

1.5: Reliability of a Scale

Reliability vs Validity

In criminal justice, we often aim to study variables or concepts that are symbolic or abstract in nature. For instance, it's not possible to precisely measure the levels of disadvantage in society as we would measure weight using a scale. Instead, we rely on multiple indicators, such as poverty, low education, or employment status, to capture the degree of disadvantage. Ensuring the accuracy and consistency of these measures is critical, as invalid and unreliable measures can undermine the quality of our analysis.

The reliability of a scale refers to its consistency and freedom from measurement error. Essentially, it reflects how stable or consistent a measure, test, or observation is internally and across repeated uses. When multiple measurements of the same subject under the same conditions produce highly similar results, we consider the instrument reliable. On the other hand, the validity of a scale is about whether an instrument or test accurately measures what it's supposed to, or how free it is from systematic error. Validity requires reliability, meaning that a measure must be consistent to be valid. However, it's important to note that reliability alone does not guarantee validity (Vogt & Johnson, 2011).

Let me provide an example to illustrate reliability. Imagine you step on your weight scale, and it shows 160 pounds. You step off and on again, and it reads 150 pounds this time. Repeating the process again, it shows 140 pounds. In this case, your scale wouldn't be considered reliable because it gives inconsistent readings.

However, your scale would be reliable if you stepped on the scale multiple times and consistently got the same reading, say 170 pounds each time. Even if the reading isn't accurate (maybe you weigh 160 pounds), the scale is reliable if it consistently gives the same result.

Determining the validity of a scale is often more challenging than estimating its reliability because validity assessment requires more nuanced judgments about the accuracy and appropriateness of those measurements. Reliability can be assessed using statistical methods and quantifiable measures of consistency, but validity is a multifaceted concept involving various dimensions (e.g., content, criterion, and construct validity). In this chapter, we will only focus on the reliability of a scale.

Many social constructs in social science are abstract, and it is very challenging to measure consistently, unlike our physical weight, which is an objective physical property. Thankfully, there is a statistical way to examine the reliability of a scale for crime analysts. In this chapter, we will use the data from the National Crime Victimization Survey (NCVS) to compute the reliability of a scale.

National Crime Victimization Survey

I noted that crimes that go unreported to the police are often called the "dark figure of crime" because they remain undiscovered and unreported, making them hidden from official records. Estimating unreported crimes is crucial for the criminal justice system to function properly. To address the limitations of official crime measurement methods, the National Crime Victimization Survey (NCVS) was initiated. Conducted by the United States Census Bureau for the Bureau of Justice Statistics, the NCVS surveys over 200,000 individuals from 150,000 randomly selected households. The goal is to generate findings representative of the entire U.S. population. The survey, conducted every six months for three years, asks respondents about their experiences with crime, including whether they have been assaulted and whether they have reported the crime to the police. The survey aims to uncover the true extent of unreported and undiscovered crimes.

Despite its benefits, the NCVS has limitations. One major challenge is that respondents may not always be forthcoming about their victimization experiences. For example, individuals may feel uncomfortable discussing sexual offenses openly. Additionally, the accuracy of respondents' memories may be affected since they are asked about victimization events every six months, leading to potential confusion about timelines. Furthermore, the survey design primarily focuses on street crimes, with limited coverage of other offenses. This narrow focus may result in the underrepresentation of certain types of crimes in the data, such as Part 2 offenses. Nonetheless, the NCVS provides valuable insights into unreported crime.

Test-Retest and Internal Consistency Methods

Two commonly used indicators of a scale's reliability are test-retest reliability and internal consistency. Test-retest reliability involves administering a scale to the same individuals on two separate occasions and calculating the correlation between the scores obtained each time. A high test-retest correlation indicates greater reliability of the scale.

On the other hand, internal consistency assesses the extent to which the items covering the scale all measure the same underlying attribute. There are various methods to measure internal consistency, with Cronbach's alpha coefficient being the most widely used statistic which is also available in R.

Cronbach's Alpha Coefficient

Cronbach's alpha coefficient is one of the most commonly used indicators of internal consistency. Ideally, the Cronbach alpha coefficient of a scale should be above 0.7 (DeVellis, 2012). However, Cronbach's alpha values are influenced by the number of items in the scale. For short scales, such as those with fewer than ten items, it is common to observe relatively low Cronbach values (e.g., below 0.7). In such cases, reporting the scale's mean inter-item correlation may be appropriate. An optimal range for the inter-item correlation is between 0.2 and 0.5 (Pallant, 2016).

Importing the Data in Stata Format

First, you will download the revised data (rNCVS2016.dta) from [the shared Google Drive folder containing the rNCVS2016.dta data](#). I removed numerous cases and variables that are less pertinent to the focus of this chapter. Next, we need to import this data into R. The data is in Stata format, indicated by the .dta extension. You can read the data from the haven package discussed in a previous chapter.

```
library(haven)

rNCVS2016 <- read_dta("C:/Users/75JCH0I/OneDrive - West Chester
University of PA/WCU Research/R/data/rNCVS2016.dta")

View(rNCVS2016)
```

Guardianship

We would like to examine the internal consistency of capable guardianship against identity theft in this dataset. Capable guardianship, the presence of a measure that deters criminal activity, involves a potential target's ability to protect themselves. There are seven items considered to create the capable guardianship scale. Respondents were asked if they had taken seven specific self-protection measures in the past 12 months, including:

- Checking their credit report
- Changing passwords on financial accounts
- Purchasing credit monitoring services or identity theft insurance
- Destroying documents containing personally identifying information
- Reviewing banking or credit card statements for unfamiliar charges
- Using a security software program on their computer to protect against credit card loss or theft
- Purchasing identity theft protection from a company offering protection services

These items were labeled as Guardian1, Guardian2, Guardian3, Guardian4, Guardian5, Guardian6, and Guardian7. The response options for each item ranged from 0 (no) to 1 (yes).

Psych Package

The “psych” package in R is a comprehensive toolbox for psychometric and psychological research. It offers various functions for conducting various types of analyses, including factor analysis, reliability analysis, and item response theory.

```
install.packages("psych")

library(psych)
```

Alpha Function

Now let's compute Cronbach's alpha.

```
# Select the columns corresponding to the Guardian variables and store them in a
separate dataframe

guardian_data <- rNCVS2016 %>%
```

```
select(starts_with("Guardian")) # Select columns starting with "Guardian"
# Calculate Cronbach's alpha
alpha_result <- alpha(guardian_data, check.keys = TRUE)
# Print the results
print(alpha_result)
```

I used the tidyverse package to select from the dataset “rNCVS2016” the columns whose internal consistency I wanted to examine. Then, the alpha function from the psych package was used to calculate the Cronbach’s alpha for the selected columns of the dataset. The parameter check.keys = TRUE was included to ensure that the function checks for missing data and calculates alpha accordingly.

The syntax print(alpha_result) printed the results of the Cronbach’s alpha analysis to the console. The output includes various statistics related to internal consistency reliability, such as the Cronbach’s alpha value, item statistics, and frequencies.

Reporting the Results Regarding the Internal Consistency

Here is how you report the results regarding internal consistency. While the reliability of the Low Self Control (LSC) scale (Cronbach $\alpha = .67$) was slightly lower than the widely accepted standard ($\alpha = .70$, DeVellis, 2012), this alpha is sensitive to the small number of items in the scale. Briggs and Cheek (1986) suggested that as the mean inter-item correlation (.22) for the items is above .2, this scale can be regarded as reliable, indicating that the seven items are reliable sources to capture the levels of guardianship.

Conclusion

In this chapter, we covered the concept of reliability and explored various methods to measure it, including calculating internal consistency using R. The next chapter will delve into cross-tabulation and the chi-square test. These statistical tools are widely utilized for analyzing the relationship between two categorical variables.

References

- DeVellis, R. F. (2012). Scale development: Theory and applications (Vol. 26). Sage.
- Pallant, J. (2016). SPSS survival manual (6th ed.). McGraw Hill.
- Vogt, W. P., & Johnson, R. B. (2011). Dictionary of statistics & methodology: A nontechnical guide for the social sciences (4th ed.). Sage.

This page titled [1.5: Reliability of a Scale](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jaeyong Choi](#) ([The Pennsylvania Alliance for Design of Open Textbooks \(PA-ADOPT\)](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.