calculates the size of an association it is always a smaller value than gamma (neither statistic is preferred to the other)

Somer's D a PRE **asymmetric measure for discrete ordered variables**; due to how it calculates the size of an association it is always a smaller value than gamma; because it is asymmetric **the DV should always be specified** when reporting the statistic