

The Unknown Equal Variance Model of Bivariate Meta-Analysis Based on the Difference Mean

Pengfei Guo^{1,2,3}, Shuangyin Liu^{2,3,4}, Gang Li^{5*}, Lei Zhang^{2,3,4}, Liang Cao^{2,3,4}, Xiangru Sun⁶, Longqin Xu^{2,3,4}, Qian Wang¹

¹College of Computational Science, Zhongkai University of Agriculture and Engineering, Guangzhou Guangdong

²Intelligent Agriculture Engineering Research Center of Guangdong Higher Education Institutes, Zhongkai University of Agriculture and Engineering, Guangzhou Guangdong

³Guangdong Province Key Laboratory of Waterfowl Healthy Breeding, Zhongkai University of Agriculture and Engineering, Guangzhou Guangdong

⁴College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou Guangdong

⁵Institute of Reproductive Medicine, Panyu Hexian Memorial Hospital of Guangzhou, Guangzhou Guangdong

⁶Guangzhou E-Government Center, Guangzhou Guangdong

Email: 450799063@qq.com, *24715926@qq.com

Received: Jan. 24th, 2019; accepted: Feb. 6th, 2019; published: Feb. 13th, 2019

Abstract

This paper introduces the unknown equal variance model of bivariate Meta-analysis based on the difference mean. We obtain the maximum likelihood estimator of the difference mean effect and unknown variances; then we find the combining overall difference mean effect D_{overall} by the fixed-effect model of bivariate Meta-analysis based on the difference mean. Moreover, we obtain the corresponding covariance matrix and the $(1-\alpha)100\%$ confidence interval for the overall difference mean effect D_{overall} .

Keywords

Meta-Analysis, The Maximum Likelihood Estimate, Fixed-Effect Model

基于均差估计二变量Meta-分析的未知相等方差模型

郭鹏飞^{1,2,3}, 刘双印^{2,3,4}, 李刚^{5*}, 张垒^{2,3,4}, 曹亮^{2,3,4}, 孙向茹⁶, 徐龙琴^{2,3,4}, 王前¹

*通讯作者。

文章引用: 郭鹏飞, 刘双印, 李刚, 张垒, 曹亮, 孙向茹, 徐龙琴, 王前. 基于均差估计二变量 Meta-分析的未知相等方差模型[J]. 统计学与应用, 2019, 8(1): 134-142. DOI: 10.12677/sa.2019.81015

¹仲恺农业工程学院计算科学学院, 广东 广州

²仲恺农业工程学院广东省高校智慧农业工程研究中心, 广东 广州

³仲恺农业工程学院广东省水禽健康养殖重点实验室, 广东 广州

⁴仲恺农业工程学院信息科学与技术学院, 广东 广州

⁵广州市电子政务服务中心, 广东 广州

⁶番禺何贤纪念医院生殖医学科, 广东 广州

Email: 450799063@qq.com, *24715926@qq.com

收稿日期: 2019年1月24日; 录用日期: 2019年2月6日; 发布日期: 2019年2月13日

摘要

本文介绍基于均差估值的二变量Meta-分析的未知相等方差模型, 通过极大似然估计法给出效应量及方差估计, 进而得到各研究均值效应量协方差矩阵的估计量。然后, 通过基于均差效应量二变量Meta-分析的固定效应模型给出了合并均差统计量 D_{overall} 的具体形式和权重。进而得到了 D_{overall} 的协方差矩阵和两个变量的 $(1-\alpha)100\%$ 置信区间。

关键词

Meta-分析, 极大似然估计, 固定效应模型

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在科学研究的进程中, 有这样一类问题: 不同的研究者会对同一个问题进行研究, 在研究实验过程中, 总是通过实验组和对照组的比较给出一些问题的因果关系。由于不同的研究者所在的区域不同, 研究的对象和实验条件有一定的差异, 在实验中的样本量大小也各有不同, 这就导致同一项研究得到的研究结论有一定的差异。同一科学研究问题结论的差异性会产生这样的疑问: 如何对这一问题给出一个综合的评价, 得到一个具有统计学意义的因果关系呢? 对于这一问题, 我们需要对每一个研究者的研究进行合并, 给出每一个研究在综合评价研究中的权重。并且要保证我们得到的综合评价是具有统计学意义的。

在上述问题中, 一个科学问题需要客观评价时, 将所有这一问题的各个研究的数据放在一起, 这首先体现了数据的大量性; 由于对同一个问题研究有着很大的差异性, 因此得到的数据是多样的真实的。从这一问题的特点来看, 这应该是最早的大数据问题。如何从这些复杂多样的数据中获得我们想要的信息, 解决这一问题的方法称为 Meta-分析法。

Meta-分析是针对同一问题的不同研究进行定量合并的综合评价方法。虽然 Meta-分析在很多应用领域都有较好的应用, 但对 Meta-分析本身的统计方法研究相对少了很多。从广义上来讲, Meta-分析是以估计量标准误差的逆为权重的点估计加权平均。固定效应模型 Meta-分析依赖于各个研究具有相等效应量, 最终的效应量是以估计量方差的逆作为权重得到的加权平均效应量。虽然固定效应模型变量之间不

允许有协变量，但是对固定响应模型的处理相对来说是很标准的。

随着数据科学的快速发展，各行各业数据维度和数据量急速增加导致单变量 Meta-分析无法满足现实应用的需求。多变量 Meta-分析方法应运而生并具有许多单变量 Meta-分析不具备的优点，如整体性，相关性及优化参数等。针对多变量 Meta-分析的理论研究也在不断深入。1988 年，为了研究教练在 SAT 中的作用，Raudenbush SW 等人通过广义最小二乘法建立了多效应量合并的模型[1]；2002 年，van Houwelingen HC 等人在广义多元混合线性模型的框架下，通过似然估计方法给出了多元 Meta-回归模型，并将此模型扩展到了非正态分布情形[2]。2008 年，Riley RD 等人通过极大似然估计法给出了一些特殊相关系数情形下的二变量 Meta-分析的协方差矩阵估计[3]；2008 年，Ritz J 等人通过极大似然估计和估计方程给出了协方差矩阵已知情形下的多元效应量回归参数，并将此模型应用到肺癌发病率的临床推断中[4]。2010 年，Paul M 等人通过基于可积嵌套拉普拉斯近似的贝叶斯方法给出了多元 Meta-分析的合并效应量估计，这种方法得到的方差估计偏移量更小且稳定[5]。

本文通过极大似然估计法给出效应量及方差估计得到各研究均值效应量协方差矩阵的估计量。由基于均差效应量二变量 Meta-分析的固定效应模型给出了合并均差统计量 D_{overall} 的具体形式和权重，构建了基于均差估值的二变量 Meta-分析的未知相等方差模型。进而得到了 D_{overall} 的协方差矩阵和两个变量的 $(1-\alpha)100\%$ 置信区间。

2. 问题描述

在协方差矩阵的基于均差估值的二变量 Meta-分析的未知相等方差模型中，我们假设个体量 X_{ij}^C ， X_{ij}^T ， X_{i2j}^C 和 X_{i2j}^T 是独立的并且是正态分布，其均值分别为 μ_1^C ， μ_1^T ， μ_2^C 和 μ_2^T ，假设两个变量研究的方差分别为 σ^2 ， τ^2 。因此 \bar{X}_{i1}^C ， \bar{X}_{i1}^T ， \bar{X}_{i2}^C 和 \bar{X}_{i2}^T 是独立的且服从正态分布，其均值分别为 μ_1^C ， μ_1^T ， μ_2^C 和 μ_2^T ，方差分别为 $\frac{\sigma^2}{n_{i1}^C}$ ， $\frac{\sigma^2}{n_{i1}^T}$ ， $\frac{\tau^2}{n_{i2}^C}$ ， $\frac{\tau^2}{n_{i2}^T}$ 。假设两个变量之间的关系是相互独立的。

由上述假设，可以得到均差效应量 D_i 的分布为：

$$D_i = \begin{pmatrix} D_{i1} \\ D_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix} \right)$$

其中 $\mu_1 = \mu_1^T - \mu_1^C$ ， $\mu_2 = \mu_2^T - \mu_2^C$ 为均值， $w_{i1} = \frac{1}{n_{i1}^T} + \frac{1}{n_{i1}^C}$ ， $w_{i2} = \frac{1}{n_{i2}^T} + \frac{1}{n_{i2}^C}$ 及 ρ_0 是非随机的。一般地， μ_1 ， μ_2 为未知参数。接下来通过极大似然估计法来估计 μ_1 ， μ_2 。

3. 均值效应量的极大似然估计

根据之前的假设效应量 D_i 服从正态分布，因此关于 μ 的似然函数为：

$$L(\mu, D_i, S_i) = \prod_{i=1}^k \frac{1}{2\pi \det(S_i)^{\frac{1}{2}}} \exp \left\{ -\frac{(D_i - \mu)^T S_i^{-1} (D_i - \mu)}{2} \right\}$$

其中 $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ ， $S_i = \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix}$ 。其所对应的对数似然函数为：

$$\ln L(\mu, D_i, S_i) = -k \ln(2\pi) - \frac{1}{2} \sum_{i=1}^k \ln[\det(S_i)] - \frac{1}{2} \sum_{i=1}^k (D_i - \mu)^T S_i^{-1} (D_i - \mu)$$

定理 3.1 设效应量 D_i 服从正态分布 $D_i = \begin{pmatrix} D_{i1} \\ D_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix} \right)$ ，则

- 1). $\boldsymbol{\mu}$ 的极大似然估计量的分量为: $\hat{\mu}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}$, $\hat{\mu}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}$ 。
- 2). σ^2, τ^2 的极大似然估计量分别为: $\sigma_{ML}^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i1} - \hat{\mu}_1)^2}{w_{i1}}$, $\tau_{ML}^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i2} - \hat{\mu}_2)^2}{w_{i2}}$ 。

证明: 因为关于 $\boldsymbol{\mu}$ 的似然函数为:

$$L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i) = \prod_{i=1}^k \frac{1}{2\pi \det(\mathbf{S}_i)^{\frac{1}{2}}} \exp \left\{ -\frac{(\mathbf{D}_i - \boldsymbol{\mu})^T \mathbf{S}_i^{-1} (\mathbf{D}_i - \boldsymbol{\mu})}{2} \right\}$$

其所对应的对数似然函数为:

$$\begin{aligned} \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i) &= -\frac{1}{2} \sum_{i=1}^k \ln[\det(\mathbf{S}_i)] - \frac{1}{2} \sum_{i=1}^k (\mathbf{D}_i - \boldsymbol{\mu})^T \mathbf{S}_i^{-1} (\mathbf{D}_i - \boldsymbol{\mu}) - k \ln(2\pi) \\ &= -\frac{1}{2} \sum_{i=1}^k \ln[w_{i1} w_{i2} \sigma^2 \tau^2] - \frac{1}{2} \sum_{i=1}^k \left[\frac{(D_{i1} - \mu_1)^2}{\sigma^2 w_{i1}} + \frac{(D_{i2} - \mu_2)^2}{\tau^2 w_{i1}} \right] - k \ln(2\pi) \end{aligned}$$

1). 上述对数似然函数对 μ_1, μ_2 的偏导数为:

$$\frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \mu_1} = \sum_{i=1}^k \frac{D_{i1} - \mu_1}{\sigma^2 w_{i1}}, \quad \frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \mu_2} = \sum_{i=1}^k \frac{D_{i2} - \mu_2}{\tau^2 w_{i2}}$$

令 $\frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \mu_1} = 0, \frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \mu_2} = 0$ 即得:

$$\sum_{i=1}^k \frac{D_{i1} - \mu_1}{\sigma^2 w_{i1}} = 0, \quad \sum_{i=1}^k \frac{D_{i2} - \mu_2}{\tau^2 w_{i2}} = 0.$$

因此上述关于 μ_1, μ_2 的线性方程组的解为: $\hat{\mu}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}$, $\hat{\mu}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}$ 。

2). 上述对数似然函数对 σ, τ 的偏导数为:

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \sigma} &= -\frac{k}{\sigma} + \sum_{i=1}^k \frac{(D_{i1} - \mu_1)^2}{\sigma^3 w_{i1}}, \\ \frac{\partial \log L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \tau} &= -\frac{k}{\tau} + \sum_{i=1}^k \frac{(D_{i2} - \mu_2)^2}{\tau^3 w_{i1}}. \end{aligned}$$

令 $\frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \sigma} = 0, \frac{\partial \ln L(\boldsymbol{\mu}, \mathbf{D}_i, \sigma^2, \tau^2, \mathbf{S}_i)}{\partial \tau} = 0$ 即得:

$$-\frac{k}{\sigma} + \sum_{i=1}^k \frac{(D_{i1} - \mu_1)^2}{\sigma^3 w_{i1}} = 0, \quad -\frac{k}{\tau} + \sum_{i=1}^k \frac{(D_{i2} - \mu_2)^2}{\tau^3 w_{i1}} = 0.$$

因此上述关于 σ, τ 的方程组的解为: $\sigma^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i1} - \mu_1)^2}{w_{i1}}, \tau^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i2} - \mu_2)^2}{w_{i2}}$ 。

将 $\hat{\mu}_1, \hat{\mu}_2$ 代入上式即得: $\sigma_{ML}^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i1} - \hat{\mu}_1)^2}{w_{i1}}, \tau_{ML}^2 = \frac{1}{k} \sum_{i=1}^k \frac{(D_{i2} - \hat{\mu}_2)^2}{w_{i2}}$ 。

注: 由上述定理可知, $\mu_{ML} = \left(\sum_{i=1}^k W_i^{-1} \right)^{-1} \sum_{i=1}^k W_i^{-1} D_i$, 其中 $W_i = \begin{pmatrix} w_{i1} & 0 \\ 0 & w_{i2} \end{pmatrix}, i=1, \dots, k$; 且

$$S_{iML} = \begin{pmatrix} \sigma_{ML}^2 w_{i1} & 0 \\ 0 & \tau_{ML}^2 w_{i2} \end{pmatrix}, i=1, \dots, k。$$

由上述效应量 D_i 服从正态分布的合并效应量均值估计量分量的具体形式, 可以得到合并效应量均值估计量的如下性质:

性质 3.1 设效应量 D_i 服从正态分布 $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix}\right)$, μ 的极大似然估计量的分量为:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}, \hat{\mu}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}。则 \hat{\mu}_{ML} 对于 \mu 是无偏的。$$

证明: 要证明估计量的无偏性, 需证等式 $E(\hat{\mu}_{ML}) = \mu$ 成立。根据我们给出的效应量 D_i 服从正态分布的合并效应量均值估计量分量的具体形式, 需要证明 $E(\hat{\mu}_1) = \mu_1, E(\hat{\mu}_2) = \mu_2$ 。因为 $E(D_{i1}) = \mu_1, E(D_{i2}) = \mu_2$, 所以

$$E(\hat{\mu}_1) = E\left\{ \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}} \right\} = \frac{\sum_{i=1}^k \frac{E(D_{i1})}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}} = \frac{\sum_{i=1}^k \frac{\mu_1}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}} = \mu_1,$$

$$E(\hat{\mu}_2) = E\left\{ \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}} \right\} = \frac{\sum_{i=1}^k \frac{E(D_{i2})}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}} = \frac{\sum_{i=1}^k \frac{\mu_2}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}} = \mu_2。$$

综上所述, $\hat{\mu}_{ML}$ 对于 μ 是无偏的。

4. 均值效应量极大似然估计量的协方差矩阵

假设所有的研究都提供所有的均值效应。由多变量统计学的大数定理, 合并效应量均值估计量可以近似为一个多元正态分布, 其对应的协方差矩阵可由下面的定理给出:

定理 4.1 设效应量 D_i 服从正态分布 $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix}\right)$, 那么 μ 的合并均值效应量极大似然估

$$\text{量为: } \hat{\mu}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}, \hat{\mu}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}。则 \hat{\mu}_{ML} 所对应的协方差矩阵 Var(\hat{\mu}_{ML}) 为: Var(\hat{\mu}_{ML}) = \begin{pmatrix} \sigma^2 & 0 \\ \sum_{i=1}^k \frac{1}{w_{i1}} & \\ 0 & \tau^2 \\ \sum_{i=1}^k \frac{1}{w_{i2}} & \end{pmatrix}$$

证明: 根据多元随机变量可知: 如果随机变量 $\mathbf{D} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $Var(\mathbf{MD}) = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}^T$, 其中 T 表示矩阵的转置。

因为合并均值效应量极大似然估计量的矩阵形式为: $\boldsymbol{\mu}_{ML} = \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \sum_{i=1}^k \mathbf{W}_i^{-1} \mathbf{D}_i$, 且 $\mathbf{S}_i = \mathbf{W}_i \boldsymbol{\Sigma}_1$, 其中 $\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}$ 由 $\mathbf{W}_i = \mathbf{W}_i^T$ 及 $\mathbf{W}_i, \boldsymbol{\Sigma}_1$ 为对角矩阵, 通过计算可得: $\mathbf{W}_i^{-1} = (\mathbf{W}_i^{-1})^T$, $\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} = \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \right]^T$, $\mathbf{W}_i \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1 \mathbf{W}_i$, $\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1 \mathbf{W}_i^{-1}$, $\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1 \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1}$ 。则

$$\begin{aligned} Var(\hat{\boldsymbol{\mu}}_{ML}) &= Var \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \sum_{i=1}^k \mathbf{W}_i^{-1} \mathbf{D}_i \right] \\ &= Var \left[\sum_{i=1}^k \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \mathbf{D}_i \right] \\ &= \sum_{i=1}^k Var \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \mathbf{D}_i \right] \\ &= \sum_{i=1}^k \left\{ \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \right] \mathbf{S}_i \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \right]^T \right\} \\ &= \sum_{i=1}^k \left\{ \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \right] \mathbf{W}_i \boldsymbol{\Sigma}_1 \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \right]^T \right\} \\ &= \boldsymbol{\Sigma}_1 \sum_{i=1}^k \left\{ \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \mathbf{W}_i^{-1} \mathbf{W}_i (\mathbf{W}_i^{-1})^T \left[\left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \right]^T \right\} \\ &= \boldsymbol{\Sigma}_1 \left\{ \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right) \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \right\} \\ &= \boldsymbol{\Sigma}_1 \left(\sum_{i=1}^k \mathbf{W}_i^{-1} \right)^{-1} \end{aligned}$$

由 $\mathbf{W}_i = \begin{pmatrix} w_{i1} & 0 \\ 0 & w_{i2} \end{pmatrix}$ 及 $\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}$ 可得 $\hat{\boldsymbol{\mu}}_{ML}$ 所对应的协方差矩阵 $Var(\hat{\boldsymbol{\mu}}_{ML})$ 为:

$$Var(\hat{\boldsymbol{\mu}}_{ML}) = \begin{pmatrix} \frac{\sigma^2}{\sum_{i=1}^k \frac{1}{w_{i1}}} & 0 \\ 0 & \frac{\tau^2}{\sum_{i=1}^k \frac{1}{w_{i2}}} \end{pmatrix}。$$

通过合并均值效应量极大似然估计量 $\hat{\boldsymbol{\mu}}_{ML}$ 所对应的协方差矩阵 $Var(\hat{\boldsymbol{\mu}}_{ML})$ 可以给出 $\hat{\boldsymbol{\mu}}_{ML}$ 的如下性质:

性质 4.2 设效应量 \mathbf{D}_i 服从正态分布 $N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix} \right)$, 那么 $\boldsymbol{\mu}$ 的极大似然估计量的分量为:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}, \hat{\mu}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}. \text{ 则 } \hat{\mu}_{ML} \text{ 对于 } \mu \text{ 是一致的。}$$

证明: $\hat{\mu}_{ML}$ 对于 μ 的一致性的充分条件是证明 $\lim_{k \rightarrow \infty} E(\hat{\mu}_{ML}) = \mu$ 并且 $\lim_{k \rightarrow \infty} V(\hat{\mu}_{ML}) = 0$ 。

由性质 3.1 可知, 对于任意的 k 都有 $E(\hat{\mu}_{ML}) = \mu$ 成立, 因此可得 $\lim_{k \rightarrow \infty} E(\hat{\mu}_{ML}) = \mu$ 。又因为在合并均值效应量极大似然估计量 $\hat{\mu}_{ML}$ 所对应的协方差矩阵 $Var(\hat{\mu}_{ML})$ 中有:

$$\lim_{k \rightarrow \infty} Var(\hat{\mu}_{ML})_{(1,1)} = \lim_{k \rightarrow \infty} \frac{\sigma^2}{\sum_{i=1}^k \frac{1}{w_{i1}}} = 0,$$

$$\lim_{k \rightarrow \infty} Var(\hat{\mu}_{ML})_{(2,2)} = \lim_{k \rightarrow \infty} \frac{\tau^2}{\sum_{i=1}^k \frac{1}{w_{i2}}} = 0.$$

即得 $\lim_{k \rightarrow \infty} V(\hat{\mu}_{ML}) = 0$ 。综上所述, $\hat{\mu}_{ML}$ 对于 μ 是一致的。

5. 整体均值合并统计量的估计

在上述研究中, 效应量用的是均差。因此我们可以给出均值效应量 D_i 的协方差估计为: $\hat{S}_{iML} = W_i \hat{\Sigma}_1$,

$$\text{其中 } \hat{\Sigma}_1 = \begin{pmatrix} \sigma_{ML}^2 & 0 \\ 0 & \tau_{ML}^2 \end{pmatrix}.$$

在 Meta-分析中, 合并统计量是由具体研究表现统计量的加权平均给出的。根据均差估计二变量 Meta-分析的固定效应模型可知, 通过 \hat{S}_{iML} 可以给出每一个研究的权重及均差合并统计量。具体如下:

1). 对于均值合并统计量, 每个研究均差的权重为:

$$\mathbf{v}_i = \left(\sum_{i=1}^k \hat{S}_{iML}^{-1} \right)^{-1} \hat{S}_{iML}^{-1}$$

$$= \left[w_{i1} w_{i2} \sum_{j=1}^k \frac{1}{w_{j1}} \sum_{j=1}^k \frac{1}{w_{j2}} \right]^{-1} \begin{pmatrix} \sum_{j=1}^k \frac{w_{j2}}{w_{j2}} & 0 \\ 0 & \sum_{j=1}^k \frac{w_{j1}}{w_{j1}} \end{pmatrix}.$$

注意到权重 \mathbf{v}_i 只与个体研究的样本量有关并且是非随机的。

$$2). \text{ 整体均值合并统计量的估计量为: } \mathbf{D}_{\text{overall}} = \sum_{i=1}^k \mathbf{D}_i \mathbf{v}_i = \begin{pmatrix} \sum_{i=1}^k \frac{D_{i1}}{w_{i1}} \\ \sum_{i=1}^k \frac{1}{w_{i1}} \\ \sum_{i=1}^k \frac{D_{i2}}{w_{i1}} \\ \sum_{i=1}^k \frac{1}{w_{i2}} \end{pmatrix}$$

从上式我们注意到, $\mathbf{D}_{\text{overall}}$ 与合并均值的极大似然估计量 $\hat{\mu}_{ML}$ 的结果是一致的。因此 $\mathbf{D}_{\text{overall}}$ 的协方差矩阵的估计量为:

$$\text{Var}(\mathbf{D}_{\text{overall}}) = \begin{pmatrix} \frac{\hat{\sigma}_{ML}^2}{\sum_{i=1}^k \frac{1}{w_{i1}}} & 0 \\ 0 & \frac{\hat{\tau}_{ML}^2}{\sum_{i=1}^k \frac{1}{w_{i2}}} \end{pmatrix} = \begin{pmatrix} \frac{\frac{1}{k} \sum_{i=1}^k \frac{(D_{i1} - \hat{\mu}_1)^2}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}} & 0 \\ 0 & \frac{\frac{1}{k} \sum_{i=1}^k \frac{(D_{i2} - \hat{\mu}_2)^2}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}} \end{pmatrix}.$$

通过 $\mathbf{D}_{\text{overall}}$ 所对应的协方差矩阵 $\text{Var}(\mathbf{D}_{\text{overall}})$ 可以给出单变量和联合变量的置信区域, 其具体形式为:

性质 5.1 设效应量 D_i 服从正态分布 $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 w_{i1} & 0 \\ 0 & \tau^2 w_{i2} \end{pmatrix}\right)$, 那么合并均值效应量 $\mathbf{D}_{\text{overall}}$ 的估计量

为: $\hat{D}_1 = \frac{\sum_{i=1}^k \frac{D_{i1}}{w_{i1}}}{\sum_{i=1}^k \frac{1}{w_{i1}}}$, $\hat{D}_2 = \frac{\sum_{i=1}^k \frac{D_{i2}}{w_{i2}}}{\sum_{i=1}^k \frac{1}{w_{i2}}}$ 。则合并均值效应量 $\hat{\mu}_{ML}$ 所对应各个分量的 $(1-\alpha)100\%$ 置信区间为:

$$\left(\mu_1 - Z_{\frac{\alpha}{2}} \sqrt{\text{Var}(\mathbf{D}_{\text{overall}})_{(1,1)}}, \mu_1 + Z_{\frac{\alpha}{2}} \sqrt{\text{Var}(\mathbf{D}_{\text{overall}})_{(1,1)}} \right) \text{ 和 } \left(\mu_2 - Z_{\frac{\alpha}{2}} \sqrt{\text{Var}(\mathbf{D}_{\text{overall}})_{(2,2)}}, \mu_2 + Z_{\frac{\alpha}{2}} \sqrt{\text{Var}(\mathbf{D}_{\text{overall}})_{(2,2)}} \right),$$

其中 $Z_{\frac{\alpha}{2}}$ 表示正态分布的 $\frac{\alpha}{2}$ 分位数, $\text{Var}(\mathbf{D}_{\text{overall}})_{(i,j)}$ 表示矩阵 $\text{Var}(\mathbf{D}_{\text{overall}})$ 第 i 行第 j 列的分量。

6. 结束语

本文通过极大似然估计法给出效应量及方差估计, 进而得到各研究均值效应量协方差矩阵的估计量。然后, 由基于均差效应量二变量 Meta-分析的固定效应模型给出了合并均差统计量 $\mathbf{D}_{\text{overall}}$ 的具体形式和权重, 构建了基于均差估值的二变量 Meta-分析的未知相等方差模型, 进而得到了 $\mathbf{D}_{\text{overall}}$ 的协方差矩阵和两个变量的 $(1-\alpha)100\%$ 置信区间。本文的结果对流行病学的研究有着重要的统计学意义。

声明

本项目由如下基金支持: 国家自然科学基金(61871475, 61471133, 61571444, 61473331), 广东省科技计划(2017B010126001, 2017A070712019, 2016A040402043, 2015A070709015, 2015A020209171, 2016B010125004, 2014B040404070, 2015A040405014, 2016A070712020), 广东省教育厅科技计划(2017GCZX001, 2016GCZX001, 2017KTSCX094, 2017KTSCX095, 2017KQNCX098), 广州市科技计划(201707010221)。

参考文献

- [1] Raudenbush, S.W., Becker, B.J. and Kalaian, H. (1988) Modeling Multivariate Effect Sizes. *Psychological Bulletin*, **103**, 111-120. <https://doi.org/10.1037/0033-2909.103.1.111>
- [2] van Houwelingen, H.C., Zwinderman, K.H. and Stijnen, T. (1993) A Bivariate Approach to Meta-Analysis. *Statistics in Medicine*, **12**, 2273-2284. <https://doi.org/10.1002/sim.4780122405>
- [3] Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J. and Thompson, J.R. (2007) Bivariate Random Effects Meta-Analysis and the Estimation of Between-Study Correlation. *BMC Medical Research Methodology*, **7**, 3. <https://doi.org/10.1186/1471-2288-7-3>
- [4] Ritz, J., Demidenko, E. and Spiegelman, D. (2008) Multivariate Meta-Analysis for Data Consortia, Individual Patient Meta-Analysis, and Pooling Projects. *Journal of Statistical Planning and Inference*, **138**, 1919-1933.

<https://doi.org/10.1016/j.jspi.2007.07.004>

- [5] Paul, M., Riebler, A., Bachmann, L.M., Rue, H. and Held, L. (2010) Bayesian Bivariate Meta-Analysis of Diagnostic Test Studies Using Integrated Nested Laplace Approximations. *Statistics in Medicine*, **29**, 1325-1339.
<https://doi.org/10.1002/sim.3858>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org