DOCTORAL THESIS

New Developments in Meta-analysis with Five-number Summary

SHI, Jiandong

Date of Award:
2021

Abstract

Meta-analysis is a statistical method for synthesizing multiple studies to achieve more comprehensive and reliable conclusions. This thesis mainly focuses on the meta-analysis with continuous outcomes, where some studies are reported with the whole or part of the five-number summary including the minimum and maximum values, the first and third quartiles, and the median. Given that most existing meta-analysis models can only handle the studies reported with the sample mean and standard deviation (SD), it is often desired to convert the five-number summary back to the sample mean and SD before synthesis, as otherwise the studies reported with the five-number summary have to be excluded from further analysis. To tackle this problem, some data transformation methods have emerged recently, and they are getting more and more popular in practice.

We note, however, that most popular methods for data transformation are built on the normality assumption, which may not always hold in practice. In particular, when a study chooses to report the five-number summary, it can be an indication that the underlying data may not be normal or symmetric. For such data, if still applying the normal-based methods for data-transformation, the final results can be misleading or even wrong in meta-analysis. Motivated by this, we propose to further enhance the meta-analysis literature on data transformation with the five-number summary. Specifically, this thesis consists of four important projects including (1) the hypothesis tests for skewness and normality, (2) the mean and variance estimation from the five-number summary of a log-normal distribution, (3) new effect size estimation methods, and (4) a new paradox in random-effects meta-analysis.

In Chapter 2, we propose three skewness tests and a normality test with the whole or part of the five-number summary. Despite of the limited data available, the information in the five-number summary, together with the sample size, is surprisingly sufficient enough to conduct the skewness tests. Moreover, when the five-number summary is fully available, we further incorporate the kurtosis information in the test for testing the normality beyond skewness. Simulation studies demonstrate that the type I error rates are well controlled and the newly proposed tests also provide

good statistical power.

In Chapter 3, we propose to estimate the mean and variance from the reported five-number summary of a log-normal distribution. For normal data, some well-performed methods are established. However, when the data are significantly skewed, there are few methods that could properly handle the problem. Motivated by this and noting that many skewed medical data are modeled with the log-normal distribution, we provide two types of estimators for the mean and variance with the five-number summary of a log-normal distribution. Their performance is demonstrated by the simulation studies and real data analysis.

In Chapter 4, we develop new methods that estimate the mean difference and standardized mean difference from the five-number summary. This is motivated by the fact that, even though the normal-based methods can be readily applied, the estimated sample mean and SD are unlikely to be the same as the true values. As a consequence, if one directly applies them as the true sample mean and SD and then applying the classical methods including the Cohen's d or Hedges' g to estimate the effect size, it may yield biased estimates so that the final meta-analytical results can be unreliable. Our new methods, as demonstrated by simulation studies, achieve a better accuracy and a higher coverage probability than the existing methods.

In Chapter 5, we introduce a new paradox in random-effects meta-analysis. Once the new paradox appears, the individual studies and the meta-analytical result are contradictory, which leads to a dilemma for clinical decision making. As found, the key reason for the paradox is the between-study heterogeneity involved in random-effects meta-analysis. In particular when the heterogeneity is large and the number of studies is small in meta-analysis, the probability of the new paradox appearing is not ignorable. Moreover, the new paradox only appears in random-effects meta-analysis but it does not exist in the common-effect and fixed-effects models. It thus raises an interesting question whether the current random-effects model is reasonable and tenable for meta-analysis, or it needs to be abandoned or further improved.

Keywords: Five-number summary, Log-normal distribution, Mean difference, Meta-analysis, Normality test, Paradox, Skewness test, Standardized mean difference

Table of Contents

List of Figures

List of Tables