

Encyclopedia of Social Science Research Methods

Generalizability Theory

Contributors: Robert L. Brennan

Editors: Michael S. Lewis-Beck & Alan Bryman & Tim Futing Liao

Book Title: Encyclopedia of Social Science Research Methods

Chapter Title: "Generalizability Theory"

Pub. Date: 2004

Access Date: April 06, 2015

Publishing Company: SAGE Publications, Inc.

City: Thousand Oaks

Print ISBN: 9780761923633

Online ISBN: 9781412950589

DOI: <http://dx.doi.org/10.4135/9781412950589.n366>

Print pages: 419-421

©2004 SAGE Publications, Inc. All Rights Reserved.

This PDF has been generated from SAGE knowledge. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412950589.n366>

Generalizability theory provides an extensive conceptual framework and set of statistical machinery for quantifying and explaining the consistencies and inconsistencies in observed scores for objects of measurement. To an extent, the theory can be viewed as an extension of classical test theory through the application of certain ANALYSIS OF VARIANCE (ANOVA) procedures. Classical theory postulates that an observed score can be decomposed into a TRUE SCORE and a single undifferentiated random error term. By contrast, generalizability theory substitutes the notion of *universe score* for true score and liberalizes classical theory by employing ANOVA methods that allow an investigator to untangle the multiple sources of error that contribute to the undifferentiated error in classical theory.

Perhaps the most important and unique feature of generalizability theory is its conceptual framework, which focuses on certain types of studies and universes. A *generalizability* (G) study involves the collection of a sample of data from a *universe of admissible observations* that consists of *facets* defined by an investigator. Typically, the principal result of a G study is a set of estimated random-effects variance components for the universe of admissible observations. A D (*decision*) study provides estimates of universe score variance, error variances, certain indexes that are like RELIABILITY COEFFICIENTS, and other statistics for a measurement procedure associated with an investigator-specified *universe of generalization*.

In univariate generalizability theory, there is only one universe (of generalization) score for each object of measurement. In multivariate generalizability theory, each object of measurement has multiple universe scores. Univariate generalizability theory typically employs random-effects models, whereas MIXED-EFFECTS MODELS are more closely associated with multivariate generalizability theory.

[p. 419 ↓]

Historical Development

The defining treatment of generalizability theory is the 1972 book by Lee J. Cronbach, Goldine C. Gleser, Harinder Nanda, and Nageswari Rajaratnam titled *The Dependability*

of *Behavioral Measurements*. A recent extended treatment of the theory is the 2001 book by Robert L. Brennan titled *Generalizability Theory*. The essential features of univariate generalizability theory were largely completed with technical reports in 1960–1961 that were revised into three journal articles, each with a different first author (Cronbach, Gleser, and Rajaratnam). Multivariate generalizability theory was developed during the ensuing decade. Research between 1920 and 1955 by Ronald A. Fisher, Cyril Burt, Cyril J. Hoyt, Robert L. Ebel, and E. F. Lindquist, among others, influenced the development of generalizability theory.

Example

Suppose an investigator, Smith, wants to construct a measurement procedure for evaluating writing proficiency. First, Smith might identify or characterize essay prompts of interest, as well as potential raters. Doing so specifies the facets in the universe of admissible observations. Assume these facets are viewed as infinite and that, in theory, any person p (i.e., object of measurement) might respond to any essay prompt t , which in turn might be evaluated by any rater r . If so, a likely data collection design might be $p \times t \times r$, where “ \times ” is read as “crossed with.” The associated linear model is

$$X_{ptr} = \mu + v_p + v_t + v_r + v_{pt} + v_{pr} + v_{tr} + v_{ptr},$$

where the v are uncorrelated score effects. This model leads to

$$\begin{aligned} \sigma^2(X_{ptr}) = & \sigma^2(p) + \sigma^2(t) + \sigma^2(r) + \sigma^2(pt) \\ & + \sigma^2(pr) + \sigma^2(tr) + \sigma^2(ptr), \quad (1) \end{aligned}$$

which is a decomposition of the total observed score variance into seven *variance components* that are usually estimated using the expected MEAN SQUARES in a random-effects ANOVA.

Suppose the following estimated variance components are obtained from Smith's G study based on n

t

essay prompts and n

r

raters:

$$\begin{aligned} \hat{\sigma}^2(p) &= 0.25, & \hat{\sigma}^2(t) &= 0.06, & \hat{\sigma}^2(r) &= 0.02, \\ \hat{\sigma}^2(pt) &= 0.15, & \hat{\sigma}^2(pr) &= 0.04, & \hat{\sigma}^2(tr) &= 0.00, \\ & \text{and } \hat{\sigma}^2(ptr) &= 0.12. \end{aligned}$$

Suppose also that Smith wants to generalize persons' observed mean scores based on

n'

t

essay prompts and n'

r

raters to these persons' scores for a universe of generalization that involves an infinite number of prompts and raters. This is a verbal description of a D study with a $p \times T \times R$ random-effects design. It is much like the $p \times t \times r$ design for Smith's G study, but the sample sizes for the D study need not be the same as the sample sizes for the G study, and the $p \times T \times R$ design focuses on *mean* scores for persons. The expected score for a person over the facets in the universe of generalization is called the person's universe score.

If n'

t

= 3 and n'

r

= 2 for Smith's measurement procedure, then the estimated random-effects variance components for the D study are

$$\begin{aligned}\hat{\sigma}^2(p) &= 0.25, & \hat{\sigma}^2(T) &= \frac{\hat{\sigma}^2(t)}{n'_t} = 0.02, \\ \hat{\sigma}^2(R) &= \frac{\hat{\sigma}^2(r)}{n'_r} = 0.01, & \hat{\sigma}^2(pT) &= \frac{\hat{\sigma}^2(pt)}{n'_t} = 0.05, \\ \hat{\sigma}^2(pR) &= \frac{\hat{\sigma}^2(pr)}{n'_r} = 0.02, & \hat{\sigma}^2(TR) &= \frac{\hat{\sigma}^2(tr)}{n'_t n'_r} = 0.00, \\ & \text{and } \hat{\sigma}^2(pTR) &= \frac{\hat{\sigma}^2(ptr)}{n'_t n'_r} = 0.02,\end{aligned}$$

with $\hat{\sigma}^2(p) = .25$ as the estimated universe score variance. The other variance components contribute to different types of error variance.

Absolute error, Δ
 p

, is simply the difference between a person's observed mean score and the person's universe score. For Smith's design and universe, the variance of these errors is

$$\begin{aligned}\sigma^2(\Delta) &= \sigma^2(T) + \sigma^2(R) + \sigma^2(pT) + \sigma^2(pR) \\ &\quad + \sigma^2(TR) + \sigma^2(pTR),\end{aligned}$$

which gives $\hat{\sigma}^2(\Delta) = 0.02 + 0.01 + 0.05 + 0.02 + 0.00 + 0.02 = 0.12$. Its square root is $\sigma_{\#}(\Delta) = 0.35$, which is interpretable as an estimate of the "absolute" standard error of measurement for a randomly selected person.

Relative error, δ
 p

, is defined as the difference between a person's observed deviation score and his or [p. 420 ↓] her universe deviation score. For Smith's design and universe,

$$\sigma^2(\delta) = \sigma^2(pT) + \sigma^2(pR) + \sigma^2(pTR),$$

which gives $\hat{\sigma}^2(\delta) = 0.05 + 0.02 + 0.02 = 0.09$. Its square root is $\sigma_{\#}(\delta) = 0.30$, which is interpretable as an estimate of the "relative" standard error of measurement for a randomly selected person.

Two types of reliability-like coefficients are widely used in generalizability theory. One coefficient is called a *generalizability coefficient*, which is defined as

$$E\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}.$$

It is the analog of a reliability coefficient in classical theory. The other coefficient is called an *index of dependability*, which is defined as

$$\Phi = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\Delta)}.$$

Because $\sigma^2(\delta) < \sigma^2(\Delta)$, it necessarily follows that $E\rho^2 > \Phi$. For example, using Smith's data, $E\rho^2 = 0.25/(0.25 + 0.09) = 0.74$, and $\Phi = 0.25/(0.25 + 0.12) = 0.68$.

Robert L. Brennan

<http://dx.doi.org/10.4135/9781412950589.n366>

References

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.