Universe Scientific Publishing Volume 8 Issue 04

Hierarchical Bayesian and meta-analysis

Yong Wang[1]*, Ying Zhou*[2]*, Minghui Zhang*[1*]
1. Guangzhou Institute of Navigation, Guangzhou 510000, China
2. Guangdong Institute of Technology, Zhaoqing 526000, China

Abstract: Meta analysis is a method to extract the data analysis results from previous empirical studies for secondary analysis. It was first proposed by American educational psychologist Douglas Yu. Since then, the method has been widely used in scientific fields such as education, psychology, management and epidemiology. Bayesian network is a directed acyclic graph, and Bayesian probability analysis for data analysis is the statistical foundation of machine learning.

Key words: Meta; Bayesian network; Hierarchical linear model; EM algorithm

1.Meta analysis

1.1 Effective, coding

In meta-analysis, effect size refers to the coefficient of correlation between two variables. For survey studies, reporting the correlation matrix, you can get the direct correlation coefficient r; For experimental studies, instead of directly reporting the correlation coefficient between two variables, the mean M and standard deviation SD, the correlation coefficient r is calculated.

Encoding refers to collating the qualitative and quantitative data in the study according to predetermined rules. In the coding phase, all the information needed for the research needs to be sorted out.

1.2 Publication bias, heterogeneity

Publication bias means that the calculated effect size is biased because the study sample is biased. The main reasons are as follows: (1) significant findings are more likely to be published; And (2) published studies are more likely to be included in meta-analysis studies. Although publication bias cannot be completely eliminated, specific research indicators can be used to demonstrate that the publication bias of the study is not severe.

Heterogeneity refers to significant differences in effect sizes between study samples. When heterogeneity is high, it is particularly important to explore the causes of the differences (the process of finding moderators).

2. Advantages of meta analysis

2.1 Objectivity

As a quantitative literature review, meta-analysis needs to cover all literatures related to interest, and the research conclusions drawn from it will be more objective. On the other hand, in terms of the selection of moderating variables, the moderating variables of qualitative literature review also depend on the subjective judgment of the researcher and will not be verified. The selection of moderating variables in meta-analysis is generally based on the heterogeneity of variable relationships and obtained through data observation, and the moderating variables are verifiable. Accuracy is also related to qualitative literature review. Through standardized literature search, meta-analysis increased the sample size and statistical power, overcame the limitation of small sample size, and improved the accuracy of effect size estimation.

2.2 Economy

In meta-analysis, researchers' ideas do not need additional sample testing, and can only be verified by using existing research results; On the other hand, since any variable relationship in the correlation matrix can be used as the basis of meta-analysis, it provides researchers with a large amount of data that can be used for analysis, thus improving the economy of meta-analysis.

3. Meta analysis process

3.1. Problems and hypotheses

Determine research values: whether repeatable research can be achieved, whether controversial and inconsistent issues can be resolved, and whether future research directions can be indicated.

3.2 Collect and evaluate literature screening

3.2.1 Develop search strategies and screening criteria: Identify search terms, databases and relevant screening criteria

3.2.2 Methods: Research systematic reviews of the same or similar topics

3.2.3 Consistency in search strategies and screening criteria.

3.3 Analysis of data calculation

3.3.1 Combinatorial statistics: showing the combined effect of multiple studies

3.3.2 Indicators are dichotomous variables: odds ratio (OR), relative risk (RR), risk spread (RD), mean spread (MD or WMD), standard mean spread (SMD).

3.3.3 Null hypothesis for different statistical models: Fixed-effect models: There is no difference between each independent effect size (unless heterogeneity is evident). Random effect model: Each independent effect size is based on the aggregation of multiple true effect sizes, there is some degree of difference between independent effect sizes, and the choice of random effect model and fixed effect model should depend on the inferential process made by the meta-analyst. Summarize the results when conducting correlation analysis: Choose either a fixed or a random model. A large number of studies these days are randomized models.

3.3.4 Calculation of effects: Forest plots show effect sizes and confidence intervals for each study. If there is little overlap between the confidence intervals of the studies, this indicates that there may be a preliminary determination of heterogeneity.

3.3.5 Post-event heterogeneity test According to the statistical principle, only homogeneous data can be combined or compared for statistical analysis, and vice versa. Heterogeneity testing is a key step in synthesizing the effect size of individual studies into the overall effect size, essentially testing whether each study belongs to the same distribution. There are two common ways to distinguish: Q test and Chi-square test. If the results of the heterogeneity test indicate that the study is homogeneous, the combined effect values are calculated using the fixed-effect model. Instead, choose a random effects model.

3.3.6 Sensitivity analysis: Studies with abnormal effect sizes can be found by culling each of the included studies one by one and then combining the effect sizes.

3.4 Discussion

The discussion ideas are as follows: the report should explain the comparison between important findings and existing meta-analysis conclusions, the relationship with hypotheses, and analyze possible causes to indicate future research directions.

To be specific:

4.4.1 Analyze the source of variation. If heterogeneity exists among the studies included in the meta-analysis, the outliers should be examined and the source of the heterogeneity and its effect on the summary value of the effects should be discussed. Sources of heterogeneity mainly include inconsistent inclusion criteria among studies and different baseline levels between studies.

3.4.2 Discuss the identification and control of various biases. Meta-analysis is essentially an observational study, so there will inevitably be various biases in the meta-analysis process. Therefore, the bias and its possible causes should be determined by calculating the fail-safe factor or drawing a funnel plot.

3.4.3 It must be noted that the meta-analysis is an observational study and therefore its results must be interpreted with great caution. When discussing the results of meta-analysis, the discussion should be combined with the research background and practical significance, and if necessary, the consistency of the results of a single study with a large sample can be considered for comparison with the results of meta-analysis.

4. Questions suitable for meta-analysis

4.1 Validation Studies

Researchers have a preliminary expectation of the outcome, but only use meta-analysis methods to further test their conjectural assumptions, such as antecedents and consequences studies, which usually focus only on simple bivariate relationships.

4.2 Exploratory research

It means that the researcher already has the research question, but does not know the answer to the question, such as the exploration of mediating mechanisms, the exploration of regulatory effects. This type of research usually involves three or more variables: (1) mediating: exploring the mediating mechanism of the association between two variables (2) mediating: reconciling inconsistent findings. It is mainly divided into the following seven categories of research situations; Research design: experiment and investigation; Sample characteristics: type of participants, sample recovery rate, regional differences of samples, etc. Task characteristics: routine task and non-routine task; Variable measurement methods: self-assessment and other assessment; Document type: journal paper, conference paper, thesis; In the meta-analysis process, the control variables involved in the literature included in the meta-analysis can also be used as important data sources, thus providing a large number of raw materials for our research meta-analysis.

5. Bayesian networks

A single meta-analysis can usually only evaluate unit relationships if one wants to compare the impact of multiple factors on the same problem, in addition many problem phenomena may be multifactorial or have multiple risk factors. Therefore, network analysis is needed to compare multiple factors of the same topic in network meta-analysis. Construct network meta-analysis based on Bayesian posterior probability method to realize the common analysis of all studies.

In network meta-analysis, it is very difficult to sample from prior distributions. At this time, the Markov chain Monte Carlo method is needed to calculate the prior distribution, and the transfer calculation through the steady-state Markov chain is equivalent to sampling from the P(x) distribution, so as to realize the calculation of the prior distribution, and the increase of the problem dimension will not slow down the convergence speed or complicate its convergence.

After the prior probability is obtained, the relative effect of each study can be further calculated, and the corresponding model can be selected according to the characteristics of the effect value variables to be calculated, and the heterogeneity between studies can be determined. A Bayesian network is a directed acyclic graph with nodes representing random variables, which can be observable measurements, hidden variables, unknown parameters or assumptions; The directed edges between the nodes represent conditional dependencies, with arrows pointing to the nodes. The point depends on the node from which the arrow emanates (the parent node), and each node is associated with a probability function.

Bayesian network diagrams require a large amount of raw datas to be input for computation. When Bayesian networks are structurally fixed as standardized inputs, they can only be linearly combined with variables first, then nonlinear transformation of variables and estimation of the parameters of the probabilistic model at the end.

6. Hierarchical Linear model

6.1 Hierarchical Linear model,

The linear regression model studies the effect of X on Y, and the HLM (hierarchical linear model) model also studies the effect of X on Y, but it takes into account the clustering factor of the group (i.e. the intra-group correlation problem is not independent). Without considering the group, i.e., without considering the “aggregation” problem, then linear regression can be used directly without using the HLM model. The HLM model is an advanced way to deal with the “aggregation” problem; If the HLM model is used, the individual effect is considered in the analysis, and the group-level effect is also considered, that is, the fixed effect term and the random effect term are included.

6.2 EM Algorithm

It is troublesome when there are hidden variables. If I know which samples belong to the same class, I can estimate the distribution parameters of the samples from the maximum likelihood function; Similarly, only if I know the distribution parameters of samples of different classes can I determine which class the sample is more likely to belong to.

We first assign an initial distribution to each class (i.e., the hidden variable) based on experience, which is equivalent to assuming the distribution parameters. Then, based on the parameters of the distribution, we can get the expectation of the hidden variable for each data tuple; (2) Calculate the maximum likelihood value of the distribution parameter according to the classification result, and then recalculate the expectation of the hidden variable of each tuple according to the maximum likelihood value of the classification result. So the cycle repeats, and finally if the expectation of the hidden variable and the maximum likelihood value of the parameter are stable, then the EM algorithm is executed.

It consists of two steps. The first step is step E, which is seeking expectations; The second step is step M, which is maximization: Step E (expectation) : calculate the expectation of the sample hidden variable based on the current parameter values; M step (maximum) : based on the current sample’s hidden variable, solve for the maximum likelihood estimate of the parameter.

Numerical methods used to estimate the parameters of a hierarchical model: weighted generalized least square method and Fisher scoring algorithm. The Fisher scoring algorithm is a log-likelihood function that uses Newton’s method to maximize logistic regression. Stata and SAS are used to deal with complex multidimensional estimates in hierarchical models, such as mixed effects and random effects models in biometrics; Stochastic coefficient regression models in econometrics; And component variance models in statistics. But there are many differences between the modeling groups.

Bayesian layered models are extensions of general layered models, allowing for additional sources of uncertainty. Examples include uncertainty of the model, which is not usually represented in traditional models. The biggest feature of Bayesian models is the use of prior distributions to deal with parameters of position. By using Markov chain Monte Carlo (MCMC) methods, especially GibbsSampling and MetropolisHastings’ algorithms, Bayesian computation becomes easier to implement.

References:

[1] Shuangli Xu,Zhaohe Lv. Development and application of Meta analysis method in economics [J]. Journal of statistics and information BBS, 2019, 34 (8) : 8. DOI: CNKI: SUN: TJLT. 0.2019-08-006.

[2] Kuo Liu,Dianqin Sun,Xing Liao, etal. [J]. Chinese Journal of Evidence-based Cardiovascular Medicine, 2019, 11(3):8.

[3] Yonggang Zhang,Letian Yang,Xin Yang, etal. Interpretation of Systematic Review/Meta-Analysis Reporting Standard (PRISMA-DTA) for diagnostic accuracy tests [J]. Chinese Journal of Evidence-Based Medicine, 2018, 18(9):10.

[4] Tiansong Zhang. Rational use of funnel plot in traditional meta-analysis [J]. Chin J Hospital Statistics, 2023, 30(4):304-308. (in Chinese)

[5] Pengxiang Zhou,Yingying Yan,Suodi Zhai. Methodology and report quality evaluation for systematic evaluation/meta-analysis of hospital pharmaceutical staff in China [J]. Chinese Journal of Evidence-Based Medicine, 2017, 17(2):7.

[6] Zhuoya Ju,Zhihai Wang. Bayesian Classification Algorithm based on Selective Pattern. Journal of Computer Research and Development, 2020, 57(8):12.

* Corresponding author: ZHANG Minghui